

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/59503>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Kant, Nietzsche, and the Moral Agent

by

Thomas W. Bailey

A thesis submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy in Philosophy

University of Warwick, Department of Philosophy

December 2003

Table of contents

Acknowledgements	v
Declaration	vi
Abstract	vii
Abbreviations	viii
Introduction	1
I. Choice, Reasons, and Kant: Nietzsche's Criticism of the Moral Agent	
Introduction to part I	7
1. Choice	8
<i>Willing</i>	9
<i>'I' and 'free will'</i>	18
<i>Conclusions</i>	31
2. Reasons	33
<i>Against practical reason</i>	34
<i>The 'revaluation of values'</i>	39
<i>Conclusions</i>	51
3. Prevailing approaches	53
<i>Five prevailing strands</i>	54
<i>Responsibility</i>	60
<i>Conclusions</i>	62

4.	Nietzsche on Kant	63
	<i>Standard interpretations</i>	64
	<i>Nietzsche's remarks about Kant</i>	67
	<i>Conclusions</i>	84
	Conclusions of part I	86
II.	The Good Will: Kant's Analysis of the Moral Agent	
	Introduction to part II	91
5.	The argument of the <i>Groundwork</i>, sections I and II	92
	<i>From a good will to the formula of universal law</i>	94
	<i>'Stepping out into metaphysics'</i>	119
	<i>Kant's method</i>	139
	<i>Conclusions</i>	143
6.	Moral reasons	144
	<i>The formula of universal law</i>	146
	<i>Kant's further formulas</i>	170
	<i>The object of moral judgement</i>	187
	<i>Conclusions</i>	201
7.	Spontaneity in choice	202
	<i>Intelligible and empirical choice</i>	203
	<i>An argument from rational choice</i>	209
	<i>Doubt regarding autonomy: ambitious and modest responses</i>	214
	<i>An argument from responsibility</i>	229

	<i>Choice and reasons</i>	232
	<i>Conclusions</i>	237
8.	Moral anthropology	238
	<i>Conclusions</i>	250
	Conclusions of part II	251
III.	The Animal That May Promise: Nietzsche's Rival Analysis of the Moral Agent	
9.	Nietzsche's rival analysis	255
	<i>The 'animal that may promise'</i>	257
	<i>Nietzsche's Kantian ethics</i>	261
	<i>History and common sense</i>	276
	<i>Conclusions</i>	281
	Bibliography	282

Acknowledgements

I thank Stephen Houlgate for his exemplary supervision and extreme patience. For discussion of aspects of the thesis, I also thank Darren Ambrose, Keith Ansell-Pearson, Christine Battersby, Thomas Brobjer, Frank Cameron, Tsarina Doyle, Randall Havas, Lee Kerckhove, Peter Poellner, Dennis Schulting, and William Smith. Finally, I thank my family and friends for their support.

Declaration

I declare that this thesis is my own work and that it has not been submitted for another degree. The final chapter of the thesis includes, in modified form, material included in my 'Nietzsche's Kantian Ethics', in *International Studies in Philosophy*, v.35, no.3, 2003, pp.5-27.

Abstract

This thesis examines Kant's and Nietzsche's treatments of the moral agent. It argues for three broad conclusions. Firstly, it argues that, although Nietzsche's explicit criticisms of Kant's conception of the moral agent can be understood only in the context of Nietzsche's broader moral philosophy, neither these criticisms nor their context are well understood by the prevailing literature. The thesis thus engages with existing scholarship on the nature of Nietzsche's moral philosophy and with the scanty literature on the relationship between Kant's and Nietzsche's moral philosophies. Secondly, the thesis argues that Kant's conception of the moral agent is not undermined by the criticisms which Nietzsche explicitly levels at it, or, indeed, by others which are commonly made in Nietzsche's name. In doing so, the thesis combines original interpretations of Kant with elements of recent Kant scholarship. Finally, however, the thesis argues that neglected elements of Nietzsche's own moral philosophy provide for a more sophisticated, telling, and, indeed, original critical engagement with Kant's conception of the moral agent. Thus the thesis defends an original interpretation of Nietzsche's moral philosophy and its critical relation to Kant's, and demonstrates the pertinence of a certain neglected critical approach to Kant's conception of the moral agent. On the basis of these conclusions, the thesis ultimately defends a conception of the moral agent which, although Kantian, owes something to both Kant and Nietzsche.

Abbreviations

With the exception of the *Kritik der reinen Vernunft*, *Reflexionen*, and correspondence, Kant's texts are referred to by page number in the appropriate volume of the *Akademie* edition (*Königlich Preußische* [now *Deutsche*] *Akademie der Wissenschaften* ed., *Kants gesammelte Schriften*, Berlin: Georg Reimer [now De Gruyter], 1900-). The appropriate volume numbers are given below, following the year of publication of the first edition and that of any significant later edition of each text. In the case of texts which are included in more than one volume of the *Akademie* edition, references also provide the volume number immediately after the abbreviation of the title, with page numbers following after a colon. References to the *Kritik der reinen Vernunft* are to page numbers in the first ('A') and/or second ('B') editions, and references to *Reflexionen* and correspondence are to numbers in the appropriate volume of the *Akademie* edition.

Nietzsche's texts are referred to by section number, with the following exceptions: *Also sprach Zarathustra*, *Zur Genealogie der Moral*, *Götzen-Dämmerung*, and *Ecce Homo* are referred to by part, chapter, or essay number, or chapter title, followed by section number, Nietzsche's unpublished notebooks are referred to by part, volume, notebook, and manuscript numbers in Giorgio Colli and Mazzino Montinari eds., *Werke: Kritische Gesamtausgabe*, Berlin: De Gruyter, 1967- , and Nietzsche's correspondence is referred to by numbers in Colli and Montinari eds., *Briefwechsel*, Berlin: De Gruyter, 1978- . The additional abbreviation 'V' indicates a reference to the text's 'Vorrede' or 'Vorwort'. The year of publication of the first edition and that of any significant later edition of

each text are given below, followed in each case by the appropriate part and volume numbers of the Colli and Montinari edition.

Kant

- KrV* *Kritik der reinen Vernunft* (1st ed., 1781; 2nd ed., 1787; 3, 4)
- P* *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft
wird auftreten können* (1783; 4)
- RS* 'Rezension von Johann Heinrich Schulz's *Versuch einer Anleitung zur
Sittenlehre für alle Menschen*' (1783; 8)
- WA* 'Beantwortung der Frage: Was ist Aufklärung?' (1784; 8)
- I* 'Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht'
(1784; 8)
- G* *Grundlegung zur Metaphysik der Sitten* (1785; 4)
- MAM* 'Mutmaßlicher Anfang der Menschengeschichte' (1786; 8)
- MAN* *Metaphysische Anfangsgründe der Naturwissenschaften* (1786; 4)
- KpV* *Kritik der praktischen Vernunft* (1788; 5)
- RU* [Kraus] 'Rezension von Johann August Heinrich Ulrichs *Eleutheriologie*'
(1788; 8)
- KU* *Kritik der Urteilskraft* (1790; 5)
- R* *Die Religion innerhalb der Grenzen der bloßen Vernunft*
(1st ed., pt. 1, 1792, pts. 1-4, 1793, 2nd ed., 1794; 6)
- TP* 'Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber
nicht für die Praxis' (1793; 8)
- E* 'Das Ende aller Dinge' (1794; 8)

- EF* *Zum Ewigen Frieden: Ein philosophischer Entwurf*
(1st ed., 1795; 2nd ed., 1796; 8)
- MS* *Die Metaphysik der Sitten* (1st ed., 1797; 2nd ed., 1798; 6)
- SF* *Der Streit der Fakultäten* (1798; 7)
- ApH* *Anthropologie in pragmatischer Hinsicht* (1798; 7)
- VL* *Immanuel Kants Logik: Ein Handbuch zu Vorlesungen* (1800; 9)
- Re* *Reflexionen* (–; 14-19)
- EKU* ‘Erste Einleitung in der Kritik der Urteilskraft’ (–; 20)
- VE* *Vorlesungen über Ethik* (–; 27, 29)
- B* *Briefwechsel* (–; 10-13)

Nietzsche

- GT Die Geburt der Tragödie aus dem Geiste der Musik*
(1st ed., 1872; 2nd ed., 1878, 1886; III:1)
- H Vom Nutzen und Nachtheil der Historie für das Leben* (1874; III:1)
- RW Richard Wagner in Bayreuth* (1876; IV:1)
- MA Menschliches, Allzumenschliches: Ein Buch für freie Geister*
(1st ed., 1878; 2nd ed., 1886; IV:2)
- MS Vermischte Meinungen und Sprüche* (1st ed., 1879; 2nd ed., 1886; IV:3)
- WS Der Wanderer und sein Schatten* (1st ed., 1880; 2nd ed., 1886; IV:3)
- M Morgenröte: Gedanken über die moralischen Vorurteile*
(1st ed., 1881; 2nd ed., 1886; V:1)
- FW Die fröhliche Wissenschaft* (1st ed., 1882; 2nd ed., 1887; V:2)
- Z Also sprach Zarathustra: Ein Buch für Alle und Keinen*
(pts.1 and 2, 1883, pt.3, 1884, pt.4, 1885; VI:1)
- JGB Jenseits von Gut und Böse: Vorspiel Einer Philosophie Der Zukunft*
(1886; VI:2)
- GM Zur Genealogie der Moral: Ein Streitschrift* (1887; VI:2)
- W Der Fall Wagner: Ein Musikanten-Problem* (1888; VI:3)
- GD Götzen-Dämmerung, oder Wie man mit dem Hammer philosophiert*
(1889; VI:3)
- UB Unzeitgemässe Betrachtungen* (1893; III:1, IV:1)
- A Der Antichrist* (1895; VI:3)
- EH Ecce Homo: Wie man wird, was man ist* (1908; VI:3)

KGW Werke: Kritische Gesamtausgabe

KGB Briefwechsel

Introduction

At first glance, the prospects for a telling critical engagement between Kant's and Nietzsche's moral philosophies might not seem good. Nietzsche's few explicit references to Kant's moral philosophy rarely extend beyond a few words of scathing derision for a caricatured 'Kant', while Kant would probably have anticipated many by dismissing Nietzsche's empirical claims and informal techniques as he dismisses those of his 'popular' contemporaries – namely, as 'a disgusting mishmash of thrown-together observations and half-rationalized principles, on which shallow heads feast because it is something useful for everyday prattle'.¹ Nor does the literature since suggest that a critical engagement between Kant's and Nietzsche's moral philosophies might progress beyond this impasse. Despite the rehabilitation of Nietzsche's moral philosophy and the quality and volume of the literature on Kant's moral philosophy, Kant scholars generally maintain a stony silence towards Nietzsche, and the paltry literature on the relation between Kant's and Nietzsche's moral philosophies is written almost exclusively by Nietzsche sympathisers who, like Nietzsche himself, rarely offer more than derision and caricature in their treatments of Kant.

The purpose of this thesis is to take issue with this impasse. Specifically, by examining Kant's and Nietzsche's treatments of the moral agent, the thesis argues for three broad conclusions. Firstly, it argues that, although Nietzsche's explicit criticisms of Kant's conception of the moral agent can be understood only in the context of Nietzsche's broader moral philosophy, neither these criticisms

¹ *G* 409. See also *G* 410. Compare, for instance, Russell's remarks about Nietzsche in his *History of Western Philosophy*, pp.728-39.

nor their context are well understood by the prevailing literature. The thesis thus engages with existing scholarship on the nature of Nietzsche's moral philosophy and with the scanty literature on the relationship between Kant's and Nietzsche's moral philosophies. Secondly, the thesis argues that Kant's conception of the moral agent is not undermined by the criticisms which Nietzsche explicitly levels at it, or, indeed, by others which are commonly made in Nietzsche's name. In doing so, the thesis combines original interpretations of Kant with elements of recent Kant scholarship. Finally, however, the thesis argues that neglected elements of Nietzsche's own moral philosophy provide for a more sophisticated, telling, and, indeed, original critical engagement with Kant's conception of the moral agent. In particular, Nietzsche presents an alternative, but nonetheless Kantian, conception of the moral agent which raises substantial questions for Kant's own conception. Thus the thesis defends an original interpretation of Nietzsche's moral philosophy and its critical relation to Kant's, and demonstrates the pertinence of a certain neglected critical approach to Kant's conception of the moral agent. On the basis of these conclusions, the thesis ultimately defends a conception of the moral agent which, although Kantian, owes something to both Kant and Nietzsche.

These three broad conclusions are defended in the three parts of the thesis, respectively. The first part is occupied with Nietzsche, and is primarily expository. It begins with Nietzsche's critical treatment of a common sense conception of agency, defined by a claim regarding choice and another regarding reasons. The first and second chapters present Nietzsche's critical treatment of each claim, in turn. The third chapter then argues that, despite its salient place in Nietzsche's works, this critical treatment is poorly represented by the pertinent strands of the

literature. Finally, the fourth chapter turns to Nietzsche's explicit critical remarks regarding Kant's moral philosophy, and argues, in particular, that these remarks are strongly informed by Nietzsche's underappreciated concern with the common sense conception of agency, and that they are also poorly represented by the relevant literature. The first part of the thesis thus presents and defends an account of Nietzsche's explicit criticisms of Kant's conception of the moral agent by examining these criticisms and their place in Nietzsche's broader critical treatment of a common sense conception of agency, and by critically engaging with the relevant literature.

The second part of the thesis defends Kant's conception of the moral agent against the criticisms identified in the first part. The first chapter of this part, the fifth of the thesis, offers an interpretation of Kant's derivation of the basic features of his conception. In this light, the following chapter defends Kant's formulas and their place in his conception of moral judgement against Nietzsche's criticisms and those commonly made in his name. Nonetheless, in so doing, this chapter also argues that Nietzsche's criticisms of the common sense conception of agency require certain accommodations on Kant's part. The seventh chapter then argues that Nietzsche's criticisms of Kant's insistence on spontaneity in choice are misplaced, but that Kant's arguments for this spontaneity are equally unconvincing. Finally, the eighth chapter demonstrates that, despite Nietzsche's allegation to the contrary, Kant neither claims that there is evidence of human progress nor conceives of human progress in moral terms. The second part of the thesis thus defends the second broad conclusion, that Kant's conception of the moral agent is not undermined by the criticisms which Nietzsche explicitly levels at it, or by others which are commonly made in Nietzsche's name.

The third, and final, part of the thesis consists of a single chapter. This chapter defends the third broad conclusion of the thesis, that neglected elements of Nietzsche's moral philosophy provide for a more sophisticated, telling, and original critical engagement with Kant's conception of the moral agent than the criticisms considered in the first two parts of the thesis would suggest. In particular, it argues that Nietzsche articulates an alternative conception of the moral agent which, although Kantian, differs from Kant's own in revealing ways. In the light of these differences, the chapter defends a Kantian conception of the moral agent that draws from both Kant and Nietzsche.

In interpreting and discussing Kant's and Nietzsche's moral philosophies, I attempt to maintain the integrity of textual and argumentative contexts, individual texts, and chronological periods. The method, particularly common in the literature on Nietzsche, of selecting passages apparently by whim or, at best, by theme serves only to obscure the grounds, nature, implications, and developments of the claims made in them. In those cases in which I concentrate on a particular text, part of a text, or passage, I attempt to justify this by appeal to the author's own indications of its significance, along with its maturity, extensiveness, or explanatory power. I also take it that the coherence of the position which I attribute to Nietzsche demonstrates that, at least as far as this position extends, those commentators who suppose his texts to resist coherent formulation are mistaken.

I refer to Kant's and Nietzsche's notes and, occasionally, their correspondence only in footnotes, and only to indicate cases in which claims made in their published works are also made in notes or correspondence. My interpretative claims therefore do not rely on such evidence. This contrasts with the previously common practice of referring freely to Nietzsche's notebooks, as if

they had a status equal to his published works. This practice has been criticised, primarily on the grounds that only Nietzsche's published works can be presumed to contain his considered and polished thoughts, and that he intended his notebooks neither to be published nor preserved posthumously.² Besides such reasonable, principled considerations, in presenting and defending an interpretation of Nietzsche's published moral philosophy, I have neither found any need to refer to his notes or correspondence for clarification or explanation, nor discovered many radically divergent claims in these other sources.³

Finally, translations of Kant's and Nietzsche's texts are my own, although I have consulted the available translations. Insertions and omissions are indicated by square brackets, emphasis and punctuation are original, and original German terms are italicised. References are given in footnotes to the first instance of each quotation, according to the scheme of abbreviations which immediately precedes this introduction.

² See, in particular, Magnus, 'The Use and Abuse of *The Will to Power*', Clark, *Nietzsche on Truth and Philosophy*, pp.25-7, and Leiter, *Nietzsche on Morality*, pp.xvii-iii.

³ The treatment of Nietzsche's theoretical philosophy might require a different strategy, however. See Poellner's remarks in his *Nietzsche and Metaphysics*, pp.10-11.

Part I

Choice, Reasons, and Kant: Nietzsche's Criticism of the Moral Agent

Introduction to part I

This part of the thesis demonstrates that, although Nietzsche's explicit criticisms of Kant's conception of the moral agent can be understood only in the context of Nietzsche's broader moral philosophy, neither these criticisms nor the pertinent context are well represented by the prevailing literature.

This part begins by demonstrating that Nietzsche's critical claims and arguments regarding moral agency, throughout his texts, are directed at a conception of agency which he considers to strongly inform common sense. This conception makes two distinctive claims. Firstly, it considers an agent's action to be sufficiently explained by the agent's 'willing' of it, in the sense that the action is, given the circumstances, causally necessitated in the agent's conscious choice of it. Secondly, this conception considers the agent to consciously choose, or 'will', an action for normative reasons, in the sense that her cognitive acknowledgement of reasons motivates her choice. Nietzsche's critical treatment of these two claims is presented by the first and second chapters of this first part of the thesis. The third chapter then demonstrates that this treatment is poorly represented by pertinent strands in the prevailing Nietzsche literature.

The final chapter turns to Nietzsche's explicit remarks regarding Kant's moral philosophy. It argues, in particular, that these remarks are strongly informed by Nietzsche's critical treatment of the common sense conception of agency, and that they offer little support to the claims standardly made in the literature on the relationship between Kant's and Nietzsche's moral philosophies.

1.

Choice

Nietzsche's basic critical object with regard to choice is a particular common sense claim. This claim is that an agent's action is sufficiently explained by her 'willing' of it, in the sense that the action is, given the circumstances, causally necessitated in the agent's conscious choice of it. However, Nietzsche's critical discussions of this claim are divided between those which consider the claim as such, and those which consider one of two particular instances of it. This chapter is therefore divided into two sections, corresponding to Nietzsche's own divided concerns.

I.

Nietzsche's critical treatment of the common sense claim regarding choice is first and most clearly aired in a section of the first edition of *The Gay Science*. There he writes the following.

Aftereffects of the most ancient religiosity. – Every thoughtless person thinks that the will alone is effective [*Wirkende*]; that willing is something simple, simply given, underivable, in-itself-intelligible. He is convinced that when he does something, for example, strikes something, it is he who strikes, and that he struck because he *willed* [*wollte*] to strike. He notices no problem at all here, the feeling of *will* is enough for him not only for the assumption of cause and effect, but also for the belief that he *understands* [*verstehen*] their relation. He knows nothing of the mechanism of the event and of the hundredfold fine work that must be done for the strike to happen, or of the incapacity of the will in itself to do even the slightest part of this work. The will is for him a magically effective [*wirkend*] force: the belief in the will, as the cause of effects, is the belief in magically effective forces. Now man believed originally that everywhere he saw an event, a will [had] to be effective in the background as cause and personal willing being – the concept of mechanics lay far off. But because man believed for an immensely long time only in persons (and not in matter, forces, things, and so on), the belief in cause and effect became for him the basic belief, which he uses wherever something happens – also still now instinctive and a piece of atavism of the most ancient origin. The propositions, 'no effect without cause', 'every effect in turn a cause' appear as generalisations of much more

limited propositions: 'where [something] is effected, there [something] has been willed', 'there can be effects only on beings that will', 'there is never a pure, consequence-less suffering of an effect, but all suffering is an excitement of the will' (to action, resistance, revenge, retribution) – but in the prehistoric ages of humanity these propositions were identical, the first were not generalisations of the second, but the second explanations of the first.¹

Nietzsche's critical object here is the common sense claim regarding choice. He first states this claim as the supposition that an agent performed an action simply 'because he *willed* to', and that this 'willing' is 'something simple, simply given, underivable, in-itself-intelligible'. However, Nietzsche proceeds to analyse this supposition in terms of a particular causal account of an action. In particular, he maintains that common sense considers 'willing' to be something 'effective', or an 'effective force', such that, with the 'feeling' of 'willing', one assumes that 'willing' and action are 'cause and effect' and that one '*understands* their relation'. Thus what is 'felt' in 'willing' is supposed to be a causal power which, given the circumstances, necessitates a succeeding event – namely, the agent's action.

This passage clearly indicates that Nietzsche considers it misguided to account for an event by identifying a necessitating causal power. Indeed, he also maintains here that it is precisely the common sense claim regarding choice which misleads us to 'instinctively' account for other events in this manner, whether by the primitive supposition of a 'personal willing being' behind each event or in modern explanations of changes of state among 'matter, forces, things, and so on'. However, Nietzsche's objection to such causal accounts is clearest in an

¹ FW 127.

earlier section of the same book of *The Gay Science*, entitled 'Cause and effect'.

There he states his position as follows.

We describe better – we do not explain any more than our predecessors. We have uncovered a multiple one-after-another where the naive man and inquirer of older cultures saw only two separate things, 'cause' and 'effect' is what one says; we have merely perfected the image of becoming, but without reaching beyond the image, behind the image. In every case the series of 'causes' stands before us much more completely, we infer: this and that must first precede in order that this may then follow – but this does not involve any *comprehension* [*begriffen*]. In every chemical process, for example, quality appears as a 'miracle', as ever, also every locomotion; nobody has 'explained' a push.²

With this, Nietzsche endorses the 'Humean' claim that we do not experience causal powers which would necessitate a succeeding event, and which would thus provide for the event's 'explanation'. In Nietzsche's terms here, we have no 'comprehension' of the event. Nonetheless, Nietzsche also admits the equally 'Humean' claim that we experience regular correlations, or 'a multiple one-after-another', between distinct events of the same types, according to which we justify causal universals of the form 'this and that must first precede in order that this may then follow'. For Nietzsche, we can thus causally 'describe' an event by subsuming it under an appropriate causal universal, and one such causal 'description' can be 'better' or more 'complete' than another, in the sense that it acknowledges universals which justifiably encompass and exceed those which the

² *FW* 112. See also *M* 121, *JGB* 21, and *KGW* VII:1 24 [9] (Winter 1883-1884), VII:3 34 [53] (April-June 1885), VII:3 36 [28] (June-July 1885), VIII:1 2 [83, 89, and 139] (Autumn 1885-Autumn 1886), and 7 [56] (End 1886-Spring 1887), and VIII:3 14 [98] (Spring 1888).

other acknowledges.³ Further, as the remark regarding our not ‘reaching beyond the image’ intimates and as he proceeds to elaborate in the succeeding, unquoted part of the section, Nietzsche parts company from the ‘Humean’ by insisting that our experience of regular correlations between distinct events of the same types is a reflection of human beings’ particular experiential apparatus.⁴ However, since the interpretation and plausibility of this further claim is notoriously controversial and not directly pertinent to Nietzsche’s critical treatment of the common sense conception of agency, I will not consider it here.

In ‘*Aftereffects...*’, Nietzsche invokes the ‘Humean’ position articulated in ‘*Cause and effect*’ to dispose of the common sense claim regarding choice. Regarding the assumption that ‘willing’ provides for the ‘understanding’ of the ‘relation’ between ‘cause and effect’, he refers to his denial of causal ‘explanation’ when he insists on ‘the incapacity of the will in itself to do even the slightest part of this work’. Regarding the assumption that ‘willing’ and action are ‘cause and effect’, on the other hand, he complains that one thus ‘knows nothing of the mechanism of the event and of the hundredfold fine work that must be done for [...it] to happen’. With this, he insists on the inadequacy, and the insufficient complexity, of the common sense claim as a causal ‘description’ of choice. As he also puts it later in the section, ‘willing is only a mechanism which is so well practised that it almost escapes the observing eye’.⁵

³ My interpretation of how Nietzsche understands ‘explanation’ and ‘description’ here agrees with Poellner’s discussion, in his *Nietzsche and Metaphysics*, pp.36-46, 53-5, and 267-8. In the light of the ‘Humean’ echoes of Nietzsche’s position, it is also worth emphasising that, unlike Hume, Nietzsche attributes the misconception of causal accounts of events to the common sense claim regarding ‘willing’, rather than to habit.

⁴ Besides *FW* 112, see *WS* 11, *FW* 99, 121, and 357, and *JGB* 21.

⁵ *FW* 127.

II.

Only in *Twilight of the Idols* does Nietzsche restate, if only in part, the criticism of the common sense claim regarding choice which he offers in the first edition of *The Gay Science*. In a section of the chapter, 'The four great errors', he again holds this claim responsible for our general commitment to causal 'explanation' and criticises it precisely for pretending to provide such unavailable 'explanations'. Regarding this claim he writes the following.

Error of a false causality. – In every age one has believed that one knows what a cause is: but from where did we get our knowledge, more precisely, our belief that we know here? From the realm of the famous 'inner facts', none of which have until now proved factual. We believed ourselves to be causal in the act of will: we thought there, at least, we were *catching causality in the act*. [...] Today we don't believe a word of all that anymore. The 'inner world' is full of illusions and will-o'-the-wisps: the will is one of them. The will no longer moves [*bewegt*], consequently also no longer explains [*erklärt*] – it merely accompanies [*begleitet*] processes, it can also be absent.⁶

The critical object of these remarks is the common sense belief that in choosing an action an agent experiences a causal power which necessitates the action, given the circumstances, and which thus provides sufficient explanation of it. As Nietzsche puts it here, the agent is supposed to be 'causal in the act of will', or to succeed in 'catching causality in the act' as she wills. Against this, Nietzsche insists that '[t]he will no longer moves, consequently also no longer explains', and

⁶ GD VI 3.

that it ‘merely accompanies processes’ or is entirely ‘absent’ from them.⁷ He thus denies that choosing serves to account for an event in the sense of providing a necessitating causal power, that which ‘moves’ the event. That this denial rests, as in ‘*Aftereffects...*’, on Nietzsche’s denial that we experience any such causal power is indicated by the remainder of the section. There he denies that there is any “‘empirical evidence’” for the common sense supposition of causality in ‘willing’, and laments that we nonetheless extend this notion of causality to every event. This extension is, he maintains, merely ‘[t]he most ancient and longest-lived psychology [...]: every event [...] an action, every action the consequence of a will’.⁸

III.

Although in ‘*Error of a false causality*’ Nietzsche states that ‘willing’ either ‘merely accompanies processes’ or is entirely ‘absent’ from them, there he does not emphasise the other criticism of the common sense claim regarding choice which he makes in ‘*Aftereffects...*’, that it is inadequate as a causal ‘description’ of choice. However, he arguably raises this criticism again in a section of *Beyond Good and Evil*, although the continuity between this section and his critical discussions elsewhere is not immediately obvious. That the common sense claim regarding choice is his critical object in this section is indicated by his insistence that he is there concerned to dispense with the ‘*popular prejudice*’ that

⁷ In *GD* III 5, Nietzsche similarly refers to ‘the error that the will is something that *effects* [*wirkt*], – that will is an *ability* [or faculty, *Vermögen*]’, and he makes a similar remark at A 14.

⁸ *GD* VI 3. See also *KGW* VII:1 24 [9] (Winter 1883-1884), VIII:1 2 [83] (Autumn 1885-Autumn 1886), VIII:1 5 [9] (Summer 1886-Autumn 1887), and VIII:3 14 [81-2, 98, 125, 129, 146, and 152] (Spring 1888).

‘the will is the most familiar thing in the world’, and, in particular, the belief ‘that willing *suffices* for action’.⁹ However, rather than raising his ‘Humean’ doubts regarding causal powers or directly dismissing the common sense claim as an inadequate causal ‘description’, Nietzsche here provides a phenomenology of choice which is intended to show how, considering choice phenomenologically, the common sense claim is unsurprising, yet false.

In outline, Nietzsche’s phenomenology here considers ‘every willing’ to include phenomenological elements of ‘feeling’, ‘thought’, and ‘affect’. The elements of feeling include feelings of one’s original condition, of leaving this condition and arriving at another, and of the condition at which one arrives, and accompanying muscular feelings which arise ‘through a sort of habit, as soon as we “will”’. Besides such feelings, Nietzsche proposes that ‘willing’ involves ‘a commanding thought’ and ‘the affect of command’. He does not elaborate on the ‘commanding thought’ here, but he is presumably referring to the intentional content of an agent’s choice. The ‘affect of command’, on the other hand, Nietzsche describes as the affect of directing attention to the performance of the chosen action – in his terms, as ‘that unconditional evaluation “now necessarily do this and nothing other”’. Corresponding to the thought and affect of command, Nietzsche further insists, is ‘something in [the agent] himself which obeys or which he believes obeys’, ‘the successful executive instruments, the serviceable “under-wills” or under-souls’. Thus, Nietzsche writes, ‘as the obedient one we know the feelings of constraint, compulsion, pressure, resistance, movement, which tend to begin immediately after the act of willing’.¹⁰

⁹ *JGB* 19.

¹⁰ *JGB* 19.

Given this phenomenology, Nietzsche accounts for the common sense belief in the sufficiency of choice for action, and the falsity of this belief, as follows. Encouraged by the affect of command, he claims, what (phenomenologically) ‘obeys’ in agency tends to be obscured by, and confused with, what (phenomenologically) ‘commands’, such that the successful performance of the action tends to be mistakenly attributed simply to the latter. Thus, he writes, “Freedom of the will” – that is the word for the multifaceted state of pleasure of the willer, who commands and at the same time identifies himself with the executor [of the command], – who, as such, enjoys the triumph over resistances, but by himself judges, [that] it is his will itself which really overcomes the resistances’.¹¹ In other words, in the light of his phenomenology of choice Nietzsche argues that the common sense belief in choice as providing sufficient explanation of an action gains phenomenological plausibility from the ‘affect of command’, and its encouragement of an over-interpretation of ‘command’ in ‘willing’.

Arguably, however, Nietzsche’s primary intention in thus considering choice phenomenologically is to better, on phenomenological grounds, the common sense claim as a causal ‘description’ of choice. That is, he intends his phenomenology of choice to better this claim as any causal ‘description’ betters another, namely, by acknowledging regular empirical correlations which encompass and exceed those acknowledged by the other. This intention is indicated particularly by the manner in which Nietzsche introduces his phenomenology: he states that, in contrast to common sense and its philosophical advocates, ‘[w]illing appears to me above all as something *complicated*’, and then

¹¹ *JGB* 19. See also *KGW* VII:1 24 [9] (Winter 1883-1884), VII:3 38 [8] (June-July 1885), and VIII:3 14 [98] (Spring 1888).

proposes, '[l]et us therefore be careful for once, let us be "unphilosophical"'.¹² These remarks suggest that Nietzsche presents his phenomenology on empirical grounds, as an account of how choice 'appears to' him, and that he considers it to better common sense by showing how choice is more 'complicated' than common sense supposes, or how common sense is not 'careful' in its description of choice. This section of *Beyond Good and Evil* can thus be treated as continuous with the critical strategy which Nietzsche adopts towards the common sense claim regarding choice in the first edition of *The Gay Science* and, later, in *Twilight of the Idols*.¹³

¹² *JGB* 19. See also *KGW* VII:3 38 [8] (June-July 1885).

¹³ A similar interpretation might be offered for Nietzsche's statement at *A* 14: 'The old word "will" only serves to designate a resultant, a kind of individual reaction, which necessarily follows a host of partly contradictory, partly harmonious stimuli'.

With the exception of his phenomenology of choice in *Beyond Good and Evil*, Nietzsche's critical discussions of choice between the first edition of *The Gay Science* and *Twilight of the Idols* make no reference to the common sense claim as such, but are rather occupied with two particular instances of it. His critical discussions before the first edition of *The Gay Science* are yet more restricted, being almost exclusively concerned with one of these instances and a single, later abandoned objection to it. The two instances are distinguished by their manner of conceiving of choice: one attributes it to a substantial agent, or 'I', and the other to 'free will'. Nietzsche's considered critical account of these two instances is most clearly presented in a series of six consecutive sections in the first part of *Beyond Good and Evil*, one of which presents his phenomenology of choice. I will focus on this series, and mention only secondarily Nietzsche's discussions of these two instances in *On the Genealogy of Morality* and *Twilight of the Idols*, to demonstrate that the latter discussions make, and, in minor respects, clarify or obscure, claims which are already made in the sections of *Beyond Good and Evil*.

I.

Nietzsche's critical concern in the series of sections in *Beyond Good and Evil* is with the common sense claim regarding the sufficiency of choice for action, as it is in the section in which he presents his phenomenology of choice. In

that section, besides insisting that common sense misinterprets the phenomenology of 'command' in choice, Nietzsche also attributes this misinterpretation to 'the synthetic concept "I"', with which, he claims, we obscure the phenomenological 'duality' of 'command' and 'obedience' in choice.¹⁴ With this, he proceeds beyond his phenomenology, and invokes his discussion in the first and second sections of this series. These sections describe an instance of the common sense treatment of choice which is distinguished by its attributing choice to a substantial agent – that is, to an agent considered as a 'substratum' which persists beyond the observable, changing properties of choice. Thus, regarding what he calls the apparent 'immediate certainty' of the statement 'I will', in these two sections Nietzsche raises three sceptical questions: namely, why must there be a 'something' that wills, and why must this 'something' be an 'I'?; why must willing be thought of as 'an activity and effect on the part of a being thought of as a cause'?; and, how does one identify this activity, and distinguish it from other activities? Nietzsche maintains that even to replace 'I' with 'it', so as to state 'it wills' or 'there is willing', is still to presuppose an actor for the activity, a substantial agent which wills. 'One infers here according to the grammatical habit "[...] to every activity belongs one that is active [...]"', Nietzsche maintains. In other words, he insists that to attribute choice to a substantial agent is to misinterpret the 'process [*Vorgang*]' of choice in terms of a subject that 'does not belong' to it, under the influence of the 'grammatical habit' of treating activity as a predicate with a subject.¹⁵

¹⁴ *JGB* 19.

¹⁵ *JGB* 16, 17. Nietzsche presents his remarks in these sections in terms of the statement 'I think', but indicates in the first sentence that he is also concerned with the statement 'I will'. See also *JGB* 34.

Nietzsche's insistence on the unfortunate effects of this 'grammatical habit' on the interpretation of basic activities might appear to complete his critical account of apparent 'immediate certainties' such as the statement 'I will'. But in the fifth in the series of sections in *Beyond Good and Evil*, he states otherwise. Referring to the grammatical function of 'the subject-concept' in particular, he writes, 'the spell of certain grammatical functions is in the last analysis the spell of *physiological* value judgements and racial conditions', and, under such spells, philosophers' 'thinking is in fact much less a discovering than a recognising, a remembering, a return and home-coming to a far-off, primeval total household of the soul out of which [...their philosophical] concepts once grew: – philosophising is to that extent a species of atavism of the first rank'.¹⁶ Thus Nietzsche insists on the primeval nature of the 'grammatical habit', or 'spell', which misleadingly suggests that choice must be attributed to a substantial agent. It is plausible to suppose, then, that he is thus referring to his account of the misleading influence of the common sense claim regarding choice on the primitive and modern conceptions of causality, particularly regarding choice itself. Indeed, as I will demonstrate below, he makes this reference clear in his later discussions.¹⁷

In the last of the series of sections in *Beyond Good and Evil*, Nietzsche proceeds to identify a further instance of the common sense claim regarding the sufficiency of choice for action. This instance is distinguished by treating choice as

¹⁶ *JGB* 20. See also *JGB* V, and *KGW* VII:3 38 [3] (June-July 1885), and VIII:2 10 [158] (Autumn 1887). Note that, although the first part of *JGB* is entitled 'On the prejudices of philosophers' and Nietzsche refers to philosophers' errors throughout the series of sections which discuss 'I' and 'free will', in these sections he does not draw a strict distinction between the errors of philosophers and those of common sense. Regarding choice, his position is presumably that expressed in *JGB* 19, namely, that philosophers 'always tend to' have simply 'adopted and exaggerated a *popular prejudice*'.

¹⁷ Note also that Nietzsche's claim that primeval errors regarding nature persist in subject-predicate grammar echoes Lubbock, in his *The Origin of Civilisation*, pp.219-20. In making this claim, Lubbock follows Max Müller, a volume of whose essays Nietzsche also possessed.

‘free will’ – that is, as undetermined by antecedent causes. In this section, Nietzsche refers to this instance as “‘freedom of will”, in that metaphysical superlative sense’, according to which one is supposed ‘to be [...] *causa sui* and, with more than Münchhausen temerity, to pull oneself into existence out of the swamp of nothingness by one’s own hair’. He objects to this instance with the simple statement that ‘[t]he *causa sui* is the best self-contradiction hitherto imagined’.¹⁸ By this, he presumably means that to conceive of choice as a ‘cause of itself’ is to conceive of it as both a cause and the effect of that cause, which generates a ‘self-contradiction’ on the plausible assumption that a cause and its effect are not identical. However, Nietzsche also here objects that ‘free will’ underwrites an over-inflated sense of responsibility. ‘The desire for’ such undetermined choice is, he writes, ‘the desire to bear the whole and ultimate responsibility for one’s actions and to absolve God, world, ancestors, chance, society from responsibility for them’.¹⁹ Nietzsche apparently considers such a sense of responsibility to be self-evidently absurd.

Notably, Nietzsche’s objections to ‘free will’ in this section differ from that which he endorses in passages of *Human, All Too Human*, *Assorted Opinions and Maxims*, *The Wanderer and His Shadow*, and *Daybreak*. In these earlier texts, his discussions of agency almost exclusively conceive of choice as ‘free will’ and object to it on the grounds that ‘freedom’ in this sense is incompatible with the principle of causal determinism – namely, that every event, or set of events, has a sufficient antecedent cause.²⁰ Now, although Nietzsche’s objection to ‘free will’ as

¹⁸ *JGB* 21. See also *WS* 23, and *FW* 110 and 345.

¹⁹ *JGB* 21.

²⁰ See *MA* 39, 70, 99, 106-7, 133, and 376, *MS* 33, *WS* 12, 24, and 28, *M* 120, 124, and 128, and even *KGW* I:2 13 [6 and 7] (April-October 1862). This feature of these texts might reflect Nietzsche’s reading of Schopenhauer, who articulates this objection to ‘free will’ in, for example, *Über die Freiheit des Menschlichen Willens*, ch.3, *Die Welt als Wille und Vorstellung*, §23. See

a ‘self-contradiction’ in *Beyond Good and Evil* might follow from the principle of causal determinism, he need not, and in this section does not, also endorse this principle.²¹ Furthermore, in the remainder of the section, he effectively dismisses his earlier objection, by denying that, in the absence of ‘free will’, determination by antecedent causes precludes choice. In his terms here, he asks that one also ‘erase from [...one’s] head the reversal of the non-concept “free will”: I mean the “unfree will”, which amounts to an abuse of cause and effect’.²² As this remark suggests, his argument here once again calls upon his ‘Humean’ position regarding causation. He writes the following.

One should not mistakenly *objectify* [*verdinglichen*] ‘cause’ and ‘effect’, as [...when one] has the cause press and push, until it ‘effects’; one should help oneself to ‘cause’ and ‘effect’ only as pure *concepts*, that is, as conventional fictions for the purpose of description, of communication, not of explanation. In the ‘in-itself’ there is nothing of ‘causal associations’, of ‘necessity’, of ‘psychological unfreedom’, there ‘the effect from the cause’ does *not* follow, there no ‘law’ rules. It is *we* alone who have invented causes, succession, for-each-other, relativity, compulsion, number, law, freedom, ground, purpose; and if we write, mix this symbol-world into things as ‘in-itself’, we thus carry on as we always have carried on, namely, *mythologically*. The ‘unfree will’ is mythology [...]. – It is almost always a

also n.21 below. It might also reflect the influence of Rée, whose friendship with Nietzsche extended from the mid-1870’s until late 1882 and who wrote his *Der Ursprung der moralischen Empfindungen*, while Nietzsche wrote *MA*, during their stay in Sorrento in the winter of 1876-7. Indeed, *MA* 39 corresponds precisely – even, in two sentences, almost word-for-word – with Rée’s views on the history of morality and on the implications of causal determinism in *Der Ursprung der moralischen Empfindungen*. See esp. pp.17, 39, and 61-3.

²¹ Nonetheless, it is worth noting that Nietzsche’s terms in *JGB* 21, including the reference to Münchhausen, echo those of Rudolf von Ihering, in the first paragraphs of his *Der Zweck im Recht*, and those of Schopenhauer, in his *Über die vierfache Wurzel des Satzes vom zureichenden Grunde*, ch.2, §8, and *Die Welt als Wille und Vorstellung*, §7, and Suppls., ch.8. In these passages both writers object to the conception of agency as ‘*causa sui*’ on the grounds that it is incompatible with the principle of causal determinism. Some of Nietzsche’s notes on Ihering’s book can be found at *KGW* VII:1 7 [69] (Spring-Summer 1883).

²² *JGB* 21.

symptom, of what is lacking in him, when a thinker senses in every 'causal connection' and 'psychological necessity' something of compulsion, need, having-to-obey, pressure, unfreedom.²³

Thus Nietzsche again denies that we experience causal powers which would necessitate a succeeding event, and which would thus provide for the 'explanation' of it. Indeed, he also raises his further, un-'Humean' doubts about our experiences of regular correlations between distinct events of the same types, according to which we 'describe' an event by subsuming it under an appropriate causal universal. In his terms here, we should neither have 'the cause press and push, until it "effects"', nor 'write, mix [...] into things' the descriptive concepts of causal universals, such as 'succession', 'number', 'law', and 'ground'. Despite the latter doubts regarding causal 'description', however, Nietzsche also proposes that we continue to practise it, and his objection to 'unfree will' does not concern these further doubts regarding it. Specifically, his objection to 'unfree will' is that, however complete the description of an event as 'necessary' given antecedent conditions, this 'necessity' consists neither of a causal power, since experience of such a power is unavailable to us, nor of a command, since the 'necessity' concerned is not of this, normative kind.²⁴ Nietzsche concludes that there is thus nothing of 'compulsion, need, having-to-obey, pressure, unfreedom' in the causal determination of choice, and that it is therefore mistaken to suppose, as proponents of 'unfree will' do, that determination by antecedent causes precludes choice, by making it merely a passive, obedient effect of other causes.

²³ *JGB* 21. See also *WS* 61, *A* 15, and *KGW* VIII:2 9 [91] (Autumn 1887).

²⁴ Nietzsche also objects to the misconception of the 'rule' of 'law' as a command at *MS* 9, *FW* 59 and 109, *JGB* 9 and 22, and *KGW* VII:3 36 [18] (June-July 1885), VIII:1 1 [44] (Autumn 1885-Spring 1886), 2 [139 and 142] (Autumn 1885-Autumn 1886), and 7 [14] (End 1886-Spring 1887), and VIII:3 14 [79 and 98] (Spring 1888).

Finally, Nietzsche also remarks that, just as ‘free will’ underwrites an over-inflated sense of responsibility, so ‘unfree will’ underwrites an under-inflated sense, by reducing responsibility to nothing. In particular, he writes that ‘unfree will’ is embraced by those who ‘want [...] not to be responsible, to be guilty of nothing[,] and desire, out of an inner self-contempt, to be able to *shift the blame* for themselves to somewhere else’, while it is resisted by those who ‘want at no price to let go of their “responsibility”, the belief in *themselves*, the personal right to *their* merit’.²⁵ Nietzsche apparently does not object to the latter sense of responsibility. In particular, he does not suggest that it is, or that it presupposes, the over-inflated ‘whole and ultimate responsibility’ which ‘free will’ underwrites. However, here he provides little indication of how responsibility is to be judged rightly, remarking only that ‘in real life it is only a matter of *strong* and *weak* wills’.²⁶

II.

The two instances discussed, and some of the claims and criticisms made of them, in the series of sections in *Beyond Good and Evil* recur in Nietzsche’s two other critical discussions of agency in his later texts – namely, the thirteenth section of the first essay of the *Genealogy*, and certain sections of the chapter, ‘The four great errors’, in *Twilight of the Idols*. These discussions provide little novelty, however, and are perhaps most notable for their modest elaboration of Nietzsche’s remarks regarding the misleading influence of ‘free will’ on the

²⁵ JGB 21.

²⁶ JGB 21.

determination of responsibility. Here, then, I will first note the other, recurrent claims which Nietzsche makes in these discussions, before considering his elaboration of his objection to ‘free will’.

Nietzsche’s discussion in the thirteenth section of the first essay of the *Genealogy* is distinguished, and obscured, by its unfortunate conflation of the two instances discussed separately in *Beyond Good and Evil*. He refers to the instance which affirms a substantial agent, or ‘I’, when he refers to the supposition of a “‘being” behind the doing, effecting, becoming’ of agency. Regarding this, he maintains that “‘the doer” is merely imagined into the doing, – the doing is everything’, and that this imaginary “‘doer”” arises ‘only through the seduction of language (and the fundamental errors of reason petrified in it), which understands and misunderstands every effecting as conditioned by an effective thing, by a “subject”’. Nietzsche thus again maintains that the attribution of choice to a substantial agent, or ‘I’, derives from certain errors of common sense, which are expressed in subject-predicate grammar. However, here he also describes the notion of a substantial agent as one with which ‘popular morality [...] detaches strength from the expressions of strength as if there were behind the strong an indifferent substratum, which is *free* to express strength or not to’. With this, he writes, ‘[t]he people basically double the doing, [...] this is a doing-doing: it posits the same event first as cause and then once again as its effect’. Nietzsche thus refers to the other instance of the common sense treatment of choice, that which regards choice as ‘free’ in the sense of not determined by antecedent causes, objects that it conceives of choice as both a cause and an effect of that cause, and suggests that it derives from certain ‘popular’ errors. Given his earlier discussions, then, this section is distinguished by Nietzsche’s conflation of the two instances,

and his objections to them – his critical object here is, as he expresses it towards the end of the section, ‘the belief in the indifferent freely choosing “subject”’.²⁷

In sections of the chapter, ‘The four great errors’, in *Twilight of the Idols*, on the other hand, Nietzsche clearly distinguishes the two instances, by treating them as two distinct ‘errors’. In ‘*Error of a false causality*’, he refers to the belief in ‘the I (the “subject”) as cause’ and rejects it as ‘a fiction, a play on words’, thus referring to the instance which attributes agency to a substantial agent, and again rejecting it as a fabrication deriving from certain errors expressed in grammar. Notably, in this section he also makes clear what he merely suggests in *Beyond Good and Evil*, namely, that he considers these errors in the light of his account of the primitive and modern misconceptions of causality and their origins in the common claim regarding choice. While in the *Genealogy* he notes only that these errors are ‘petrified in’ grammar, in ‘*Error of a false causality*’ he indicates that their origins are not merely grammatical, and that they arise from common sense’s basic error regarding the experience of a necessitating causal power in choice. In particular, he writes that the attribution of choice to a substantial agent, along with the belief that an agent chooses for reasons, ‘were merely born afterwards, after causality had been firmly established by the will as given, as *empirical evidence*’.²⁸

Nietzsche’s discussion of ‘free will’ in ‘The four great errors’, in two sections under the title ‘*The error of free will*’, is more lengthy. It is also explicitly

²⁷ GM I 13. A clearer formulation can be found at KGW VIII:1 2 [84] (Autumn 1885-Autumn 1886). Nietzsche’s example of ‘the lightning’ and ‘its flash’ in this section might derive from Lubbock, who, in discussing subject-predicate grammar and the primitive conception of nature, maintains that Aryan languages have tended to confuse the subject of a sentence like ‘the sky thunders’ with a proper name, and that this confusion partly explains the idea that a spirit is responsible for the thunder – that ‘Zeus thunders’, for example. See Lubbock, *The Origin of Civilisation*, pp.219-220.

²⁸ GD VI 3. See also GD III 1-3 and 5, and KGW VIII:1 2 [139] (Autumn 1885-Autumn 1886), VIII:1 7 [1] (End 1886-Spring 1887), VIII: 2 9 [98] (Autumn 1887), and VIII:3 14 [98] (Spring 1888).

restricted to ‘the psychology of every making-responsible’ – that is, to the role of ‘free will’ in determining an agent’s moral responsibility.²⁹ This element of Nietzsche’s criticism of ‘free will’ is first introduced in the last of the series of sections in *Beyond Good and Evil*, and provided with some elaboration in both of his later discussions. In the thirteenth section of the first essay of the *Genealogy*, the terms in which he elaborates on this element echo his remark regarding ‘strong and weak wills’ in *Beyond Good and Evil*. For there he writes that an agent’s ‘strength’, or ‘force’, is ‘a quantum of drive, will, effect’, and constitutes ‘his *essence*, his effect, his whole unique inevitable, undetachable reality’. In this light, he insists that those who uphold the ‘belief [...] that *the strong one is free* to be weak, and the bird of prey to be a lamb[, ...] with it [*damit*] gain for themselves the right to *hold accountable* the bird of prey, for being a bird of prey’. Equally, Nietzsche continues, this belief permits ‘weakness’ to be treated as ‘a voluntary achievement, something willed, chosen, an *act*, a *merit*’. Among the ‘weak’, then, Nietzsche claims that it is ‘no wonder if the suppressed, hiddenly glowing affects of revenge and hate exploit this belief for themselves’ – that is, for the promulgation of supposed moral requirements which serve the interests of the ‘weak’ in constraining the ‘strong’.³⁰ In this section of the *Genealogy*, therefore, he maintains that ‘free will’ legitimates the consideration of an agent’s ‘strength’ or ‘weakness’ as something for which the agent is responsible, and that this legitimation is exploited by the ‘*ressentiment*’-fuelled ‘slave revolt’ with which the first essay is primarily concerned.

In the two sections under the title ‘*The error of free will*’ in *Twilight of the Idols*, on the other hand, Nietzsche makes no reference to ‘strength’ or ‘weakness’,

²⁹ *GD* VI 7.

³⁰ *GM* I 13. See also *GM* II 7, and *EH* I 6.

but rather refers simply to an agent's 'nature', or her 'being this or that'. In particular, he writes that, with 'free will', 'being this or that is traced back to will, to intentions, to responsible acts', or the agent is supposed to be 'responsible for being at all, for being constituted as this or that, for being in these circumstances, in these surroundings'. This treatment of choice, Nietzsche maintains, 'has its presupposition in this, that its authors, the priests at the head of ancient communities, wanted to create for themselves a *right* to impose punishments – or wanted to create a right for God to do so'. Against this, Nietzsche's 'teaching' is simply that '[n]o one *gives* a human being his qualities, neither God, nor society, nor his parents and ancestors, nor *he himself*'. 'The fatality of [...one's] nature', Nietzsche continues, 'is not to be detached from the fatality of all that was and will be'.³¹

Nietzsche's only hint of what one's 'nature', or one's 'being this or that', might consist of is a remark which he makes at the beginning of 'The four great errors', in a discussion under the title, '*The error of confusing cause and consequence*'. There his example is Luigi Cornaro, a sixteenth-century Venetian writer who attributed his longevity to a paltry diet. To this, Nietzsche counters that Cornaro's slow metabolism was the 'precondition' of his longevity, and made necessary his paltry diet. Thus, Nietzsche remarks, Cornaro's 'frugality was *not* a "free will": he became ill when he ate more'. Generalising this, Nietzsche insists that, just as Cornaro's metabolism required him to restrict his diet, so the happy functioning of 'the order which [...one] physiologically represents', or the extent of one's 'instinct degeneration, [...] disgregation of will', requires one to do or refrain from certain things. Nietzsche concludes that for supposed moral

³¹ *GD* VI 7, 8. For similar remarks, see *GD* V 6, A 14-15 and 38, *EH* IV 8, and *KGW* VIII:3 15 [30] (Spring 1888).

requirements to ignore these conditions is simply for them to be instances of 'unreason'.³² Assuming that the 'free will' to which his remark regarding Cornaro refers is precisely the 'free will' which he elsewhere associates with an over-inflated sense of responsibility, therefore, Nietzsche can be supposed to include an agent's physiological constitution, and the state of her 'instincts' and 'will', within the 'fatality' for which she cannot reasonably be held responsible.³³

This and Nietzsche's corresponding account of 'strength' and 'weakness' in the *Genealogy* serve to express a conception of responsibility as restricted to what is within the agent's control, and the claim that 'free will' underwrites a sense of responsibility which oversteps this restriction. However, since these accounts provide only the broadest intimations of precisely how the restriction is thus contravened, Nietzsche's commitment to it remains largely formal, and his understanding of the rightful scope of an agent's responsibility remains largely indeterminate.

Finally, although it is plausible to suppose that 'free will' would logically underwrite a sense of responsibility that would overstep Nietzsche's restriction, less convincing is his allegation that, in fact, proponents of unfulfilable requirements have in appealed to 'free will' to underwrite them. Important counter-examples are provided by Christianity, at which he generally directs this allegation.³⁴ Calvin, for instance, held the agent responsible for those predetermined events of which she is the most immediate or relevant instrumental

³² *GD VI 1, 2.*

³³ This is also implied by *GM II 22*, in which Nietzsche criticises the denial of certain natural noncognitive motivating states as 'a kind of madness of the will' for the impossibility of its demands. At *EH 'M' 2*, he also maintains that the 'sense' of 'free will', among other concepts, is 'to ruin humanity physiologically'.

³⁴ For Nietzsche's concern for Christianity in this regard, see *GM I 13*, and *GD VI 2, 7*, and *8*.

cause.³⁵ Even Augustine, who is often considered the first to conceive of the responsible agent in terms of a spontaneous causal ‘will’, held that the human will could do good only with divine grace.³⁶ Similarly, other philosophers that deny ‘free will’ do not necessary satisfy Nietzsche’s restriction of responsibility to what is within the agent’s control – consider the compatibilisms of Hobbes, Locke, Hume, or Leibniz, for example.³⁷ Indeed, despite Nietzsche’s claim in *Beyond Good and Evil* that the notion of ‘free will’ ‘still rules in the heads of the half-educated’, common sense arguably makes rare recourse to ‘free will’, but nonetheless endorses at least some unfulfilable requirements.³⁸ Such counterexamples, of which Nietzsche was surely aware, leave somewhat puzzling his confidence that to dispel ‘free will’ is to dispel over-inflated senses of responsibility – his confidence that in dispelling ‘free will’, ‘the *innocence* of becoming is restored’, and ‘we redeem the world’, in the emphatic terms with which he concludes ‘*The error of free will*’.³⁹

³⁵ See, for example, Calvin, *Institution de la religion chrestienne*, bk.1, ch.15, §8, and ch.17, §§3-6.

³⁶ See, for instance, Augustine, *De libero arbitrio*, bk.3, ch.17, *De spiritu et littera*, ch.30, and *De civitate Dei*, bk.5, chs.8-10. For a typical interpretation of Augustine as the originator of the modern conception of the will, see Dihle, *The Theory of Will in Classical Antiquity*, ch.6.

³⁷ See, for instance, Hobbes, *Leviathan*, ch.21, Leibniz, ‘On Freedom’, and *Discours de Métaphysique*, §§30-2, Locke, *An Essay Concerning Human Understanding*, bk.2, ch.11, §§5-30, and Hume, *Enquiry Concerning Human Understanding*, §8.

³⁸ JGB 21.

³⁹ GD VI 8.

Conclusions

Nietzsche's critical treatment of the common sense claim that an agent's action is sufficiently explained by her 'willing' of it, in the sense that, given the circumstances, it is causally necessitated in the agent's conscious choice of it, can be summarised as follows. By denying that we experience necessitating causal powers, with which a succeeding event could be 'explained', he denies that we experience such a power in an agent's choosing. Indeed, he also holds the common sense claim itself responsible for our general commitment to such causal 'explanation'. Considered simply as a causal 'description' according to experienced regular correlations between types of events, Nietzsche also suspects that the common sense claim obscures significant causes of chosen action other than conscious choice itself.

Nietzsche also objects to two particular instances of the common sense claim. With regard to an instance which attributes choice to a substantial agent, or 'I', he insists that such an agent is simply not part of choice. It is added to choice, he maintains, only under the misleading influence of subject-predicate grammar, which is ultimately merely a particular expression of our mistaken general commitment to causal 'explanation'. With regard to an instance of the common sense claim which conceives of choice as undetermined by antecedent causes, Nietzsche objects that it is incoherent, since it conceives of choice as both a cause and the effect of that cause, and that it underwrites an over-inflated sense of responsibility. Indeed, at least in his later texts, Nietzsche argues that the contrary notion that causal determination precludes choice, by making it merely a passive,

obedient effect of other causes, is equally mistaken. This notion relies, he maintains, on our misconceiving causal determination in the 'explanatory' terms of necessitating causal powers or on our misconceiving as normative commands the regular correlations according to which we causally 'describe' events.

Reasons

The common sense conception of agency which Nietzsche criticises not only considers an agent's conscious choice to sufficiently explain her action, but also considers this choice to be made for reasons, in the sense that it is motivated by the cognitive acknowledgement of normative reasons for action. Nietzsche consistently denies this claim and treats actions as instead motivated by noncognitive states informed by reasons. However, in only one passage does he provide an argument for this position, being overwhelmingly concerned with its consequences for the practice of making moral judgements about reasons for action, and the moral philosophy parasitic on this practice. In this chapter, then, I elucidate the grounds of Nietzsche's consistent denial of practical reason, before showing how it informs the 'revaluation of values' with which he proposes to replace the moral philosophy of practical reasons.

Like that of the common sense claim regarding choice, Nietzsche's critical treatment of the common sense claim regarding reasons is most clearly aired in a section of the first edition of *The Gay Science*. The relevant section in this case, entitled '*Long live physics!*', begins with the common sense claim regarding reasons, stated with regard to moral reasons by Nietzsche's common sense 'friend': "when the human being judges '*this is right*', [and] when he concludes from this '*hence it must happen!*' and then *does* what he has thus recognised as right and described as necessary, – then the nature of his action is *moral!*". For Nietzsche, the prevalence of this moral version of the claim best 'testifies to' common sense's general lack of understanding of the role of reasons in agency. Regarding it, he objects that the judgement 'this is right' is generally made uncritically, particularly by appeal to 'conscience', and that it thus generally merely symptomises the judge's particular inclinations and experiences.¹ However, Nietzsche's more fundamental criticism is that no agent is motivated by such a judgement per se. Rather, he insists, an agent can always ask regarding such a judgement, "what actually drives me to give a hearing to it?", and this motivation consists of her 'drives, inclinations, aversions', in contrast to the 'belief' of the judgement itself. In the case of a moral judgement, Nietzsche suggests that what motivates might resemble discipline, love, fear, or simple stupidity.² In short, then, Nietzsche here insists that an agent is motivated not by

¹ FW 335. As applied to moral judgement, this objection is frequently made elsewhere. See, for instance, M 10, 11, 21, and 24, and FW 338.

² FW 335.

reasons, considered as cognitive states ('beliefs'), but by noncognitive states ('drives, inclinations, aversions') informed by reasons.³

After making important remarks regarding Kant to which I will return in the following chapter, Nietzsche proceeds to provide grounds for his denial of practical reason. He writes the following of someone who has 'taken five steps in self-knowledge'.⁴

he would know that there neither are nor can be identical actions, – that every action which has been done was done in a completely unique and irretrievable way, and that this will be just as true of every future action, – that all rules of action (and even the most inward and subtle rules of all moralities so far) refer only to the coarse exterior, – that they can well reach an appearance of identity, *but only an appearance*, – that in looking forward to or back upon *any* action, it is and remains an impenetrable thing, – that our opinions of 'good', 'noble', 'great' can never be *proven* by our actions, because every action is unknowable, – that certainly our opinions, valuations, and tables of what is good belong among the most powerful levers in the mechanism of our actions, but that in every single case the law of their mechanism is unprovable.⁵

In this passage, I propose, Nietzsche argues that a reason could not motivate an agent because it prescriptively underdetermines action. Here he refers to a reason as a 'rule of action' – that is, a prescription of a type of action in a type of circumstance. For instance, the 'thoughtless person' to whom Nietzsche refers

³ Versions of this venerable position are adopted by, among others, Hume and Schopenhauer, with whose claims Nietzsche was acquainted, first- and second-hand. See, for instance, Hume, *A Treatise of Human Nature*, bk.1, pt.3, §10, bk.2, pt.3, §§3 and 6, and bk.3, pt.1, §1, and Schopenhauer, *Die Welt als Wille und Vorstellung*, §§23 and 55.

⁴ FW 335.

⁵ FW 335.

in *'Aftereffects...'* might have a reason to 'strike something' in the presence of the 'something' and in the absence of physical obstructions to striking it. As a reason, such a prescription also has normative grounds, if not necessarily distinctively 'moral' ones – reasons are, as Nietzsche puts it, 'opinions of "good", "noble", "great"', or 'opinions, valuations, and tables of what is good'.

In the first half of the passage, Nietzsche argues that a reason prescriptively underdetermines action because an action and its circumstances are singular, while a reason's prescription is made in terms of general types of action and circumstance. In his terms, there is no 'identity' among actions and circumstances which would allow a 'rule of action' to grasp more than their 'coarse exterior', or to express more than a mere 'appearance of identity' among them. Since he provides no further argument for this singularity here, he presumably considers it to be confirmed simply by empirical 'self-knowledge'. In the second half of the passage, Nietzsche concludes that any particular action must be motivated other than by reasons, and, furthermore, that its motivations are undeterminable. Thus he proceeds from a reason's expression of a mere 'appearance of identity' among actions to the 'impenetrable' or 'unknowable' character of an action for one 'looking forward to or back upon' it. I take it that Nietzsche considers common sense to 'look forward' to an action primarily from the perspective of reasons, which prospectively prescribe actions, and to 'look back upon' an action primarily with a view to causally 'describing' it, and particularly the reasons according to which it was chosen. From both perspectives, then, Nietzsche's claim is that, as he writes of reasons in the last phrase, 'in every single case the law of their mechanism is unprovable'.

With the first conclusion, Nietzsche effectively denies that an agent could, as common sense supposes, simply do ‘what he has [...] recognised as right and described as necessary’ with reasons. ‘*Long live physics!*’ thus provides what is lacking elsewhere in Nietzsche’s texts – namely, grounds for his consistent denial of practical reason. In *Beyond Good and Evil*, for instance, he writes of the belief in ‘[t]he intention as the whole origin and prehistory of an action’ that, in fact, ‘the intention is only a sign and symptom that first needs interpretation, moreover a sign that means too many things and consequently almost nothing by itself’. In a section of the second edition of *The Gay Science*, he similarly insists that, although ‘one is used, according to an ancient error, to seeing the *driving* force [of an action] precisely in the goals (purposes, professions, etc.)’, in fact these are ‘relatively discretionary, arbitrary, almost indifferent’. Indeed, Nietzsche here also refers to the denial of this ‘ancient error’ as ‘one of my most essential steps’. He again articulates this position in ‘*Error of a false causality*’ in *Twilight of the Idols*, where he treats the common sense claim regarding reasons as a supposed ‘inner fact’, according to which ‘all the *antecedentia* of an action, its causes, were to be sought in consciousness and could be discovered there, if one sought them – as “motives”: otherwise one would not have been free *for* it, not responsible *for* it’. This claim is simply an ‘error’, Nietzsche insists, because a reason is ‘[m]erely a surface phenomenon of consciousness, an accompaniment of the act, which conceals the *antecedentia* of an act rather than represents them’.⁶

Nietzsche’s second conclusion in this passage of ‘*Long live physics!*’ threatens further claims that he makes, however. For in numerous passages, including ‘*Long live physics!*’, Nietzsche accounts for actions as motivated not by

⁶ *JGB* 32, *FW* 360, *GD* VI 3. See also *MA* 34, and *KGW* VIII:1 7 [1] (End of 1886-Spring 1887).

reasons, but by noncognitive states informed by reasons.⁷ Presumably by elimination, then, he supposes that if reasons do not motivate an agent to act, then noncognitive states informed by reasons do. But his second conclusion in the passage of '*Long live physics!*' is that the motivations of any particular action are undeterminable. His grounds for this conclusion are presumably those of his premise, that an action and its circumstances are singular. For it follows from this premise that no causal 'description' could be provided of an action at all, since no investigation of conditions and consequences according to observed regularities could be undertaken. Thus, although the singularity of an action and its circumstances provides grounds for Nietzsche's denial of practical reason, it also undermines his consistent provision of alternative causal 'descriptions' of actions in terms of noncognitive states. In response, it may be suggested that, to articulate this denial and provide these alternative 'descriptions', Nietzsche need be committed merely to the particularity, rather than singularity, of an action and its circumstances, and hold simply that noncognitive states might capture this particularity, while reasons cannot. For particularity, unlike singularity, admits the investigation according observed regularities which causal 'description' requires. Nietzsche's argument would thus turn on the relative generality of reasons' prescriptions, noncognitive states informed by reasons, and actions and their circumstances. Nonetheless, any such modification of his argument in '*Long live physics!*' must admit that there Nietzsche himself makes a more radical claim.

⁷ See, for instance, *MA* 57, 107, 132-3, and 138, *M* 133, and *FW* 3. Nietzsche objects to uncritical causal accounts of moral and immoral actions more generally at *MA* 13, and *GD* VI 4-6, and makes related remarks at *JGB* 192 and *FW* 355.

Nietzsche's primary concern regarding his denial of practical reason is with the consequences of this denial, rather than its grounds. In particular, he takes this denial to undermine the practice of making judgements about reasons for action, and the moral philosophy which undertakes to provide moral 'grounds' for such judgements. In the place of such moral philosophy, Nietzsche proposes a 'revaluation' of obedience to moral reasons, consisting of both a symptomatology of the relevant noncognitive motivating states and an evaluation according to his own particular concern for 'humanity'. These programmatic claims are most clearly expressed in the fifth part of *Beyond Good and Evil*, 'On the natural history of morality', and in its opening few sections in particular. I will therefore first focus on these sections, before considering the 'revaluation of values' which Nietzsche proceeds to offer in the *Genealogy*.

I.

Nietzsche begins the fifth part of *Beyond Good and Evil*, 'On the natural history of morality', by raising moral philosophers' concern for 'the *grounding* of morality'.⁸ As he proceeds to make clear, he considers a 'ground' of morality to express a requirement of reason which is supposed to be unconditional and overriding – that is, a rational requirement which is supposed to bind an agent as

⁸ *JGB* 186.

such, irrespective of her contingent concerns, and to trump any other requirement. Such a 'ground' would serve to guide the judgement of reasons for action in specific circumstances, by providing for the derivation, testing, and reform of more specific reasons, which are thus gilded with an equally unconditional and overriding authority. In these opening sections, Nietzsche refers in particular to Schopenhauer's principle of compassion, Kant's categorical imperative, and the reformist undertakings of utilitarians and 'anarchists'. In later sections, he writes, regarding the 'Socratism' in Plato's moral philosophy and in 'every utilitarianism', that 'rationality [...] wants to know according to grounds, according to a "why?"', according to purposiveness and utility evaluated and acted upon', and that a 'moral' requirement is one which it is considered to be an 'error' for any agent to contravene. He also remarks of 'morals' which appeal to the supposed "happiness" of the individual, among which he includes those of the Stoics, Spinoza, Aristotelianism, and religion, 'all of them speaking unconditionally, taking themselves unconditionally'.⁹

In the opening section, Nietzsche first criticises moral philosophers for, particularly in their choice of 'ground', having 'known moral *facta* only crudely, in an arbitrary extraction or as chance abbreviation' – for instance, as the requirements which prevail in a particular environment.¹⁰ His more fundamental objections, however, derive from his denial of practical reason. Presumably on the grounds that agents are motivated to act not by reasons, but by noncognitive states informed by reasons, he objects, firstly, that what should occupy judgement is not the reasons according to which an agent acts, but rather the noncognitive states which motivate the agent; and, secondly, that an agent can be meaningfully

⁹ *JGB* 190, 191, 198. See also *JGB* 186-8, 199, and 200-202.

¹⁰ *JGB* 186.

considered to have a reason to do or refrain from an action only relative to her particular concerns, as determined by her particular motivating states, and that an unconditional and overriding reason is therefore incoherent. He insists on the first objection when he writes in the opening section, ‘what the philosophers called “grounding of morality” [...] was, seen in the right light, only an erudite form of good *faith* in the prevailing morality, a new means of its *expression*’. In the following section, he also provides a list of different motivations for obedience to moral reasons and ‘grounds’, before concluding with the general statement that ‘even morals are only a *sign language of affects*’. He emphasises his second objection in the later sections. For instance, he writes of ‘Socratism’ that the supposition of an unconditional and overriding rational requirement constitutes ‘the irrational in moral judgements’, because ‘reason is only a tool’ – and a ‘tool’ of an agent’s ‘faith’, or ‘instincts’, in particular. Of moral reasons directed at the individual’s supposed “‘happiness’”, he similarly writes, ‘all of them baroque and unreasonable in form – because they are directed at “everyone”, because they generalise what may not be generalised’. At best, he suggests, these reasons might hold relative to the particular ‘*dangerousness*’ of an agent’s circumstances, or to ‘his passions, his good and bad tendencies’, or as ‘small or large clevernesses and artifices’.¹¹

In place of the practice of judging reasons for action and the provision of ‘grounds’ for such judgement, Nietzsche proposes to provide a symptomatology of

¹¹ *JGB* 186, 187, 191, 198. See also *JGB* 6, 193, 194, and 200. Elsewhere, Nietzsche dismisses brands of rational egoism on related grounds. In *M* 99-105, he prefaces a statement of these grounds by dismissing the claim, which he attributes to La Rochefoucauld, that moral requirements are justified only with regard to an agent’s own interests, despite agents’ deceptive, and self-deceptive, claims to other motivations. See also *WS* 20 in this regard. In *GM* I 1-3 he similarly dismisses the ‘English’ position that moral requirements are justified only with regard to their consequences for an agent’s own ‘utility’, despite the processes of forgetfulness, habit, or association which account for other, spurious moral concepts and motivations. In this regard, see also *FW* 4, 21, 345, and 373, *JGB* 228, *GM* V 4 and 7, and *GD* I 12, and IX 5, 37, and 38.

the noncognitive states which motivate obedience to moral reasons and an evaluation of this obedience according to his own particular concern for 'humanity'. This genuine "science of morality" requires, as he puts it in the opening section of 'On the natural history of morality', the 'collection of material, conceptual formulation and ordering together of an enormous realm of tender value feelings and value distinctions, which live, grow, procreate, and are destroyed [...] – as preparation for a *doctrine of types* of morality'. He indicates in the second section that he intends to distinguish 'types' of morality by reading the '*sign language of affects*' expressed in obedience to moral reasons. Then, in the lengthy third section, he indicates that he also intends to provide a particular kind of evaluation of such obedience. He begins this section with the following statement: 'Every morality is, in opposition to *laisser aller*, a piece of tyranny against "nature", also against "reason": but that is still no objection to it, for which one would have to again decree from some other morality that every kind of tyranny and unreason is forbidden'.¹² He thus refuses to evaluate obedience to moral reasons according to either a 'ground' or to '*laisser aller*', the contingent concerns of particular individuals or 'types' of individuals. He considers the former incoherent and the latter, apparently, simply of dubious promise or use. He proposes instead to evaluate obedience to moral reasons as a 'means' of developing practices which 'make [...] it worthwhile to live on earth, for example virtue, art, music, dance, reason, intellectuality'.¹³ He concludes this section as follows.

¹² JGB 188.

¹³ JGB 188.

this tyranny, this arbitrariness, this strict and grandiose stupidity [of obedience to supposed moral reasons] has *trained* the spirit; slavery in the cruder and the more refined senses is, it seems, the indispensable means even for spiritual discipline and breeding. Consequently one may regard every morality [as follows]: the ‘nature’ in it teaches hatred of the *laissez aller*, the all-too-great freedom [...] ‘You should obey, someone, and for a long time: *otherwise* you will be destroyed and lose the last respect for yourself’ – this seems to me to be the moral imperative of nature, which is of course neither ‘categorical’, as the old Kant asked of it (hence the ‘otherwise’ –), nor directed at the individual [...], but rather to peoples, races, ages, classes, but above all to the whole ‘human’ animal, to *the* human.¹⁴

Nietzsche thus proposes to evaluate obedience to a moral reason by considering it as if it were the prolonged obedience of a group, and even humanity in general, to a reason whose justification is neither unconditional and overriding nor relative to the contingent concerns of particular individuals or ‘types’, but rather might lie in upholding the ‘self-respect’ of the group, and of humanity in general. However, that he, or anyone else, is committed to such an evaluation Nietzsche considers to be a matter of noncognitive motivating states appropriately orientated to such a normative perspective. Thus when he returns to this proposal in the concluding section of ‘On the natural history of morality’, he addresses it to ‘[w]e who are of another faith’.¹⁵ He also there refers to it by the name which he employs throughout his succeeding texts, namely, a ‘revaluation of values’.¹⁶

¹⁴ *JGB* 188.

¹⁵ *JGB* 203. Arguably, Nietzsche also addresses himself to this ‘we’ in his earlier texts. See, for instance, *GT* 23, *H* 4, and *RW* 4.

¹⁶ *JGB* 203.

II.

Nietzsche provides brief examples of such a ‘revaluation of values’ throughout ‘On the natural history of morality’ and elsewhere in *Beyond Good and Evil*, and he commits himself to it again in the second edition of *The Gay Science*, and, at some length, in the preface to the *Genealogy*.¹⁷ The three essays of the *Genealogy* can be considered as his most considered attempt to provide a ‘revaluation of values’, given his commitment in the preface, his note, on the reverse side of the title page, that the text is intended to provide a ‘supplement and clarification’ of *Beyond Good and Evil*, the relative length and sophistication of the essays themselves, and the absence of conflicting treatments in his later texts. Of course, this is not to say that Nietzsche considers the *Genealogy* to provide a conclusive ‘revaluation of values’. On the contrary, he suggests that the ‘revaluation of values’ which the *Genealogy* provides is speculative and incomplete.¹⁸ Nor is it to deny that he also employs other critical strategies in this text and elsewhere – I have considered other such strategies above, and will consider more below. Nonetheless, I propose that the *Genealogy* is devoted in large part to a ‘revaluation of values’, the basic elements of which are the following.

In the first essay, Nietzsche considers two instances of the same kind of moral reason, a kind concerned with the morally ‘good’ agent in particular.¹⁹ He refers to the two instances as reasons of ‘good and bad’ and reasons of ‘good and

¹⁷ See *JGB* 187-9, 195-202, and 260, *FW* 345, and *GM* V 3, 5, 6, and 7, I 17n, and III 12, and also *KGW* VIII:2 7 [6 and 8] (End 1886-Spring 1887).

¹⁸ See *GM* I 17 and 17n, and III 27, *EH* ‘GM’, and *KGB* 971 (4th January 1888). Nietzsche’s continued commitments to a ‘revaluation of values’ in his succeeding texts also suggest that he hardly considers *GM* to provide the conclusive such ‘revaluation’. See *GD* V, VI 2, VII 1, and X 5, A 13, 61, and 62, and *EH* I 1, II 9, ‘MA’ 6, ‘M’ 1, ‘JGB’ 1, ‘GD’ 3, ‘W’ 4, and IV 1.

¹⁹ See Nietzsche’s remarks on ‘good’ at *GM* I 2-5, 7, 11, and 13.

evil', respectively. First, he presents reasons of 'good and bad' as ultimately justified only by appeal to a distinguishing characteristic of exemplary 'good', or 'bad', agents, a characteristic which is supposed to bestow 'goodness', or 'badness', on their actions. Thus an agent is required to do, or not do, *x* ultimately because agents with a distinguishing characteristic *y* do, or do not do, *x*. Nietzsche illuminates this with an account of the '*concept transformation*' with which "“noble”, “aristocratic” in the social sense' became "“good” in the sense of “noble soul”, “aristocratic”, of “soul of a high order”, “privileged soul”", and the 'subjective turn' with which this 'good' was often then identified with 'a *typical character trait*' – with 'the blond-headed', 'the warrior', or 'the truthful', for instance. By simple negation, Nietzsche claims, the opposite of "“noble”, “aristocratic” in the social sense', and then the opposites of 'good' character traits, became correspondingly conceived as 'bad'.²⁰

However, Nietzsche's primary concern in the first essay is with a second instance of this kind of reason. Indeed, he 'revalues' only this second instance, by distinguishing its corresponding motivating state from that of the first. That is, reasons of 'good and evil' ostensibly differ from those of 'good and bad' only by their 'inversion' of the latter's 'goodness'- and 'badness'-bestowing characteristics.²¹ Thus, for instance, Nietzsche has the proponents of 'good and evil' requirements say, "“[...] good is everyone who does not violate, who injures no one, who does not attack, who does not requite, who leaves revenge to God, who keeps himself hidden as we do, who avoids all evil and in general desires little from life, like us, the patient, humble, just”".²² However, Nietzsche also argues that, while obedience to reasons of both 'good and bad' and 'good and evil'

²⁰ *GM I* 4, 5. See also *JGB* 260, and *GM I* 2, 4, 5, 10, and 11.

²¹ *GM I* 7.

²² *GM I* 13. See also *GM I* 7, 8, and 11.

is motivated by a certain affirmation of one's 'self' as possessing 'goodness'-bestowing characteristics, in obedience to reasons of 'good and evil' this affirmation takes a distinct form. As Nietzsche presents it, this form is distinguished by '*ressentiment*', which he treats as an agent's basic feeling that natural limits of, or threats to, his agency – 'his enemies, his accidents, his *misdeeds*', as Nietzsche puts it – ought not to be.²³ Thus in obeying reasons of 'good and evil', Nietzsche claims, one affirms one's 'self' as possessing 'goodness'-bestowing characteristics negatively, against such limits or threats.²⁴ With this claim, then, Nietzsche supposes to dispel the plausibility of professed alternative motivations, such as Christian love or those which would, as in cases of reasons of 'good and bad', refer simply to the 'goodness'-bestowing characteristics themselves. He undertakes to substantiate his claim with an account of how the 'inversion' of 'goodness'- and 'badness'-bestowing characteristics originated in a '*ressentiment*'-fuelled '*slave revolt*' against reasons of 'good and bad'.²⁵

In the second essay, Nietzsche 'revaluation' is a little more straightforward, and a substantial part of the essay is occupied with other, extraneous matters which I will consider in the third part of the thesis. The object of Nietzsche's 'revaluation' in this essay is a religious kind of reason which requires the denial of certain natural noncognitive motivating states. As Nietzsche expresses this kind of reason, it 'interprets' man's 'actual and inescapable animal instincts [...] themselves as guilt against God'. The motivation to obey such reasons, Nietzsche argues, can lie only in instincts of cruelty – and, in particular, in instincts of cruelty

²³ *GM* I 10.

²⁴ See *GM* I 10-11, and III 14-15, *EH* I 6, and *KGW* VIII:1 8 [2] (Summer 1887) and VIII:3 14 [29] (Spring 1888).

²⁵ *GM* I 7. See also *JGB* 195 and 260, *GM* I 8, 10, 11, and 13-6, *A* 24, 40, 45, and 51, and *EH* 'GM'. Note that Nietzsche also refers to the '*slave revolt*' itself as a 'revaluation' of the requirements of 'good and bad'. See *JGB* 46, *GM* I 7 and 8, *GD* VI 4, and *EH* IV 7.

which, denied outer satisfaction by political constraints, are forced to 'turn themselves inwards' on the agent. Such motivations are, he claims, distinctive of 'bad conscience'. The further appeal to divine grounds against 'instincts' themselves, and the notion of 'guilt' associated with this, Nietzsche attributes to a further 'entanglement' of 'bad conscience' with the supposed debts of a community to its ancestors, once these supposed debts have been transformed, under the influence of the relation between 'creditor' and 'debtor', into demanding obligations to deities.²⁶ Rather than a rational recognition of 'guilt against God', then, Nietzsche insists that obedience to these reasons is motivated by instincts of cruelty, turned against the agent herself.

Finally, in the third essay of the *Genealogy*, Nietzsche 'revalues' what he calls 'ascetic ideals'. Again, this 'revaluation' is relatively straightforward, although Nietzsche provides a relatively thorough presentation and discussion of it. 'Ascetic ideals' are requirements to deny necessary features of human life for the sake of another 'existence', supposed to lie beyond these features. As Nietzsche puts it, these ideals express a particular '*evaluation* of our life [...]: it (together with that to which it belongs, "nature", "world", the whole sphere of becoming and transitoriness) is put in relation to an entirely different kind of existence, to which it acts contradictorily and exclusively, *unless* it turns against itself, *negates itself*: in this case, [...] life is held to be a bridge for that other existence'.²⁷ Obedience to such reasons, Nietzsche argues, can be motivated only by a '*ressentiment*' associated with the inevitable 'suffering' of much of human life. He refers, in particular, to 'suffering' at political constraints and at the 'undetermined' nature of human life. By evaluating this 'suffering' in terms of

²⁶ *GM* II 22, 16, 24, 21. See also *GM* II 17-20 and 23, and *EH* 'GM'.

²⁷ *GM* III 11.

another 'existence', Nietzsche argues, an ascetic ideal gives this 'suffering' a 'meaning' and directs the basic 'meaning' of the associated '*ressentiment*' – that “[s]omeone must be to blame for my feeling bad” – onto the sufferer herself.²⁸ Indeed, he further argues that insofar as modern science, which does without appeals to 'an entirely different kind of existence', is practiced by appeal to the unconditional value of truth, it is equally an ascetic ideal.²⁹

With each of his 'revaluations' in the *Genealogy*, then, Nietzsche does not undertake to judge agents' reasons for actions according to a 'ground'. Nor, indeed, does he undertake to show that agents' professed reasons for actions differ from the reasons which motivate them. Rather, in each case he intends to identify the kind of *noncognitive* state which motivates obedience to a particular kind of reason. In particular, he refers to certain feelings of '*ressentiment*' and to certain instincts of cruelty. Notably, this ought not to imply that Nietzsche considers a kind of reason or a kind of noncognitive motivating state to exist in isolation. That he rather considers each such kind as part of a complex, from which each 'revaluation' abstracts a distinct strand, is strongly suggested by the interrelations between his 'revaluations' in the *Genealogy*. Among these interrelations, one might note, for instance, that '*ressentiment*' and the character of the 'priest' are salient in the 'revaluations' of both the first essay and the third, that the objects of 'revaluation' in both the second and the third essay are both modes of religious self-denial, that a particular instance of 'bad conscience' is considered as a particular 'ascetic ideal' in the third essay, and that the implications of political constraints play a role not only in the argument of the second essay, but also in

²⁸ *GM* III 13, 15. See also *GT* V 5, *GM* III 14 and 16-22, and II 7, and *GD* V 1-5, and IX 34.

²⁹ See *GM* III 23-5.

those of the first and third.³⁰ Regarding ‘ascetic ideals’, Nietzsche even admits that obedience to such reasons can be ‘meaning’-less in the sense of being motivated by noncognitive states which are, he argues, inessential to such obedience.³¹

Besides identifying a noncognitive state which motivates obedience to a kind of reason, each ‘revaluation’ in the *Genealogy* also evaluates this obedience in the manner which Nietzsche proposes in ‘On the natural history of morality’. That is, it evaluates whether obedience to a particular kind of reason, on a relatively large scale, serves to uphold the ‘self-respect’ of ‘humanity’. This is particularly pronounced in the first and third essays. In the first, Nietzsche denounces obedience to reasons of ‘good and evil’ on the grounds that it serves to produce ‘the “tame man”, the hopelessly mediocre and uninspiring’. At least, Nietzsche complains, there is something to ‘admire’ in the communities who, rather than being ‘tamed’, allowed themselves a fearful violence towards those ‘outside’. He nonetheless admits that his disappointment, or ‘nihilism’, is his own. ‘What is it that is utterly unbearable to me in particular?’, he asks regarding ‘the “tame man”’, and he expresses his answer, that ‘the “tame man”’ is not ‘a human being that justifies *the* human being’, exclusively in terms of ‘I’ and ‘we’.³² In the third essay, however, he is less modest, and more positive, in his evaluation of the obedience concerned. He begins the essay by insisting on ‘the basic fact of the human will, its *horror vacui: it needs a goal*’, and proceeds to treat this supposed ‘fact’ or ‘need’ in terms of a ‘goal’ which would answer the question, “‘what is humanity in general for?’”. In particular, Nietzsche treats ascetic ideals as providing an answer to this question which has prevailed primarily for the lack of

³⁰ For these interrelations, see, in particular, *GM* I 6-7, II 16-8, III 13, 15, and 19-21, and *KGB* 971 (4th January 1888).

³¹ See *GM* III 4-10.

³² *GM* I 11, 12. See also *GM* I 17, and, on ‘taming’, *GM* II 15, *GD* VII 2-5, and *KGW* VIII:3 15 [55 and 72-3] (Spring 1888).

an alternative answer – as he puts it, the human will ‘would rather will *nothingness* than *not* will’.³³ With regard to the modern candidate for an alternative, the practice of science, he maintains that insofar as it does not appeal to the unconditional value of truth, and thus simply constitute another ascetic ideal, it should be dismissed for failing to provide ‘a goal, a will, an ideal, a passion of great faith’ at all.³⁴ In the third essay Nietzsche thus, I suggest, attributes to ‘the human will’ as such his own commitment to ‘humanity’, while elsewhere he carefully indicates that this commitment, like every other, ultimately rests on noncognitive motivating states which need not be possessed by every agent. Indeed, he concludes the second essay precisely by warning against the criticism of one ‘ideal’ by another, and rather hoping for the ‘*redemption*’ of the ‘reality’ which every ‘ideal’ obscures.³⁵

³³ *GM* III 1, 28. See also *GM* III 23, and *EH* ‘GM’.

³⁴ *GM* III 23.

³⁵ *GM* II 24. See also *EH* V 2-3.

Conclusions

Nietzsche, then, consistently denies the common sense claim that an agent chooses for reasons, such that she is motivated by her cognitive acknowledgement of reasons, and instead considers actions to be motivated by noncognitive states informed by reasons. However, he provides grounds for this position only in a section of the first edition of *The Gay Science*, entitled '*Long live physics!*'. There he argues that an agent could not be motivated to act by reasons because a reason prescriptively underdetermines action, and that a reason prescriptively underdetermines action because it prescribes a general type of actions in a general type of circumstances, while an action and its circumstances are singular. He also concludes from this singularity that no causal 'description' of an action can be provided at all, a conclusion which undermines his own provision of such 'descriptions' in terms of noncognitive states. In response, I suggested that he need be committed merely to the particularity, rather than singularity, of an action and its circumstances, and hold simply that noncognitive states capture this particularity, while reasons do not.

Nietzsche considers his denial of practical reason to undermine the practice of making judgements about reasons for action, and to reveal the incoherence of supposed unconditional and overriding reasons, such as the 'grounds' which moral philosophy often undertakes to provide. He proposes that moral philosophy should instead be occupied with the 'revaluation' of obedience to moral reasons, by means of a symptomatology of noncognitive motivating states and an evaluation of such obedience in the light of his own particular concern for 'humanity'. The three

essays of the *Genealogy* provide his most considered attempt at such a 'reevaluation', by providing 'reevaluations' of obedience to three kinds of moral reasons.

Prevailing approaches

In this brief chapter, I contrast the preceding interpretation of Nietzsche's critical treatment of the moral agent with five pertinent strands in the prevailing Nietzsche literature, and emphasise one notable implication of his criticism of the common sense claim regarding choice in particular. Since I am concerned simply to indicate that, for the most part, the literature overlooks or misappropriates this significant element of Nietzsche's moral philosophy, I do not provide the thorough discussions which these strands, and their broader contexts and relations, would otherwise deserve.

Five prevailing strands can be distinguished among current interpretative approaches to Nietzsche's moral philosophy, I suggest. The first strand presents Nietzsche as accounting for and evaluating behaviour according to unconscious, 'power'-orientated drives. This project is sometimes, but not always, considered as an instance of a broader metaphysics according to which every being and process must be accounted for in such terms. The extensive and systematic interpretations provided by Müller-Lauter and, more recently, John Richardson provide representative examples.¹ Others have defended the project on more modest metaphysical grounds – Peter Poellner, for instance, has Nietzsche appeal only to an agent's correct interpretation of her own mental states.² However, this project clearly has no place in Nietzsche's critical treatment of the common sense conception of agency, as I have presented this treatment above. For his criticism of the common sense claim regarding choice consists of certain doubts regarding the claim's adequacy as a causal account of choice, and certain arguments that two particular instances of this claim either misrepresent choice or make it incoherent, while his criticism of the claim regarding reasons consists of exposing the motivation of action by noncognitive states. In neither case does Nietzsche appeal to determination by unconscious, 'power'-orientated drives. In particular, his

¹ See Müller-Lauter, *Nietzsche*, esp. pp.(11-22, 31-7, 43-9, 61-3, and 66-72), 'Nietzsches "Lehre" vom Willen zur Macht', pp.(130-41, 147-53, and 156-9), and 'Nietzsches Auf-lösung des Problems der Willensfreiheit', pp.34-5 and 50-63, and Richardson, *Nietzsche's System*, esp. pp.18-72, 157-60, 203-19, and 262-90. For other examples of this strand in the literature, see Deleuze, *Nietzsche et la philosophie*, esp. pp.1-82 (1-72), Danto, *Nietzsche as Philosopher*, pp.68-99 and 214-28, Foucault, 'Nietzsche, la Généalogie, l'Histoire', and Blondel, *Nietzsche*, esp. pp.275-346 (201-59).

² See Poellner, *Nietzsche and Metaphysics*, esp. pp.214-57. Similarly modest treatments include Kaufmann, *Nietzsche*, pp.179-207, and Clark, *Nietzsche on Truth and Philosophy*, pp.205-44.

symptomatology of the noncognitive states which motivate professed obedience to moral reasons betray no general commitment regarding the character of noncognitive states at all – that is, no general commitment regarding, for instance, their origins, objects, locus, structure, or ends, or their conscious or unconscious status.

In this regard, one piece of textual evidence is also relevant. In a section of *Beyond Good and Evil* that is often seized upon by advocates of this strand in the literature, Nietzsche argues for ‘the right to clearly determine *all* effective force as: *will to power*’. However, his argument is premised precisely on the common sense claim regarding choice: he writes, ‘The question is ultimately whether we acknowledge the will as really *effective*, whether we believe in the causality of the will [...] and the belief in *this* is basically just our belief in causality itself’. If this belief were admitted, Nietzsche argues, then the ‘right’ to which he refers would follow from certain further premises – in particular, the premise that right ‘method’ requires that kinds of causality be kept to a minimum, the premise that “will” can [...] effect only “will”, and the premise that every human drive can be reduced to ‘one basic form of will’, ‘the will to power’.³ Nietzsche thus offers a hypothetical argument for a metaphysics of the kind commonly attributed to him, but premises it on the claim regarding choice which he consistently and emphatically rejects.⁴ Furthermore, he presents his last premise, that regarding the

³ *JGB* 36. Nietzsche makes related remarks at *JGB* 13, 22, and 23, and *KGW* VII:3 35 [15] (May-July 1885).

⁴ This feature of the argument at *JGB* 36 is sometimes noted, particularly by those proposing metaphysically modest interpretations. See, for example, Clark, *Nietzsche on Truth and Philosophy*, pp.212-8, and, for a more cautious conclusion, Poellner, *Nietzsche and Metaphysics*, pp.46 and 267-70.

reduction of drives to 'one basic form of will', as a hypothesis which he is merely 'assuming', rather than as even a metaphysically modest commitment.⁵

Two further strands in the literature treat Nietzsche as conceiving of moral agency primarily, or even exclusively, in terms of the agent's 'nature'. In particular, one strand treats Nietzsche as considering moral agency to be a matter of the agent's motivating states, while the other treats him as considering moral agency to be a matter of the agent's constitution. Lester Hunt, for instance, provides an extensive instance of the first of these approaches. According to him, Nietzsche is occupied with a conception of a 'virtue' as a certain relation between an agent, a passion, and its purpose, such that the agent gives the passion a purpose which would integrate her particular self, as defined by her particular drives and purposes.⁶ Brian Leiter, on the other hand, has recently taken the second approach. He considers Nietzsche to hold that each individual has a fixed psycho-physical constitution which causally, and non-trivially, circumscribes her possible actions and beliefs, and which justifies normative judgements about her interests. This position is supposed to dispose of certain descriptive claims regarding agency which some kinds of moral judgement logically presuppose. However, according

⁵ Nietzsche expresses a similar hesitancy at the end of *JGB* 19, and when he reduces particular instances of human behaviour to 'power'-orientated drives, he consistently refrains from making more general claims. See, for instance, *MA* 50 and 103, *WS* 6, *M* 146, 189, and 281, *JGB* 23, 44, and 259, *GD* IX 11 and 20, and X 3, and *A* 6, 16, and 17. In *GM* in particular, he draws very limited conclusions from a 'power' metaphysics at *GM* I 13, and II 11-13, and even when he claims at *GM* III 7 that '[e]very animal [...] strives instinctively [...] to release its force completely and reach its maximum in feeling of power', he does not deny other 'instincts' or motivations. Finally, his redefinition of 'good', 'bad', and 'happiness' in terms of 'power' at *A* 2-6 is intended to express his particular evaluative concern for 'humanity', rather than an account of human behaviour.

⁶ See Hunt, *Nietzsche and the Origin of Virtue*, pp.70-144. More common, perhaps, is the simple view that Nietzsche treats the 'power' characteristic of an agent's noncognitive motivating states as basic to moral value. See, for instance, Slote, 'Nietzsche and Virtue Ethics', esp. pp.24-5. Among other variants of this strand are the following: Solomon, in his 'A More Severe Ethics', considers Nietzsche to endorse a basic 'morals' of practices articulated in terms of 'virtues' of Aristotelian form; Brobjer, in his *Nietzsche's Ethics of Character*, esp. pp.15-24, 31-40, and 54-97, maintains that Nietzsche's basic moral concern is with an agent's 'character', determined by her unconscious, physiological drives; and Swanton, in her 'Outline of a Nietzschean Virtue Ethics', esp. pp.31-5, differs from the common variant by doubting that Nietzsche considers 'virtue' to be basic to moral value.

to Leiter, Nietzsche's primary concern is to promote the interests of individuals with a certain, 'higher' constitution, a normative concern that, since it is not relative to individuals' interests, lacks justification and mere reflects his own particular constitution.⁷

Again, however, these two strands bear little relation to Nietzsche's critical treatment of the moral agent, as I have presented it above. Regarding the common sense claim regarding choice, his criticisms concern the claim's credentials as a causal account of choice and how two particular instances of the claim misrepresent choice or make it incoherent. His criticism of the claim that an agent chooses for reasons equally makes no references to agents' motivating states or constitutions, being rather concerned with reasons' underdetermination of action. Regarding the evaluation of agents, Nietzsche is admittedly overwhelmingly concerned with identifying the noncognitive states which motivate professed obedience to reasons, and also concerned that rational requirements appreciate the particularity of agents' motivating states, and agents' particular 'strengths' or physiological constitutions. But he does not propose that agents should be evaluated according to such states, 'strengths', or constitutions, or, indeed, that particular agents should be evaluated at all. Even what he considers to be the proper object of evaluation – namely, a group's obedience to moral reasons – he refuses to evaluate according to the contingent concerns of particular individuals or types of individuals, apparently because he considers such an evaluation to be of little promise or use. Rather, he proposes an evaluation according to his particular concern for 'humanity', a concern which he articulates in terms of the

⁷ See Leiter, *Nietzsche on Morality*, esp. pp.3-11, 26-8, 58-63, 78-112, and 136-56, and also 'Morality in the Pejorative Sense', pp.122-42, 'The Paradox of Fatalism and Self-Creation in Nietzsche', pp.219-37, and 'Nietzsche's Metaethics'. For another example of the second approach, see May, *Nietzsche's Ethics and his War on 'Morality'*, esp. pp.9-22 and 26-38.

‘self-respect’ or ‘justification’ of humanity, and not in terms of its ‘nature’ or the ‘natures’ of particular individuals or types within it.

Two other common strands in the literature present Nietzsche’s treatment of moral agency as concerned with an agent’s self-relation. One such strand conceives of this self-relation as a kind of ‘self-creation’. A notable example is Alexander Nehamas’s interpretation, according to which proper evaluation concerns whether the agent’s ‘interpretation’ of herself – and, in particular, her behaviour, thoughts, and desires – satisfies aesthetic standards such as unity and complexity.⁸ Others, however, conceive of the pertinent self-relation as a kind of ‘self-determination’. Robert Guay, for instance, has recently presented Nietzsche as occupied with how the ultimate authority of a norm could lie in its being essential to an agent’s self-determination. Such authority, Guay argues, can be substantiated neither by subjectivism, which would be arbitrary, nor by norms ‘external’ to the ‘self’, which would be inadequate. These constraints are understood to preclude any *a priori* account of such authority, and to make such authority rather one internal to practices which provide extra-subjective meanings of, for instance, ‘self’, ‘freedom’, or ‘agency’.⁹

Once again, however, such concerns fail to surface in Nietzsche’s critical treatment of the moral agent, as I have presented it above. Nietzsche might display a certain concern for the ‘self’ with his rejection of the notion of a substantial

⁸ See Nehamas, *Nietzsche*, esp. pp.76-8, 89-91, and 155-91. For other instances of this strand, see Berkowitz, *Nietzsche*, esp. pp.1-21, and Lackey, ‘Killing God, Liberating the “Subject”’, esp. pp.751-3.

⁹ See Guay, ‘Nietzsche on Freedom’, esp. pp.302-18. Havas’s interpretation, articulated in terms of linguistic ‘responsibility’, bears some comparison with Guay’s. See Havas, *Nietzsche’s Genealogy*, esp. pp.182-238. Another notable instance is Ridley’s account of *GM* as occupied with the configurations of ‘conscience’, a capacity to act according to normative reflection on oneself. See Ridley, *Nietzsche’s Conscience*, esp. pp.15-22 and 39-40. Other instances of ‘self-determination’ interpretations include Jenkins, ‘Morality, Agency, and Freedom in Nietzsche’s *Genealogy of Morals*’, pp.73-9, and the ‘autonomy’ interpretations of Nietzsche’s criticism of Kant which I outline in the first subsection of the following chapter.

agent, or 'I', his undertaking to investigate agents' noncognitive motivating states, or his insistence that rational requirements appreciate these states and agent's particular 'strengths' or physiological constitutions. But even these elements of his critical treatment of the moral agent are decidedly distant from any concern for an agent's relation to herself. The perspective from which he proposes to evaluate professed obedience to moral reasons – namely, whether it serves humanity's 'self-respect' or 'justification' when practiced on a relatively large scale – is even further removed from such a concern. Furthermore, insofar as they tend to attribute to Nietzsche a conception of agency which, particularly in its commitment to practical reason, echoes precisely that of common sense, these strands conflict with his critical treatment of the latter.

The above account of Nietzsche's criticism of the common sense claim regarding choice has one broad implication for prevailing interpretative approaches to Nietzsche that I shall note here: it reveals that, with this criticism, Nietzsche does not deny responsibility as such. Some commentators claim the opposite, on the grounds either that Nietzsche considers 'free will' to be a necessary condition of responsibility, or that he objects equally to 'free will' and compatibilist alternatives. Müller-Lauter provides an example of the former approach, by interpreting Nietzsche as affirming a metaphysics of unconscious, 'power'-orientated drives that precludes 'free will' and thus responsibility.¹⁰ Leiter, however, takes the latter approach, by arguing that, for Nietzsche, conscious states are causally effective only by virtue of the agent's fixed psychophysical constitution, the psychological elements of which are inaccessible to consciousness. Nietzsche is thus supposed to deny the effectiveness of conscious states as such, as required by both incompatibilist and compatibilist conceptions of responsible agency.¹¹ However, although Nietzsche treats 'free will' as a necessary condition of responsibility in passages from *Human, All Too Human* to *Daybreak*, he explicitly dismisses this position in his later texts, and there merely dismisses 'free will' and the over-inflated sense of responsibility which, he claims, it logically underwrites. He thus leaves open the possibility of conceiving of responsibility differently, perhaps on the grounds of a compatibilist conception of

¹⁰ See the references in n.1 above. For an example from another perspective, see Ridley, *Nietzsche's Conscience*, pp.26-30.

¹¹ See Leiter, *Nietzsche on Morality*, esp. pp.78-81 and 87-112, and also 'Morality in the Pejorative Sense', pp.122-6, and 'The Paradox of Fatalism and Self-Creation in Nietzsche', pp.226-37.

freedom. Indeed, this possibility is effectively, if not always explicitly, admitted by much of the literature, insofar as Nietzsche is often attributed a positive conception of responsibility.¹²

¹² Thus some commentators consider Nietzsche to propose or presuppose a standard sense of responsibility, according to which an agent is responsible for what she consciously chooses, given standard qualifications regarding circumstantial and psychological constraints. See, for example, Hunt, *Nietzsche and the Origin of Virtue*, esp. pp.146-52, and Williams, 'Nietzsche's Minimalist Moral Psychology', esp. pp.241-5. However, a variety of reformed senses of 'responsibility' have also been attributed to Nietzsche. In *Nietzsche's System*, pp.207-16, Richardson considers Nietzsche to identify an agent's 'freedom' and 'responsibility' with certain qualities of her unconscious, 'power'-orientated drives. More modestly, perhaps, May maintains that, for Nietzsche, an agent's 'freedom' and 'responsibility' consists of a certain expression and affirmation of natural facts about herself. See May, *Nietzsche's Ethics and his War on 'Morality'*, pp.21-2 and 29-30, and for a similar account, Solomon, 'Nietzsche as Existentialist and as Fatalist', esp. pp.44-7, and 'Nietzsche on Fatalism and "Free Will"', esp. pp.74-83. Others present Nietzsche as conceiving of responsibility as a kind of achieved 'self-mastery'. See, for instance, Schacht, *Nietzsche*, pp.296-316, and Jenkins, 'Morality, Agency, and Freedom in Nietzsche's *Genealogy of Morals*', pp.69-79. In a similar vein, Oaklander maintains that, for Nietzsche, responsibility rests on an agent's achieved ability to realize her self-determined goals. See Oaklander, 'Nietzsche on Freedom', pp.215-8.

Conclusions

Despite its substantial character and its salience throughout his texts, then, Nietzsche's critical treatment of the common sense conception of agency is not well represented by pertinent strands in the prevailing Nietzsche literature. In particular, this critical treatment displays no concern to account for or evaluate agency in terms of unconscious, 'power'-orientated drives, in terms of motivating states or natural constitutions, or in terms of the agent's 'self-creation' or 'self-determination'. Nor does this treatment deny responsibility as such.

I have, of course, not engaged thoroughly with these strands. Advocates of them might therefore respond with indifference to my account, on the grounds that these strands find confirmation in passages other than those which I have considered, and that, for textual or other reasons, these passages ought to take priority. This would be to marginalise a substantial collection of significant passages in Nietzsche's texts, however. Furthermore, I will argue in the following chapter that Nietzsche's concern with common sense, and not the prevailing strands, strongly informs his explicit critical remarks regarding Kant's moral philosophy. Then, in the second part of the thesis, I will argue that, thus understood, Nietzsche's explicit criticisms are, if not conclusive refutations of Kant, at least more pertinent to Kant's conception of the moral agent than the prevailing strands in the Nietzsche literature. If I am right, then, to uphold these strands is not only to marginalise substantial textual evidence. It is also to misunderstand Nietzsche's explicit criticisms of Kant and to credit Nietzsche with a less promising critical perspective on Kant than he in fact offers.

Nietzsche on Kant

In this chapter I consider the relationship between Kant's and Nietzsche's moral philosophies as this is represented in the literature and in Nietzsche's own explicit critical remarks regarding Kant. I first present, again in schematic form, the literature's standard claims regarding this relationship, and then consider Nietzsche's explicit critical remarks regarding Kant's moral philosophy. I argue that, although these remarks provide some warrant for one kind of standard interpretative claim, these remarks more often raise Nietzsche's criticisms of the common sense conception of agency and further criticisms specific to Kant. I thus defend an account of Nietzsche's explicit criticisms of Kant's conception of the moral agent.

The literature on the relationship between Kant's and Nietzsche's moral philosophies is relatively small, and presents a surprisingly uncontested consensus. It standardly claims that Nietzsche rightly reveals Kant's invoking of 'other-worldly' entities and qualities as a shameful ruse to avoid critical reflection upon his conception of moral goodness, its theoretical requirements, and its supposed authority. By revealing the untenability of such 'other-worldly' entities and qualities, so this literature claims, Nietzsche admits the required critical reflection and proceeds to reveal the untenability of Kant's conception of moral goodness. The literature typically concludes that Nietzsche thus provides for a moral philosophy which is more critically sound, and richer in its insights and uses, than Kant's.

The literature differs over the kind of critical reflection with which Nietzsche effects this critical coup, however. Three kinds can be distinguished. Firstly, some commentators, such as Ridley, present Nietzsche as simply insisting that, although Kant dispels the rational credentials of certain traditional authorities, his conception of 'autonomy' or 'reason' is not yet 'autonomous' or 'rational' enough, since by invoking an authority beyond particularity, it is ascetic and uncritical.¹ Others, such as Müller-Lauter, emphasise Nietzsche's metaphysics and suppose that, with his criticism of 'other-worldly' entities and qualities, Nietzsche overcomes the dualisms of Kant's moral philosophy and urges the affirmation of

¹ See, in particular, Ridley, *Nietzsche's Conscience*, pp.1-11 and 69-72. See also Owen, *Maturity and Modernity*, chs.1-4, *Nietzsche, Politics, and Modernity*, pp.87-90, and 'Nietzsche, Enlightenment, and the Problem of Noble Ethics', pp.3-11, and May, *Nietzsche's Ethics and his War on 'Morality'*, p.13.

what allegedly remains, a 'this-worldly' realm of drives or 'becoming'.² However, the literature most commonly considers Nietzsche's criticism of Kant to be based on a commitment to 'autonomy', conceived as an individual's choice of, or according to, her particularity or, more often, her singularity. This particularity or singularity is generally defined in terms of the individual's drives or desires and supposed to be created, discovered, selected, or organised by the individual herself. Accordingly, Nietzsche is supposed to criticise Kant for betraying such 'autonomy', particularly by characterising moral goodness as universal.³ In Will Dudley's recent, relatively sophisticated version, for instance, Nietzsche is considered to provide an 'autonomy'-based criticism of conceptions of freedom as absence of constraint, but also to supersede Kant's uncritical, 'morality'-based conception of 'autonomy' with an individualistic, instinct-based 'noble' alternative, which is qualified by an individual's openness to the 'other' of her chosen particularity.⁴ Although not articulated explicitly in terms of 'autonomy', Hunt's account provides another representative instance. For Hunt similarly considers Nietzsche to endorse an individual's giving of purposes to her singular passions for the sake of integrating her particular self, and to direct this conception of 'virtue' against a Kantian conception of 'morality' as requiring a spontaneously

² See Deleuze, *Nietzsche et la philosophie*, pp.102-8 (89-94), Du Plessis, 'Nietzsche's Use and Abuse of Kant's Philosophy', Strong, *Friedrich Nietzsche and the Politics of Transfiguration*, pp.42-5, Ansell-Pearson, 'Nietzsche's Overcoming of Kant and Metaphysics', Müller-Lauter, 'Nietzsches Auf-lösung des Problems der Willensfreiheit', esp. pp.25-7, and Simon, 'Moral bei Kant und Nietzsche'.

³ See, for example, Olafson, 'Nietzsche, Kant, and Existentialism', pp.196-200, Cartwright, 'Kant, Schopenhauer, and Nietzsche on the Morality of Pity', p.83, Warren, *Nietzsche and Political Thought*, pp.172-4, White, 'Nietzsche contra Kant and the Problem of Autonomy', and *Nietzsche and the Problem of Sovereignty*, esp. pp.36-44, Ansell-Pearson, 'Nietzsche and the Problem of the Will in Modernity', pp.174-87, 'Nietzsche on Autonomy and Morality', pp.273-80, and 'The Significance of Michel Foucault's Reading of Nietzsche', pp.25-6, Gerhardt, 'Selbstbegründung', esp. pp.40-4 (293-7), Brobjer, *Nietzsche's Ethics of Character*, pp.103-4 and 206-21, and Kerckhove, 'Nietzsche's Critique of Kantian Moral Autonomy', pp.15-20, 24-6, and 28-30.

⁴ See Dudley, *Hegel, Nietzsche, and Philosophy*, pp.3-8, 123-212, and 227-30.

free agent to act for the sake of universal reasons.⁵ Also typical is Guay's remark that Kant's identification of 'self-determination' with conformity to the moral law is, according to Nietzsche's conception of 'self-determination', insufficiently critical, since it appeals to something 'external' to the 'self'.⁶ Finally, an atypically critical instance is provided by J.M. Bernstein, who argues that, by rightly pursuing 'autonomy' beyond all determination, Nietzsche proceeds beyond Kant to reveal the emptiness of 'autonomy' as a normative ideal.⁷

⁵ See Hunt, *Nietzsche and the Origin of Virtue*, chs.2 and 5-8, esp. pp.22-3.

⁶ See Guay, 'Nietzsche on Freedom', p.310.

⁷ See Bernstein, 'Autonomy and Solitude'. Several exceptions to this tripartite division of the literature should be noted here. Kittmann's *Kant und Nietzsche* is devoted to thematic comparison, rather than critical engagement, between Kant's and Nietzsche's moral philosophies. Solomon, in his 'A More Severe Morality', considers Nietzsche to dismiss Kant's 'morality' of unconditional rational requirements in the name of a 'morals' of practices articulated in terms of 'virtues'. In *Nietzsche and Metaphysics*, pp.259-65, Poellner distinguishes Nietzsche from Kant on the grounds that Nietzsche's 'values', determined by his own unconscious drives, are predominantly concerned with a certain kind of 'self-overcoming' for its own sake, rather than for a further end, as in Kant's case. Finally, in his 'Nietzsche's Response to Kant's Morality', pp.206-12, Williams argues that Nietzsche's individualistic conception of 'autonomy' reveals a better appreciation of moral psychology than Kant's alternative conception, particularly regarding how general rules of action tend to obstruct individuals' practice of critical judgement.

With the exception of the 'autonomy'-based interpretations, support for the standard interpretative claims is difficult to find among Nietzsche's explicit remarks regarding Kant's moral philosophy. Indeed, I will argue, firstly, that, insofar as these remarks specify any critical grounds at all, they generally place Kant's moral philosophy in the context of either Nietzsche's criticisms of the common sense conception of agency, or the 'revaluation of values' which he proposes and practices in his later texts. Admittedly, however, in some of these remarks Nietzsche also makes emphatic use of the terms of an individualistic conception of 'autonomy'. Nonetheless, I will argue, secondly, that this use is unspecific, thus failing to support the particular kind of gloss standardly provided in the literature, and that it is problematised by Nietzsche's critical treatment of the common sense claim regarding reasons.

Before considering these remarks, it is worth noting that Nietzsche does not make them on the grounds of a merely superficial acquaintance with Kant's moral philosophy, despite common claims to the contrary.⁸ Nietzsche's voracious reading included numerous works which comment on, or are informed by, Kant's moral philosophy or his philosophy more generally, and there is evidence that he also read first-hand a number of Kant's works on moral philosophy. Most notably, Nietzsche's own collection of books, his published works, his notebooks and letters, and his borrowings from libraries reveal his reading of Arthur Schopenhauer's *Die Welt als Wille und Vorstellung* and *Über das Fundament der*

⁸ An influential example is Heidegger, *Nietzsche*, v.1, p.130 (v.1, p.111). For other examples, see the list in Dickopp, 'Aspekte zum Verhältnis Nietzsche-Kant und ihre Bedeutung für die Interpretation des „Willens zur Macht“', n.5.

Moral, Friedrich Lange's *Geschichte des Materialismus*, Kuno Fischer's *Immanuel Kant und seine Lehre*, and Friedrich Überweg's *Grundriss der Geschichte der Philosophie*. They also reveal that he read Kant's *Critique of Judgement* at the turn of 1868, and suggest that he read that text again, along with Kant's *Critique of Practical Reason*, *Religion within the Boundaries of Mere Reason*, and *The Conflict of the Faculties*, in the library at Chur, Switzerland, between May and June 1887 – that is, immediately before he composed the *Genealogy*.⁹

I.

Despite his quite conscientious study of Kant's moral philosophy, however, the majority of Nietzsche's explicit remarks regarding it simply insist that it betrays an unforgivable lack of critical reflection.¹⁰ They often do so by presenting Kant as invoking 'other-worldly' entities and qualities which preclude, and are even intended to preclude, critical reflection on his conception of moral requirements. In the preface to *Daybreak*, for instance, Nietzsche writes of Kant that, 'to create room for his "moral realm", he saw himself necessitated to posit an indemonstrable world, a logical "Beyond", – it was for just this that he needed his critique of pure reason!'. 'In other words', Nietzsche continues, Kant needed to

⁹ For the evidence of Nietzsche's reading of these particular texts, and more extensive discussions of the sources of his knowledge of Kant's moral philosophy, see, in particular, Brobjer, *Nietzsche's Ethics of Character*, pp.198-201 and tables 2-4 to ch.7, and Hill, *Nietzsche's Critiques*, pp.9-20.

¹⁰ See *MS* 27, *WS* 216, *M* 142, 197, and 481, *FW* 193 and 335, *JGB* 5, *GM* III 12, *W* 7, *GD* III 6, IV, and IX 1, 16, 29, and 42, *A* 10, 12, 55, and 61, *EH* 'UB' 3, and 'W' 2-3, and *KGW* III:4 19 [34, 53, and 136] (Summer 1872-Beginning 1873), V:1 6 [135] (Autumn 1880), VII:1 7 [21] (Spring-Summer 1883), VII:3 35 [32] (May-July 1885), VIII:1 2 [165 and 190] (Autumn 1885-Autumn 1886) and 7 [4] (End 1886-Spring 1887), VIII:2 9 [3 and 178] (Autumn 1887), and 10 [11] (Autumn 1887), and VIII:3 14 [108 and 163] (Spring 1888).

posit a theoretically ‘indemonstrable world’ precisely in order ‘to make the “moral realm” unassailable, better still incomprehensible to reason, – he felt too strongly the assailability of a moral order of things to reason!’.¹¹

Of the many passages in which he make such remarks, however, in only six does Nietzsche indicate precisely how he considers Kant’s moral philosophy to be critically lacking. Furthermore, the exclusive concern of four of these six passages is with epistemology. In particular, these four passages, found in Nietzsche’s later texts, articulate his denial of the sceptical notion that knowledge claims which satisfy ordinary standards of justification might nonetheless be false. Besides adopting this sceptical notion himself in earlier works and notes, he consistently attributes it to Kant. Thus Nietzsche writes in perhaps the clearest of the four passages, the chapter, ‘How the “True World” Finally Became a Fable’, in *Twilight of the Idols*, that, according to the ‘Königsbergian’ conception, the ‘true world’ is ‘unattainable, indemonstrable, unpromisable, but merely as thought a consolation, an obligation, an imperative’. However, Nietzsche argues, the ‘unattained’ status of the ‘true world’ renders it ‘also *unknown*’, and ‘[c]onsequently also not consoling, redeeming, obligating’, and thus renders it entirely ‘useless, superfluous’.¹² This passage concisely expresses the objections to the sceptical notion which Nietzsche expresses elsewhere – in particular, his objections to its epistemological insignificance, its mischaracterisation of the ‘true’, and its suspiciously convenient serving of certain practical interests. In conclusion, he writes, ‘*with the true world we have also done away with the*

¹¹ *M V 3*. Of the passages listed in the preceding footnote, the following echo the claims of *M V 3* regarding Kant’s metaphysics: *MS 27*, *M 142*, 197, *FW 335*, *GM III 12*, *GD III 6*, IV, and IX 1, 16, and 29, A 10 and 55, *EH ‘W’ 2-3*, and *KGW III:4 19 [34 and 53]* (Summer 1872-Beginning 1873), V:1 6 [135] (Autumn 1880), VIII:1 2[165] (Autumn 1885-1886), and VIII:2 9 [3 and 178] (Autumn 1887).

¹² *GD IV*.

apparent!' – that is, he insists that to deny the 'true world' is also to deny the scepticism which it underwrites.¹³

Nietzsche clearly considers this epistemological denial of scepticism to dispose of the obstacle to critical reflection which, he maintains, Kant's invoking of such entities and qualities raises. In the terms of the preface to *Daybreak*, he considers Kant's "moral realm" to manifest an eminent 'assailability [...] to reason', if denied recourse to a theoretically 'indemonstrable world'. In this respect, then, there is some credibility to the literature's standard claim that in refusing 'other-worldly' entities and qualities, Nietzsche intends to admit critical reflection on Kant's moral philosophy. However, being exclusively concerned with epistemology, these four passages provide no indication of how Nietzsche proposes that this critical reflection ought to proceed. In particular, in these passages he notably fails to insist on the particularity of normative authority, urge the affirmation of the 'this-worldly', or invoke a conception of 'autonomy', as the standard interpretative claims would have him do.

II.

There remain, then, only two passages in which Nietzsche both dismisses Kant's metaphysics for obstructing critical reflection on his moral philosophy and

¹³ *GD IV*. My interpretation of this passage, I maintain, applies equally to *GM III 12*, *GD III 6*, and *A 10*, the other three passages in which Nietzsche expresses his epistemological concerns regarding Kant's moral philosophy. For a thorough defence of this interpretation of these and other passages, see my 'Two Interpretations of Nietzsche's Epistemology', and 'Friedrich Nietzsche and the "Real World"'. Broadly-speaking, my interpretation corresponds with Clark's, in her *Nietzsche on Truth and Philosophy*, pp.29-158, and 'On Knowledge, Truth, and Value'. In contrast, much of the literature on the relation between Kant's and Nietzsche's theoretical philosophies remains wedded to treating them both as brands of scepticism.

indicates how this critical reflection ought to proceed. These two passages also provide Nietzsche's most extensive discussions of Kant's moral philosophy, and in both he apparently endorses an individualistic conception of 'autonomy'. I will consider these two passages in some detail, therefore, before briefly summarising Nietzsche's other explicit critical remarks regarding Kant's moral philosophy.

The first of these two passages is part of the section entitled '*Long live physics!*' in *The Gay Science*, the section which, I have argued, is unique in providing grounds for Nietzsche's consistent denial of practical reason. Immediately before providing these grounds, and in the context of his complaints regarding the uncritical nature of many moral judgements and the failure of common sense to appreciate that actions are motivated by noncognitive states, Nietzsche raises the issue of Kant's conception of moral requirements and its associated metaphysics. He writes the following.

And now don't talk about the categorical imperative [...]: with it I remember the old Kant, who, as punishment for having *helped* himself to the 'thing in itself' – also a very ridiculous thing! –, had the 'categorical imperative' creep in and with it in his heart *lost his way back* to 'God', 'soul', 'freedom', and 'immortality' again, like a fox who loses his way back to his cage: – and it was *his* strength and cleverness which had *broken open* this cage! – What? You admire the categorical imperative in you? This 'firmness' of your so-called moral judgement? This 'unconditionality' of the feeling 'in this everyone must judge as I do'? Rather admire your *selfishness* in that! And the blindness, pettiness and undemandingness of your selfishness! For it is selfish to feel *one's* judgement as universal law; and on the other hand [it is] a blind, petty and undemanding selfishness, because it betrays that you have not yet discovered yourself, [that you have] created no particular, most particular ideal for

yourself: – for this could never be that of another, let alone of everyone, everyone! –
– Whoever still judges, ‘in this case everyone must act in this way’, still has not
taken five steps in self-knowledge [...].¹⁴

In this passage, Nietzsche first articulates his denial of the sceptical notion of the ‘true world’, as a denial which originates in Kant’s own ‘strength and cleverness’, but which Kant also contravenes with the ‘ridiculous’ concept of ‘the “thing in itself”’ and then again with the concepts of ““God”, “soul”, “freedom”, and “immortality”’ implicated in his conception of moral requirements. Unlike the typical passages in which he insists on Kant’s lack of critical reflection, however, here Nietzsche also provides the critical reflection that he considers to be required. This reflection concerns Kant’s conception of a moral judgement as a judgement which, in the terms which Nietzsche borrows from Kant, distinctively holds ‘as universal law’ or as ‘categorical imperative’. Nietzsche also expresses this supposed feature of moral judgement as the ““unconditionality” of the feeling “in this everyone must judge as I do””, and as that with which one ‘judges, “in this case everyone must act in this way”’. As Nietzsche understands it, then, the kind of judgement concerned is distinguished by holding that every agent should do or refrain from a type of action in a type of circumstance.¹⁵ By discussing such judgement while criticising the common sense claim regarding reason, which he considers to underpin the ordinary moral practice of judging reasons for actions, Nietzsche also implies that Kant intends such judgement to be made of specific actions in specific circumstances, retrospectively or prospectively.

¹⁴ *FW* 335.

¹⁵ This understanding of a Kantian moral judgement is also manifested at *MA* 25.

Regarding this conception of a moral judgement, Nietzsche first complains that 'it is selfish to feel one's judgement as universal law'.¹⁶ Presumably, he thus alleges that one could justify such a judgement only on 'selfish' grounds. However, that Nietzsche employs descriptive, rather than normative, terms in this passage and, in particular, that he employs 'to feel [*empfinden*]' and 'feeling [*Gefühl*]' suggest that he is not only concerned with the normative grounds of Kantian moral judgements, but also, as elsewhere in the section, with identifying noncognitive states which motivate professed obedience to certain reasons. Thus his complaint need not be only that one could make a Kantian moral judgement only for 'selfish' reasons, but also that the noncognitive states which motivate professed obedience to such a judgement are 'selfish'.

Although Nietzsche provides no argument for this complaint, he does proceed to further specify it, by insisting that it is 'a blind, petty, and undemanding selfishness' which makes or obeys a Kantian moral judgement. Presumably, the 'selfishness' is 'blind' because it betrays a lack of 'self-knowledge', a failure to have 'discovered' oneself. Nietzsche immediately proceeds to identify the 'self-knowledge' required as that which disposes of the common sense belief in practical reason – namely, the knowledge that an agent could not act simply for reasons because, given the singularity of actions and circumstances, reasons inevitably underdetermine action. Thus Nietzsche effectively implicates Kant's conception of moral judgement in the mistaken common sense claim regarding reasons.

That the 'selfishness' which makes or obeys a Kantian moral judgement is 'petty' and 'undemanding', on the other hand, is due to its betraying a lack of a

¹⁶ Nietzsche makes a similar complaint at *KGW* IV:2 23 [154] (End 1876-Summer 1877).

‘particular, most particular ideal’ which ‘could never be that of another’. This allegation strongly suggests a conception of ‘autonomy’ of the kind standardly attributed to Nietzsche, and this suggestion is reinforced by the conclusion which he draws immediately after dismissing the common sense belief in practical reason. There he writes the following.

Let us therefore *limit* ourselves to the purification of our opinions and valuations and to the *creation of our own new tables of what is good*: – but let us no longer want to brood over the ‘moral worth of our actions’! [...] To sit in judgement morally should go against our taste! Let us leave this talk and this bad taste to those who have nothing more to do than to drag the past some small way further through time and who are never themselves present, – the many therefore, the majority! We, however, *want to become who we are*, – the new, the unique, the incomparable, those who give themselves laws, those who create themselves! And for that we must become the best students and discoverers of everything lawful and necessary in the world: we must be *physicists* in order to be able to be *creators* in this sense, – while hitherto all valuations and ideals have been constructed on *ignorance* of physics or in *contradiction* to it. So: long live physics! And even more that which *compels* us to it, – our honesty!¹⁷

With this, Nietzsche first concludes that his preceding claims, and particularly his immediately preceding denial of practical reason, dispose of the practice of making moral judgements about reasons for action, which is his critical object throughout the section. He here expresses this practice in Kantian terms as the practice of judging ‘the “moral worth of our actions”’. Instead, Nietzsche proposes, we ought to create ‘our own new tables of what is good’, and thus

¹⁷ FW 335.

become 'the new, the unique, the incomparable, those who give themselves laws, those who create themselves'. These remarks undeniably suggest that Nietzsche endorses a conception of 'autonomy' as an individual's singular and creative 'self-legislation', and considers the practice of, or obedience to, Kantian moral judgement to betray a failure to achieve such 'autonomy'. However, these remarks provide no more specification of the conception than this, and Nietzsche's immediately preceding and succeeding claims serve to problematise the specifications standardly made of it. That is, they problematise the standard claim that Nietzsche endorses a conception of 'autonomy' as an individual's choosing, or choosing according to, her particularity or singularity, defined in terms of her drives or desires and to be created, discovered, selected, or organised by the individual herself.

Firstly, although Nietzsche expresses his proposal as the 'creation of our own new tables of what is good' and thus echoes his preceding reference to reasons as 'tables of what is good', he is presumably not proposing that an agent act according to particular or singular reasons, and thus 'give' these reasons to herself. For in the immediately preceding passage he denies that an agent is motivated by reasons at all, on the grounds that an action and its circumstances are singular, while reasons are general. He concludes from this not just that we ought not to make Kantian moral judgements, but that we ought not to make any moral judgements about reasons for action at all. It would therefore be surprising, to say the least, if he should proceed to articulate an ideal in terms of reasons for action, and particularly if he should do so in terms of singular reasons.

It seems likely, then, that the 'tables of what is good' and 'laws' concerned refer to something other than reasons. They might refer to desires or drives, say, or

other noncognitive states, and an agent might be supposed to 'give' or 'create' the required 'tables' or 'laws' by being motivated by, or otherwise expressing, particular or singular noncognitive states in action. Alternatively, the 'tables of what is good' and 'laws' might refer to other kinds of normative considerations regarding behaviour – the 'opinions of "good", "noble", "great"' to which Nietzsche also refers, for instance – which an agent is supposed to 'give' or 'create' particularly or singularly for herself, and to express or fulfil with her behaviour. However, such interpretative speculations not only founder on Nietzsche's failure to specify the referents of these crucial terms. They also sit uncomfortably with the other conclusion which he draws in the immediately preceding passage. For in that passage Nietzsche argues that the inevitable singularity of any particular action entails not only that it could not be motivated by reasons, but also that its motivations are undeterminable. In the second chapter, I suggested that this conclusion applies as much to desires, drives, or other noncognitive states as to reasons, since it is effectively a preclusion of any causal 'description' of an action and its circumstances. It is therefore unlikely that Nietzsche would proceed to propose that actions be judged according to any causal 'descriptions' of them at all, whether these 'descriptions' would concern their motivation by reasons or noncognitive states, their expression of such states, or their fulfilment of other normative terms.

Equally puzzling in this regard, although it might give some hope to those who present him as urging the affirmation of the 'this-worldly', is Nietzsche's further insistence that, in order to become 'the new, the unique, the incomparable', we must also become good 'physicists', in the sense of 'students and discoverers of everything lawful and necessary in the world'. For if, as I suggested,

Nietzsche's conclusion that the motivations of any particular action are indeterminable is effectively a preclusion of causal 'descriptions' of actions and their circumstances according to observed regularities, then he could not consistently include actions and their circumstances within 'everything lawful and necessary in the world'. Indeed, in the concluding part of the section he asks precisely that we no longer 'drag the past some small way further through time', as the provision of causal 'descriptions' of actions and circumstances would have us do. This leaves profoundly mysterious the supposed practical value of studying and discovering 'everything lawful and necessary in the world'.

Arguably, however, these difficulties lend support to the modification of Nietzsche's argument against practical reason which I suggested in discussing that argument. According to this modification, in order to deny practical reason without precluding his own provision of alternative causal 'descriptions' of actions, Nietzsche need be committed to the mere particularity, rather than singularity, of an action and its circumstances, and hold simply that noncognitive states and beliefs capture this particularity, while reasons do not. If this were his position, then he might also reasonably include actions and their circumstances among the possible objects of causal 'descriptions', and therefore consider some such 'descriptions' to be of practical value. Furthermore, insofar as such 'descriptions' could be considered to capture the relevant variables, they might be employed for the sake of the judgement of actions. In particular, such a modification might admit a conception of 'autonomy' as the relative or maximal particularity, but not singularity, of the noncognitive states which motivate an agent's actions, or are otherwise expressed in them, or as the fulfilment of other particular, but not

singular, normative terms. Again, however, it should be emphasised that in '*Long live physics!*' Nietzsche manifestly fails to offer such a modified argument.

III.

Nietzsche's most extensive single discussion of Kant's moral philosophy, and the second of the two passages in which he rejects Kant's metaphysics and offers the critical reflection which he considers it to obstruct, is found in a series of three sections in *The Antichrist*, notably composed in the year after he apparently read a number of Kant's works on moral philosophy in the library at Chur. This discussion begins with Nietzsche once again complaining that, as he expresses it here, Kant's 'crafty-clever scepticism' serves to ensure that an epistemologically inaccessible "*true world*" to which moral entities and qualities can be attributed is 'if not demonstrable yet no longer *refutable*'. With this, Nietzsche unsurprisingly insists, one simply inverts the order of 'reality'. However, he also proceeds to object to Kant's conception of moral goodness, with '[a] word against Kant as *moralist*'.¹⁸ He expresses his objections, somewhat repetitively, as follows.

A virtue must be *our* invention, *our* most personal self-defence and need: in any other sense it is merely a danger. What does not condition our life *harms* it: a virtue merely from a feeling of respect for the concept 'virtue', as Kant wanted it, is harmful. 'Virtue', 'duty', 'good in itself', good with the character of impersonality and universality – phantoms, in which decline, the final debilitation of life,

¹⁸ A 10, 11. See also A 17.

Königsbergian Chinadom expresses itself. The profoundest laws of preservation and growth demand the reverse: that each should invent *his* virtue, *his* categorical imperative. A people perishes if it confuses *its* duty with the concept of duty in general. [...] Kant's categorical imperative should have been felt as *dangerous to life!* [...] An action compelled by the instinct of life has in pleasure its proof of being a *right* action: and that nihilist with Christian-dogmatic bowels understands pleasure as [an] *objection* ... What destroys more quickly than to work, to think, to feel without inner necessity, without a deep personal choice, without *pleasure*? as an automaton of 'duty'? It is virtually the *recipe* for *décadence*, even for idiocy ... [...] Did Kant not see in the French Revolution the transition from the inorganic form of the state to the *organic*? Did he not ask himself whether there was an occurrence which could not be explained other than by a moral predisposition of humanity, so that with it the 'tendency of humanity towards the good' would be *proved* once and for all? Kant's answer: 'The Revolution is that.' The erring instinct in all and everything, anti-naturalness as instinct, German *décadence* as philosophy – *that is Kant!*¹⁹

This passage reveals two important developments in Nietzsche's understanding of Kant's conception of moral goodness, with respect to the earlier discussion in '*Long live physics!*'. Firstly, this passage presents a Kantian moral judgement not as a judgement which holds that every agent should do or refrain from a type of action in a type of circumstance, but as a judgement made according to the "good in itself", good with the character of impersonality and universality'. Nor does the passage suggest that such judgements are supposed to be made of specific actions in specific circumstances. Secondly, this passage attributes to Kant himself an account of the motivation for obedience to such a

¹⁹ A 11.

judgement, as ‘merely [...] a feeling of respect for the concept “virtue”’. Nor is Nietzsche’s objection that of ‘*Long live physics!*’, namely, that the making or obeying of a Kantian moral judgement betrays ‘a blind, petty, and undemanding selfishness’. Rather, Nietzsche’s objection here is that to make or obey a Kantian moral judgement is ‘harmful’. In particular, he maintains that Kant ‘understands pleasure as [an] objection’ in the making or obeying of a moral judgement, and that thus ‘to work, to think, to feel without inner necessity, without a deep personal choice, without *pleasure*’ is to ‘destroy’ oneself, to court ‘*décadence*’ and ‘*idiocy*’.²⁰ As in ‘*Long live physics!*’, however, Nietzsche articulates his counter-proposal in the terms of an individualistic conception of ‘autonomy’: ‘each should invent *his* virtue, *his* categorical imperative’, he writes, and thus serve ‘the profoundest laws of preservation and growth’. In conclusion, and presumably in allusion to the second part of Kant’s *The Conflict of the Faculties*, Nietzsche refers dismissively to Kant’s alleged claim that the French Revolution ‘*proved*’, or can be ‘explained’ only by, a ‘moral predisposition’ or ““tendency [...] towards the good”” on the part of humanity.²¹

Nietzsche hardly elaborates here on how an agent should ‘invent *his* virtue, *his* categorical imperative’, or should judge and act from an ‘inner necessity’ or ‘deep personal choice’. With this lack of specificity regarding the terms of an individualistic conception of ‘autonomy’, this passage echoes ‘*Long live physics!*’. However, here Nietzsche also articulates the ‘harm’ and ‘preservation and growth’ with which he is concerned in terms of ‘life’, ‘nihilism’, ‘*décadence*’, and ‘*anti-naturalness*’. He thus refers to his broader concerns in *The Antichrist*, and provides some context for his affirmation of individual ‘autonomy’. In the opening sections,

²⁰ Nietzsche remarks similarly on Kant’s supposed ‘objection’ to pleasure at *M* 339, *GM* II 6, *GD* IX 49, and *KGW* IV:1 9 [1] (Summer 1875), and VIII:2 10 [11] (Autumn 1887).

²¹ Compare, in particular, *SF* 84-6.

he writes that his ‘problem’ in this text concerns ‘which type of human being one should *breed*, should *will*, as the more valuable, more worthy of life, more certain of the future’. In this light, he defines ‘life’ as ‘instinct for growth, for persistence, for accumulation of forces, for *power*’, or ‘will to power’, and ‘*décadence*’ or ‘nihilism’ as the absence of such an ‘instinct’ or ‘will’ in values.²² By criticising Kant in terms of ‘life’, ‘nihilism’, and ‘*décadence*’, then, Nietzsche is pursuing that element of a ‘revaluation of values’ which consists of evaluating professed obedience to moral reasons by ‘humanity’, a concern whose particularity he once again emphasises by using ‘I’ and ‘we’.²³ His claim that Kant’s conception expresses ‘anti-naturalness’, on the other hand, is illuminated by one of the following sections, in which Nietzsche defines ‘*hatred of the natural* (– actuality! –)’ in terms which identify it with the ‘*ressentiment*’ which he claims motivates obedience to ‘ascetic ideals’ in the third essay of the *Genealogy*. In particular, he writes that the Christian belief in another, ‘*pure world of fiction*’ is ‘the expression of a profound discontent with the actual’.²⁴

IV.

For the most part, Nietzsche’s other explicit critical remarks regarding Kant’s moral philosophy also place Kant’s moral philosophy in the context of Nietzsche’s ‘revaluation of values’, as I have presented this above. Thus, as I have

²² A 3, 6. See also A 36-8.

²³ For this emphasis, see, in particular, A V, 1, 3, 6, 8, 13, and 62. In criticising Kant’s conception of moral judgement at MA 25, Nietzsche’s primary concern is also with ‘humanity’s needs’, and particularly ‘*the conditions of culture*’. See also KGW V:1 6 [148] (Autumn 1880).

²⁴ A 15. See also A 16 and 18. Nietzsche also refers to Kant’s conception of moral goodness as ‘ascetic’ at M 339, GD IX 49, and A 20, and as an expression of ‘*anti-nature*’ at EH IV 7.

already mentioned, in the opening sections of ‘On the natural history of morality’ in *Beyond Good and Evil* Nietzsche considers Kant to exemplify moral philosophers’ misguided pursuit of an unconditionally and overridingly rational ‘ground’ for the judgement of reasons for action. Here and elsewhere, Nietzsche insists that the value of Kant’s moral philosophy instead lies merely in its articulation of certain moral conceptions which prevailed in Kant’s particular environment, a fraction of the wealth of such conceptions which Nietzsche insists must be articulated and subjected to ‘revaluation’.²⁵ In the third essay of the *Genealogy*, Nietzsche is a little more specific, with his remarks that Kant’s invoking of ‘other-worldly’ entities and qualities expresses ‘something’ of a typically ascetic epistemology, and is implicated in the ascetic commitment of some modern scientists to the unconditional value of truth.²⁶ In the last of the three sections in *The Antichrist*, Nietzsche also alleges that, in conceiving of moral goodness as obedience to an unconditional and overriding reason, Kant uncritically echoes the ascetic belief in an ‘imperative of the beyond’. In particular, Nietzsche writes, ‘with the concept “practical reason”’: [...Kant] invented a reason specifically for the case in which one need not concern oneself with reason, namely, when morality, when the lofty demand “you should”, becomes known’.²⁷ Thus Nietzsche clearly commits himself to the particularity of normative authority, and considers Kant to contravene this in a quasi-‘ascetic’ manner.

Finally, Nietzsche also makes a few passing references to Kant in criticising the common sense claim regarding choice. In particular, he twice

²⁵ Besides *JGB* 186-8, see *M* 207, *FW* 193, *JGB* 11 and 210-12, and *KGW* IV:3 30 [188] (Summer 1878), VII:2 25 [437] (Early 1884), VII:2 25 [351] (Spring 1884), VII:3 34 [36, 79, and 82] (April-June 1885), VIII:1 7 [4] (End 1886-Spring 1887), VIII:2 10 [11] (Autumn 1887), and VIII:3 14 [116] (Spring 1888).

²⁶ See *GM* III 12 and 25, respectively.

²⁷ A 12. Nietzsche makes related remarks at *FW* 5 and *JGB* 220.

includes Kant's conception of the 'thing in itself' among the theoretical entities erroneously posited under the influence of subject-predicate grammar.²⁸ More pertinently for the criticism of Kant's moral philosophy, Nietzsche parenthetically remarks on Kant in the sections under the title '*The error of free will*' in the chapter, 'The four great errors', in *Twilight of the Idols*. There Nietzsche accuses Kant of propounding the conception of choice as 'free will', as follows: 'No one gives a human being his qualities, neither God, nor society, nor his parents and ancestors, nor *he himself* (– the nonsense of the idea last rejected here was taught as "intelligible freedom" by Kant [...]).'²⁹ Thus Nietzsche concisely invokes against Kant, firstly, the 'nonsense' of 'free will', which Nietzsche elsewhere identifies as the incoherence of conceiving of choice as both a cause and the effect of that cause; secondly, the implication of 'free will' in an over-inflated sense of responsibility, according to which an agent could be responsible for her 'qualities' and which Nietzsche rejects in '*The error of free will*' and elsewhere; and, thirdly, the objections which Nietzsche raises to the common sense claim regarding choice, of which he considers 'free will' to be a particular instance.

²⁸ See *GM I* 13, and *GD VI* 3.

²⁹ *GD VI* 8. In 1885 Nietzsche also made a corresponding modification to *MA* 39 in his own copy. See Handwerk's translation, p.325.

Conclusions

The standard interpretations of the relationship between Kant's and Nietzsche's moral philosophies therefore finds little support in Nietzsche's explicit critical remarks regarding Kant's moral philosophy. Nietzsche certainly objects that Kant's sceptical epistemology would obstruct critical reflection on his moral philosophy, insists on the particularity of normative authority, and in important passages proposes an individualistic conception of 'autonomy' as an alternative to Kant. However, he neither restricts his critical reflection on Kant to his insistence on the particularity of normative authority nor specifies his conception of 'autonomy'. Indeed, the standard specifications of his conception of 'autonomy' sit uncomfortably with his critical treatment of the common sense conception of agency, particularly insofar as they tend to attribute to Nietzsche the common sense claim regarding reason.

Besides these points of contact with the standard interpretations, Nietzsche's explicit remarks regarding Kant's moral philosophy also allege that it is committed to the mistaken common sense claims regarding reasons and choice, and to 'free will' and the associated, over-inflated sense of responsibility in particular. Nietzsche also offers two critical accounts of making and obeying a Kantian moral judgement, both of which are accompanied by affirmations of individual 'autonomy'. Firstly, in *'Long live physics!'*, he conceives of a Kantian moral judgement as a judgement of specific actions in specific circumstances, according to which every agent should do or refrain from a type of action in a type of circumstance. There he objects that such a judgement can be made or obeyed

only on 'selfish' grounds. Later, in a section of *The Antichrist*, he conceives of a Kantian moral judgement differently, as a judgement made on 'universal' grounds, and objects that to make or obey such a judgement is, particularly in its exclusion of 'pleasure', self-destructive, characteristic of 'ascetic ideals', and fails to satisfy Nietzsche's particular evaluative concern for 'humanity'. There, in passing, he also dismisses Kant's alleged claim that the French Revolution reveals 'the "tendency of humanity towards the good"'.

Whatever its textual credentials, and like Nietzsche's critical treatment of the common sense conception of agency, this account of his explicit criticisms of Kant might appear deflationary to some Nietzsche partisans. Once again, then, it is worth emphasising that in the following part of the thesis I will attempt to show that the prevailing approaches to Nietzsche's criticism of moral agency and his critical relation to Kant in fact have little bearing on Kant's own concerns, while my account credits Nietzsche with, if not conclusive criticisms of Kant, then at least some concerns that engage with Kant's own.

Conclusions of part I

In this part of the thesis, I have argued that Nietzsche's explicit criticisms of Kant's conception of the moral agent are strongly informed, although not exhausted, by Nietzsche's critical treatment of a certain common sense conception of agency, and that neither these criticisms nor this critical treatment are well represented by the prevailing literature.

I began with Nietzsche's critical treatment of a common sense conception of agency, defined by two claims. The first claim is that an agent's action is sufficiently explained by the agent's 'willing' of it, in the sense that the action is, given the circumstances, causally necessitated in the agent's conscious choice of it. In the first chapter, I demonstrated that Nietzsche objects to this claim on the grounds that, as a causal account of choice, it founders on the unavailability of experience of necessitating causal powers or, if not, at least on a failure to appreciate other significant causes of action in the experience of regular correlations between pertinent kinds of events. With regard to two particular instances of the claim, Nietzsche also insists that no substantial agent, or 'I', is part of choice, and that to conceive of choice as undetermined by antecedent causes is to conceive of it incoherently, as both a cause and the effect of that cause, and to underwrite an over-inflated sense of responsibility.

The common sense conception's second claim is that an agent consciously chooses, or 'wills', an action for normative reasons, in the sense that her choice is motivated by her cognitive acknowledgement of reasons. In the second chapter, I demonstrated that Nietzsche's objection to this claim is that since a reason

prescribes a general type of action in a general type of circumstances, while an action and its circumstances are singular, a reason prescriptively underdetermines action, and therefore could not motivate an agent to act. Nietzsche consistently accounts for actions as motivated not by reasons, but by noncognitive states informed by reasons, and thus presumably concludes that if actions are not motivated by reasons, then, by elimination, they are motivated by noncognitive states informed by reasons. Given that his accounts are causal accounts resting on investigations of regular correlations between types of events, however, I suggested that the premise of his objection ought to be reduced to the particularity, rather than singularity, of an action and its circumstances. Beyond this objection, I demonstrated that Nietzsche considers his denial of practical reason to undermine the practice of making moral judgements about reasons for action, and to reveal the incoherence of supposed unconditional and overriding reasons, such as the rational 'grounds' for moral judgement which moral philosophy has hitherto offered. In the place of such moral philosophy, Nietzsche proposes the 'revaluation' of obedience to moral reasons, by means of a symptomatology of noncognitive motivating states and an evaluation of such obedience according to his own particular concern for 'humanity'.

In the third chapter, I demonstrated that Nietzsche's critical treatment of the common sense conception of agency is not well represented by prevailing strands in the literature. In particular, I argued that this critical treatment displays no concern to account for or evaluate agency in terms of unconscious, 'power'-orientated drives, in terms of motivating states or natural constitutions, or in terms of the agent's 'self-creation' or 'self-determination'. I also noted that this

treatment does not deny responsibility as such, as Nietzsche is occasionally supposed to do.

In the final chapter of this part of the thesis, I turned to Nietzsche's explicit criticisms of Kant's moral philosophy. With these criticisms, Nietzsche confirms standard interpretative claims that he objects to Kant's sceptical epistemology as an obstruction to critical reflection on his moral philosophy, insists on the particularity of normative authority, and proposes an individualistic conception of 'autonomy' as an alternative to Kant. However, Nietzsche also neither restricts his critical reflection on Kant to his insistence on the particularity of normative authority nor specifies his conception of 'autonomy'. Indeed, I noted that the standard specifications of his conception of 'autonomy' tend to attribute to Nietzsche precisely the common sense claim regarding reason which he rejects.

Rather than the standard interpretative claims, I argued, Nietzsche's explicit criticisms of Kant's moral philosophy are strongly informed by his critical treatment of the common sense conception of agency. Thus Nietzsche accuses Kant of endorsing both of the common sense claims which constitute this conception, along with 'free will' and the associated, over-inflated sense of responsibility in particular. Furthermore, Nietzsche offers two critical accounts of making and obeying a Kantian moral judgement, both of which are accompanied by affirmations of individual 'autonomy'. Firstly, in *'Long live physics!'*, he conceives of a Kantian moral judgement as a judgement of specific actions in specific circumstances, according to which every agent should do or refrain from a type of action in a type of circumstance. There he objects that such a judgement can be made or obeyed only on 'selfish' grounds. Later, in a section of *The Antichrist*, he conceives of a Kantian moral judgement differently, as a judgement

made on 'universal' grounds, and objects that to make or obey such a judgement is, particularly in its exclusion of 'pleasure', self-destructive, characteristic of 'ascetic ideals', and fails to satisfy Nietzsche's particular evaluative concern for 'humanity'. There, in passing, he also dismisses Kant's alleged claim that the French Revolution reveals 'the "tendency of humanity towards the good"'.

These, then, are the criticisms of Kant which I will evaluate in the following part of the thesis.

Part II

The Good Will: Kant's Analysis of the Moral Agent

Introduction to part II

This part of the thesis defends Kant's conception of the moral agent against the criticisms identified in the first part. In other words, it defends the second broad conclusion of the thesis, that Kant's conception of the moral agent is not undermined by the criticisms which Nietzsche explicitly levels at it, or by others which are commonly made in Nietzsche's name.

My defence of Kant's conception of the moral agent begins with an interpretation of his derivation of the basic features of this conception in the first and second sections of the *Groundwork of the Metaphysics of Morals*. This interpretation also reveals, in particular, his underappreciated method and the cogency of his derivations of formulas. On the basis of this interpretation, the sixth chapter then defends Kant's formulas and their place in his conception of moral judgement against Nietzsche's criticisms and those commonly made in his name. The most telling of these criticisms, I argue, are Nietzsche's criticisms of the common sense conception of agency, since, while they do not undermine Kant's conception of the moral agent, these criticisms require certain accommodations on Kant's part. The seventh chapter then argues that Nietzsche's criticisms of Kant's insistence on spontaneity in choice are misplaced, but that Kant's arguments for this spontaneity are equally unconvincing. Finally, the eighth chapter demonstrates that, despite Nietzsche's allegation to the contrary, Kant neither claims that there is evidence of human progress nor conceives of human progress in moral terms.

The argument of the *Groundwork*, sections I and II

The first section of Kant's *Groundwork of the Metaphysics of Morals* is often treated with suspicion. Even sympathetic commentators often consider the grounds and progress of its argument to be obscure, and allege that its infamous discussion of 'duty', its presentation of three 'propositions', and its derivation of the formula of universal law are faulty or inadequate. In the light of this, commentators generally avoid detailed textual examination of the section by offering creative interpretations of isolated claims or passages, or by simply noting the section's broadest claims and proceeding swiftly to the second section, which is generally considered to provide a more substantial argument and to refer, at most, only to the broadest claims of the first. This chapter takes issue with this common interpretative and critical approach. In particular, it argues that the first section of the *Groundwork* provides a sophisticated argument which is not vulnerable to the criticisms which are persistently levelled at it, and that the argument of the second section proceeds from, and depends upon, that of the first. The following chapters undertake to show, further, that the argument of the first and second sections of the *Groundwork* substantially informs other branches of Kant's moral philosophy, and his conception of the moral agent in particular.

This chapter is divided into three sections. The first two sections concern the first and second sections of the *Groundwork*, respectively, while the final section considers how Kant's argument illuminates his method.

From a good will to the formula of universal law

Kant's argument in the first section of the *Groundwork* is a little complicated, and his presentation of it is not always clear. Before considering the details of the argument and Kant's presentation of it, then, it is worth outlining the argument schematically. As the schematic outline on the following page shows, Kant's argument proceeds from a conception of the goodness of a good will to two distinctive features of moral goodness, which the formula of universal law is intended to express.

Kant's argument in the first section of the *Groundwork*

Kant's conception of the goodness of a good will

According to Kant's conception of the goodness of a good will, a good will is:

(a) the only good which is good in all contexts,

and,

(b) a condition of the goodness of every other good.

Kant draws two conclusions from this conception

From (a), Kant concludes:

(a') a good will does what is morally good because it is morally good ('from duty').

From (b), Kant concludes:

(b') moral goodness is a goodness which only a will can achieve.

On the basis of these conclusions, Kant identifies two distinctive features of moral goodness

From (b'), Kant identifies the first distinctive feature of moral goodness:

(i) moral goodness is a goodness of a will's reason for action alone (the 'second proposition').

(Note that (i) does not follow from (a'), and Kant does not hold that it does.)

From (a') and (i), Kant identifies the second distinctive feature of moral goodness:

(ii) moral goodness is good for a will as such (the 'third proposition').

Kant then derives the formula of universal law as an expression of these two features

Kant derives the formula of universal law as an expression of features (i) and (ii), but intends this formula to emphasise only feature (ii).

I. Kant's analysis of the goodness of a good will

Kant's argument in the first section of the *Groundwork* begins with two basic claims regarding a good will, which he makes in the opening two paragraphs. These two basic claims are identified as (a) and (b) in the outline above. The first claim is famously expressed in the first sentence: 'It is not possible to think of anything in the world, or even out of it, which could be considered good without limitation, except a **good will**'. Kant's following remarks regarding other goods indicate that by 'good without limitation', he means 'good in all contexts'. Indeed, towards the end of the second section, he states precisely that a good will 'cannot be evil'. The second claim, as Kant expresses it a few pages into the first section, is that a good will is 'not [...] the sole good and the complete good, but [...] the highest good and the condition of every other'. In the opening paragraphs, he makes this claim by referring to other goods not only as goods which in some contexts are not good, but also as goods which can be good only in the context of a good will.¹

According to the opening two paragraphs of the first section, then, a good will is (a) the only good which is good in all contexts, and (b) a condition of the goodness of every other good. Kant attributes this conception of the goodness of a good will to 'common moral rational cognition', according to which, he writes, 'good and desirable' things other than a good will 'can also be extremely evil and harmful, if the will which is to make use of [...them] is not good'.² It is important to note, however, that the two basic claims, (a) and (b), are logically distinct. Kant occasionally refers to a good will as 'unconditionally' good, and it is perhaps

¹ G 393, 437, 396. See also G 394 and 454-5, KpV 62, and KU 443.

² G 393. See also KU 443.

tempting to suppose that what is ‘unconditionally’ good in the sense of being ‘good in all contexts’ must also be ‘unconditionally’ good in the sense of being ‘a condition of the goodness of every other good’, and vice versa. However, there is no such logical entailment between the two claims, or between the two senses of ‘unconditional’ goodness, and Kant’s references to a good will’s ‘unconditional’ goodness are better understood as a succinct way of referring to the two, logically distinct claims.³

Kant begins his analysis of these two claims with a particular analysis of the first, (a). From this analysis he draws conclusion (a’), that a good will is one which does what is morally good ‘from duty’. That is, he concludes that a good will does what is morally good precisely because what is morally good is morally good, rather than because it happens to coincide with the satisfaction of inclination, and so irrespective of whether it does so coincide. Kant’s well-known examples are intended to reveal this, by revealing that what a will is inclined to do is only contingently related to what it is morally good to do, and that the moral goodness of a will which, according to claim (a), is good in all contexts therefore cannot be conditional upon its inclinations.

As Kant presents it, this analysis proceeds by employing the concept of ‘duty’ to ‘explicate [or expand, *entwickeln*]’ the concept of a good will. This is so, Kant insists, because ‘the concept of **duty** [...] contains that of a good will, although under certain subjective limitations and hindrances, which, far from concealing it and making it unrecognisable, rather bring it out by contrast and let it shine forth all the more brightly’. Kant’s examples – preserving one’s life in the face of adversity and grief, being beneficent to others despite one’s indifference to

³ For Kant’s references to the ‘unconditional’ goodness of a good will, see *G* 394, 401, and 437. Note that in later texts, he clearly distinguishes the two claims. See *KpV* 62 and *R* 3.

their suffering or one's preoccupation with one's own, and, in the case of the gout sufferer, refusing the temptation to secure more determinate and certain satisfactions than those of one's health – indicate that the 'subjective limitations and hindrances' to which he refers are provided by inclinations, and reasons for action which concern them, insofar as they *fail* to coincide with doing what is morally good. This has led critics since Schiller to suppose that Kant adopts the objectionable position that a will can act 'from duty' *only* against such inclinations.⁴ However, Kant's concern in this analysis is simply to reveal that an inclination, and so any reason for action which concerns an inclination, is only contingently related to what it is morally good to do. As Kant expresses it, an inclination, and so any reason which concerns one, 'lacks moral content', and so either 'fortunately hits upon' what is morally good or unfortunately provides 'subjective limitations and hindrances' to doing what is morally good. In contrast, Kant holds that to act 'from duty' is to act precisely because the action is morally good, and thus for a reason which is necessarily related to such action. He concludes from this that the goodness of a will which is good in all contexts must 'shine forth' only in action 'from duty'. Kant thus does not hold that acting 'from duty' presupposes the 'subjective limitations and hindrances' of countervailing inclinations. Rather, he holds simply that acting 'from duty' is acting because the action is morally good, irrespective of whether the action also coincides with the satisfaction of inclination.⁵

However, as he explains before considering his examples, Kant does hold that it is 'difficult to notice' whether a will which does what is morally good acts

⁴ For early statements of this persistent criticism of Kant, see, for instance, Schiller, *Xenien*, 'Die Philosophen', and Schopenhauer, *Über das Fundament der Moral*, pt.II, §6, and *Die Welt als Wille und Vorstellung*, v. 1, App..

⁵ *G* 397, 398. See also *G* 390 and 399, *KrV* B28-9, *KpV* 93, and *TP* 278-9.

‘from’ an inclination or ‘from duty’ in a certain kind of case – namely, when an inclination succeeds in ‘hitting upon’ what is morally good and, unlike the inclination of the ‘prudent’ shopkeeper not to exploit inexperienced customers, does so with regard to effects which are ‘immediately’ related to the action.⁶ Kant insists that the goodness of a good will therefore ‘shines forth all the more brightly’ when no inclination ‘hits upon’ what is morally good than it does when inclinations ‘immediately’ coincide with what is morally good. It is this, and not an objectionable notion of action ‘from duty’, which explains Kant’s insistence on the presence of ‘subjective limitations and hindrances’ in his examples.⁷

Kant’s analysis of his first claim regarding the goodness of a good will thus provides his first substantial conclusion, (a’), that a good will is one which does what is morally good precisely because it is morally good, rather than ‘from’ a concern for inclination. However, this conclusion alone can establish nothing about what first makes an action morally good to do, a ‘duty’ – that is, this conclusion alone can establish nothing about moral goodness itself. For a will can do what is morally good ‘from duty’ according to *any* account of what first makes the action morally good to do. For instance, a will might do ‘from duty’ what is determined to be morally good by a consequentialist theory, or a divine command

⁶ *G* 397. See also *G* 425, *TP* 278-9, and *MS* 382n. At *G* 397, Kant insists that in the presence of a coinciding inclination which is not ‘immediately’ related to the action, such as the prudent shopkeeper’s inclination to his own ‘advantage’, it is ‘easy to distinguish’ between action ‘from’ this inclination and action ‘from duty’. He contrasts this inclination of the shopkeeper with the possibility of the shopkeeper having ‘an immediate inclination towards the customers, in order from love, as it were, to give no one preference over another in price’.

⁷ I take it that the passages to which I have referred also discredit Latham’s and Wood’s recent defences of the notion of action ‘from duty’ which Kant’s critics have persistently found objectionable. In particular, these passages render problematic the response which Latham and Wood offer to these critics, namely, that Kant admits that a will which does what is morally good ‘from’ a coincident inclination, rather than ‘from duty’, can nonetheless be a good one. See Latham, ‘Causally Irrelevant Reasons and Acting Solely from the Motive of Duty’, pp.613-7, and Wood, *Kant’s Ethical Thought*, pp.26-35, 120, and 133. I also take it that these passages discredit the now well-discussed claim that Kant allows actions ‘from duty’ to be ‘overdetermined’ in the sense that, although ‘duty’ is ‘sufficient’ for action, the will also acts ‘from’ inclination. See, in particular, Henson, ‘What Kant Might Have Said’, and, for a thorough critical discussion, Baron, *Kantian Ethics Almost Without Apology*, pp.146-87.

theory. It is therefore mistaken to suggest, as some commentators' remarks do, that Kant's discussion of action 'from duty' in the first section already expresses the grounds from which he intends to derive a substantial position regarding moral goodness.⁸ Of course, this is not to say that *after* arguing for his substantial position, Kant does not then express it in terms of action 'from duty'.⁹

It is appropriate, then, that Kant rests the claims which he proceeds to make about moral goodness itself not only upon this first conclusion, (a'), drawn from his analysis of his first claim regarding the goodness of a good will, (a), but also upon his analysis of his second claim regarding a good will's goodness, (b). From this second claim (b), that a good will is a condition of the goodness of every other good, Kant draws the conclusion (b'), that moral goodness is a goodness which only a will can achieve. His argument for this conclusion is simply that if a good will is a condition of the goodness of every other good, then its goodness, which he equates with moral goodness, must be such that only a will can achieve it, since otherwise this condition could be fulfilled by other causes.

Kant presents this analysis of his second claim regarding the goodness of a good will in the paragraphs which follow his consideration of examples of action 'from duty' and which precede his first expression of the formula of universal law. However, these paragraphs are primarily concerned with presenting the two distinctive features of moral goodness which Kant proceeds to derive from the

⁸ See, for instance, Paton, *The Categorical Imperative*, pp.58, 61-2, and 69-71, Korsgaard, 'Kant's Analysis of Obligation', pp.47, 55, and 60-3, Reath, 'The Categorical Imperative and Kant's Conception of Practical Rationality', pp.389-90 and nn.10 and 13, Allison, *Kant's Theory of Freedom*, pp.120-1, and 'On a Presumed Gap in the Derivation of the Categorical Imperative', pp.144-5, and Baron, *Kantian Ethics Almost Without Apology*, pp.181-4. For criticism on the grounds mentioned in the text above, see Teale, *Kantian Ethics*, pp.98-101, Hill, 'Kant's Utopianism', p.69, Johnson, 'Expressing a Good Will', pp.161-6, Wood, *Kant's Ethical Thought*, p.48, and Gaut and Kerstein, 'The Derivation without the Gap', p.25. Nonetheless, it should be noted that those commentators whose remarks suggest such a questionable interpretation generally also suppose what I take to be a further, substantial element in Kant's argument, his identification of moral goodness as (i).

⁹ See, for instance, *G* 400, 401n, and 402.

conclusions of his two analyses. Indeed, his conclusion that moral goodness must be such that it can be achieved only by a will is made explicit only in the last of these paragraphs, in which he summarises the argument which is made in them. Before considering how Kant presents his analysis of his second claim regarding a good will's goodness, then, it is worth outlining his consequent derivation of two distinctive features of moral goodness.

Kant derives the first of the distinctive features of moral goodness, (i), from the conclusion which he draws from his analysis of his second claim regarding the goodness of a good will, (b'), that moral goodness is a goodness which only a will can achieve. He argues that if this is so, then moral goodness must be a goodness of a will's reason for action alone, rather than of actions' effects or of reasons for action insofar as they concern actions' effects. He calls this the 'second proposition' (the 'first proposition' is not explicitly identified), and he expresses it as the claim that 'an action from duty has its moral worth *not in the intention [Absicht]* to be achieved by it, but in the maxim according to which it is decided upon, [and] therefore depends not upon the actuality of the object of the action, but merely upon the *principle of willing* according to which the action happens without respect for any object of the ability of desire'. Kant then derives a second distinctive feature of moral goodness, (ii), from this 'second proposition' and his conclusion regarding a good will's action 'from duty'. Given these claims, namely (i) and (a'), Kant argues that moral goodness must also be good for a will as such, rather than good for a will only for satisfying certain of its inclinations. Kant calls this the 'third proposition', and expresses it as the claim that '*duty is the necessity of an action from respect for law*'. In presenting his derivation, he also refers to this 'law' as 'the mere law for itself', and expresses

this feature as the claim that what must ‘determine’ the will in acting ‘from duty’ is ‘the maxim of complying with such a law even with the demolition of all my inclinations’.¹⁰ By a ‘law’, then, Kant means a reason which concerns what is good for a will as such, rather than only for satisfying certain of its inclinations.

Before considering how Kant presents this argument, it is illuminating to compare it with Christine Korsgaard’s well-known interpretation. According to Korsgaard, in his discussion of action ‘from duty’, Kant identifies the reason why a good will performs a morally required action (or has a morally required purpose) with the reason why the action (or purpose) is morally required, and argues that, since the former concerns the reason’s ‘capacity to express a [normative] demand made on us’, the latter is also this capacity, which Korsgaard also calls a reason’s ‘legal character’.¹¹ For Korsgaard, it follows from this that the capacity of a reason to express a normative demand must derive from a quality of the reason itself, rather than from a source which is external to the reason. This is so, she maintains, because only in this way can this capacity itself make an action (or purpose) morally required, and a good will perform the action (or have the purpose) because a reason has this capacity. Finally, according to Korsgaard, Kant derives the formula of universal law on the grounds that the capacity to express a normative demand can derive from the quality of a reason itself only if the reason and its ‘becom[ing] a universal law’ can both be willed, which Korsgaard also calls the reason’s having ‘lawlike form’.¹²

Korsgaard’s interpretation bears some similarity to that presented above, particularly in emphasising a concern with identifying moral goodness as a goodness of a will’s reason for action alone – that is, a concern with the first

¹⁰ *G* 399-400, 400-1.

¹¹ Korsgaard, ‘Kant’s Analysis of Obligation’, p.61.

¹² *Ibid.*. See also pp.47, 55, and 60-5.

distinctive feature of moral goodness, (i). However, Korsgaard fails to provide adequate grounds for the first premise of the argument which she attributes to Kant, namely, his identification of a good will's reason for performing a morally required action (or having a morally required purpose) with the reason why the action (or purpose) is morally required. Korsgaard simply appeals to Kant's discussion of action 'from duty', and is therefore vulnerable to the criticism raised above, that a good will's reason for doing what is morally good can be that it is morally good, or 'duty', according to any account of why it is morally good. Korsgaard's interpretation thus leaves unexplained Kant's concern with identifying moral goodness as a goodness of a will's reason for action alone.¹³ Furthermore, her interpretation also attributes to Kant an inadequate derivation of the formula of universal law, since, as critics have noted, the requirement that a reason's normative demand derive from a quality of the reason itself is satisfied not only by the formula of universal law, but equally by other, weaker principles. For example, it is satisfied by the principle that any reason which is held to be required, reasonable, or permissible for a will in certain circumstances must be held to be required, reasonable, or permissible for any will in relevantly similar circumstances.¹⁴ Finally, Korsgaard's interpretation is a relatively creative one, providing little extended textual examination of the first section of the *Groundwork*. In particular, Korsgaard fails to relate her account to Kant's opening claims regarding a good will, except insofar as she supposes these claims

¹³ Cummiskey closely follows Korsgaard's interpretation of the first section of *G*, but Cummiskey also holds that Kant's identification of moral goodness as a goodness of a will's reason for action is a consequence of his presupposition of 'internalism', understood as the claim that moral requirements provide motivating reasons for action and overriding justifying reasons for action. However, Cummiskey treats this as a mere presupposition, and thus, like Korsgaard, does not explain why Kant should insist on this feature of moral goodness. See Cummiskey, *Kantian Consequentialism*, pp.25-35 and 164-9.

¹⁴ See Gaut and Kerstein, 'The Derivation without the Gap', pp.24-5.

to already express the first premise of the argument which she attributes to Kant. The interpretation presented above, and defended below, suggests that attending more carefully to Kant's opening claims and to how he presents his following argument provides an explanation for his identification of moral goodness as a goodness of a will's reason for action alone, and, as will be explained below, also a more satisfactory derivation of the formula of universal law.¹⁵

As mentioned above, it is only after having presented his derivation of two distinctive features of moral goodness, (i) and (ii), that Kant makes explicit his analysis of his second claim regarding the goodness of a good will, (b), and the role which the conclusion of this analysis, (b'), plays in his derivation. The paragraph which follows, and summarises, his presentation of this derivation reads as follows.

The moral worth of an action therefore does not lie in the effect expected from it, [and] therefore also not in any principle of action which needs to borrow its motive from this expected effect. For all these effects (agreeableness of one's condition, indeed even the promotion of others' happiness) could also have been

¹⁵ Besides Korsgaard, a few other commentators suggest aspects of the interpretation which I present, although none provides textual evidence for the suggestions or relates Kant's argument to his opening claims regarding the goodness of a good will. Robinson and Harré briefly suggest that, for Kant, since '[a]ny state of affairs achieved through human activity *qua* activity could be brought about in non-moral ways', the moral worth of an action must instead lie in the 'principle' of action and its being 'answerable to a moral law', which is 'possible [...] only for an entity able to comprehend justificatory principles' ('The Demography of the Kingdom of Ends', p.6). Guyer also interprets Kant's argument for the 'third proposition', and his general denial that inclination is the source of moral value, as based on the 'assumption that moral worth can attach only to what is an expression solely of the activity of the agent', an assumption which, Guyer holds, 'functions in the *Groundwork* as a first principle that cannot itself be argued for because it cannot be derived from anything more fundamental' ('Duty and Inclination', pp.345-6; see also pp.349-50). Engstrom also holds that, for Kant, the intrinsic goodness of a good will, along with the unconditional nature of 'duty', entails that 'the will's relation to what duty requires can be understood only as a relation to what is required by its own inner law', and Johnson briefly suggests that Kant's argument is based simply on the intrinsic goodness of a good will, such that this goodness entails that a good will's 'principle cannot be one that makes it only of extrinsic value – as it would be were the principle simply a recipe for producing this or that end' (Engstrom, 'Kant's Conception of Practical Wisdom', p.30, Johnson, 'Expressing a Good Will', p.165; see also Engstrom, 'Kant's Conception of Practical Wisdom', p.31).

achieved by other causes, and for this the will of a rational being, in which, nevertheless, the highest and unconditional good can alone be found, would therefore not have been necessary. This is why it can be nothing other than the *representation of the law* in itself, *which of course occurs only in a rational being*, insofar as it, not the hoped-for effect, is the determining ground of the will, which constitutes the preeminent good, which we call moral, [and] which is already present in the person himself who acts according to it, but may not first be expected from the effect.¹⁶

Kant thus derives the first distinctive feature of moral goodness, (i) – which he expresses in the first sentence and the last phrases of this paragraph – from the claim that a will must be ‘necessary’ for the achievement of moral goodness, in the sense that ‘other causes’ cannot achieve it. This claim is the conclusion, (b’), which Kant draws from his analysis of his second claim regarding the goodness of a good will, (b), that a good will is a condition of the goodness of every other good. This is indicated, slightly less clearly, by his statement in the second sentence that ‘the highest and unconditional good can alone be found’ in a will, a succinct statement of the two claims which comprise his conception of the goodness of a good will, (a) and (b).

This interpretation of Kant’s derivation of the first distinctive feature of moral goodness also illuminates the argument which he presents two paragraphs earlier, immediately after first expressing this feature as the ‘second proposition’. There he writes that the negative aspect of this feature, that moral goodness is not a goodness of actions’ effects or of reasons for action insofar as they concern these effects, ‘is clear from what has gone before’. He then argues that if this is so, then, by elimination, moral goodness ‘can lie nowhere else *than in the*

¹⁶ G 401.

principle of the will without respect for the ends which can be produced through such an action'.¹⁷ The grounds of this argument would be mysterious were it not for the conclusion which Kant draws from his claim (b) that a good will is a condition of the goodness of every other good, namely, the conclusion (b') that moral goodness is a goodness which only a will can achieve. This conclusion explains why Kant claims that moral goodness is not a goodness of actions' effects or of reasons for action insofar as they concern these effects, and why he writes that this 'is clear from what has gone before'. It also explains why he considers actions' effects, reasons for action insofar as they concern these effects, and reasons for action insofar as they do not concern these effects to exhaust the possible basic objects of moral goodness, since these possibilities exhaust the objects of goodness which a will can achieve. Finally, and most importantly, it explains why he insists that the basic object of moral goodness must be the last of these possibilities, since it is only the goodness of this object which *only* a will can achieve.

Kant's claim that the negative aspect of the first distinctive feature of moral goodness 'is clear from what has gone before' can therefore be taken simply to refer to the conception of the goodness of a good will with which he opens the section, and to his second claim regarding this goodness in particular. However, it might instead be taken to refer to the expression of, and argument for, the first distinctive feature of moral goodness which Kant provides immediately after presenting his conception of the goodness of a will at the beginning of the section. There, he expresses this first feature as follows: 'The good will is not good through what it effects or achieves, through its suitability for the attainment

¹⁷ G 400.

of some proposed end, but only through its willing, i.e. [it is] good in itself, and, regarded for itself, is to be valued incomparably higher than everything which through it could only be achieved in favour of some inclination, [and] indeed, if one will, the sum of all inclinations'.¹⁸ It is often supposed either that this claim is simply a further claim which Kant attributes to 'common moral rational cognition', without analytical relation to his preceding claims regarding the goodness of a good will, or that it follows from the first of these claims, on the grounds that what is good in all contexts simply is 'good in itself' in the sense which Kant identifies.¹⁹ However, Kant's following sentence indicates that, although he does understand his identification of the first feature of moral goodness to follow from his conception of the goodness of a good will, he does not understand it to follow from his claim that a good will is good in all contexts. This sentence reads, 'Even if, through a special disfavour of fate, or through meagre provision of a stepmotherly nature[,] this will would wholly lack in ability to carry out its intention; if with its greatest endeavours it would yet achieve nothing, and only the good will (of course not as a mere wish, but as the summoning of all means, insofar as they are in our power) remain left: then it would, like a jewel, still shine for itself, as something that has its full worth in itself'.²⁰ This sentence indicates that, for Kant, moral goodness is a goodness of a will's 'willing' alone, and not of actions' effects or of 'willing' insofar as it

¹⁸ G 394.

¹⁹ For the interpretation of Kant's claim at G 394 as simply a further claim attributed to 'common moral rational cognition', see, for example, Wood, *Kant's Ethical Thought*, pp.23-4, and Rawls, *Lectures on the History of Moral Philosophy*, pp.155-6. For the interpretation of it as derived from the claim that a good will is good in all contexts, see, for example, Paton, 'Analysis of the Argument', p.17, Potter, 'The Argument of Kant's *Groundwork*, Chapter 1', p.31, and Korsgaard, 'Two Distinctions in Goodness', pp.257-8, 'Kant's Formula of Humanity', p.117, and 'Kant's Analysis of Obligation', p.55. The latter interpretation is also implied, without explanation of the entailment, in Allison, *Kant's Theory of Freedom*, p.107, Höffe, *Immanuel Kant*, p.141, and Herman, 'Leaving Deontology Behind', p.213.

²⁰ G 394.

concerns these effects, because moral goodness cannot be a goodness which a will may lack the 'ability to carry out'. Although Kant again does not make the grounds of his argument explicit here, the implicit premise is that moral goodness must be a goodness which a will as such can achieve, a premise which is included in the conclusion which he draws from his second claim regarding the goodness of a good will, namely, the conclusion (b') that moral goodness is a goodness which only a will can achieve.²¹

Kant's derivation of the second distinctive feature of moral goodness, on the other hand, is not made clear by the summarising paragraph, which indicates only that his derivation rests upon the first distinctive feature and upon his conception of the goodness of a good will. However, his derivation is presented more clearly in the preceding paragraph, in which he expresses this feature as the 'third proposition'. There he again argues by elimination. He first appeals to the first distinctive feature of moral goodness, (i), which he here expresses in terms of the notion of 'respect'. That is, he states that there can be respect only for 'activity of a will', and not for what is 'merely an effect', and therefore only for 'that which is connected with my will merely as ground and never as effect'.²² He then appeals to his claim regarding action 'from duty', the claim (a') that a good will does what is morally good precisely because it is morally good, rather than because it coincides with the satisfaction of inclination. He concludes that if moral goodness lies in a will's reason for action alone and a good will's reason for doing what is morally good is not concerned with its inclinations, then 'there

²¹ This interpretation of Kant's argument at *G* 394 also suggests an interpretation of the argument which follows at *G* 394-6, and which is generally ignored or dismissed, since it appears to rest upon an unsupported and questionable teleological assumption. This 'teleological assumption' is better interpreted as (b'), I would suggest.

²² *G* 400. Kant makes similar remarks in the paragraph at *G* 399 which precedes his statement of the 'second proposition', and in the footnote attached to the paragraph at *G* 401. See also *E* 337-8.

remains nothing left for the will which could determine it, other than objectively the *law* and subjectively *pure respect* for this practical law, [and] therefore the maxim of complying with such a law even with the demolition of all my inclinations'.²³ That is, Kant concludes that (ii) moral goodness must be good for a will as such, a 'law' for it, because it must be a goodness of a will's reason for action alone and a good will's reason for doing what is morally good is not concerned with its inclinations, which are contingent for it.

This derivation is also indicated by Kant's statement that the 'third proposition' is 'a conclusion of the two gone before'.²⁴ Kant does not explicitly identify the first 'proposition', and his statement regarding the relationship between the propositions is often dismissed.²⁵ However, Kant's statement is accurate if, as I have suggested, his derivation of the second distinctive feature of moral goodness, (ii), as expressed by the 'third proposition', rests upon the first distinctive feature of moral goodness, (i), as expressed by the 'second proposition', and his claim (a') regarding action 'from duty', which is commonly supposed to be the 'first proposition'.²⁶

To summarise, then, Kant argues from his conception of the goodness of a good will to two distinctive features of moral goodness. He begins by analysing

²³ G 400.

²⁴ G 400.

²⁵ For the dismissal of Kant's statement that the 'third proposition' is 'a conclusion of the two gone before', on the grounds of differing interpretations of the first two 'propositions', see, for example, Paton, 'Analysis of the Argument', p.21, Teale, *Kantian Ethics*, p.101, and Wood, *Kant's Ethical Thought*, p.43.

²⁶ For this common supposition, see, for example, Paton, 'Analysis of the Argument', p.19, Teale, *Kantian Ethics*, p.98, Potter, 'The Argument of Kant's *Groundwork*, Chapter 1', p.31, Aune, *Kant's Theory of Morals*, p.9, Allison, 'On a Presumed Gap in the Derivation of the Categorical Imperative', p.144 and n.6, Guyer, 'The Strategy of Kant's *Groundwork*', p.216, and 'Kant on Common Sense and Scepticism', p.23, Wood, *Kant's Ethical Thought*, p.40, and Gaut and Kerstein, 'The Derivation without the Gap', p.27. However, my interpretation is also consistent with identifying the first 'proposition' as Kant's first claim regarding the goodness of a good will, (a), as expressed in the opening sentence of the section, since this is the basis of his claim regarding action 'from duty', (a'), and with identifying the first 'proposition' as Kant's conception of a good will's goodness as a whole, (a) and (b), since my interpretation treats this as the basis of all of his following argument.

his first claim regarding a good will's goodness, (a), that a good will is the only good which is good in all contexts, by means of examples of action 'from duty'. From this analysis, he concludes that (a') a good will does what is morally good precisely because it is morally good, and thus irrespective of whether what is morally good coincides with its inclinations. He then employs his analysis of his second claim regarding a good will's goodness, (b), that a good will is a condition of every other good, to derive the first distinctive feature of moral goodness, (i), as expressed by the 'second proposition'. That is, from his analysis of his second claim, he concludes that (b') moral goodness is a goodness which only a will can achieve, and from this conclusion, he derives the first distinctive feature of moral goodness, (i), that moral goodness is a goodness of a will's reason for action alone, rather than of actions' effects or of reasons for action insofar as they concern actions' effects. Then, by appealing to this first distinctive feature of moral goodness, (i), and to his conclusion regarding action 'from duty', (a'), Kant derives the second distinctive feature of moral goodness, (ii), as expressed by the 'third proposition'. That is, he argues that if moral goodness is a goodness of a will's reason for action alone and a good will's reason for doing what is morally good is not concerned with its inclinations, then moral goodness must be good for a will as such, a 'law' for it, rather than good for satisfying its inclinations, which are contingent to it.

Unlike Korsgaard's interpretation, therefore, this interpretation explains why Kant should insist that moral goodness is a goodness of a will's reason for action alone, and places this insistence in the context of his opening claims regarding a good will and the sophisticated argument which he bases upon these claims. As will now be explained, this interpretation also provides the grounds of

a persuasive account of Kant's derivation of the formula of universal law, a derivation which Korsgaard's interpretation, like many others, fails to make persuasive.

II. *The first derivation of the formula of universal law*

Kant first expresses the formula of universal law immediately after having derived the two distinctive features of moral goodness, (i) and (ii). He expresses the formula in a number of ways here: he writes that what 'should serve the will as its principle' is 'the conformity of actions as such with universal law [or the universal conformity with law of actions as such, *die allgemeine Gesetzmäßigkeit der Handlungen überhaupt*]' ; he then identifies this principle with the principle, 'I should never act other than in such a way *that I could also will that my maxim should become a universal law*'; and after considering the example of a false promise, he expresses the formula as the requirement that a maxim 'can fit [*passen*] as principle in a possible giving of universal law [or universal giving of law, *allgemeine Gesetzgebung*]' .²⁷

These brief remarks provide little indication of the intended derivation – or, indeed, the meaning – of the formula of universal law. However, Kant's surrounding remarks indicate that he derives the formula by combining the two distinguishing features of moral goodness which he has already identified, (i) and (ii). That is, he claims that the formula of universal law expresses what is required for a goodness to be a goodness of a will's reason for action alone and good for a

²⁷ G 402, 403.

will as such. He indicates this derivation in a number of ways, but he indicates its grounds most clearly by the question with which he introduces the formula, immediately after the paragraph which summarises his argument for the two distinctive features of moral goodness. There he asks, 'what kind of law can that be, the representation of which, even without regard for the effect expected from it, must determine the will, so that this [will] can be called good absolutely and without limitation?'.²⁸ With this, Kant indicates that his derivation rests upon the two claims which comprise his conception of the goodness of a good will: the claim (a) that only a good will is good in all contexts, or good 'without limitation', and the claim (b) that it is a condition of the goodness of every other good, from which he derives, via (b'), the first distinctive feature of moral goodness, (i), its being a goodness of a will's reason for action alone, such that a good will is good 'absolutely'.²⁹ By referring to 'law', he indicates that the formula also rests upon the second distinctive feature of moral goodness, (ii), as derived from the first such feature and the conclusion which he draws from his claim (a) that a good will is good in all contexts, namely, the conclusion (a'), that a good will does what is morally good 'from duty'. His appeal to this second feature is also indicated by his immediate answer to the question asked, which begins with another argument by elimination: 'Since I have deprived the will of every impulse which could arise for it from obeying some law, there remains nothing but the conformity of actions as such with universal law, which alone should serve the will as its principle'. After re-expressing the formula, he then remarks that this 'mere conformity to law as such' does not have 'at its basis some law determined for certain actions', and that

²⁸ G 402.

²⁹ For Kant's employment of the term 'absolute' to describe the goodness of a will which has this first distinctive feature of moral goodness, see his expression of this feature at G 394. There he also describes a will with this goodness as 'good in itself', and at G 403, he writes that a reason's being a possible 'universal law' is 'the condition of a will good *in itself*'.

this must be so, ‘if duty is not to be everywhere an empty illusion and chimerical concept’.³⁰

Besides referring to the grounds of his derivation of the formula in these ways, Kant indicates how he derives the formula from these grounds in his discussion of the example of a false promise. There he writes of the difference between promising truthfully because so promising is morally good and promising truthfully because so promising is prudent, ‘in the first case the concept of the action in itself already contains a law for me, in the second I must first look around elsewhere, [to see] which effects on me might be connected with it’.³¹ Thus Kant holds that the formula expresses a goodness for which one need not ‘look around elsewhere’ – that is, beyond a reason for action alone, to the action’s effects – for its goodness, and which ‘already contains a law’ for a will – that is, is good for a will as such, rather than for its inclinations, which are contingent for it. In other words, Kant holds that the formula of universal law expresses the two distinctive features of moral goodness which he derives from his conception of the goodness of a good will.

This interpretation of Kant’s derivation of the formula of universal law reveals that allegations of ‘gaps’ in this derivation are misplaced – that is, he does not undertake to derive this formula simply from the claim that a good will does what is morally good ‘from duty’, or simply from a claim that the moral ‘law’ must express a requirement which is necessary for a will as such. As mentioned above, the first undertaking could not succeed, for the simple reason that Kant’s claim regarding action ‘from duty’ cannot itself establish anything about what first makes an action morally good to do, a ‘duty’. It is fortunate, then, that this claim is

³⁰ *G* 402.

³¹ *G* 402. Kant uses similar terms at *R* 4 and 42.

simply a conclusion which Kant draws from one of his two claims regarding a good will's goodness, and that his derivation of the formula combines this claim with conclusions drawn from the other of these claims.

The more commonly alleged 'gap' is the second, however. As noted above, a 'gap' can be found in Korsgaard's interpretation of Kant's derivation, since this appeals only to a requirement close to the first distinctive feature of moral goodness, (i), a requirement that a reason's normative demand must derive from a quality of the reason itself. This requirement is equally satisfied by principles other, and weaker, than the formula of universal law, such as the principle that any reason which is held to be required, reasonable, or permissible for a will in certain circumstances must be held to be required, reasonable, or permissible for any will in relevantly similar circumstances. However, as commonly expressed, the allegation of a 'gap' in Kant's derivation is directed at a different interpretation of Kant's derivation. According to this interpretation, Kant appeals only to a claim closer to the second distinctive feature of moral goodness, (ii), a claim that the moral 'law' must express a requirement which is necessary for a will as such, rather than conditional on contingent inclinations. The problem of a 'gap' is raised here because this claim suffices only to justify the weak principle which also satisfies Korsgaard's requirement, namely, the principle that any reason which is held to be required, reasonable, or permissible for a will in certain circumstances must be held to be required, reasonable, or permissible for any will in relevantly similar circumstances. This principle is much 'weaker' than the formula of universal law, in the sense that the former principle provides less, and, indeed,

minimal, guidance for reasons – in particular, the former principle can be satisfied by a reason concerned with the satisfaction of an inclination.³²

This allegation of a ‘gap’ is usually directed at Kant’s identification of ‘the conformity of actions as such with universal law’, interpreted as an expression of the weak principle, with the formula of universal law, expressed as the principle, ‘I should never act other than in such a way *that I could also will that my maxim should become a universal law*’. It is also often directed at the passage in the second section in which Kant identifies the conformity of maxims to ‘nothing other than the universality of a law as such’, which the allegation interprets as an expression of the weak principle, with the formula of universal law.³³ In response to this allegation, some commentators maintain that Kant can justify expressing the formula of universal law here only by making a premature appeal to the requirement of ‘autonomy’, understood as the requirement that reasons not concern anything other than a will and as a requirement which he explicitly introduces and justifies only in later passages of the *Groundwork* or in the *Critique of Practical Reason*.³⁴

However, the interpretation provided above reveals that Kant effectively introduces the requirement of ‘autonomy’ in the first section of the *Groundwork*,

³² For this allegation as directed at the first section of *G*, and in many cases also at the second section of *G* and at *KpV*, see Paton, *The Categorical Imperative*, pp.71-2, Potter, ‘The Argument of Kant’s *Groundwork*, Chapter 1’, pp.35-6, Aune, *Kant’s Theory of Morals*, pp.29-34, Hill, ‘Kant’s Argument for the Rationality of Moral Conduct’, p.268, ‘A Kantian Perspective on Moral Rules’, pp.38-9, and ‘Is a Good Will Overrated?’, n.14 to p.42, Wood, *Kant’s Ethical Thought*, p.48 and 81-2, and Allison, ‘On a Presumed Gap in the Derivation of the Categorical Imperative’, pp.143-7 and 150-1. The allegation is occasionally directed only at the second section of *G* or at *KpV*, however, as in Wood, *Hegel’s Ethical Thought*, pp.135 and 161-9, Cummiskey, *Kantian Consequentialism*, p.57, and Pippin, ‘Kant’s Theory of Value’, p.249.

³³ *G* 421.

³⁴ For the response that Kant must make a premature appeal to the requirement of ‘autonomy’, see Aune, *Kant’s Theory of Morals*, pp.86-9, Allison, *Kant’s Theory of Freedom*, pp.207-10 and 212-3, and ‘On a Presumed Gap in the Derivation of the Categorical Imperative’, pp.150-4, and Wood, *Kant’s Ethical Thought*, pp.48-9 and 81-2. These commentators differ, however, regarding whether Kant explicitly introduces this requirement in the second section of *G*, the third section of *G*, or *KpV*.

as the requirement that moral goodness be a goodness which only a will can achieve, so that it is a goodness of a will's reason for action alone. That is, he introduces the requirement of 'autonomy' as the conclusion, (b'), which he draws from his second claim regarding a good will's goodness, (b), and as the first distinctive feature of moral goodness, (i), which he derives from this conclusion. Thus the common allegation of a 'gap' in Kant's derivation can be dismissed, as resting on a failure to appreciate his appeal to 'autonomy'. Furthermore, Kant's derivation also does not suffer from a 'gap' similar to that found in Korsgaard's interpretation. For it appeals not only to 'autonomy', in the form of the first distinctive feature of moral goodness, (i), but also to the second distinctive feature of moral goodness, (ii), the requirement that moral goodness be good for a will as such. Indeed, even the latter requirement alone entails that moral goodness must be expressed by a principle stronger than the weak principle which the allegation identifies. For this requirement is not just the requirement that the moral 'law' must be 'necessary' in the sense that it is valid irrespective of the contingent inclinations of a will, a validity which is possessed by both the weak principle and the formula of universal law. Rather, the requirement is that the moral 'law' must express a goodness which is good for a will as such. It therefore excludes the goodness of satisfying a will's contingent inclinations, and is thus stronger than the weak principle which the allegation identifies.

Kant's derivation of the formula of universal law in the first section is therefore not vulnerable to allegations of various kinds of 'gaps'. Nonetheless, and despite the significance which sympathisers and critics have traditionally attributed to it, Kant does not intend the formula of universal law to be a comprehensive expression of the nature of moral goodness. At the conclusion of the first section,

he writes that, with this formula, ‘we have reached [*gelangt*], in the moral cognition of common human reason, its supreme principle’, but insists that ‘*common human reason* is driven [...] to go out of its sphere and to take a step into the field of *practical philosophy*, in order to obtain on account of it information and clear instruction regarding the source of its principle and correct determination of it’.³⁵ Kant also makes similar remarks elsewhere in the first and second sections.³⁶ He therefore does not intend the formula of universal law to provide the complete ‘determination’ of ‘*the supreme principle of morality*’, and the ‘examination’ of it and its ‘sources’, which he promises in the preface.³⁷ In this sense, then, Kant can be said to part ways with many of his sympathisers, and to concur with many of his critics, over the persistent issue of whether this formula provides a comprehensive expression of the nature of morality. The precise sense in which he holds the formula not to be comprehensive, however, is indicated by his argument in the second section. As the following section will demonstrate, Kant there continues his analysis precisely in order to emphasise the first distinctive feature of moral goodness – namely, (i), that this goodness is a goodness of will’s reason for action alone, because (b’) it is a goodness that only a will can achieve. This is understandable, given that, although Kant derives the formula of universal law as an expression of his two distinctive features of moral goodness, its expression of this goodness in terms of ‘universal law’ emphasises only the second such feature – namely, (ii), that moral goodness is good for a will

³⁵ *G* 403, 405.

³⁶ After discussing the formula of universal law in the first section, Kant writes at *G* 403 that, with it, ‘I do not yet have insight into what this respect [that is, a reason which can also be willed to be “a universal law”] is grounded upon (which the philosopher may investigate)’. He begins the second section, ‘we have drawn our concept of duty thus far from the common use of our practical reason’, and, a few pages later, he writes that, without the developments of the second section, it would be possible ‘to determine precisely for speculative judgement the moral [element] of duty in everything that conforms with duty’, but ‘impossible to ground moral on their genuine principles even in merely common and practical use’ (*G* 406, 412).

³⁷ *G* 392.

as such, a 'law' for it. For Kant, therefore, the formula of universal law is intended to express both of the two distinctive features of moral goodness which he identifies in the first section, but emphasises only one of these features. As we will see, this formula also provides Kant with a compass as he 'takes a step into practical philosophy' in pursuit of expressions of moral goodness which emphasise its other distinctive feature.

I. Practical reason and the second derivation of the formula of universal law

In the second section, Kant introduces a philosophical account of the will as 'practical reason' in order to provide expressions of moral goodness which emphasise the first feature which he identifies in the first section, namely, that (i) moral goodness is a goodness of a will's reason for action alone, since (b') it is a goodness which only a will can achieve. He first presents what he calls the 'general rules of determination' of the will – that is, an account which is indifferent to whether a will is a good one. A 'will', he begins, is 'the ability to act *according to the representation* of laws, i.e. according to principles'. He continues, 'Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason'. Kant thus understands willing as action according to reasons, or principles.³⁸

Kant supplies some elaboration of these claims in footnotes. In a footnote in the third section, he writes that 'reason becomes practical, i.e. a will-determining cause' only by 'tak[ing] an interest in an action', and that this distinguishes 'a rational being' from 'non-rational creatures', which 'feel only sensible impulses'. In a footnote to his account in the second section, he maintains that this 'taking of interest' might concern either a contingent 'inclination' determined by 'feelings', in which case 'reason supplies [...] the practical rule as

³⁸ G 412.

to how to remedy the need of inclination', or 'principles of reason *in themselves*'.³⁹ For Kant, then, a rational agent acts according to reasons, rather than being determined by 'impulses', whether she pursues her 'inclinations' or independent rational requirements.⁴⁰

In the text, however, Kant simply states that reasons 'say that to do or to refrain from something would be good', and so are reasons 'of a will which is good in some way'. In order to identify a good will's reasons in particular, Kant first distinguishes two kinds of reasons according to the sense in which a reason represents an action as 'good': a 'hypothetical' reason represents an action as contingently good, 'as [a] means to something other that one wills (or [that] it is possible for one to will)', while a 'categorical' reason represents an action 'as *in itself* good, hence as necessary in a will in itself in conformity with reason, as its principle'.⁴¹ Although Kant's expression of the distinction here has led some commentators to interpret a 'categorical' reason as distinguished by its not being concerned with an 'end' at all, his following account of the 'end' of such a reason and his expressions of the distinction elsewhere make clear that he holds both 'hypothetical' and 'categorical' reasons to represent actions as good as 'means' to an 'end', and distinguishes them according to whether they represent actions, and so 'ends', as contingently good or as good for a will as such.⁴²

Kant then argues that since a categorical reason represents an action as good for a will as such, what is required for a reason to be categorical can be expressed by the formula of universal law. Kant's derivation of the formula of

³⁹ G 460n, 413n.

⁴⁰ Kant makes these claims in a notably clearer manner in the introduction to *MS*, at *MS* 211-4.

⁴¹ G 413, 414. On the distinction between hypothetical and categorical reasons, see also G 415-6.

⁴² For the misinterpretation of Kant's distinction here, see, for example, Paton, *The Categorical Imperative*, pp.114-5. Kant expresses the distinction elsewhere at *KpV* 11n, 19-21, and 58-9, *MS* 222, and *VL* 86-7. Even at G 414-6, though, he repeatedly distinguishes categorical reasons by their not being concerned with 'other' ends, rather than their not being concerned with 'ends' at all.

universal law in the second section thus mirrors that in the first section. That is, with this formula, he expresses moral goodness as (b') a goodness which can be achieved only by a will, so that it is (i) a goodness of a will's reason for action alone, and as (ii) a goodness which is good for a will as such. He expresses the former feature – namely, (i) – by providing the derivation in the context of his account of the will as practical reason, but again the formula, and here even the derivation, emphasise only the latter feature – namely, (ii). Thus, before providing the derivation, Kant emphasises that 'law brings with it the concept of an *unconditioned* and indeed objective and hence universally valid *necessity*', and that a categorical reason 'alone sounds like [*als...laute*] a practical **law**' because it 'leaves the will no discretion with respect to the opposite'. He then proceeds to derive the formula of universal law by arguing that 'the mere concept' of a categorical reason, unlike that of a hypothetical one, thus 'contains' what is required for a reason to be categorical, namely, that a reason satisfies the requirement expressed by the formula of universal law.⁴³ This is so, Kant argues here, because a categorical reason must be good for a will as such – that is, it must express 'an unconditioned and indeed objective and hence universally valid necessity' for a will, or must 'leave the will no discretion'. Kant also puts this as follows.

If I think of a *hypothetical* imperative in general, I do not know beforehand what it will contain: until the condition is given to me. But if I think of a *categorical* imperative, I know at once what it contains. For since, except the law, the imperative contains only the necessity that the maxim be in conformity with this law, but the law contains no condition to which it would be limited, there remains nothing other than

⁴³ *G* 416, 420.

the universality of a law as such with which the maxim should be in conformity, and the imperative represents this conformity alone as necessary.

The categorical imperative is therefore only singular and [is] actually this: *act only according to that maxim through which you can at the same time will that it become a universal law.*⁴⁴

Given that the derivation of the formula of universal law in the second section corresponds with that provided in the first section, then, it no more suffers from a 'gap' between a weaker principle and the formula than the derivation in the first section, despite the claims of some critics. In particular, like the derivation in the first section, that in the second section invokes the requirement of 'autonomy' by appealing to conclusion (b'), and the feature of moral goodness which Kant derives from it, namely, (i). It consequently makes no premature appeal to such a requirement.⁴⁵

Having already, in the first section, expressed the goodness of a good will by the formula of universal law, Kant can identify a will acting according to a categorical reason with a good will. As he writes after deriving and discussing the formula of universal law in the second section, 'We have therefore demonstrated at least that, if duty is a concept which should contain meaning and real lawgiving for our actions, it can be expressed only in categorical imperatives, not at all in hypothetical [ones]'. This remark and others indicate that, for Kant, the formula of universal law completes the part of his argument in the second section which is concerned to 'follow and clearly present the practical ability of reason from its general rules of determination to where the concept of duty arises from it', as he

⁴⁴ G 420-1.

⁴⁵ For references regarding this allegation, see nn.32 and 34 above.

puts it before beginning to present these ‘general rules’ of willing.⁴⁶ By again deriving the formula of universal law from the notion of a categorical reason, then, Kant demonstrates that the concept of duty ‘arises from’ his account of the will with this notion in particular.

II. *The derivation of the formula of the end in itself*

Kant’s identification of a will acting according to a categorical reason with a good will, by means of the formula of universal law, allows him to proceed to use the notion of a categorical reason to guide his search for expressions of duty, or moral goodness, which emphasise what the formula of universal law does not.⁴⁷ This next part of the argument in the second section requires, as Kant notes, that he ‘take a step out [...] into the metaphysics of morals’ – that is, that he no longer consider the ‘general rules of determination’ of a will, with a view to identifying the kind of reason according to which a good will acts, but instead consider ‘the relation of a will to itself, insofar as it determines itself merely through reason’. Kant explains the meaning and necessity of this ‘step’ in terms of the notion of ‘the *end*’ of a will, as that which ‘serves the will as [the] objective ground of its self-determination’. A will’s ‘end’, then, is what a will takes to make an action ‘good’, either contingently or for a will as such, and to thus provide a reason for action. For Kant, it is necessary to consider how a will could ‘determine itself merely through reason’ because an end, ‘if it is given through mere reason, must hold equally [or identically, *gleich*] for all rational beings’ – that is, it must make

⁴⁶ *G* 425, 412. The other remarks which indicate that the formula of universal law plays this role in the argument of the second section are found at *G* 413n, 421, and 431.

⁴⁷ Kant notes that his argument employs this notion as a guide at *G* 440.

actions good for a will as such.⁴⁸ To identify an end which is 'given through mere reason' would therefore explain how a will could act according to a categorical reason, and so be a good will.

In Kant's terms, such an end would provide a synthetic and *a priori* 'connection' between a will and a categorical reason. For, he writes, acting according to a categorical reason is 'something that is not contained in' the concept of a will, and a categorical reason, by representing an action as good for a will as such, 'connects the deed with the will without a presupposed condition of some inclination', given contingently and *a posteriori*.⁴⁹ It is the latter, *a priori* nature of the 'connection' which requires Kant to 'take a step out [...] into the metaphysics of morals', since to consider how a will could 'determine itself merely through reason' is precisely to consider how a will's reason for action could not concern an inclination given contingently and *a posteriori*, but rather be 'given' by reason, or the ability to act according to reasons, alone. As Kant writes as he takes this 'step', if the formula of universal law expresses a valid requirement, then 'it must already be connected (completely *a priori*) with the concept of the will of a rational being as such', and 'if *reason for itself alone* determines conduct', then 'everything that has relation to the empirical of itself no longer applies'.⁵⁰

However, Kant does not identify the end which would provide a categorical reason for action simply by identifying an end which holds 'equally [or identically] for all rational beings'. This would be to appeal only to the feature of moral goodness which the formula of universal law already emphasises, namely (ii) that moral goodness is good for a will as such. Rather, Kant identifies the end

⁴⁸ *G* 426-7, 427.

⁴⁹ *G* 420n. See also *G* 420, 428-9, 429n, 440, 444-5, and 447.

⁵⁰ *G* 427. See also *G* 387-8 and 425-6, and Kant's criticism of Wolff at *G* 390-1.

which would provide a categorical reason for action by appealing to both of the features of moral goodness which he identifies in the first section, just as he appeals to both of these features in order to derive the formula of universal law. In other words, he holds that for a will to act according to a categorical reason, and so be a good will, what it must take to provide its reason for action must not only 'hold equally [or identically] for all rational beings', but also be such that the will achieves a goodness of its reason for action alone, a goodness which only a will can achieve. Indeed, in deriving the formulas of the end in itself and autonomy, Kant is concerned to emphasise this latter feature of moral goodness – that is, feature (i) – precisely because this feature is not emphasised by the formula of universal law.

Kant first derives the formula of the end in itself. This derivation is often interpreted either as an inference of the unconditional goodness of will itself from a will's capacity to be unconditionally good, or, following Korsgaard, as 'a regress upon the conditions' of a will's taking its 'ends' to be unconditionally good, such that a will must take itself to 'confer' this goodness upon its ends by way of its own unconditional goodness.⁵¹ There is little to support either interpretation in Kant's text, however, and both attribute to him questionable reasoning: the former employs a questionable inference from the goodness of x to the goodness of the capacity for x , and the latter rests upon an implausibly strong claim that a will takes each of its 'ends', including non-moral ones, to be unconditionally good, and

⁵¹ For the former interpretation, see, for example, Cooper, 'The Formula of the End in Itself', pp.401-2, Pogge, 'Kant's Theory of Justice', n.6 to p.409, and 'The Categorical Imperative', pp.197-9, and O'Neill, 'Universal Laws and Ends-in-Themselves', p.137. For the latter interpretation, see, in particular, Korsgaard, 'Kant's Formula of Humanity', pp.114-24, 'The Right to Lie', pp.143-4 and 151-2, 'Aristotle and Kant on the Source of Value', p.241, and 'Motivation, Metaphysics, and the Value of the Self', pp.53-4, Cummiskey, *Kantian Consequentialism*, pp.62-83 and 169-71, and Wood, 'Humanity as End in Itself', pp.173-7, and *Kant's Ethical Thought*, pp.124-32.

the equally suspect claim that what 'confers' unconditional goodness must itself be unconditionally good.⁵² Alternatively, the formula of the end in itself is commonly supposed to derive from the unconditionality required by the formula of universal law, on the grounds that for a will to act according to a reason which is good for a will as such is already for it to adopt the 'end in itself'.⁵³ Such a derivation is also questionable, however, since the notion of a reason which is good for a will as such provides, in itself, no indication of what end such a reason must concern. Nor, finally, is it promising to suppose, as Guyer does, that the formula of the end in itself is intended to identify, independently, the teleological end which justifies the unconditionality that the formula of universal law requires.⁵⁴ For with regard to what precedes it, this would mark a profound, and unremarked, discontinuity in Kant's argument, in contrast with his apparently cumulative intentions.

It is fortunate, then, that Kant in fact derives the formula of the end in itself by appealing to the two features of moral goodness which he identifies in the first section, and, in particular, to the feature which the formula of universal law does not emphasise. That is, his derivation is primarily concerned to identify what a will must take to provide its reason for action if the will's goodness is to be (i) a goodness of its reason alone, and therefore (b') a goodness which only a will can achieve. His conclusion is that a will can be good only by taking 'rational nature'

⁵² On the questionable inference, see Jones, *Kant's Principle of Personality*, p.18, and Pogge, 'The Categorical Imperative', n.25 to p.199. For criticism of Korsgaard's strong claim regarding 'ends', see Gaut, 'The Structure of Practical Reason', Schneewind, 'Korsgaard and the Unconditional in Morality', pp.43-8, Pippin, 'Kant's Theory of Value', pp.252-5, Kerstein, 'Korsgaard's Kantian Arguments for the Value of Humanity', pp.28-9 and 30-3, Hill, 'Personal Values and Setting Oneself Ends', pp.251 and 258-68, and nn.31 and 71, and Melnick, 'Kant's Formulations of the Categorical Imperative', pp.291-2 and 305-7. For criticism of Korsgaard's claim regarding what 'confers' unconditional goodness, see Gaut, 'The Structure of Practical Reason', and Kerstein, 'Korsgaard's Kantian Arguments for the Value of Humanity', pp.33-4.

⁵³ See, for example, Paton, 'Analysis of the Argument', pp.32-3, and *The Categorical Imperative*, pp.167-71 and 175-8, Sullivan, *Immanuel Kant's Moral Theory*, pp.193-4, and Rotenstreich, 'On the Formalism of Kant's Ethics', pp.55-6.

⁵⁴ See, in particular, Guyer, 'Kant's Morality of Law and Morality of Freedom', pp.138-47, 'The Possibility of the Categorical Imperative', pp.191-200, and 'Ends of Reason and Ends of Nature', pp.163-4.

– that is, a will, or the ability to act according to reasons – ‘*as an end, never merely as a means*’ – that is, as providing a reason for action.⁵⁵ This is so, he argues, because to take anything else, such as an inclination or something without the ability to will, as providing a reason for action would be for a will to be good in a sense which other causes could achieve, and so for the will not to be a good one. For Kant, therefore, a will’s ‘end’ can be ‘given through mere reason’ in the sense that the ability to act according to reasons *itself* can provide a reason for action, and only with such an ‘end’ can a will’s goodness be (i) a goodness of its reason for action alone, and therefore (b’) a goodness which only a will can achieve.

Kant again indicates his derivation in a number of ways, but its grounds are clearest in his summary at the end of the second section. There, he writes that the ‘end’ of a good will ‘can be nothing other than the subject of all possible ends [that is, a will, or the ability to act according to reasons] itself, because this is at the same time the subject of a possible absolutely good will; for this [good will] cannot without contradiction be subordinated to [or follow, *nachsetzen*] any other object’.⁵⁶ Thus Kant appeals to his claim (b), that a good will is a condition of – or, as he puts it here, ‘cannot be subordinated to [or follow]’ – the goodness of any other good, in order to argue that a will’s goodness must consist of its taking a will itself to provide a reason for action. The implicit intermediary steps in this argument are those which Kant provides in the first section, where he concludes from his claim that a good will is a condition of the goodness of any other good that moral goodness must be (b’) a goodness which only a will can achieve, and therefore (i) a goodness of a will’s reason for action alone.

⁵⁵ G 428-9, 429.

⁵⁶ G 437.

However, it is in the paragraph which follows his ‘step out [...] into the metaphysics of morals’ and his introduction of the notion of a will’s ‘end’ that Kant provides his fullest, if not his clearest, presentation of his derivation of the formula of the end in itself. This paragraph begins with his conclusion, that ‘the human being and in general every rational being *exists* as end in itself, *not merely as means* to discretionary uses for this or that will, but must in all his actions, directed both to himself and also to other rational beings, always be regarded *at the same time as an end*’. Kant then argues, first, that, ‘[a]ll objects of inclinations have only a conditioned worth’, since their worth is conditional on the existence of the appropriate inclinations. He then states that inclinations themselves, on the existence of which the worth of their objects is conditional, do not have absolute worth. He concludes from this that, ‘The worth of every object *to be acquired* through our action is therefore always conditional’. He then proceeds to consider ‘[b]eings the existence of which is based not on our will, but on nature’. He states of such beings that, ‘if they are beings without reason [that is, without the ability to will], [they] still have only a relative worth, as means’, but that ‘rational beings [that is, beings with the ability to will]’ are such that ‘their nature already distinguishes them as ends in themselves’.⁵⁷

Taken in isolation, these claims, and their roles in Kant’s derivation, are obscure. In particular, Kant here fails to explain why he denies that inclinations themselves have ‘absolute worth’, why he concludes from this that the worth of objects ‘to be acquired’ by action is ‘always conditional’, why he claims that beings without the ability to will have ‘only a relative worth’, and, finally, why he

⁵⁷ G 428.

holds that the 'nature' of beings with the ability to will 'distinguishes them as ends in themselves'.

However, a number of Kant's remarks in this paragraph and in other relevant parts of the section indicate that his derivation here is concerned with expressing the two features of moral goodness which he identifies in the first section, and with expressing the first such feature in particular. These indications thus serve to illuminate the otherwise obscure claims which he makes in this paragraph. The first indication of his concerns is given at the end of the paragraph, where Kant re-states his conclusion as follows: 'These [that is, rational beings] are therefore not merely subjective ends, the existence of which as an effect of our action has a worth *for us*; but *objective ends*, i.e. beings the existence of which is in itself an end, and indeed one such that [...] without it nothing of *absolute worth* would be found anywhere; but if all worth were conditional, [and] therefore contingent, then no supreme practical principle would be found anywhere'. Immediately before presenting his derivation, Kant also states that the 'end' with which he is concerned is 'something *the existence of which in itself* has an absolute worth'. Secondly, Kant also indicates the concern of his derivation when, in the paragraph which presents it, he writes that 'an end in itself' is 'something that may not be used *merely* as a means, [and] therefore so far limits all choice [*Willkür*] (and is an object of respect)'.⁵⁸ He also expresses his conclusion in terms of a 'limiting condition' four times in the rest of the section.⁵⁹ Finally, in his summary at the end of the section, Kant refers to an end in itself as an end which 'must be

⁵⁸ G 428.

⁵⁹ See G 430-1, 431, 436, and 438.

thought not as an end to be effected, but as an *independent* [or self-sufficient, *selbstständig*] end'.⁶⁰

These terms are precisely those which Kant employs in the first section to refer to the first distinctive feature of moral goodness – namely, (i) – and they therefore indicate that his derivation of the formula of the end in itself is primarily concerned with expressing this feature. In the first section, he refers to this feature in terms of ‘absolute’ goodness and ‘respect’, and to the claims from which he derives it – namely, claims (b) and (b’) – in terms of a goodness which is a ‘condition’ of the goodness of every other good and a goodness other than that which can be achieved by an action’s ‘effect’.⁶¹ Nonetheless, by referring to moral goodness as not a ‘conditional, [and] therefore contingent’ goodness, Kant’s re-statement of his conclusion at the end of the paragraph which presents his derivation indicates that this derivation is not only concerned with expressing the first distinctive feature of moral goodness, but is also concerned with expressing the second such feature, namely, (ii).

These remarks, then, imply that Kant’s derivation, as presented in the paragraph which follows his ‘step out [...] into the metaphysics of morals’ and his introduction of the notion of a will’s ‘end’, should be interpreted as follows. His ultimate concern in this paragraph is to express what a will must take to provide a reason for action – that is, what its ‘end’ must be – if it is to achieve moral goodness, as distinguished by the two features (i) and (ii) which are identified in the first section, and, in particular, to emphasise the first such feature. He begins by denying that the ‘end in itself’ can be an object of inclination, on the grounds that the goodness of such an object is conditional on the existence of the

⁶⁰ *G* 437.

⁶¹ For the notion of ‘absolute worth’, see *G* 394 and 402, and on this notion and the notion of ‘respect’, see the discussion in the first subsection of this chapter.

appropriate inclination. This claim might be taken to rest on the second feature of moral goodness, according to which moral goodness is not 'conditional, [and] therefore contingent' for a will. However, Kant proceeds to appeal to the first feature of moral goodness, expressed as the requirement of 'absolute' goodness, when he denies that an inclination itself, on the existence of which the goodness of its objects is conditional, could be the 'end in itself', and concludes that no 'object to be *acquired* through our action' could be such an end. His claim, then, is that the goodness of objects of inclination is conditional on the existence of appropriate inclinations, and that inclinations themselves and, indeed, any 'effect' of an action can be good only in a sense which causes other than a will can achieve. Such goodness fails to satisfy his claim (b'), that moral goodness is a goodness which only a will can achieve.

Kant then proceeds to consider the goodness of beings the existence of which cannot be the 'effect' of an action. As he puts it here, the existence of such beings 'is based not on our will, but on nature'. Regarding such beings, he states first that those which do not have the ability to will cannot be ends at all, but only means, and then that those which do have the ability to will are 'marked [or distinguished]' by this as ends in themselves. Kant's concern with emphasising the first distinctive feature of moral goodness implies that, with these statements, he intends to claim that the goodness of beings without the ability to will is not such that it can be achieved by a will at all, but that the goodness of beings with the ability to will is such that it can be so achieved and, crucially, such that it can *only* be so achieved. For Kant, then, to take a will itself to provide a reason for action is for a will to achieve a goodness which only a will can achieve, and therefore a goodness of its reason for action alone. In other words, Kant holds that a will thus

achieves a goodness which satisfies his claim (b'), and therefore also satisfies his first distinctive feature of moral goodness, (i).

III. *The derivations of the formulas of autonomy and the realm of ends*

Kant, however, also derives another formula which emphasises his claim that, since (b) a good will is a condition of the goodness of every other good, its goodness must be such that (b') only a will can achieve it, and so (i) lie in its reason for action alone, rather than in actions' effects or in reasons for action insofar as they concern actions' effects. Kant expresses this formula, that of autonomy, as '*the idea of the will of every rational being as a will giving universal law*'. According to Kant, this idea 'follows' from the formula of universal law and the formula of the end in itself, a remark which has led some commentators to claim that the idea is derived simply by combining the preceding two formulas.⁶² But Kant also claims that this idea expresses what the preceding formulas 'only assumed', namely, 'the renunciation of all interest in willing from duty' as 'the mark distinguishing categorical from hypothetical imperatives'. This suggests that, for Kant, neither the notion of the possible willing of a reason as 'a universal law' nor the notion of taking 'rational nature' as an 'end' makes sufficiently explicit the features of moral goodness which they are intended to express, and that the formula of autonomy is intended to remedy these failings. This suggestion is reinforced by Kant's following discussion, which explains why the formulas of

⁶² G 431. For this claim regarding the derivation of this formula, see, for example, Paton, 'Analysis of the Argument', pp.33-4, and *The Categorical Imperative*, pp.180-1, Korsgaard, 'An Introduction to the Ethical, Political, and Religious Thought of Kant', p.22, and Wood, *Kant's Ethical Thought*, pp.158-9.

universal law and the end in itself fail to make sufficiently clear even the features which they are intended to emphasise, rather than merely express. In this discussion, Kant argues that the notion of '*a will giving universal law through all of its maxims*' – that is, the formula of autonomy – is '*very well suited*' to expressing moral goodness, because '*a will that stands under law* may still be bound to this law by means of an interest'. Kant claims that the notion simply of '*a will that stands under law*' thus fails to make sufficiently explicit that moral goodness is '*based on no interest*', or '*unconditional*', and encourages the misunderstanding that a will's '*conformity with law* would be necessitated by *something other*' than a will itself.⁶³ Kant thus warns against misinterpreting the notion of '*law*' in the formula of universal law as implying that moral goodness is a goodness which is contingent for a will as such, and against misinterpreting either this notion of '*law*' or the notion of a will itself '*giving*' a reason for action in the formula of the end in itself as implying that moral goodness is a goodness of and for something other than a will alone. In other words, Kant warns against misinterpreting the formula of universal law such that it would not express his second distinctive feature of moral goodness, (ii), and against misinterpreting either this formula or the formula of the end in itself such that the formula would not express the first such feature, (i).

Kant claims that the notion of '*a will giving universal law*' avoids these dangers. As he puts it here, for a will's reason to be categorical is for it to be such that '*the will is not merely subject to the law, but so subject that it must be viewed as also giving law to itself* [or itself lawgiving, *selbstgesetzgebend*] and just because of this [as] first subject to the law (of which it can regard itself as the

⁶³ G 431, 432, 433.

author)'. Or, as Kant also writes, for a will to act according to a categorical reason, it must 'have as its object itself as giving universal law', and its reason must 'arise from' will itself. It is this quality of a will's reason which Kant calls 'the **autonomy** of the will', and which expresses the first feature of moral goodness in particular, as he makes especially clear slightly later in the section. There he writes, '*Autonomy* is [...] the ground of the dignity' of a will, its 'dignity' indicating that its goodness cannot 'be replaced by something other', as its '*equivalent*' or '*price*'. As Kant also puts it there, after again referring to two of his examples of moral goodness, 'Nature, as well as art, contains nothing which, lacking these [that is, lacking instances of moral goodness], it could put in their place; for their worth does not consist in the effects which arise from them, [...] but in [...] maxims of the will [...], even if success does not favour them'.⁶⁴ For Kant, therefore, 'the autonomy of the will' is a goodness which only a will can achieve, and thus has 'dignity', since it is a goodness of a will's reason for action alone.

By 'autonomy', then, Kant means a goodness of a will's reason for action such that it is (ii) good for a will as such, and thus 'gives universal law', and such that it (i) lies in the will's reason for action alone, and is thus (b') a goodness which only a will can achieve, or which a will alone 'gives to itself'. Kant's formula of autonomy therefore emphasises both of the two distinctive features of moral goodness which he identifies in the first section, and, in doing so, is intended to warn against misinterpreting his preceding formulas in ways which would fail to express these distinctive features.

⁶⁴ G 431, 432, 433, 436, 434, 435. See also G 439-41, 444, 447, 449, 450, 453, and 458.

Kant concludes his analysis of 'common moral rational cognition' regarding a good will by claiming that the formula of autonomy 'leads to a very fruitful concept dependent on it, namely that *of a realm of ends*'.⁶⁵ Like the formula of autonomy, this concept is often supposed to be derived simply by combining the formulas of universal law and the end in itself.⁶⁶ However, Kant's brief discussion of this concept indicates that it is 'dependent' on, or that one is 'led' to it by, the formula of autonomy in the sense that it introduces what is bracketed by even this most adequate of his preceding formulas – namely, the plurality of agents. Given a plurality of agents, he insists, the quality of a reason's being provided by mere will must extend not only to *a* will or *some* wills, but to *every* will. Kant expresses this by defining 'a realm' as 'a systematic connection [or association, *Verbindung*] of various rational beings through common [or communal, *gemeinschaftlich*] laws', and 'a realm of ends' as such a 'connection' insofar as its common laws are 'objective' – that is, insofar as it is governed by the requirements of moral goodness. Later, he also writes that reasons must 'harmonize into a possible realm of ends, as [if it were] a realm of nature', and thus constitute 'the *allness* or totality of the system' of mere wills as ends. Revealingly, in his summary, he also emphasises that the goodness of this 'complete determination' is not only a goodness of 'universal laws', as (ii) good for a will as such, but is also (b') a goodness which only a will can achieve, and therefore (i) a goodness of its reason for action alone. As he puts it here, 'such a realm would actually be achieved through maxims whose rule the categorical imperative prescribes to every rational being, *if they were universally followed*'.

⁶⁵ *G* 433.

⁶⁶ See, for example, Sullivan, *Immanuel Kant's Moral Theory*, p.212, and O'Neill, 'Universal Laws and Ends in Themselves', pp.141-4. Others claim that this concept derives simply from the formula of universal law. See, for instance, Robinson and Harré, 'The Demography of the Kingdom of Ends', p.8.

but neither divine command nor fortunate natural causes could achieve it, since it can be achieved 'only according to maxims, i.e. self-imposed rules', of a will.⁶⁷

IV. *The equivalence and diversity of Kant's formulas*

With the formulas of universal law, the end in itself, autonomy, and the realm of ends, then, Kant's analysis of 'common moral rational cognition' regarding a good will achieves the 'determination' of 'the supreme principle of morality' which he intends his analysis to provide. Indeed, he begins his summary of the results of his analysis by stating, quite appropriately, that with these formulas, 'We can at this point end where we started out from at the beginning, namely the concept of an unconditionally good will'.⁶⁸ Furthermore, immediately before making this statement, Kant presents his formulas in a systematic way which makes their roles in his analysis particularly clear. As a means of summarising the preceding interpretation of Kant's argument in the first two sections of the *Groundwork*, therefore, I will briefly consider this systematic presentation.

In this presentation, Kant writes of the formulas of universal law, the end in itself, and the realm of ends that these 'three ways of representing the principle of morality are fundamentally only so many formulas of the very same law, one of which of itself unites the other two in it'. However, he then states that 'there is still a difference in them, which is indeed subjectively rather than objectively practical,

⁶⁷ *G* 433, 436, 438. See also *G* 434, 439, and 462-3. Pogge and Wood also consider the derivation of the formula of the realm of ends to introduce a notion of system among a plurality of agents. See Pogge, 'The Categorical Imperative', pp.201-4, and Wood, *Kant's Ethical Thought*, pp.166-7.

⁶⁸ *G* 437. At *G* 447, Kant also identifies moral goodness with the maxim of 'an absolutely good will'.

in order namely to bring an idea nearer to intuition (according to a certain analogy) and thereby to feeling'. The 'subjectively practical' difference between these three formulas, Kant claims, consists in their referring to the '*form*', '*matter*', and '*complete determination*' of morally good reasons, respectively.⁶⁹ Thus Kant holds that the formulas are fundamentally equivalent, but also that each provides a different 'analogy' of moral goodness, by referring to a different aspect of this goodness.

This is precisely what the preceding interpretation of Kant's derivations of formulas would lead one to expect. That is, Kant's formulas are intended to be equivalent insofar as each of them is intended to express, and to be derived from, the two distinctive features of moral goodness which are identified in the first section, namely, (i) and (ii). The later formulas therefore do not express the 'ground' of the formula of universal law, or otherwise make substantial new claims regarding moral goodness, as some commentators have supposed.⁷⁰ The formulas differ, however, because they each emphasise – or, as Kant puts it here, 'bring [...] nearer to intuition (according to a certain analogy) and thereby to feeling' – different aspects of the two features of moral goodness. Thus the formula of universal law emphasises the second feature, and the formula of the end in itself emphasises the first. The formula of autonomy, on the other hand, emphasises both features and warns against certain possible misinterpretations of the preceding formulas, and the formula of the realm of ends introduces a significant further qualification which is bracketed by the preceding formulas. The formulas therefore do not differ merely in terms of 'perspective', as has also been

⁶⁹ G 437, 436. See also his remarks at G 462-3.

⁷⁰ For examples, see n.54 above. See also the discussion in the second section of ch.6 below.

supposed.⁷¹ Nor, indeed, is any formula alone quite adequate. As Kant writes regarding the three formulas to which he refers in his systematic presentation, ‘to provide *access* to the moral law [...] it is very useful to lead one and the same action through the three mentioned concepts and thereby, as far as can be done, to approach it to intuition’.⁷² Nonetheless, despite their differences, each formula also implies the aspects which it fails to emphasise, and thus can be understood to ‘unite the other two in it’. In particular, the formula of universal law is concerned specifically with reasons for action, although it does not emphasise that moral goodness is a goodness of a will’s reason for action alone, while the formula of the end in itself identifies the end which provides a categorical reason, although it does not emphasise that moral goodness is good for a will as such, a ‘law’ for it. These mutual implications are unsurprising, given the common grounds of the derivations.

⁷¹ O’Neill and Rawls claim that the formula of universal law considers an action from the perspective of the agent, while the formula of the end in itself considers it from the perspective of those affected by the action. See O’Neill, ‘Universal Laws and Ends in Themselves’, pp.141-4, and Rawls, ‘Themes in Kant’s Moral Philosophy’, pp.89-90, and *Lectures on the History of Moral Philosophy*, pp.183, 191, 194-5, and 200.

⁷² *G* 437.

It is notable that, thus understood, Kant's argument in the first and second sections of the *Groundwork* is a continuous one, and one ultimately based upon the conception of the goodness of a good will which he claims to find in 'common moral rational cognition'. The argument of the second section is therefore not, as commentators often maintain, largely independent of, and more substantial than, that of the first, and intended to be, at most, merely consistent with the broadest, and supposedly less substantial, results of the first.⁷³ Rather, the second section completes the argument which the first section begins. Specifically, in the first section Kant identifies two distinctive features of moral goodness by analysing the conception of the goodness of a good will expressed in the opening pages. There he succeeds in providing a formula, the formula of universal law, which emphasises only the second feature. Then, in the second section he proceeds to provide formulas which emphasise the second feature of moral goodness, by introducing a general account of the will as 'practical reason' and employing the formula of universal law to guide this procedure – in particular, to identify a reason which this account characterises as 'categorical' with the reason according to which a good will acts. As he writes in the last paragraph of the second section, 'the metaphysics of morals' with which the second section provides formulas beyond the formula of universal law 'showed only through the expansion of the

⁷³ For prominent examples, see Korsgaard, 'An Introduction to the Ethical, Political, and Religious Thought of Kant', pp.12-4, Reath, 'The Categorical Imperative and Kant's Conception of Practical Rationality', pp.391-3 and n.19, Allison, *Kant's Theory of Freedom*, pp.85-6 and 107-8, and Wood, *Kant's Ethical Thought*, pp.17-20 and 48-50. For other examples, see Levin, 'Kant's Derivation of the Formula of Universal Law as an Ontological Argument', pp.60-1, Baker, 'Counting Categorical Imperatives', pp.393-4, O'Neill, 'Reason and Autonomy in *Grundlegung* III', n.6 to p.54, and Melnick, 'Kant's Formulations of the Categorical Imperative', p.304.

universally accepted concept of morality: that an autonomy of the will unavoidably depends on it, or rather lies in its ground'.⁷⁴ Or, as he concludes after providing this 'metaphysics', 'We can at this point end where we started out from at the beginning, namely the concept of an unconditionally good will'.

Kant's method in these two sections is therefore simply to analyse the conception of the goodness of a good will which he attributes to 'common moral rational cognition', and not to appeal to further, independent commitments to justify, modify, or qualify the account of moral goodness which this analysis reveals. In particular, he does not appeal to an independent conception of 'rationality', deontological constraints, or 'the good'. This conforms precisely to the 'method' which he proposes for the first and second sections in the preface, namely, 'to proceed analytically from common cognition to determination of its supreme principle'. He claims that this method dictates the division and functions of the sections, such that, as their titles indicate, the first section concerns the 'transition from common moral rational cognition to philosophical', and the second the 'transition from popular moral philosophy to metaphysics of morals'. Kant's analysis in the first section is therefore intended to be 'philosophical' in the sense which he also identifies in the preface, when he writes, 'that which distinguishes [philosophy] from common rational cognition is that it presents in isolated [*abgesondert*] science what the latter comprehends only confusedly'.⁷⁵ The second section, however, is intended to continue this analysis in a particular manner, by employing a general account of the will as practical reason to provide a 'metaphysics of morals' – that is, an account of how a will might take its reason

⁷⁴ *G* 444-5.

⁷⁵ *G* 392, 390. On the relation between 'philosophy' and 'common moral rational cognition', see also *KrV* A830-1/B858-9, *G* 403-4, 409, and 444-5, *KpV* 8n and 27, and *MS* 216, and on Kant's understanding of 'philosophy', see *KrV* A712-38/B740-66 and A835-40/B863-8, *P* 255, *KU* 174, and *EKU* 195.

for action to be provided by a will alone. The ‘transition’ to this account from ‘popular moral philosophy’ involves simply excluding what Kant calls the ‘disgusting mishmash’ of empirical claims and informal techniques which many of his contemporaries promoted and practised as ‘philosophy’.⁷⁶ Unsurprisingly, Kant excludes this manner of ‘philosophy’, and supports his alternative, on the grounds that only his alternative can account for the two features of moral goodness identified in the first section.⁷⁷

This method is also followed in the *Critique of Practical Reason*, which, Kant writes in the preface to it, ‘presupposes the *Groundwork of the Metaphysics of Morals* [...] insofar as this constitutes preliminary acquaintance with the principle of duty and provides and justifies a determinate formula of it’. This ‘formula’, he continues in a footnote, does not express ‘a new basic principle of morality [...] as if [...] the world had been ignorant of what duty is, or in thoroughgoing error about it’, but might be better considered as corresponding to the analytical formulas of a ‘mathematician’.⁷⁸ Thus, in the body of the *Critique* Kant makes little reference to a good will, but begins by expressing the two distinctive features of moral goodness, (i) and (ii), which the first and second sections of the *Groundwork* identify.⁷⁹ Here and in his following works, he proceeds to express moral goodness in ways which correspond with the formulas presented in the *Groundwork*.⁸⁰

⁷⁶ G 409. See also G 410, and, for a helpful account of ‘popular’ philosophy, see Beiser, *The Fate of Reason*, pp.76 and 165-9.

⁷⁷ Besides G 404-12, see G 388-90, 419-20, and 424-6, and also *KrV* Axvii-xviii, Bxxviii, and Bxxxv-v, P 375-6, R 62-3, and *MS* 206, 216, and 447.

⁷⁸ *KpV* 8, 8n.

⁷⁹ See *KpV* 19-21, and also *KpV* 36-7, and 45-6. Kant refers to a good will only at *KpV* 62 and 79.

⁸⁰ Beyond those in the first and second sections of *G*, I would suggest that Kant’s expressions of moral goodness can be divided into: echoes of the formula of universal law, or the version of it which draws the analogy with ‘laws of nature’; re-expressions of that formula in terms of a distinction between a reason’s ‘form’ and its ‘matter’; echoes of the formula of autonomy; and

Despite this substantial, and underappreciated, continuity of method, however, three other argumentative methods that Kant employs should also be noted. Firstly, in the third section of the *Groundwork* Kant effectively provides an independent vindication of the analysis provided in the first and second sections. As I will show in the seventh chapter, however, his argument is profoundly unconvincing, and he rescinds it in the *Critique of Practical Reason*. Secondly, Kant does not undertake to account for more particular moral propositions by analysing more particular conceptions of ‘common moral rational cognition’. Rather, as I will show in the following chapter, he attempts to derive such propositions systematically from one or both of the two distinctive features of moral goodness, combined with more particular conditions. Finally, Kant provides independent arguments for claims that he considers to be independent of his analysis of moral goodness. Most notably, as I will show in the seventh chapter, he provides three independent arguments for the claim that human choice must be considered as undetermined by antecedent causes. One of these arguments is that of the third section of the *Groundwork*, which thus exhibits a singular combination of the first and the third of these three argumentative methods, and which is replaced in the *Critique* by an equally unconvincing argument intended to depend on Kant’s analysis of moral goodness.

echoes of the formula of the end in itself. However, in many passages he employs, or even combines, two or more of the first three of these kinds.

The first and second sections of the *Groundwork* are, then, intended to provide an analysis of the two claims regarding the goodness of a good will which Kant attributes to 'common moral rational cognition' in the opening paragraphs of the first section. In particular, this analysis issues in a number of formulas, each of which is intended to express the two distinctive features of moral goodness (i) and (ii) which Kant draws from these two claims, and to emphasise different aspects of one or both of them. Thus understood, the argument is not vulnerable to many objections which are persistently levelled at it, and at the argument of the first section in particular. By analysing a basic conception of goodness to which 'common moral rational cognition' is supposed to be committed, the argument also displays an underappreciated method.

Moral reasons

This chapter is occupied with the interpretation of Kant's formulas and their place in his conception of moral judgement, and with the evaluation of Nietzsche's relevant criticisms. The first section considers the formula of universal law, and argues that, rather than an algorithmic 'test' of reasons for action in specific circumstances, this formula is intended to emphasise feature (ii) of moral goodness, and to be employed in Kant's proposed derivation of a system duties of varying generality from features (i) and (ii). In this light, Nietzsche's criticisms of this formula are shown to be misplaced. The second section turns to Kant's further formulas, and argues that their intended meanings as expressions of (i) and (ii) are not captured by prevailing interpretations, and that they are, again, intended to be employed in Kant's proposed derivation of a system of duties. With respect to the formula of autonomy in particular, I also briefly argue that, although Nietzsche sympathisers often insist that Kant betrays the individual 'autonomy' which Nietzsche affirms, in fact such 'autonomy' does not, and need not, concern Kant. The final section considers Kant's conception of moral judgement in the light of Nietzsche's broader criticisms of the common sense conception of agency, criticisms which he also directs at Kant. It argues that these criticisms are somewhat more telling than Nietzsche's other criticisms, but that, nonetheless,

Kant could accommodate this success without surrendering his conception of moral judgement and the place of his formulas in it.

Until quite recently, the interpretation, and thus the criticism and defence, of Kant's account of moral judgement focussed on his formula of universal law, and supposed that he intends this formula to provide an algorithmic 'test' of reasons for action in specific circumstances, according to the consequences of a reason's universal adoption. Even now, when consideration generally admits other interpretations of the formula and extends to his other formulas, Kant is often supposed to treat moral judgement as the 'testing' of reasons for action in specific circumstances, through the bringing of such reasons under his formulas. Although Nietzsche displays no interest in the mechanics of such 'testing', he exemplifies the traditional focus on the formula of universal law, and, at least in *'Long live physics!'*, also the supposition that this formula is to be applied directly to reasons for action in specific circumstances, 'testing' them for moral goodness.

In this section, I will first explain the difficulties which, as critics have endlessly noted, arise from treating the formula of universal law as a 'test' regarding the consequences of a reason's universal adoption. I will then, in the second subsection, argue that Kant's claims regarding the formula in the *Groundwork* provide little support for this traditional approach, and better accord with the account of the formula's derivation which I provided in the preceding chapter. In the third subsection, I will argue for a corresponding conclusion regarding Kant's employment of the formula in his derivation of a general duty of beneficence in *The Metaphysics of Morals*, a derivation which reveals the formula's place in his considered understanding of moral judgement. Finally, in

the last subsection I will argue that the preceding claims show Nietzsche's criticisms of the formula of universal law to be misplaced.

I. *The 'emptiness' of the traditional approach*

Regarding Kant's formulas, critical attention has traditionally focussed on the passage of the second section of the *Groundwork* in which he discusses four examples in the light of the formula of universal law – or, more specifically, in the light of a version of this formula which draws an analogy between 'a universal law' and 'a **universal law of nature**'.¹ There he introduces a provisional distinction between 'perfect' and 'imperfect' duties, or 'strict or narrow (unremitting)' and 'wide (meritorious)' duties, according to which a duty of the former kind, unlike one of the latter kind, 'permits no exception for the advantage of inclination'. He employs this distinction and the distinction between duties to oneself and to others in dividing his examples, and, crucially, claims that 'perfect' and 'imperfect' duties are considered differently in the light of the formula of universal law: he writes that a 'maxim' which contravenes a 'perfect' duty 'cannot even be *thought* without contradiction' as a universal law, and therefore cannot be 'willed' as such a law, but that a 'maxim' which contravenes an 'imperfect' duty can be so thought, but is 'still impossible to *will*' as a universal law. He makes similar remarks in his discussion of each example, claiming that particular maxims of suicide and false promising 'contradict' themselves, or could not 'exist as

¹ Kant states that the formula of universal law can be expressed as 'act as if the maxim of your action were to become through your will a **universal law of nature**', because, he claims, 'the existence of things insofar as it is determined according to universal laws [...] constitutes what is properly called *nature* in the most general sense' (G 421).

nature', if considered as universal laws, and that particular maxims of not cultivating one's natural predispositions and not assisting others in great need could 'exist' as such laws, but are 'impossible to will', given what 'a rational being [...] necessarily wills'.² These remarks have traditionally been taken to indicate that Kant intends the formula to provide two different 'tests' for reasons for action, one for 'perfect' duties and one for 'imperfect' duties, and these two 'tests' have become known as the 'contradiction in conception' and 'contradiction in the will' tests, respectively. Both 'tests' have traditionally been supposed to concern the consequences of the universal adoption of a 'maxim' in specific circumstances, a 'maxim' being commonly understood as a proposition stating both the intended action and the reason for doing it.

Regarding the supposed 'contradiction in conception' test for 'perfect' duties, there are three major interpretations. One is found primarily in some of the less recent literature, and identifies a maxim which contravenes the test either as one which is inconsistent with, or fails to fulfil, a relevant natural purpose, or as one which, if it were universally adopted, would undermine a system of purposes.³ However, this interpretation has been unpopular for some time, since it relies on

² G 421, 421n, 424, 422, 423.

³ Paton endorses this interpretation in the latter version, and Aune attributes the latter version to Kant's application of the 'formula of the law of nature'. Harrison and Kemp attribute the former version to Kant's treatment of the maxim of suicide, and Sullivan attributes the former version to Kant's treatment of the maxim of false promising, while also claiming that the two other major interpretations are 'equally correct'. Beck suggests both of the two other major kinds of interpretation as negative tests of the permissibility of maxims, but also endorses both versions of the teleological interpretation as providing a stronger, positive test. See Paton, 'Analysis of the Argument', p.30, and *The Categorical Imperative*, pp.149-57 and 162-3, Harrison, 'Kant's Examples of the First Formulation of the Categorical Imperative', pp.233-4 and 240-1, Kemp, 'Kant's Examples of the Categorical Imperative', pp.249-50, Beck, *A Commentary on Kant's Critique of Practical Reason*, pp.159-61, Aune, *Kant's Theory of Morals*, pp.59-60, and Sullivan, *Immanuel Kant's Moral Theory*, pp.169-70.

the additional and questionable presupposition of the necessity of relevant natural purposes and/or a system of purposes.⁴

The two other major interpretations of this test have been more persistently popular. The first identifies a contravening maxim as one which could not exist if it were also universally adopted, while the second kind identifies a contravening maxim as one whose purpose could not be achieved if it were also universally adopted.⁵ The first seems best suited to maxims whose universal adoption would undermine a practice which they depend on, such as Kant's example of the maxim of false promising. However, this interpretation is unable to explain any 'contradiction' concerning a maxim which is not dependent on a practice, such as a maxim of humiliation or injury. It also excludes maxims whose concern with modifying or undermining certain practices might seem morally permissible or laudable, such as a maxim of buying, but not selling, certain things, or Hegel's example of a maxim of helping the poor.⁶

⁴ For such criticism, see Harrison, 'Kant's Examples of the First Formulation of the Categorical Imperative', p.241, O'Neill, *Acting on Principle*, pp.64-5, Korsgaard, 'Kant's Formula of Universal Law', pp.87-92, Herman 'Moral Deliberation and the Derivation of Duties', n.5 to p.136, Wood, *Kant's Ethical Thought*, n.16 to p.91, and Steinberger, 'The Standard View of the Categorical Imperative', pp.92-3. Korsgaard argues that this interpretation also presents difficulties for distinguishing the 'contradiction in conception' test from the 'contradiction in the will' test, since, on this interpretation, the former seems to ultimately refer to the latter. See Korsgaard, 'Kant's Formula of Universal Law', p.96.

⁵ For examples of the first of these two kinds of interpretation, see Kemp, 'Kant's Examples of the Categorical Imperative', pp.251-3, Henrich, 'Das Problem der Grundlegung der Ethik bei Kant und im spekulativen Idealismus', p.(99), Wood, 'Kant on False Promises', pp.615-9, and *Hegel's Ethical Thought*, pp.156-8, Herman, 'Mutual Aid and Respect for Persons', p.47, and 'Murder and Mayhem', pp.117-8 and n.5, Timmons, 'Contradictions and the Categorical Imperative', pp.305-11, O'Neill, 'Consistency in Action', pp.96-7, and 'Universal Laws and Ends-In-Themselves', p.132, and Doore, 'Contradiction in the Will', p.141. For examples of the second kind of interpretation, see Harrison, 'Kant's Examples of the First Formulation of the Categorical Imperative', pp.234-6 and 241-2, Singer, *Generalization in Ethics*, pp.251-60 and 275-9, O'Neill, *Acting on Principle*, pp.69-71, 72-3, and 78-81, Korsgaard, 'Kant's Formula of Universal Law', pp.78 and 92-4, 'Kant's Formula of Humanity', pp.126-7, 'The Right to Lie', pp.135-6, and 'Kant's Analysis of Obligation', pp.63-4, Pogge, 'Kant's Theory of Justice', p.409 and n.5, and 'The Categorical Imperative', pp.192-3 and 199-200, and Allison, 'On a Presumed Gap in the Derivation of the Categorical Imperative', pp.149 and 153.

⁶ On these points, see the discussions in O'Neill, *Acting on Principle*, pp.67-8, Korsgaard, 'Kant's Formula of Universal Law', pp.81-5, and Steinberger, 'The Standard View of the Categorical Imperative', pp.96-7. See also n.9 below.

The second popular interpretation of the ‘contradiction in conception’ test, according to which a contravening maxim is one whose purpose could not be achieved if it were also universally adopted, also succeeds in accounting for ‘contradictions’ between maxims which depend on a practice and their universal adoption, if such universal adoption would undermine the relevant practice. Unlike the first popular interpretation, however, the second is also suited to explaining some cases of ‘contradiction’ concerning maxims which do not depend on a practice, such as maxims of humiliation or injury.⁷ The explanatory capacity of the test can also be extended on both interpretations if the possibility of ‘contradiction’ is understood to involve not only the agent concerned, but also any other agent, such that a ‘contradiction’ arises if a maxim’s universal adoption would entail either that the maxim could not exist for any agent or that the maxim’s purpose could not be achieved by any agent.⁸ Nonetheless, the explanatory capacity of the second interpretation is still limited to those maxims whose purpose would be undermined by the universal adoption of the maxim, and can explain no ‘contradiction’ regarding maxims whose purpose would not be so undermined, such as maxims of humiliation or injury for the purpose of revenge.⁹

⁷ See Korsgaard, ‘Kant’s Formula of Universal Law’, p.97 and pp.97-100.

⁸ McNair introduces this extension to both interpretations in his ‘Universal Necessity and Contradictions in Conception’, pp.32-6.

⁹ On this limitation, see Korsgaard, ‘Kant’s Formula of Universal Law’, pp.82-5 and 97-101, and Herman, ‘Murder and Mayhem’, pp.118-9. In his ‘Universal Necessity and Contradictions in Conception’, pp.36-40, McNair argues that the first popular interpretation serves to remove the limitation of the second which Korsgaard identifies, that regarding maxims of killing whose purpose is not undermined by universal adoption – such as a purpose of revenge or of mercy, as in Kant’s example of the maxim of suicide – and that it removes this limitation while still allowing for maxims of killing whose purpose is self-defence. The first popular interpretation removes this limitation simply because killing makes *any* action by the victim impossible. Furthermore, in her ‘Universal Laws and Ends-In-Themselves’, pp.132-3, O’Neill argues that any maxim of violent victimisation serves to ‘undercut the agency’ of the victims, at least temporarily, and so is excluded by the first popular interpretation of the formula of universal law. However, neither McNair nor O’Neill consider cases of maxims which do not depend on practices and which also do not have the peculiar feature of maxims of violent victimisation, that they ‘undercut the agency’ of victims. Indeed, even maxims of *non-violent* victimisation need not ‘undercut the agency’ of victims in the manner of maxims of violent victimisation.

The explanatory capacity of both interpretations of the supposed ‘contradiction in conception’ test is also limited by the specificity of the maxim concerned. Specific purposes or specific time and object conditions, for example, can ensure that a maxim is not excluded when, without such conditions or with less specific ones, it would be excluded.¹⁰ To respond by proposing that every maxim ‘applicable’ to a particular action, or only the ‘underlying’ maxim, be tested not only raises the further difficulty of identifying an ‘applicable’ or ‘underlying’ maxim, but also fails to solve the original one, for the exclusion of an ‘applicable’ or ‘underlying’ maxim will equally vary according to its specificity.¹¹ Furthermore, if the maxim specifies the condition that an action be performed only if it is expected to have no adverse consequences for the future existence of actions or maxims of the same kind or for the purpose of such maxims – as in the case of secret misdemeanours, for example – then no restriction is imposed on the maxim, irrespective of its other specifications.¹² In a similar vein, according to either interpretation of the supposed test, it excludes intuitively acceptable maxims concerning the benefits of exploiting certain common kinds of behaviour by behaving exceptionally. Such maxims include, for example, maxims of buying, but

¹⁰ See Hill, ‘The Kingdom of Ends’, pp.62-3, Herman, ‘Murder and Mayhem’, p.116, and ‘Moral Deliberation and the Derivation of Duties’, pp.141-3 and 152, and Wood, *Hegel’s Ethical Thought*, pp.160-1, and *Kant’s Ethical Thought*, pp.102-4.

¹¹ These two responses have been proposed, successively, by O’Neill. For the first, see her *Acting on Principle*, pp.41-2 and 71-2, and for the second, see her ‘Consistency in Action’, pp.84 and 97, ‘Between Consenting Adults’, n.2 to p.112, and ‘Universal Laws and Ends in Themselves’, p.129. For criticism along the lines mentioned, see Herman, ‘Moral Deliberation and the Derivation of Duties’, pp.142-3, and ‘Leaving Deontology Behind’, pp. 219-20, and Wood, *Kant’s Ethical Thought*, pp.104-5.

¹² See Steinberger, ‘The Standard View of the Categorical Imperative’, pp.97-8, and, for related remarks, Wood, *Kant’s Ethical Thought*, pp.103-4.

not selling, certain things and maxims of performing an action – playing tennis, say – only at times of day when fewer other people do so.¹³

Finally, the supposed ‘contradiction in conception’ test on either interpretation is a test of the permissibility, and not the requiredness, of maxims. It is commonly supposed that the test explains requiredness at one remove, by identifying impermissible maxims, the contradictory of which is therefore required. However, in many cases it might also be permissible not to adopt any relevant maxim at all. In these cases, then, the test cannot explain any positive duty.¹⁴

Regarding the supposed ‘contradiction in the will’ test for ‘imperfect’ duties, interpretations tend to echo the interpretation of the ‘contradiction in conception’ test which concerns the achievement of a maxim’s purpose. The interpretation concerned with natural purposes is, for general reasons already mentioned, now unpopular, and Kant’s claim that maxims which contravene ‘imperfect’ duties can still be ‘thought’ as universal laws suggests that he does not consider these maxims to ‘contradict’ the existence of practices on which they depend, as the other major interpretation of the ‘contradiction in conception’ test maintains.¹⁵ However, the supposed ‘contradiction in the will’ test also cannot be straightforwardly interpreted in terms of the achievement of the tested maxim’s purpose. For Kant claims of maxims which contravene ‘imperfect’ duties that ‘contradictions’ arise in the light of what ‘a rational being [...] necessarily wills’.

¹³ For this criticism, see O’Neill, *Acting on Principle*, pp.76-7, Wood, *Hegel’s Ethical Thought*, p.157, and *Kant’s Ethical Thought*, pp.105-6 and 108, and Herman, ‘Moral Deliberation and the Derivation of Duties’, pp.138-40.

¹⁴ This is noted in Korsgaard, ‘The Right to Lie’, p.152, and, in particular, Wood, *Kant’s Ethical Thought*, pp.100-1 and 164. See also Paton, *The Categorical Imperative*, pp.141-2.

¹⁵ For examples of the interpretation of the ‘contradiction in the will’ test, see Harrison, in, ‘Kant’s Examples of the First Formulation of the Categorical Imperative’, p.241, Singer, *Generalization in Ethics*, pp.260-71 and 275-9, and Korsgaard, ‘Kant’s Formula of Universal Law’, pp.95-7.

He claims, for example, that a maxim of not cultivating one's natural predispositions is 'impossible to will' on the grounds that 'a rational being [...] necessarily wills that all of the abilities in him be developed, because they are useful and given to him for all sorts of possible intentions'.¹⁶ As is commonly noted, this suggests that a maxim which contravenes an 'imperfect' duty is to be identified as a maxim whose universal adoption threatens the achievement of another, 'necessarily willed' maxim's purpose, rather than the former maxim's own purpose. However, this raises a profound difficulty for the traditional approach to the formula of universal law, since this formula does not identify which maxims are 'necessarily willed'.¹⁷ Furthermore, like the supposed 'contradiction in conception' test, the supposed 'contradiction in the will' test concerns only the permissibility, and not the requiredness, of maxims, and therefore cannot explain positive duties in cases in which it is permissible not to adopt any relevant maxim at all.¹⁸

The traditional approach to the formula of universal law, according to which reasons for action in specific circumstances are to be 'tested' by considering the consequences of their universal adoption, is therefore decidedly vulnerable to the charges of 'emptiness' to which it has been persistently subjected. The charge in its strong form, that the formula excludes no maxims at all, is famously made by Hegel, and also made by Schopenhauer and Mill.¹⁹ Regarding the supposed

¹⁶ G 423.

¹⁷ On this difficulty, see Pogge, 'The Categorical Imperative', p.196.

¹⁸ For this criticism, see Wood, *Kant's Ethical Thought*, pp.100-2, and for an account vulnerable to it, see Doore, 'Contradiction in the Will', pp.147-50.

¹⁹ See, in particular, Hegel, *Über die Wissenschaftlichen Behandlungsarten des Naturrechts...*, pp.435-8 and 441, *Phänomenologie des Geistes*, paras.429-31 and 437, *Grundlinien der Philosophie des Rechts*, para.135 and *Zusatz*, *Enzyklopädie der philosophischen Wissenschaften*, pt.1, para.54 *Zusatz*, and *Vorlesungen über die Geschichte der Philosophie*, (v.3, pp.460-1), Schopenhauer, *Über das Fundament der Moral*, pt.II, §7, and Mill, *Utilitarianism*, pp.134 and 188. Before Hegel, the strong form of the allegation was expressed by Tittel, in his *Über Herr Kants Moralreform*, pp.14-5 and 32-6, and Fichte, in his *Das System der Sittenlehre nach den Principien*

'contradiction in conception' test, the charge can be upheld in a weaker form, insofar as this test does not exclude certain maxims which, intuitively, should be excluded (so-called 'false positives').²⁰ Regarding the supposed 'contradiction in the will' test, the charge can be upheld in the stronger form, since the supposed test excludes no maxims at all, and can exclude maxims only by presupposing certain other, contingently-related maxims. The traditional approach to the formula therefore makes it 'empty' in the sense that it cannot provide for an adequate derivation of duties. This approach is also vulnerable to the further charge, which Hegel also articulates, that the formula excludes certain maxims which, intuitively, should not be excluded (so-called 'false negatives'), and is thus too restrictive.²¹

II. Kant's claims regarding the formula of universal law in the *Groundwork*

There is, therefore, good reason to hope that the traditional approach misinterprets Kant's formula of universal law. In this subsection, I will argue that in fact the traditional approach derives little support from Kant's claims regarding the formula in the *Groundwork*, and that these claims better accord with the account of the formula's derivation which I provided in the preceding chapter.

Firstly, despite the traditional focus on Kant's presentation and discussion of his four examples, these passages in fact provide little support for the traditional approach. Kant first discusses the maxim of false promising in the first section,

der Wissenschaftslehre, p.234 (246). More recently, it has been expressed by, for instance, MacIntyre, in *After Virtue*, pp.45-6.

²⁰ For an example of this weaker form of the allegation, see Henrich, 'Das Problem der Grundlegung der Ethik bei Kant und im spekulativen Idealismus', pp.(101-2).

²¹ See Hegel, *Über die Wissenschaftlichen Behandlungsarten des Naturrechts...*, pp.438-9, and *Vorlesungen über die Geschichte der Philosophie* (v.3, p.460).

and he writes there that the purpose of his discussion is merely to show that 'common human reason also agrees with [...the formula of universal law] in its practical judgements and always has this principle before its eyes'.²² Kant does not assert this 'agreement' because he holds that common human reason employs, or should employ, the formula as a 'test' in the manner which is traditionally supposed. Instead, he writes the following, in the name of common human reason: 'I do not yet have insight into what this respect [for the formula of universal law] is grounded upon (which the philosopher may investigate), but I at least understand this much: that it is an estimation of a worth which far outweighs all worth of what is recommended through inclination, and that the necessity of my action from *pure* respect for the practical law is what constitutes duty, to which every other motive must give way, because it is the condition of a will good *in itself*, whose worth surpasses all else'.²³ Kant's discussion in the first section, therefore, is far from suggesting that he intends the formula as a 'test' of reasons for action in specific circumstances, and, indeed, also far from suggesting that he intends the formula's primary significance to lie in its referring to the consequences of a maxim's universal adoption. Rather, it suggests precisely what his derivation of the formula, as I have presented it, would lead one to expect. That is, it refers to the formula as an expression of the commitments of common human reason in its conception of the goodness of a good will, and it indicates that the formula is intended to emphasise only one of these commitments, namely, that moral goodness is good for a will as such, 'a worth which surpasses all else'.²⁴

²² G 402.

²³ G 403.

²⁴ This is also suggested by Kant's discussion of a maxim of safely increasing one's wealth at *KpV* 27-8. There Kant refers to an expression of moral goodness similar to the formula of universal law, and writes that 'the most common understanding' acknowledges that, considered alone, acting according to this maxim is not morally good because 'I cannot cite my inclination [...] as the

Kant also emphasises the inadequacy of the formula of universal law as he makes the transition from the first section to the second. As I mentioned in examining his derivations, he concludes the first section by claiming that ‘we have reached, in the moral cognition of common human reason, its supreme principle’, but that ‘*common human reason* is driven [...] to go out of its sphere and to take a step into the field of *practical philosophy*, in order to obtain on account of it information and clear instruction regarding the source of its principle and correct determination of it’. Similarly, Kant begins the second section by noting that ‘we have drawn our concept of duty thus far from the common use of our practical reason’, but, a few pages later, he writes that, without the developments of the second section, it would be possible ‘to determine precisely for speculative judgement the moral [element] of duty in everything that conforms with duty’, but ‘impossible to ground morals on their genuine principles even in merely common and practical use’.²⁵ Again, then, his remarks hardly suggest that he believes the formula of universal law to provide a ‘test’ sufficient to determine the moral goodness of reasons in specific circumstances, and instead suggest precisely what his derivation of the formula, as I have presented it, would lead one to expect.

Kant’s discussion of the formula and his examples in the second section admittedly does begin by referring to the possibility that ‘all imperatives of duty can be derived [*abgeleitet*] from this one imperative [that is, the formula of universal law] as from their principle’. But he immediately makes clear that he means by this only that, with this formula, it is possible ‘we can at least indicate what we think through [...the concept of duty] and what this concept wants to

determining ground proper to a universal practical law’. This exclusion of inclination by ‘common’ moral judgement is also Kant’s concern when he refers to examples at *KpV* 19, 35-6, 43-4, and 69-70, and in the latter two of these passages he expresses moral goodness in ways similar to the formula of universal law.

²⁵ *G* 406, 412.

say'.²⁶ With this, I suggest, Kant is simply referring to his argumentative strategy in the second section, which employs the formula of universal law to connect his account of the will with his analysis of a good will, by means of the concept of 'duty', in the first section. Kant refers to this strategy again after discussing his examples, when, in a sentence already quoted above, he writes, 'We have therefore demonstrated at least that, if duty is a concept which should contain meaning and real lawgiving for our actions, it can be expressed only in categorical imperatives, not at all in hypothetical [ones]'.²⁷

Two other remarks in Kant's discussion of the formula of universal law might be thought to support the traditional approach. One refers to the four examples: 'These are some of the many actual duties, or at least what we hold to be such, whose derivation [*Ableitung*] from the one principle cited above [namely, the formula of universal law] is clear'. The other remark concludes Kant's discussion of the formula in the second section: he states that, with this discussion, 'we have [...] presented clearly and determined for every use the content of the categorical imperative, which must contain the principle of all duty'.²⁸ However, in the four editions of the *Groundwork*, the former remark reads '*Abteilung*', 'classification', instead of '*Ableitung*', 'derivation', the latter being substituted by the editors of the *Akademie* edition (and followed, until very recently, by English translators). The latter remark is also quite consistent with a concern for 'using' the formula of universal law simply to 'classify', rather than 'derive', duties. Furthermore, reading these remarks in this manner corresponds with Kant's other remarks regarding his discussion of the examples. He claims that with the

²⁶ G 421.

²⁷ G 425. For Kant's other references to this strategy, see the discussion in the second section of ch.5 above.

²⁸ G 423-4, 425.

examples he ‘will count some duties according to the usual division of them’, and that he ‘reserve[s] the division of duties entirely for a future *Metaphysics of Morals*, [and] therefore this [division] stands here only at my discretion (in order to arrange my examples)’. After discussing the examples and explaining how ‘perfect’ and ‘imperfect’ duties are treated differently, he also claims simply that ‘thus all duties, regarding the kind of obligation (not the object of their actions) have been through these examples set out completely in their dependence on the single principle’.²⁹ At most, then, these remarks suggest that Kant intends the formula to provide for a provisional classification of given duties as ‘perfect’ or ‘imperfect’, according to the different kinds of consequences of the universal adoption of contravening maxims.

Finally, Kant is often supposed to endorse the traditional approach to the formula of universal law when, after deriving his further formulas, he writes, ‘one does better if one always proceeds in moral *judgement* by the strict method and places at its ground the universal formula of the categorical imperative: *Act according to the maxim which can at the same time make itself a universal law*’.³⁰ This ‘universal formula’ is commonly identified with the formula of universal law.³¹ But the reference to a maxim’s ‘making itself a universal law’ differs from that formula’s reference to maxims which ‘could also be willed as a universal law’, and the ‘universal formula’ might equally be identified with the formula of

²⁹ G 421, 421n, 424. At G 432n, Kant also refers to his four examples as intended merely for the ‘elucidation [*Erläuterung*]’ of his formulas, and when he returns to the first example at G 429, he writes that he ‘must here pass over a closer determination that would avoid any misunderstanding [...]; that belongs to morality proper’. Regarding the ‘division’ of duties, in the preface to *KpV* he again emphasises that ‘the *division* of all practical sciences to *completion* was not enclosed’ in G (*KpV* 8).

³⁰ G 436-7.

³¹ See, for example, Paton, *The Categorical Imperative*, p.130, O’Neill, ‘Universal Laws and Ends in Themselves’, p.127, Hill, ‘A Kantian Perspective on Moral Rules’, p.36, and ‘Hypothetical Consent in Kantian Constructivism’, pp.74 and 76, Guyer, ‘The Possibility of the Categorical Imperative’, p.175, and Melnick, ‘Kant’s Formulations of the Categorical Imperative’, pp.304 and 307.

autonomy, which Kant proceeds to express as ‘the fitness of maxims [...] to make themselves into universal law’.³² The formula of autonomy would also be a preferable candidate on my account of Kant’s derivations, since, on this account, the formula of autonomy is intended to emphasise both distinctive features of moral goodness, while the formulas of universal law and the end in itself are not. Furthermore, even if the ‘universal formula’ is identified with the formula of universal law, Kant merely claims that it is ‘better’ to place this formula at the ‘ground’ of moral judgement, and not that it provides an adequate ‘test’ of any reason in specific circumstances.

The context of Kant’s claim regarding the ‘universal formula’ also agrees better with my account of his derivations than with the traditional approach. This context is his systematic presentation of the formulas of universal law, the end in itself, and the realm of ends. As already noted, he states there that these formulas are fundamentally equivalent, but that ‘there is still a difference in them, which is indeed subjectively rather than objectively practical, in order namely to bring an idea nearer to intuition (according to a certain analogy) and thereby to feeling’. He proceeds to refer to this ‘subjectively practical’ difference as a way of providing ‘access to the moral law’, which he distinguishes from the practice of ‘moral judgement’, for which he recommends the ‘universal formula’ as a ‘ground’. To provide ‘access’, he writes, ‘it is very useful to lead one and the same action through the three named concepts’, namely, the formulas of universal law, the end

³² G 444. In the third section, Kant proceeds to employ the terms of the formula of autonomy to express moral goodness, and at G 447 he calls this formula ‘the formula of the categorical imperative’, although not the ‘universal’ such formula. See also G 449 and 458. That the ‘universal formula’ is not the formula of universal law is also suggested by a paragraph at G 434 in which Kant expresses moral goodness in a manner similar to the ‘universal law’ – as requiring ‘*that the will through its maxim could regard itself as at the same time giving universal law*’ – in the context of references to the formulas of autonomy and the realm of ends, rather than the formula of universal law. See also Wood’s comments in support of identifying ‘the universal formula’ with the formula of autonomy, in his *Kant’s Ethical Thought*, pp.187-9.

in itself, and the realm of ends.³³ Once again, then, Kant's remarks indicate that, in spite the traditional approach, he does not intend the formula of universal law to provide a comprehensive expression of the nature of moral goodness, and, indeed, that in attempting to express this goodness, or provide 'access' to it, with a series of formulas in the *Groundwork*, he is not also concerned to propose how 'moral judgement' ought to proceed.

III. An example of Kant's considered position: *Beneficence* in The Metaphysics of Morals

I mentioned at the end of the preceding chapter that the diverse expressions of moral goodness provided in the *Groundwork* are echoed throughout Kant's later texts. These texts also offer no suggestion that the formula of universal law is intended to be the basic such expression, let alone a sufficient guide for making moral judgements in specific circumstances. Furthermore, if in the *Groundwork* Kant suggests that the formula might provide for a provisional classification of given duties as 'perfect' or 'imperfect', he makes no mention of this again, including in *The Metaphysics of Morals*, with which, indeed, he also revises his distinction between 'perfect' and 'imperfect' duties. His discussion in the *Groundwork* can therefore hardly be treated as offering his considered view even of the classification of duties.

Instead, Kant's considered position regarding the place of the formula of universal law in moral judgement is best determined by turning to *The*

³³ G 437, 436. Kant makes similar remarks at G 462-3.

Metaphysics of Morals, as his late and most extensive derivation and classification of duties. A general duty of beneficence is the only duty that he there unambiguously derives from an expression of moral goodness corresponding to the formula of universal law. He classifies this duty as an 'ethical' one, since it concerns an agent's reasons for action, and as an ethical duty of 'virtue' in particular, since it is a duty to have an end.³⁴ In this case, the duty is also a duty of 'love', since the end required is a morally permitted, inclination-based end of another agent which is necessary to her 'happiness', a coherent whole of such ends, and which the benefactor agrees to be so.³⁵ Kant distinguishes this duty of beneficence from other duties of 'love', however, by specifying it as a duty to assist others in need, if one has the appropriate means.³⁶

Kant's derivation of this general duty of beneficence is concisely stated in a single, short paragraph. At first glance, furthermore, it appears to argue precisely as the traditional approach would expect – that is, from the consequences of a maxim's universal adoption – and, indeed, by elimination of the opposite maxim and on the further grounds of self-interest. For Kant writes that if an agent adopts the opposite maxim and 'makes this a universal permissive law: then everyone would likewise deny him assistance when he himself is in need[,] or at least be authorised to deny it'. From this Kant concludes that 'the self-interested maxim would conflict with itself if it were made a universal law', and that 'consequently' the general duty of beneficence holds.³⁷ Apparently, then, Kant derives the general duty of beneficence by arguing that the consequences of universal beneficence are preferable, on the grounds of each agent's self-interest, to those of universal

³⁴ On the nature of an 'ethical' duty and an ethical duty of 'virtue', see *MS* 214, 218-21, 375-6, 379-83, 384-5, 388-9, 394-6, 406-7, and 410.

³⁵ See *MS* 387-8, 393, 401-2, and 448-50, and also *E* 337-9.

³⁶ See *MS* 450-3.

³⁷ *MS* 453.

indifference. Such a derivation would, of course, raise most of the profound difficulties that I mentioned in the first subsection of this chapter. It would also raise controversial questions about the rational self-interest of a policy of beneficence, questions that would threaten the duty's intended universal scope at the very least.

However, in the last part of the paragraph's last sentence Kant strongly suggests that his derivation ought to be read differently. Firstly, he refers to the general duty of beneficence not as another maxim of 'self-interest', now enlightened as to the consequences of universal indifference, but rather as 'the common interest of beneficence towards the needy'. Secondly, he then writes that the general duty of beneficence holds 'because [...all human beings] are to be considered fellow human beings, i.e. needy rational beings, united by nature in one dwelling place for reciprocal help'.³⁸ Rather than the claim that universal beneficence is in every human agent's self-interest, then, Kant's premise is simply a purported fact that every human agent has ends which can be satisfied only with the assistance of other human agents.

By referring to 'common interest' and 'rational beings' in particular, these remarks suggest that Kant's derivation of this duty turns on what is good for a will as such, rather than on an agent's contingent self-interest. They thus suggest what his derivation of the formula of universal law in the *Groundwork* would confirm, namely, that the formula is intended to emphasise the distinguishing feature of moral goodness (ii). Furthermore, that Kant's premise concerns specifically 'human' rational beings, characterised by their 'needy' nature and their vicinity, indicates that his derivation asks what is good for a specifically human will as

³⁸ MS 453.

such, and thus admits the pertinence of conditions characteristic of a human will as such. Thus his objection to the maxim of indifference is not that it could be adopted only on the grounds of an insufficiently enlightened self-interest, but that it could be adopted only on grounds that are contingent for a human will as such. Only because these grounds are 'self-interested' in the latter sense, therefore, is the maxim supposed to 'conflict with itself' as a universal law. Equally, the maxim of beneficence suffers the same failing insofar as it is adopted on such grounds – grounds which include, notably, enlightened self-interest. However, Kant's premise about the 'needy' nature and vicinity of human wills explains how the maxim of beneficence could instead be adopted on grounds which are good for a human will as such. It thus explains how the maxim is a duty.³⁹

This derivation not only avoids the difficulties that the traditional approach raises. It also exhibits an important feature of Kant's conception of moral judgement in general. That is, it indicates that he conceives of moral judgement as judgement according to a system of duties, rather than by direct application of his formulas to specific circumstances. Each duty in this system is to be derived from the two distinctive features of moral goodness (i) and (ii) by admitting certain contingent conditions of wills, and presumably each derivation, and thus each duty, is also to be constrained by those which admit conditions of higher generality. In this regard, it is significant that Kant restricts his derivations in *The Metaphysics of Morals* to those which admit conditions that are common to all human wills, and thus of the highest human generality. As he writes in the introduction, to provide 'a metaphysics of morals', 'we must often take the particular *nature* of human beings, which can be cognised only by experience, as

³⁹ Also compare Kant's derivation of duties of 'love' in general at *KpV* 34-5, and *MS* 393 and 451.

our object, in order to show universal moral principles' conclusions for it'. After having thus completed this 'metaphysics of morals', he adds that further duties would be derived by admitting further 'difference of subjects', such as 'difference of rank, age, sex, state of health, prosperity or poverty and so on'.⁴⁰ Unlike his examples in the *Groundwork* and elsewhere, therefore, the duties derived in *The Metaphysics of Morals* are of the highest human generality, and are not intended to constitute a complete system of duties, or to sufficiently justify any moral judgement in specific circumstances.

IV. Nietzsche's criticisms of the formula of universal law

Nietzsche exemplifies the traditional focus on the formula of universal law and, at least in '*Long live physics!*', its direct application of the formula to reasons for action in specific circumstances. However, unlike the traditional critics, Nietzsche does not concern himself with the formula's success in distinguishing intuitively moral maxims from immoral ones, according to the consequences of a maxim's universal adoption. Indeed, Nietzsche does not even refer to the mechanics of such 'testing' of maxims. Rather, in '*Long live physics!*' he objects that it is 'selfish' to hold, of reasons for action in specific circumstances, that every agent should do or refrain from a type of action in a type of circumstance. He thus echoes another of Schopenhauer's criticisms, which alleges that, under the formula of universal law, an agent is to judge maxims by the consequences of their

⁴⁰ MS 218, 468. See also MS 217 and 469.

universal adoption for her selfish concerns.⁴¹ In his discussion in *The Antichrist*, on the other hand, Nietzsche alleges that to make a judgement on ‘universal’ grounds, or to obey such a judgement, is, particularly in its exclusion of ‘pleasure’, self-destructive, characteristic of ‘ascetic ideals’, and fails to satisfy his particular evaluative concern for ‘humanity’.

However, although these criticisms of the formula of universal law are somewhat different from the traditional ones, it is difficult to make a persuasive case for them. Firstly, in ‘*Long live physics!*’ Nietzsche simply misrepresents the formula. For, as I argued in the preceding chapter, the formula is not intended to express, and nor does its derivation succeed in justifying, only the weak requirement that every agent should do or refrain from a type of action in a type of circumstance.⁴² Rather, the derivation justifies the formula as it is intended, namely, as an expression of the two distinctive features of moral goodness (i) and (ii) which emphasises the latter feature, namely, that moral goodness is good for a will as such. Furthermore, Kant does not intend the formula to be applied directly to reasons for action in specific circumstances, as Nietzsche implies in ‘*Long live physics!*’. I have argued in the preceding subsections that there is no indication of such an intention in Kant’s remarks regarding the formula in the *Groundwork*, and that Kant’s considered position, expressed by his employment of the formula in *The Metaphysics of Morals*, treats moral judgement as made according to a certain system of duties, derived from features (i) and (ii).

Secondly, and more importantly, the ‘selfishness’ criticism made in ‘*Long live physics!*’ is misplaced. Even if Nietzsche’s interpretation of the formula is

⁴¹ See Schopenhauer, *Über das Fundament der Moral*, pt.II, §7, and *Die Welt als Wille und Vorstellung*, App..

⁴² This inaccuracy of Nietzsche’s discussion at FW 335 is noted in Du Plessis, ‘Nietzsche’s Use and Abuse of Kant’s Philosophy’, and Kerckhove, ‘Nietzsche’s Critique of Kantian Moral Autonomy’, pp.26 and 28.

accepted, a judgement that every agent should do or refrain from a type of action in a type of circumstance need not be made on selfish grounds. Equally, if the traditional interpretation of the formula as a 'test' is adopted, maxims' purposes need not be restricted to selfish ones, insofar as purposes are pertinent at all. Thus Schopenhauer's 'selfishness' criticism is equally misplaced. Furthermore, understood rightly, the formula emphasises feature (ii), that moral goodness is good for a will as such, and thus precludes any ultimate appeal to self-interest as contingent for a will. Contingent conditions of a will are morally pertinent, I suggested above, only as premises in systematic derivations from features (i) and (ii), and, once again, there is no reason to restrict these conditions to selfish ones. Given this, Nietzsche's apparent further claim in '*Long live physics!*' that it is selfish to obey, rather than simply to make, a Kantian moral judgement is equally implausible. For again, whether the formula is interpreted in Nietzsche's terms, Schopenhauer's terms of the traditional 'test', or rightly as an expression emphasising (ii), the motivation to obey a judgement made according to it need not be restricted to selfish ones.

Still, Nietzsche's later discussion in *The Antichrist* notably displays a better appreciation of the formula's emphasis on feature (ii), by expressing a Kantian moral judgement as one made according to the "good in itself", good with the character of impersonality and universality'. This could explain why there Nietzsche also does not restate the 'selfishness' criticism. The discussion in *The Antichrist* is also notable for its failure to present the formula as intended to be applied directly to reasons for action in specific circumstances. In this later discussion, then, Nietzsche avoids the misinterpretations and the misplaced criticism that are expressed in '*Long live physics!*'. One might speculate that these

improvements reflect Nietzsche's reading of Kant in the library at Chur, in the preceding year.

Despite these improvements in his interpretation of the formula, however, Nietzsche's criticism of it in the *The Antichrist* is unconvincing. It alleges that, with respect to the making or obeying of a judgement according to the formula of universal law, Kant 'understands pleasure as [an] objection', and thus requires an agent 'to work, to think, to feel without inner necessity, without a deep personal choice, without *pleasure*'. This is unconvincing for three reasons. Firstly, it makes the mistake of critics since Schiller, by supposing that Kant conceives of acting 'from duty' as doing what is morally good in the presence of countervailing inclinations. As I showed in the preceding chapter, the Schillerian supposition misrepresents the relevant passage of the first section of the *Groundwork*, since there Kant insists only that to act 'from duty' is to do what is morally good because it is morally good, and thus irrespective of whether it also coincides with the satisfaction of inclination. This applies equally to achieving 'a virtue merely from a feeling of respect for the concept "virtue"', as Nietzsche expresses Kantian moral motivation in *The Antichrist*.

Furthermore, Kant conceives of 'pleasure' as extending beyond the satisfaction of inclination, to the achievement of moral goodness. Indeed, he states that human willing 'is always connected with *pleasure* or *displeasure*', and distinguishes simply between (dis)pleasure at the (dis)satisfaction of an inclination and (dis)pleasure at the (lack of) achievement of moral goodness.⁴³ This position is given consistent expression from the *Critique of Judgement* onwards, and contrasts with Kant's earlier restriction of (dis)pleasure to the (dis)satisfaction of

⁴³ MS 211. See also KU 178-9, and 207-9, TP 283-4, and MS 212-3, 378, and 399-400.

inclination.⁴⁴ Thus his considered position is not simply that, contra the Schillerian supposition, the achievement of moral goodness might coincide with the ‘pleasure’ of satisfying coincident inclinations, but also that it is necessarily accompanied by the ‘pleasure’ of achieving moral goodness.

Finally, in the determination of what moral goodness requires, Kant’s formula of universal law does not preclude what provides ‘pleasure’. He conceives of judgement according to this formula as the judgement of what is (ii) good for a will as such, and as therefore ultimately unconcerned with the satisfaction of any inclination. But he does not deny that what is (ii) good for a will as such might nonetheless coincide with what satisfies an inclination, and thus provides ‘pleasure’, and, furthermore, he insists on the pertinence of such satisfactions, insofar as he includes them among the contingent conditions of a will that can be admitted as premises in systematic derivations from (i) and (ii).

If, therefore, Kant does not consider ‘pleasure as [an] objection’ to the making or obeying of a judgement according to the formula of universal law, then Nietzsche’s presentation of such judgement as ‘ascetic’ is also difficult to sustain. According to the ‘revaluation’ provided in the third essay of the *Genealogy*, ‘ascetic ideals’ deny necessary features of human life for the sake of another ‘existence’, supposed to lie beyond these features, and are obeyed from a ‘*ressentiment*’ at the ‘suffering’ of much human life. In his discussion of Kant in *The Antichrist*, I argued, Nietzsche invokes this ‘revaluation’ by alleging Kant’s ‘anti-naturalness’. But if Kant’s formula of universal law does not preclude the satisfaction of inclinations from either the achievement or the determination of moral goodness, it is implausible to describe either this achievement or this

⁴⁴ For Kant’s earlier, restrictive position, see, for instance, *KpV* 9n, 21-2, and 57-62. But note that even at *G* 396 and 401n, and *KpV* 38-9, he refers to a certain feeling at the achievement of moral goodness.

determination as 'ascetic'. This, in turn, threatens Nietzsche's negative 'revaluation' of professed obedience to Kantian moral reasons according to his particular concern for 'humanity', assuming, as seems plausible, that the formula of universal law's supposed 'asceticism' underlies this negative 'revaluation'.

However, although Kantian moral goodness does not preclude 'pleasure' in the 'ascetic' manner that Nietzsche alleges, it also does not attend to it in the manner that Nietzsche proposes. That is, unlike Nietzsche, Kant does not restrict normative authority to reasons relative to an agent's particular concerns, since, on the grounds of his commitment to (ii), he extends it to unconditional and overriding reasons. Indeed, in the immediately following section of *The Antichrist*, Nietzsche objects precisely to this extension of normative authority, albeit again on the unconvincing grounds that it is 'ascetic'. I proposed in my examination of the fifth part of *Beyond Good and Evil*, however, that this objection is better understood as deriving from Nietzsche's denial of practical reason, on the grounds that reasons must be relative to an agent's concerns, and that these concerns are particular because the noncognitive motivating states which determine them are particular. If Kant is vulnerable to Nietzsche's denial of practical reason, therefore, his commitment to (ii) is also vulnerable. I will consider this matter in the third section of this chapter.

Kant's further formulas

In this section I complete my account of the intended meaning of Kant's formulas by considering his formulas of the end in itself, autonomy, and the realm of ends. I first argue that my account of the derivation of these formulas implies that certain common treatments of them ought to be treated with some suspicion, if not rejected outright. I then consider another of Kant's derivations of general human duties in *The Metaphysics of Morals*, as an example of his considered position regarding his further formulas which also exemplifies the distinctive conception of 'community' that he associates with them. Although Nietzsche's explicit criticisms of Kant do not refer to formulas other than the formula of universal law, in the final subsection I briefly suggest that, given the preceding, Kant ought to be unmoved by Nietzsche's affirmations of individual 'autonomy', and, in particular, by Nietzsche commentators' allegations that Kant betrays such 'autonomy'.

I. Kant's further formulas

Commentators often treat Kant's further formulas as expressing moral considerations which are not expressed by the formula of universal law, and recent commentators, who generally admit that problems of 'emptiness' attach to the traditional approach to the formula of universal law, also often suppose these further moral considerations to be more substantial than those expressed by the

formula of universal law.⁴⁵ Nonetheless, although Kant again discusses or refers to his four examples in the light of two of his further formulas, commentators rarely consider these further formulas to provide a thorough ‘test’ of reasons for action in specific circumstances. Instead, these formulas are usually treated as merely expressing certain general moral considerations which, although general, provide further and perhaps more substantial guidance to moral judgement than those expressed by the formula of universal law.

The formula of the end in itself is often considered to be the most salient in this regard, since its reference to ‘rational nature’ as a supreme moral consideration has traditionally been treated as an exemplary principle of ‘respect for persons’. Such ‘respect’ has an important place in moral common sense and in much moral philosophy, and Kant’s formula has been interpreted relatively broadly in this manner – that is, as requiring not only the ‘consent’ of those affected by one’s actions, but also, for instance, the preservation, exercise, and improvement of agency generally, the appeal to reason in attempts to persuade others, the non-interference in and even beneficent support of others’ morally permissible choices, and the avoidance of actions which dishonour or degrade oneself or others.⁴⁶ However, commentators’ supposition that Kant’s further formulas express more substantial moral considerations than the formula of universal law has often instead issued in treatments of these formulas as together

⁴⁵ Among recent commentators, for instance, Korsgaard presents the formula of universal law as weaker than the later formulas. Pogge presents each further formula as providing further specifications of ‘the categorical imperative’, Hill has claimed that the formula of universal law is primarily negative, while the later formulas provide further positive content, and Wood maintains that the later formulas substantiate the otherwise ‘empty’ formula of universal law. See Korsgaard, ‘The Right to Lie’, pp.135-44 and 151-4, Pogge, ‘The Categorical Imperative’, pp.189, and 198-204, Hill, ‘Reasonable Self-Interest’, pp.155-7, and Wood, *Kant’s Ethical Thought*, pp.107-110, 164-166, and 182-190.

⁴⁶ For such broad interpretations of the formula, see, in particular, Hill, ‘Humanity as an End in Itself’, pp.93-7, and Wood, ‘Humanity as End in Itself’, pp.177-84, and *Kant’s Ethical Thought*, pp.141-2 and 147-55.

expressing a notion of the rational, procedural 'construction' of moral rules. In this manner, commentators often interpret the formula of the end in itself as requiring agents' possible or hypothetical 'consent' to moral rules under certain constraints of rationality which the other formulas express, and Kant's further formulas as thus substantiating the formula of universal law by identifying substantial constraints on what 'could also be willed as a universal law'.⁴⁷

Besides these two common interpretative approaches, some recent treatments of Kant's further formulas have been informed by doubts about the traditional classification of his moral philosophy as an exemplary 'deontological' one. Kant's further formulas have traditionally been supposed to demonstrate particularly clearly that he identifies moral requirements as constraints on the pursuit of goods, such that the moral 'right' must take precedence over the non-moral 'good'. His moral philosophy is thus supposed to be fundamentally distinct from 'teleological' moral philosophies, which treat moral requirements as requiring the pursuit of distinctively moral goods, such as general 'utility' or human 'flourishing'.⁴⁸ Recently, however, some commentators have taken issue with this classification of Kant's moral philosophy, by arguing that his further formulas identify a substantial moral good as the further 'ground' of his moral philosophy. In particular, Guyer maintains that Kant's further formulas identify a good as the teleological 'end' to which the formula of universal law expresses the 'means', and Wood holds that these further formulas identify a good as the source

⁴⁷ For Rawls's influential statement of such an interpretation, see, in particular, his 'Kantian Constructivism in Moral Theory', and also his 'Themes in Kant's Moral Philosophy', pp.82-90, and *Lectures on the History of Moral Philosophy*, esp. pp.164-76 and 237-52. O'Neill is another important example. See, for example, her 'Constructivisms in Ethics', esp. pp.212-8.

⁴⁸ For this distinction between 'deontological' and 'teleological' moral theories, see, for example, Frankena, *Ethics*, pp.14-17.

of the goodness of moral 'ends' and laws.⁴⁹ Others, notably Herman, insist that Kant's further formulas express a substantial good of rational agency itself, and thus justify a commitment to the formula of universal law.⁵⁰

For themselves, such positions regarding Kant's further formulas present fewer difficulties than the traditional approach to the formula of universal law. For instance, insofar as a principle of 'respect for persons' is not considered to provide an algorithmic 'test' of reasons for action in specific circumstances, it avoids problems in determining the specific reason to 'test', and insofar as it is not considered to provide a complete or even supreme expression of moral goodness, it avoids the implausible exclusion of other pertinent considerations.⁵¹ However, as interpretations of Kant's further formulas, the prevailing positions fail to capture the implications of his derivations, as I have presented these derivations in the preceding chapter. Most broadly, these derivations imply that Kant's further formulas are not intended to express moral considerations which are not already expressed by the formula of universal, let alone moral considerations which are more substantial or which identify an otherwise absent 'ground'. Indeed, Kant intends no formula to express moral considerations which are not also expressed by any other formula. More specifically, Kant's derivations also imply that his further formulas are intended to express the two distinctive features of moral goodness (i) and (ii), and to emphasise different aspects of these features. Thus he does not intend any formula to be interpreted only in its own terms – for example, by simply interpreting the notions of 'rational nature', or 'humanity', and of 'ends'

⁴⁹ See Guyer, 'Kant's Morality of Law and Morality of Freedom', pp.145-6, 151-4, and 159-61, 'The Possibility of the Categorical Imperative', pp.182-3 and 191-200, and 'Ends of Reason and Ends of Nature', pp.162-4, and Wood, 'Humanity as End in Itself', p.166, *Kant's Ethical Thought*, pp.111-4, 156-9, 163, and 183-6, and 'The Final Form of Kant's Practical Philosophy', pp.12, 13, and 16.

⁵⁰ See Herman, 'Leaving Deontology Behind', esp. pp.224-30.

⁵¹ On the latter point, see the discussion in Hill, 'Donagan's Kant'.

and 'means' that the formula of the end in itself employs. And although (i) and (ii) might be consistent with certain understandings of 'respect for persons' or the rational, procedural 'construction' of moral rules, they are not exhausted by them.

With regard to each particular further formula, the implications of Kant's derivations are as follows, I would suggest. Firstly, the formula of the end in itself is intended to emphasise feature (i), that moral goodness is a goodness of a will's reason for action alone, because it is (b') a goodness which only a will can achieve. It emphasises this by expressing moral goodness in terms of what a will takes to provide a reason for action, and, in particular, as the goodness of a will's taking a will itself to provide a reason for action. The formula of the end in itself is therefore not intended to express a conception of 'respect for persons' or the rational 'construction' of rules, or, indeed, intended to be interpreted on its own terms. Rather, it is simply intended to express (i) and (ii), and to emphasise (i), in the manner that Kant's derivation indicates. This is not to deny that a will's taking a will itself to provide a reason for action might coincide with its 'respect for persons', insofar as 'persons' might be conceived as 'wills', or with its acting according to rationally 'constructed' rules, insofar as wills might be conceived as providing reasons for action precisely by such rational 'construction'. But the meaning of 'a will's taking a will itself to provide a reason for action' need not be restricted to such 'respect for persons' or rational 'construction', or to any other principle that coincides with it, and is to be determined only with reference to (i) and (ii) and Kant's derivation according to them.

This understanding of the formula of the end in itself, as implied by my interpretation of the formula's derivation, also finds support in Kant's discussion of his four examples in the light of this formula in the *Groundwork*. As with the

formula of universal law, Kant suggests that the formula of the end in itself might treat examples in two different ways, but again provides no indication that he intends this to be anything more than a provisional classification. Here his suggestion is that some kinds of contravening actions are such as to ‘conflict’ with ‘rational being’ as an ‘end’, while others are such as to merely fail to ‘harmonize’ with this ‘end’.⁵² In discussing the example of a false promise as an example of the former kind, he claims that the promisee ‘cannot possibly agree’ to the false promisor’s action. This claim might suggest a narrow concern for possible or hypothetical ‘consent’, and thus a narrow principle of ‘respect for persons’ or a notion of the rational ‘construction’ of moral rules. However, Kant also treats the impossibility of the promisee’s ‘agreement’ as such that the promisee cannot ‘contain’ the action’s ‘end’, and, in discussing the example of suicide as another action which ‘conflicts’ with ‘rational being’ as an ‘end’, he also claims that the suicide’s action cannot ‘exist together with’ taking ‘rational being’ as an ‘end’. In discussing the examples concerning beneficence and the development of natural predispositions as examples of actions which merely fail to ‘harmonize’ with ‘rational being’ as an ‘end’, Kant also writes that such actions cannot be ‘consistent [...] with the promotion’ of this ‘end’, and that they are not consistent with this end’s having ‘its *full* effect in me’.⁵³ These terms are consistent with a broader principle of ‘respect for persons’ and also might accommodate a notion of the rational ‘construction’ of moral rules. However, it is significant that Kant’s discussion of each example focuses on whether ‘rational being’ is a will’s ‘end’, rather than on whether an action independently preserves or enhances ‘person’-hood or could be ‘legislated’ under certain constraints of rationality. This suggests

⁵² G 430. Here Kant also states the distinction in terms of whether the ‘agreement’ between an action and ‘rational being’ as an ‘end’ is ‘negative’ or ‘positive’.

⁵³ G 429, 430.

that Kant is concerned to express something other than 'respect for persons' or the rational 'construction' of moral rules, and implies that the formula of the end in itself can accommodate such 'respect' and 'construction' only insofar as they are conceived in terms of a will's 'end'. Furthermore, Kant's focus on a will's 'end' is better explained by his concern with emphasising (i), that moral goodness is a goodness of a will's reason for action alone, because it is (b') a goodness which only a will can achieve, than by a concern with expressing a principle of 'respect for persons' or a notion of rational 'construction'.

Similar conclusions should be drawn regarding the formulas of autonomy and the realm of ends. According to the account of its derivation which I have presented, the formula of autonomy is intended to emphasise both (i) and (ii) and, in doing so, to warn against misinterpreting the preceding formulas in ways which would fail to appreciate these two distinctive features of moral goodness. The formula of the realm of ends, on the other hand, is intended to note a further qualification which is bracketed by the other formulas. Like the formula of the end in itself, therefore, neither the formula of autonomy nor the formula of the realm of ends is intended to express a notion of the rational 'construction' of moral rules, although either formula might coincide with such 'construction', insofar as it might be conceived precisely in terms of a will's 'giving universal law', or the provision of 'common laws'. However, once again, the meanings of these formulas need not be restricted to such 'construction', and are to be determined only with reference to (i) and (ii).⁵⁴ Although he does not again discuss his four examples, Kant's remarks regarding these formulas support these conclusions,

⁵⁴ Hill, for instance, has presented the formula of the realm of ends as combining the other formulas in a notion of the perspective from which rational 'construction' ought to proceed. See his 'The Kingdom of Ends', pp.59-62, 'A Kantian Perspective on Moral Rules', pp.35-51, 'A Kantian Perspective on Political Violence', pp.223-30, and 'Hypothetical Consent in Kantian Constructivism', pp.73-6.

insofar as they reveal no concern with identifying a procedure of rational 'construction', but rather again emphasise, besides the concerns of his derivations, the will's taking 'rational beings' as 'ends'.⁵⁵

Finally, my account of the derivations of Kant's formulas supports recent suspicion of the traditional 'deontological' classification of his moral philosophy, but does not suggest that he intends to 'ground' his moral philosophy in a substantial moral good. As I have presented his derivations, Kant's formulas are intended to express the goodness of a good will, as he finds this conceived by 'common moral rational cognition' and as he analyses it in the first section of the *Groundwork*. By presenting moral requirements as the requirements of this goodness, therefore, he does not present them as 'deontological' constraints on the pursuit of goods. However, he distinguishes this moral goodness from other goods by means of its formal features – namely, (i) and (ii) – rather than in substantial terms. In particular, while Kant would deny the 'moral' status of general 'utility', human 'flourishing', or other supposed moral goods, he would not do so by invoking a competing substantial moral good. Despite the positive proposals of recent sceptics of the 'deontological' classification, then, Kant intends to express no substantial moral good, either in his further formulas or elsewhere.

II. *Vulnerabilities of agency: Political obligation in The Metaphysics of Morals*

Like that of the formula of universal law, the intended place of Kant's further formulas in moral judgement is best determined by turning to *The*

⁵⁵ See *G* 433.

Metaphysics of Morals. However, in presenting the formula of the realm of ends in the *Groundwork*, Kant makes a notable remark that anticipates his later, more extensive derivation of duties. This remark concerns what he calls a ‘sovereign [*Oberhaupt*]’ in a realm of ends, the function of which is not explicitly stated, but is presumably to judge moral goodness among a plurality of agents.⁵⁶ To fulfil a sovereign’s function, Kant claims that an agent must be ‘not subject to the will of any other’, in the sense that the agent must be ‘a completely independent being[,] without need and limitation of abilities adequate to the will’.⁵⁷ The significance of this claim lies not so much in its implications for a sovereign, but rather in its implication that other agents are not necessarily ‘independent beings’ in the sense defined, and its suggestion that the lack of such independence is pertinent to determining the requirements of moral goodness among them. Furthermore, given that, with the formula of the realm of ends, Kant emphasises that moral goodness consists of extending to every will the provision of reasons by mere will, the lack of such independence must be a matter of an agent’s will being ‘subject to’ another agent’s will – rather than, say, the achievement of an agent’s desires, needs, or other particular goods being ‘subject to’ the achievement of another’s. In other words, Kant’s claim about a sovereign suggests that what is pertinent to determining the requirements of moral goodness among a plurality of agents – and thus, indeed, what constitutes them as a ‘realm’ at all – is only and precisely the mutual vulnerabilities of their wills. Finally, however, Kant’s claim also identifies ‘need’ and ‘limitation of abilities’ as sources of such vulnerabilities. Thus he

⁵⁶ This function is suggested by *G* 439, where Kant refers to a sovereign as a ‘lawgiver’ who ‘judges the worth of rational beings’. See also *R* 95-100, and compare the usages, with which Kant was acquainted, in Leibniz, *Discours de Métaphysique*, §36, *Principes de la nature et de la grace fondés en raison*, §15, and *Principes de la philosophie ou monadologie*, §§85-90, and particularly Rousseau, *Du Contrat Social*, bk.1, chs.6-7.

⁵⁷ *G* 433, 434.

allows that factors other than wills as such, and therefore not morally pertinent in themselves, can be morally pertinent insofar as they are sources of mutual vulnerabilities of wills. Furthermore, since Kant consistently refers to human agents in general as having ‘needs’ and being limited in their practical abilities, his claim also implies that the requirements of moral goodness among human agents generally must be determined precisely according to the general mutual vulnerabilities of their wills.⁵⁸

This, I propose, is precisely what Kant undertakes in *The Metaphysics of Morals*. That is, with the exception of the general duty of beneficence, Kant derives general other-regarding duties among human agents from his further formulas – that is, those which emphasise feature (i) of moral goodness – along with premises regarding the general mutual vulnerabilities of human wills. This is well demonstrated by his derivation of the most substantial of the basic political obligations which he derives in *The Metaphysics of Morals*, those regarding the original acquisition of material things. This derivation is somewhat more complex than that of the general duty of beneficence, and is obscured by textual corruption.⁵⁹ Nonetheless, the following – arguably, highly distinctive – argument can be discerned.

A few preliminary points should first be noted. Firstly, Kant considers an agent’s political obligations to be obligations which other agents can coerce the agent to fulfil, and therefore obligations merely to do or refrain from certain kinds of actions, rather than, like ‘ethical’ obligations, obligations to do so for certain

⁵⁸ On human ‘needs’ and limited abilities, see, for instance, *G* 413n and 414. Of course, that human agents in general have ‘needs’ and limited abilities also problematises their possible fulfilment of a sovereign’s function. But Kant does not emphasise this implication, or take it up elsewhere, and he apparently introduces the ‘sovereign’ here only to illuminate, by contrast, an aspect of his conception of the ‘realm of ends’.

⁵⁹ Regarding the textual issues here, I follow Ludwig’s reconstruction of the text, as defended in his *Kants Rechtslehre*, pt.1.

kinds of reasons.⁶⁰ Nonetheless, he also insists that an ‘indirect’ ethical obligation corresponds to each political obligation, the former being the obligation to fulfil the latter for the certain kinds of reasons.⁶¹ Secondly, Kant maintains that the basic principle of political obligations is the ‘Universal Principle of Right [*Recht*]’: ‘Any action is *right* if it[,] or according to its maxim[,] the freedom of choice [*Willkür*] of each can exist together with everyone’s freedom according to a universal law’.⁶² The nature and grounds of this principle are subject to much scholarly controversy. It is variously claimed, for instance, that Kant’s principle expresses a voluntaristic, ‘social contract’ theory, a non-voluntaristic, ‘natural law’ theory, an independent analysis of the concept of ‘right’, or a commitment to a substantive moral good presented by the formula of the end in itself.⁶³ However, I will argue that the place of the principle in Kant’s derivation of the most substantial of our basic political obligations better supports an alternative interpretation – namely, that he intends this principle to express, with regard to coercible obligations, precisely the conception of a ‘realm’ of mutually vulnerable wills which he briefly expresses in the *Groundwork*.⁶⁴ Finally, the object of this derivation is the original acquisition of material things. Notably, Kant conceives of possession in general as an agent’s right to use an external object – either a material thing or another’s agency – as she

⁶⁰ See *MS* 214, 218-21, 231, 239, 375, and 382, and also *R* 95-6.

⁶¹ See *MS* 219-220.

⁶² *MS* 230. See also *R* 98, *TP* 290, and *MS* 231, 375, 382, and 396.

⁶³ For examples of each of these interpretative claims, see, respectively, Murphy, *Kant*, esp. ch.4, and Rosen, *Kant’s Theory of Justice*; Mulholland, *Kant’s System of Rights*, esp. chs.7-10, and Kersting, *Wohlgeordnete Freiheit*, both of which also flirt with ‘social contract’ interpretations; Willaschek, ‘Why the *Doctrine of Right* Does Not Belong in the *Metaphysics of Morals*’, and ‘Which Imperatives for Right?’, pp.67-9 and 75-85; Pogge, ‘Is Kant’s *Rechtslehre* a “Comprehensive Liberalism”?’’, pp.136-46, and Wood, ‘The Final Form of Kant’s Practical Philosophy’, pp.5-10, and *Kant’s Ethical Thought*, pp.321-3; and Guyer, ‘Kant’s Deductions of the Principles of Right’, esp. pp.23-7.

⁶⁴ This interpretation is also supported, I would suggest, by Kant’s presentation of his Universal Principle at *MS* 230, 232-3, and 237-8, but I will not consider these passages here.

pleases, and irrespective of whether she is physically ‘holding’ the object.⁶⁵ He also argues that the possibility of such a right follows from the Universal Principle of Right because this principle concerns mere will and willing requires the physical ability to use its objects.⁶⁶

However, Kant provides a more complex argument regarding the possibility of originally acquiring a material thing, and it is with this argument that he appeals to, and illuminates, his conception of a ‘realm’ of mutually vulnerable wills. This argument proceeds from the Universal Principle and willing’s requirement of the physical ability to use its objects, along with a crucial further claim about the possession of a material thing – namely, that a movable material thing ‘is to be regarded [normatively...] as [an] *inherence*’ in its ‘place’, the habitable ground on which it rests. From these premises, Kant first argues that ‘damage is done’ to an agent’s ‘freedom’ not by another’s taking of an un-‘held’ material thing as such, but only insofar as this taking of the thing upsets its ‘place’. He claims that the possibility of originally acquiring a ‘place’, and so the material thing which rests on it, ‘is thus based on’ willing’s requirement of the physical ability to use its objects.⁶⁷ Presumably, then, Kant’s argument here is that, although the physical ability to use a material thing is not upset by another agent’s taking the thing as such, it is upset by the thing’s thus not having a consistent ‘place’, and therefore, since willing requires the physical ability to use its objects, the Universal Principle requires the possibility of originally acquiring a material thing along with the ‘place’ on which it rests.

⁶⁵ MS 246. See also MS 247-9.

⁶⁶ See MS 246 and 257. See also MS 247, 249-50, and 255. Kant refers to the latter requirement of willing throughout his texts. See, for instance, his distinction of choice from mere ‘wish’ at G 394 and 435, MAM 122, KU 177n, MS 213, 230, 246, 356-7, and 451-2, ApH 251, and EKV 230n..

⁶⁷ MS 261, 262. Kant also appeals to this requirement in this regard at MS 263 and 267, and states his conclusion at MS 261, 263, and 269-70.

Kant proceeds to further explain this possibility, and to argue for a significant constraint on it. Specifically, he argues that a human agent can originally acquire a material thing ‘only through the united choice of all who possess it in common’. This is so, he maintains, ‘on account of the unity of all places on the earth’s surface as spherical surface: because, if it were an unending plane, human beings could be so dispersed on it that they would not come into any community with each other, [and] this [community] therefore would not be a necessary result of their existence on the earth’. He concludes that each ‘place’ on the earth must be considered to be, prior to any acquisition, possessed in common by human beings, and that the original acquisition of a material thing, along with its ‘place’, is therefore both made possible and constrained by ‘the uniting of the choice of all who can come into a practical relation with each other’.⁶⁸ Although far from unambiguous, these remarks can be taken to indicate the following, I suggest. Firstly, for Kant, it is the finitude of ‘places’ on the earth’s surface which ensures that human beings’ wills are vulnerable to each other’s taking of material things, in the sense that such taking upsets the consistency of things’ ‘places’ which the use of objects requires. Secondly, Kant takes this vulnerability to entail not only the possibility of originally acquiring a material thing, but also a substantial constraint on this possibility – namely, that an agent’s original acquisition of a material thing not upset the required consistency of things’ ‘places’. In Kant’s terms, it is as if, precisely because they are vulnerable in this way, human agents constitute a ‘community’ which first possesses each ‘place’ in common and which can grant the private acquisition of a ‘place’, and thus the private acquisition of any material thing which rests on it, to a specific agent only

⁶⁸ *MS* 262, 263. See also *MS* 264 and 267-8.

if this ensures, rather than upsets, the required consistency of things' 'places'. Furthermore, Kant holds that human agents' basic political obligations then follow from the possibility of possessing external things and originally acquiring material things, on the grounds that a civil condition is required to secure possession and to further determine the original acquisition of material things.⁶⁹

The precise meaning and ultimate success of Kant's derivation of these basic political obligations are, of course, far from obvious. In particular, the meaning of a material thing's 'place' and its intended role in the argument requires substantial further examination and support. However, I hope to have demonstrated only that this derivation appeals precisely to the conception of 'realm', or, better, 'community', which Kant briefly presents in the *Groundwork* and to which he again refers in introducing his Universal Principle of Right in *The Metaphysics of Morals* – namely, the conception according to which other-regarding obligations among a plurality of agents must be determined precisely according to the mutual vulnerabilities of their wills. I would also suggest that, besides confirming the brief intimations of the *Groundwork*, Kant's derivation makes explicit that, according to this conception, a 'community' is constituted only and precisely according to a specific mutual vulnerability of wills, and therefore extends only and precisely as far as this vulnerability extends. With regard to our basic political obligations, Kant considers a vulnerability, and thus a 'community', which concerns the consistency of material things' 'places', and which, given the 'finitude' of 'places' on the earth's surface, therefore extends across the earth's surface. He explicitly denies, however, that human willing is

⁶⁹ See MS 255-7 and 264-7. The premise regarding the earth's surface is similarly emphasised by Flikschuh and Thomson, although both mistakenly extend it to Kant's argument for the mere possibility of possession in general. See Flikschuh, *Kant and Modern Political Philosophy*, chs.4-5, and Thomson, 'Kant's Transcendental Deduction of Political Authority'.

generally vulnerable to agents' taking of un-'held' material things as such, and thus effectively denies that any other material constraint – such as material scarcity or material inequality – is pertinent to the determination of our basic political obligations, or equivalently, the general human 'community'.

Kant's derivation also has three broader implications. Firstly, although it appeals to formulas of moral goodness which emphasise (i), this derivation would be clumsily interpreted in terms of a notion of 'respect for persons' or the rational 'construction' of moral rules. To this extent, it confirms the strong suspicions of such interpretations of Kant's further formulas which I raised in the preceding subsection. Secondly, like the derivation of a general duty of beneficence, this derivation does not proceed by applying the formulas to reasons for action in specific circumstances, but rather by considering a certain general condition of human wills under features (i) and (ii) of moral goodness. It thus contributes to the system of duties that Kant proposes to derive. Thirdly, in employing his distinctive conception of 'community', this derivation exemplifies every other derivation of a general other-regarding duty in *The Metaphysics of Morals*, with the exception of the general duty of beneficence, and thus also reveals the salience of this conception as an expression of moral goodness, for Kant.

III. *Kant and Nietzsche on autonomy*

As I demonstrated in the first part of the thesis, Nietzsche's explicit criticisms of Kant's moral philosophy do not refer to Kant's further formulas, but Nietzsche does propose an unspecified individualist conception of 'autonomy' in

opposition to Kant. In the literature on the relationship between Kant's and Nietzsche's moral philosophies, this proposal is often emphasised and, further, transformed into an allegation that Kant betrays 'autonomy'. The 'autonomy' concerned is generally conceived a little more specifically than Nietzsche does, as an individual's choice of, or according to, her particularity or singularity, defined in terms of the individual's drives or desires and supposed to be created, discovered, selected, or organised by the individual herself.

For Kant, such 'autonomy' is of unlikely moral value. In particular, he conceives of moral goodness as a goodness (i) of a will's reason for action alone that is (ii) good for a will as such, and considers his own formula of autonomy to emphasise both of these distinctive features. Admittedly, these features might coincide with 'autonomy' as Nietzsche and his commentators conceive of it – just as, for instance, the formula of the end in itself might coincide with some common notions of 'respect for persons'. But, as such, to choose one's particularity or singularity, or to choose according to it, is a goodness neither (i) of a will's reason for action alone nor (ii) for a will as such, and is thus denied basic moral goodness on Kant's terms.

However, to claim not simply that Kant denies moral value to individual 'autonomy' – which is, I think, all that Nietzsche's remarks suggest – but that Kant *betrays* 'autonomy' is to suppose that he wrongly obscures, neglects, or contravenes a shared commitment to individual 'autonomy'. This, admittedly, would introduce critical bite into Nietzsche's affirmations of individual 'autonomy' against Kant. But there is nothing in Kant's conception of moral goodness to indicate any such shared commitment. He certainly identifies moral goodness with a will's 'giving itself laws' or taking will itself as an 'end', and,

indeed, with (b') a goodness that only a will can achieve, and thus shares some expressions and terminology with conceptions of individual 'autonomy'. But, on closer inspection, the goodness identified abstracts precisely from any agent's particularity or singularity, on the grounds of (i) and (ii) mentioned above. Kantian 'autonomy', then, is simply something other than the individual 'autonomy' to which Nietzsche's remarks refer, some of his commentators enthusiastically appeal, and, indeed, much contemporary discussion of 'autonomy' is directed.⁷⁰

⁷⁰ The divide between Kantian 'autonomy' and contemporary, individualist conceptions has often been emphasised by O'Neill. For a recent polemic, see her 'Autonomy', esp. pp.1-6.

Thus far, I have argued that Nietzsche's specific criticisms of Kant, and Nietzsche's often emphasised affirmations of individual 'autonomy', do not trouble Kant's conception of the moral agent, rightly understood. In this, final section of the chapter, I will argue that Nietzsche's critical treatment of the common sense conception of agency, which he implicitly directs at Kant, is more telling. I begin with Nietzsche's criticism of practical reason and how Kant's conception of moral judgement might accommodate it. I then consider Kant's definition of 'choice', and argue that, although it would allow Kant to accommodate Nietzsche's criticism of practical reason, it is also vulnerable to Nietzsche's criticism of the common sense claim regarding choice. Nonetheless, I also argue that accommodating neither criticism would threaten Kant's conception of moral goodness. Finally, I argue that, given Kant's methodological commitment to common sense, he ought to be unmoved by Nietzsche's proposed project of 'revaluing' given obedience to reasons according to his own particular concern for 'humanity', or by the project, often mistakenly attributed to both Kant and Nietzsche, of criticising moral common sense on independent and objective grounds.

I. *Reasons' underdetermination of actions*

In '*Long live physics!*', Nietzsche addresses his broad criticism of practical reason to Kant's conception of moral judgement in particular. To recall, Nietzsche's argument in this passage is that a reason prescriptively underdetermines action because it prescribes a general type of action in a general type of circumstances, while an action and its circumstances are singular. He concludes that reasons could not motivate actions, and that judgements about reasons for action – such as Kantian judgements of 'moral worth' – are therefore inappropriate. Here as elsewhere, he also consistently accounts for actions as motivated not by reasons, but by noncognitive states informed by reasons, a tendency which, I suggested, implies an argument by elimination: if actions are not motivated by reasons, then they are motivated by noncognitive states informed by reasons. In considering his criticism of practical reason, I also suggested that – on pain of not admitting any causal description of action at all, and his own descriptions in terms of noncognitive states in particular – Nietzsche ought to insist merely on the particularity, rather than the singularity, of actions and their circumstances. Thus his criticism would turn on the relative generality of reasons' prescriptions, noncognitive states, and actions and circumstances. But, even thus modified, Nietzsche's criticism presents a significant challenge to Kant's conception of moral judgement.

This kind of challenge is rarely considered in the Kant literature, although the sceptical implications of the indeterminacy of principles have been emphasised in contemporary moral philosophy more broadly. Nonetheless, as Onora O'Neill has emphasised in recent work, Kant himself responds to a challenge of this kind

by, firstly, admitting that the description of motivating reasons is unfeasible because reasons prescriptively underdetermine actions, and, secondly, insisting that the prescription of action according to reasons can nonetheless proceed.⁷¹

Kant offers the first, conciliatory element of his response most clearly at the beginning of the second section of the *Groundwork*. There he claims that any pattern of observed actions underdetermines the reasons according to which the actions were performed. In particular, he develops his claim in the first section that it is ‘difficult to notice’ whether an agent does what is morally good ‘from duty’ or from an ‘immediately’ related, coincident inclination, and that action ‘from duty’ therefore ‘shines forth all the more brightly’ in the presence of countervailing inclinations. At the beginning of the second section, Kant admits that the problem runs deeper: he states that, even if an agent does what is morally good in the apparent absence of coincident inclinations, ‘it cannot be inferred with certainty that no secret impulse of self-love [...] was not actually the real determining cause of the will, for [...] in fact we can never, even by the strictest examination, get entirely behind the secret incentives, because, when moral worth is in question, what matters is not actions, which one sees, but those inner principles of them which one does not see’. Indeed, as ‘a cold-blooded observer’, Kant’s hyperbolic doubt is ‘whether any true virtue is to be found in the world’.⁷² Kant’s claim can, of course, be extended to non-moral actions and reasons, and his terminology of ‘inference’, ‘examination’, ‘sight’, and ‘observation’ implies that he is concerned not so much for the introspection of motivating reasons as for the ascription of motivating reasons on the evidence of observed actions. Furthermore, if no pattern of actions could confirm or disconfirm a motivating reason, this is because no

⁷¹ For O’Neill’s account, see ‘Consistency in Action’, pp.83-5, ‘Universal Laws and Ends in Themselves’, pp.130 and 141, ‘Kant’s Virtues’, pp.89-97, and ‘Instituting Principles’, pp.331-43.

⁷² *G* 407.

reason's prescription is sufficiently determinate. In this sense, then, Kant anticipates Nietzsche's claim in '*Long live physics!*' that, in 'looking back upon' actions as evidence of motivating reasons, actions are 'impenetrable' or 'unknowable' because a reason underdetermines action.⁷³

However, Kant insists that, while this problem makes the descriptive ascription of motivating reasons to agents unfeasible, it does not also threaten the prescription of action according to reasons – what Nietzsche refers to as 'looking forward to' an action according to reasons. Having denied any certain 'inference' from observed actions to their motivating reasons, Kant insists that, nonetheless, 'what matters here is not whether this or that happened, but [that] reason for itself and independently of all appearances [that is, all retrospective ascriptions] commands what should happen'. Therefore, he concludes, 'e.g. pure honesty in friendship can be no less required [even] if until now there may have been no honest friend'.⁷⁴ That is, Kant insists that, although reasons underdetermine actions, reasons prescribe nonetheless, and this function is not threatened by the uncertainty of descriptively ascribing motivating reasons to agents.⁷⁵ Furthermore, as he makes clear in *The Metaphysics of Morals*, Kant conceives of moral judgement such as to reduce the indeterminacy of prescription. I have shown in the preceding sections that he conceives of moral judgement as judgement according to a system of duties of varying generality, which are to be derived by considering wills' contingent conditions, of varying generality, in the light of the features of moral goodness (i) and (ii). More importantly for the determinacy of prescription, Kant conceives of moral judgement as the mutual constraining of these multiple

⁷³ Nietzsche also refers to Kant's doubt regarding the discovery of any certain example of moral goodness at *KGW VIII:1 7* [62] (End 1886-Spring 1887), and arguably also at *M V 3*.

⁷⁴ *G* 408.

⁷⁵ The points made at *G* 407-8 are also made at *R* 20 and 95-6, *TP* 284-5 and 287, *E* 329-30, and *MS* 392-3, 441, 446-7, and 474.

duties, such that, for instance, the general duty of beneficence mutually constrains the general duty to obey the state, and each mutually constrains more specific duties.⁷⁶ This, he maintains, serves to reduce the indeterminacy of prescription to a tolerable level. Thus, in the introduction to the ‘Doctrine of Virtue’ in *The Metaphysics of Morals*, he writes that every ‘ethical’ duty ‘can command only the maxim of actions, not actions themselves, [and] this is a sign that [...it] could not indicate precisely how and how much should be done by the action’. However, he insists, this ‘is not to be understood as a permission to make exceptions to the maxim of actions, but only as a permission to limit one maxim of duty by another’.⁷⁷ Again, of course, Kant’s claim extends to non-moral reasons as well.

Kant thus demonstrates that reasons’ underdetermination of action need not undermine the prescriptive judgement of actions according to reasons, as it undermines the description of motivating reasons according to observed actions. However, much as it anticipates Nietzsche’s challenge in ‘*Long live physics!*’, Kant’s argument does not quite do justice to it. For in ‘*Long live physics!*’ Nietzsche concludes from a reason’s underdetermination of action not only that observed actions could not confirm or disconfirm a motivating reason, but also that an action could not be motivated by reasons at all, and that any judgement according to reasons is therefore inappropriate. He also consistently offers causal descriptions of actions as instead motivated by noncognitive states, and he insists that these states should therefore replace reasons as the objects of judgement. As I proposed to interpret his position, Nietzsche thus holds that a reason’s prescription is not sufficiently determinate, or particular, for its acknowledgement to motivate a

⁷⁶ Recent commentators other than O’Neill broadly appreciate this aspect of Kant’s conception of moral judgement. I discuss the examples of Hill, Baron, and O’Neill in the third part of my ‘Common Sense, Right, and Moral Judgement’.

⁷⁷ MS 390. On the ‘latitude’ of ‘ethical’ duties, see also MS 392-3 and 410-11.

particular action in particular circumstances, but that (at least some) noncognitive states informed by reasons are of sufficient determinacy, or particularity, to do so.

Although Kant does not consider this particular challenge, his account of judgement as the mutual constraining of multiple reasons allows him to accommodate it without denying its plausibility. For Nietzsche fails to appreciate that such mutual constraining yields prescriptions that are significantly more determinate, or particular, than any single reason's prescription. Insofar as an action is considered at a level of particularity no higher than that of these prescriptions, then, it could be considered as motivated by an acknowledgement of a prescription. This level of particularity need not be sufficient for observed actions to confirm or disconfirm a motivating reason. But it is presumably sufficient for actions to be generally considered as motivated by the acknowledgement of reasons' prescriptions. For example, mutually constrained reasons might prescribe that, all things considered, in certain circumstances an agent should donate a certain fraction of her income to charity. This prescription is presumably of sufficient determinacy, or particularity, for its acknowledgement to be considered as motivating her obedience to it, if not necessarily for this obedience to conclusively confirm or disconfirm such motivation. Nonetheless, considered at a higher level of particularity, Kant could admit that actions might be considered as not motivated by reasons, that their motivation might be captured by noncognitive states, and even that such motivation might be sufficiently determinate, or particular, for observed actions to confirm or disconfirm it. For instance, the all-things-considered prescription to donate to charity need not be of sufficient determinacy, or particularity, for its acknowledgement to be considered as motivating an agent's particular *manner* of obeying the prescription – her

sending the donation in a red envelope, rather than a blue one, say. At this level of particularity, her noncognitive state – her strong passion for red, say – might be considered to motivate, and, if it is sufficiently particular, such motivation might even be confirmed or disconfirmed by her observed actions. Importantly, this does not preclude her donation's also being considered as motivated by her acknowledgement of the prescription, since, at the lower level of particularity, her action can be so considered. That is, Nietzsche's challenge does not preclude practical reason, as he insists. It therefore also does not preclude prescriptive judgement according to reasons or make incoherent the notion of an unconditional and overriding reason, as he also maintains.

Nietzsche's challenge thus provides plausible grounds for supplementing Kant's conception of moral judgement with an appreciation of noncognitive motivations for obedience to reasons. Notably, such a supplementation would not threaten Kant's conception of moral goodness, since a goodness that is (i) a goodness of a will's reason for action alone and (ii) good for a will as such might be achieved equally by a will whose obedience to reasons is, above a certain level of particularity, considered as motivated by something other than them. Moreover, as I will show in the following subsection, such a supplementation can be accommodated by Kant's definition of 'choice'.

II. *Kant's definition of 'choice'*

Although Kant insists that moral judgement must regard choice as made for reasons, he does not insist that choice is made for reasons by definition or always

in fact. Insofar as he admits that choice need not be made for reasons, then, he does not propound the common sense claim regarding reasons to which Nietzsche objects, and might admit the supplementation that, I have suggested, Nietzsche's criticism of this claim provides grounds for.

This can be appreciated by attending first to a series of definitions that Kant provides in a section of the introduction to *The Metaphysics of Morals*. There, as elsewhere, he defines 'the ability of desire', or 'life', as 'the ability to be, through its representations, the cause of the objects of these representations'. He proceeds to insist that such an ability constitutes 'choice [*Willkür*]' on two conditions: firstly, this causation must be determined 'in' this ability itself, rather than 'in' its objects, making it what Kant calls 'an ability *to do or to refrain as it pleases*'; and, secondly, it must be accompanied by a consciousness of its ability to produce its objects, and therefore not be a mere '*wish*' for them.⁷⁸ I take these two conditions to mean simply that 'choice' is a particular instance of the ability to cause objects as consciously intended ('through representations'), which is distinguished, firstly, by being motivated, so that the action's 'pleasing' the agent is causally significant, and, secondly, by being constrained by the agent's theoretical beliefs regarding circumstances. Crucially, however, Kant presents 'practical reason', or 'will', as a particular instance of 'choice' which is motivated by reasons. He writes, 'The ability of desire whose inner determining ground, hence what pleases it[,] is to be found in the subject's reason is called the *will* [*Wille*]'. 'The will', he continues, 'insofar as it can determine choice, is practical

⁷⁸ *MS* 211, 213. For related remarks on 'the ability of desire', see *MAN* 544, *KpV* 9n, *KrV B* ix-x, *KU* 177n, *MS* 356-7, 381, and 384-5, *ApH* 251, and *EKU* 206 and 230n.

reason itself'.⁷⁹ Thus Kant does not equate 'choice' with choice made for reasons, but rather considers the latter a particular case of the former.

Kant here proceeds to equate 'human' choice with choice motivated by reasons, and to define it as 'free' in the sense that, insofar as it is motivated by reasons, it is '*affected*, but not *determined*, by impulses'. This distinguishes 'human' from 'animal' choice, which is 'determinable only by *inclination* (sensible impulse, *stimulus*)'.⁸⁰ However, Kant does not insist that every instance of human choice is 'human', rather than 'animal', in this sense. As he emphasises in the introduction to the 'Doctrine of Virtue' and elsewhere, he admits that human rationality can be suspended or obstructed by passing feelings such as anger, fear, or hope, and that certain persistent inclinations, such as hatred or lust for power, can eclipse other inclinations and concerns in rational reflection.⁸¹ I suggest that, although not developed by Kant, these admissions might be exploited to supplement his conception of moral judgement in the manner which I proposed in the preceding subsection – that is, to consider obedience to reasons as, above a certain level of particularity, motivated by noncognitive states.

However, while Kant's definition of 'choice' might be exploited to accommodate Nietzsche's criticism of the common sense claim regarding reasons, this definition also reveals that Kant endorses the common sense claim regarding choice to which Nietzsche objects. That is, it reveals that Kant considers choice, the agent's conscious, motivated, and sincere intention, as the sufficient cause of her action. Most clearly, it refers to choice as 'the' cause of an action. Nietzsche's objection to this claim, which he implicitly directs at Kant under the allegation

⁷⁹ MS 213.

⁸⁰ MS 213.

⁸¹ Kant refers to the former as 'affects' and the latter as 'passions', defines them at R 29n, MS 407-8, and *ApH* 251-3 and 267, and discusses various instances at *ApH* 253-75. See also *KU* 380 and *EKU* 196.

concerning 'free will', is that it offers an unavailable causal 'explanation' of chosen action, or at least an inadequate causal 'description' of it. I will show in the following chapter that Kant, like Nietzsche, denies the experience of necessitating causal powers, and therefore could not offer a causal 'explanation' of the kind to which Nietzsche objects. As I will also show in the following chapter, elsewhere in his works Kant also refers to choice as 'the' cause of an action, but offers only one, unconvincing argument that would establish that, at least in what he calls the 'intelligible' respect, choice ought to be considered as causally sufficient for action. Therefore, Kant is vulnerable to Nietzsche's demand that he admit better causal 'descriptions' of choice, and, in particular, better causal 'descriptions' which, as Nietzsche suspects, would identify salient causes of actions other than choice.

Once again, however, admitting this demand need not threaten Kant's conception of moral goodness. A will might achieve a goodness that is (i) a goodness of a will's reason for action alone and (ii) good for a will as such, although its choice is not causally sufficient for action.

III. *The 'revaluation' of common sense*

It remains to consider Nietzsche's allegation that, although Kant pursues an unconditional and overriding rational 'ground' of judgement, he succeeds only in articulating certain moral conceptions which prevailed in his own particular environment, and which ought to be subjected to further 'revaluation'. I have demonstrated that Kant conceives of moral goodness as lying in unconditional and

overriding reasons for action, and proposes to derive a system of duties from his conception of moral goodness. But, as I emphasised in the preceding chapter, his method of arriving at this conception is to analyse a basic commitment of moral common sense, rather than to identify a 'ground' beyond it. Kant thus does intentionally precisely what Nietzsche alleges that he does *unintentionally*. Still, Nietzsche's allegation is that, in doing so, Kant provides *merely* an analysis of moral common sense, in that his analysis is, firstly, not comprehensive of other common sense moral conceptions, and, secondly, not subjected to 'revaluation' on grounds independent of common sense.

With regard to the first element of this allegation, Kant admittedly displays no interest in analysing other basic commitments of common sense or more concrete common sense conceptions. However, he also does not deny or preclude such analyses. Nor does he deny or preclude the possibility that his own analysis and his derivation of duties from it might not comprehend the results of such further analyses. He thus might share Nietzsche's suspicion that the moral conceptions which prevail in different environments need not be mutually comprehensible. Given Kant's exclusive concern for a basic, abstract commitment and the implications that can be systematically drawn from it, one might speculate that he would admit mutual incomprehensibility only between such commitments and implications, on the grounds of the priority of these commitments over other moral conceptions within common sense. Still, that his method is to analyse moral common sense suggests a liberal attitude to any further analyses and to any mutual incomprehensibilities that they might reveal.

Kant would reject the second element of Nietzsche's allegation as philosophically extravagant, however. I have argued that Nietzsche provides

plausible grounds for supplementing Kant's conception of moral judgement with an appreciation of one aspect of Nietzsche's 'revaluation of values' – namely, the identification of kinds of noncognitive states which motivate obedience to kinds of reasons. But Nietzsche also undertakes to evaluate this obedience according to his particular concern for 'humanity', which he considers to be determined by his own particular noncognitive motivating states. Indeed, both Kant and Nietzsche are often supposed to invoke normative standards that hold independently of both common sense and any agent's particular concerns. With respect to Nietzsche, in the third chapter I mentioned some of the common candidates for such an 'objective' standard, and argued that his critical treatment of the common sense conception of agency invokes no such standard. In the fourth chapter, I argued for the same conclusion regarding his criticisms of Kant.

With respect to Kant's attitude towards appealing to such an 'objective' standard, the evidence is admittedly mixed. In the third section of the *Groundwork*, he raises the question of whether the conception of moral goodness which he derives in the preceding sections, by analysis of the commitment of 'common moral rational cognition' to the goodness of a good will, might be vindicated on grounds independent of 'common moral rational cognition'. In particular, he wonders how a moral judgement – that is, a judgement that a will has a morally good reason to do or refrain from an action – could be justified, given that having a morally good reason is neither conceptually included in having a will as such, nor derivable from any other reasons which a will might have, since the latter reasons would lack the unconditional character required by (ii), or, equivalently, be given *a posteriori*. In Kant's terms, he wonders how a moral judgement could be both

synthetic and *a priori*.⁸² Crucially, here he denies that this question can be answered simply by the commitment of ‘common moral rational cognition’ to such judgement. For all that the preceding analytical ‘expansion of the universally accepted concept of morality’ has shown, he writes at the end of the second section, this concept might be ‘a chimerical idea without truth’, a ‘fantasy’.⁸³ To vindicate this concept, and thus the possibility of justifying a moral judgement, Kant insists on finding independent grounds for it. Thus he describes his ‘method’ in the *Groundwork* as follows: ‘to proceed analytically from common cognition to determination of its supreme principle and in turn synthetically back from the examination of this principle and its sources to common cognition, in which its use is found’.⁸⁴ The first and second sections achieve the intended ‘determination’ of moral goodness, and thus ‘proceed analytically from common cognition’ of moral goodness. But only if Kant can identify other ‘sources’ of this moral goodness, and so ‘proceed [...] synthetically back from’ the analysis of moral goodness to its ‘use’ in a moral judgement, can he answer his question about the possibility of justifying a moral judgement.

In the third section of the *Groundwork*, therefore, Kant considers his analysis of moral common sense to require vindication on independent grounds, just as Nietzsche demands that prevailing moral conceptions be ‘revalued’ according to his particular concern for ‘humanity’, or as commentators often demand, in Kant’s or Nietzsche’s name, that such conceptions be criticised according to a normative standard that holds independently of common sense and any agent’s particular concerns. However, as I will show in the following chapter,

⁸² On the synthetic and *a priori* character of a moral judgement, see *G* 420, *G* 420n, 440, and 444-5, *KpV* 31, and *R* 26-8 and 26n.

⁸³ *G* 444-5. See also *G* 420, 428-9, 429n, 431, and 440.

⁸⁴ *G* 392.

Kant's attempt to provide such independent grounds in the third section of the *Groundwork* is unconvincing, and from the *Critique of Practical Reason* onwards he insists that his question about the possibility of justifying a moral judgement *can* be answered by the common sense commitment itself. That is, his considered position is that no independent vindication need be provided for the conception of moral goodness that he draws analytically from a basic common sense commitment.

In this light, Kant would consider the 'revaluation' of this basic commitment to be a philosophical extravagance, whether it be undertaken according to particular grounds such as Nietzsche's concern for 'humanity' or according to 'objective' grounds such as those often attributed to both Kant and Nietzsche. The difference between Kant and Nietzsche on this matter therefore remains an undecidable one.

Conclusions

In this chapter, I defended an interpretation of Kant's formulas and their place in his conception of moral judgement, on the basis of Kant's derivations and employment of the formulas and his further claims about the nature of moral judgement. In particular, I first demonstrated that Kant's formula of universal law is not intended as a 'test' of reasons for action in specific circumstances, according to the consequences of their universal adoption. Rather, the formula is intended as an expression of moral goodness which emphasises that this goodness is (ii) good for a will as such, and which, along with his other formulas, Kant proposes to employ in deriving a system of duties from (i) and (ii) and claims regarding wills' contingent conditions. Given this, I demonstrated that Nietzsche's criticisms of the formula of universal law are misplaced. I then argued that prevailing interpretations fail to capture the meanings of Kant's further formulas, and that he employs a distinctive conception of 'community' in his derivations of duties from them. With reference to the formula of autonomy in particular, I also briefly argued that Kant does not, and need not, share Nietzsche's concern for individual 'autonomy', much as Nietzsche commentators often allege that Kant betrays it. In the final section, I demonstrated that Nietzsche's criticisms of the common sense conception of agency are somewhat more telling as criticisms of Kant, since they draw certain significant accommodations from Kant's conception of moral judgement. There I also argued that, given Kant's ultimate methodological appeal to common sense, the success of Nietzsche's demand for further critical grounds is undecidable.

Spontaneity in choice

In this chapter, I examine Kant's persistent claim that human choice must be considered as undetermined by antecedent causes. In the first section, I outline his attempt to accommodate this claim by means of his account of cognitive judgement, and his restriction of the scope of the principle of causal determinism in particular. This outline is, I argue, sufficient to demonstrate that Nietzsche's allegations of the incoherence of 'free will' and the untenability of scepticism are misplaced when applied to Kant. Nonetheless, much as Nietzsche might present it as an undefended conviction, Kant offers four arguments for his claim. In the three following sections, I consider these arguments in turn, and argue that each is unconvincing, irrespective of Nietzsche's critical concerns and of the tenability of Kant's attempt to accommodate his conclusion. Although I consider along with the fourth argument Nietzsche's accusation that Kant propounds an over-inflated sense of responsibility, I reserve mentioning the other, minor ways in which Kant's arguments relate to Nietzsche's critical concerns until the final section.

Kant's commitment to the principle of causal determinism constitutes part of his account of cognitive judgement, which he presents at most length in the Transcendental Analytic of the *Critique of Pure Reason*. In outline, this account holds, firstly, that a human rational being can make a cognitive judgement only by bringing a sensible intuition of an object under a concept, and can have a sensible intuition of an object only by presupposing, *a priori*, intuitions of space and time. Kant insists that the necessity of these presuppositions for a cognitive judgement entails that the possible objects of such a judgement, which he terms 'appearances', must be distinguished from what these objects may be 'in themselves', independently of the presuppositions made in their sensible intuition.¹ Secondly, Kant holds that in order to bring an object under a concept in a cognitive judgement, certain conceptual presuppositions regarding an 'object' in general must be made.² In particular, he argues that a cognitive judgement must presuppose that its objects are 'substances' in which observable properties inhere and that these objects are subject to causality. Furthermore, on the grounds of the presupposition of time in sensible intuition, he argues that these substances must be presupposed to persist over time and this causality to be governed by the principle of causal determinism. This is so, he argues, on pain of not distinguishing cognitive representations from their objects, for without presupposing temporally persisting 'substances' and their subjection to the principle of causal determinism, basic cognitive distinctions regarding succession and coexistence could not be

¹ See *KrV* A19-49/B33-73, and also *P* 287-94.

² See *KrV* A50-130/B74-168, and also *P* 294-326.

made.³ With regard to the principle of causal determinism in particular, Kant thus admits the ‘Humean’ claim that our experience extends not to causal powers which necessitate succeeding events, but only to regular correlations between events of the same types. But, by arguing that a cognitive judgement must presuppose its object’s subjection to the principle of causal determinism, Kant argues that this principle itself is secured against such ‘Humean’ scepticism.⁴

As I will demonstrate in the following sections, Kant offers four different arguments for his consistent claim that human choice must be exempt from the principle of causal determinism. Nonetheless, his attempts to accommodate this claim all rest upon the account of cognitive judgement which he presents in the Transcendental Analytic of the *Critique*, in the manner of the conclusions regarding choice which he draws in his discussion of the third ‘antinomy’ in the Transcendental Dialectic. There Kant applies to choice a broad implication of his account of cognitive judgement, namely, that independently of the presuppositions made in its sensible intuition – that is, ‘in itself’ – a possible object of a cognitive judgement need not be subject to the principle of causal determinism.⁵ Given this, he maintains, one may distinguish the ‘intelligible’ from the ‘empirical’ choice of any particular action, each being sufficient to account causally for the action, but only the latter being a possible object of a cognitive judgement and therefore subject to the principle of causal determinism. Thus he writes that, as a possible object of a cognitive judgement, a human agent’s ‘choice has an empirical character, which is the (empirical) cause of all his actions’, and that ‘not one of [...the] conditions which determine human beings according to this character [...]

³ See *KrV* A182-211/B224-256.

⁴ See, for instance, *KrV* A195-6/B240-1. At *KU* 179-86 and *EKU* 211-6, Kant proceeds to argue, on further grounds, that a certain system among causal universals must also be presupposed.

⁵ For this implication, see *KrV* A532-58/B560-86, and also *P* 312-3 and 343-7.

would not be contained in the series of natural effects and obey the laws of nature'. One can also investigate the further causal 'sources' of this 'empirical character', and thus, again, 'one proceeds as [one does] generally in investigation of the series of determining causes of a given natural effect'. Kant writes that 'intelligible' choice, on the other hand, 'as unconditioned condition of every voluntary action, [...] allows of no condition preceding it in time, while its effect begins in the series of appearances, but can never constitute an absolutely first beginning in this series'.⁶ 'Intelligible' choice is therefore not a possible object of a cognitive judgement, and, indeed, is singular, timeless, and not itself subject to explanation.⁷

This distinction between 'intelligible' and 'empirical' choice has been persistently criticised for, among other things, having choice or its effects contravene the principle of causal determinism, making choice causally superfluous, considering any particular action to have two distinct but sufficient causes, and reducing choice to a single, ordinary instance.⁸ The interpretation of the distinction also remains highly controversial, particularly in the light of such criticisms. Some interpreters, for instance, treat the distinction as an ontological one, according to which the agent makes at least one causally significant choice which is undetermined by antecedent causes.⁹ Others hold that Kant simply considers choice under two epistemological 'aspects', that of the presuppositions of cognitive judgement and that independent of these presuppositions, and admits

⁶ *KrV* A552/B580, A554/B582. See also *KrV* A550-8/B578-86.

⁷ See *KrV* A551-7/B579-85.

⁸ For relatively recent examples, see Bennett, *Kant's Dialectic*, pp.199-204, Walker, *Kant*, pp.147-50, Beck, 'Five Concepts of Freedom in Kant', pp.42-3, Wolff, 'Remarks on the Relation of the *Critique of Pure Reason* to Kant's Ethical Theory', pp.142-9, and Sullivan, *Immanuel Kant's Moral Theory*, pp.283-6.

⁹ See, for instance, Wood, 'Kant's Compatibilism', and Barbone, 'Compatibilism in the First Critique'.

that spontaneity may be attributed to choice under the latter ‘aspect’.¹⁰ Still other interpreters maintain that Kant anticipates contemporary nonreductive materialism by accounting for choice as merely a non-causal and irreducible kind of description of events, events which are causal only under another, physical kind of description.¹¹

Nietzsche’s allegations regarding Kant’s commitment to ‘free will’ bear some similarity to these persistent critical concerns. To recall, Nietzsche’s explicit criticism of Kant with regard to choice is that, by endorsing ‘free will’, Kant conceives of choice incoherently, as both a cause and the effect of that cause. Kant is indeed vulnerable to this criticism, insofar as he either, as at the beginning of his discussion in the Transcendental Dialectic, refers to spontaneity as self-causation or, as in the body of his discussion, describes ‘intelligible’ choice as the cause of ‘empirical’ choice.¹² However, in his discussion Kant also denies that ‘intelligible’ choice causes itself, and, particularly in concluding his discussion, he presents ‘intelligible’ choice and ‘empirical’ choice as each the sufficient cause of any particular action.¹³ Nor need Nietzsche’s criticism be ultimately more telling if understood to echo persistent critical concerns regarding the coherence, given the principle of causal determinism, of considering an action’s cause as itself uncaused. For Kant does not intend ‘intelligible’ choice to contravene this

¹⁰ See, in particular, Allison, *Kant’s Transcendental Idealism*, esp. chs.1-2, *Kant’s Theory of Freedom*, chs.1-4 and 7, ‘Kant on Freedom’, pp.109-14 and 118-28, and ‘Autonomy and Spontaneity in Kant’s Conception of the Self’, pp.130-34. Comparable accounts are offered in Beck, ‘Five Concepts of Freedom in Kant’, pp.43-6, Korsgaard, ‘Morality as Freedom’, pp.171-6, and O’Neill, ‘Autonomy and Anthropology in Kant’s *Groundwork*’, pp.75-7.

¹¹ See, in particular, Meerbote, ‘Kant on the Indeterminate Character of Human Actions’, and Hudson, *Kant’s Compatibilism*, chs.2 and 3.

¹² For causal spontaneity as self-causation, see *KrV* A533-4/B561-2, and for ‘intelligible’ choice as the cause of ‘empirical’ choice, see *KrV* A534/B562, A537/B565, A538/B567, A541/B569, A544-6/B572-4, A549/B577, and A551-2/B579-80.

¹³ For the denial of self-causation, see *KrV* A541/B569, and for the equal causal sufficiency of ‘intelligible’ and ‘empirical’ choice, see *KrV* A536/B564, A538/B566, A539/B567, and A553-6/B581-4.

principle, in either the explanation or the conception of events: actions are also to be sufficiently explained by 'empirical' choice, and this in turn by other preceding causes, and 'causation' is to be conceived meaningfully without restricting it those instances of it that are subject to the principle of causal determinism.¹⁴

Nietzsche also objects, however, to the distinction between 'appearances' and 'things in themselves' on which Kant's distinction between 'intelligible' and 'empirical' choice rests. Nietzsche's objection is that the former distinction expresses a scepticism according to which knowledge claims which satisfy ordinary standards of justification might nonetheless be false, and that this scepticism relies on a notion of the 'true world' – for which, read 'things in themselves' – which is epistemologically insignificant, mischaracterises the 'true', and suspiciously serves certain practical interests, and therefore ought to be dismissed. However, although I will not consider this objection in depth, it is again not clear that it tells against Kant. For Kant does not conceive of 'things in themselves' as the possible objects of genuinely 'true' knowledge claims, which, if known, could falsify knowledge claims regarding 'appearances'. Rather, he conceives of 'things in themselves' as objects considered independently of the presuppositions under which they can be objects of a knowledge claim, or cognitive judgement, at all. 'Things in themselves' therefore do not populate a 'true world' which, although of epistemological and practical significance, is unfortunately epistemologically inaccessible to ordinary standards of justification. They rather mark the absence of any knowledge claim, or cognitive judgment, at all. If this is so, then the three concerns which Nietzsche extends from his engagement with the sceptical position to his criticism of Kant's are misplaced.

¹⁴ Kant emphasises the latter point at *KpV* 5-6, for instance.

Nietzsche's criticisms of Kant's distinction between 'intelligible' and 'empirical' choice are, therefore, not convincing. Nonetheless, in the following sections, I will argue that equally unconvincing are Kant's four arguments for the claim that this distinction is designed to accommodate – namely, the claim that human choice must be considered as undetermined by antecedent causes.

Kant's first argument for attributing spontaneity to choice is presented alongside his account of the distinction between 'intelligible' and 'empirical' choice in his discussion of the third 'antinomy' in the Transcendental Dialectic of the *Critique of Pure Reason*. The argument is presented in terms of a 'practical' sense of freedom which Kant here defines as 'the independence of choice from **necessitation** by impulses of sensibility', and which he here, as elsewhere, insists on attributing to human choice, while claiming that human choice is nonetheless also '**affected**' by such impulses.¹⁵ However, the argument's basic concern is with the rational character of choice, to which, Kant maintains, 'practical' freedom corresponds. He makes this correspondence explicit only later in the *Critique*, when, in the Transcendental Doctrine of Method, he writes that if choice is 'free' in the 'practical' sense, then it is determined 'by motives which are represented only by reason', and is therefore subject to 'laws' which 'say **what should happen**, even if it perhaps never happens'.¹⁶ In other words, Kant considers the 'practical' freedom of choice to correspond to the making of choice according to reasons – that is, the sense of rational choice that he introduces with his account of 'practical reason' in the second section of the *Groundwork*.¹⁷

In the passage from the Transcendental Doctrine of Method, Kant also writes that human choice's 'practical' freedom and corresponding rational character 'can be proved by experience', since, with regard to what 'determines

¹⁵ *KrV* A534/B562.

¹⁶ *KrV* A801-2/B829-30.

¹⁷ Kant also intimates the correspondence between 'practical' freedom and rational choice when he defines 'practical reason' at *G* 460n, and *MS* 211-4.

human choice’, ‘we have an ability to overcome the impressions on our sensible ability of desire [that is, our immediate inclinations] through representations of that which is useful or harmful in a more remote way’, and these representations are ‘based on reason’.¹⁸ However, he emphasises that such a ‘proof’ suffices only in the context of the ordinary practice of practical judgement, since in this context the question of whether choice is determined by antecedent causes does not arise.¹⁹ As his discussion in the Transcendental Dialectic makes clear, he considers this question to arise in a theoretical context, and, indeed, to provide ‘the true moment of the difficulties’ of the ‘practical’ freedom of choice itself.²⁰ This is so, Kant argues, because to propose ‘what should happen’ according to reasons, as one does in addressing human choice, is to presuppose that the choice addressed is not determined by antecedent causes. He expresses this argument most clearly in the following paragraph.

[A...] ‘should’ expresses a possible action, the ground of which is nothing other than a mere concept [that is, a reason for action ...]. Now the action certainly must be possible under natural conditions if the should is directed to it; but these natural conditions do not concern the determination of choice itself, but only the effect and its outcome in appearance. However many natural grounds, even sensible stimuli, there may be which drive me to **will**, they cannot produce the **should**, but only a still far from necessary, but always conditioned willing, against which the should that reason pronounces sets measure and goal, indeed prohibition and standing. It may be an object of mere sensibility (the agreeable) or even of pure reason (the good): reason does not give [this object] according to those grounds

¹⁸ *KrV* A802/B830.

¹⁹ See *KrV* A803/B831.

²⁰ *KrV* A533/B561.

which are given empirically, and does not follow the order of things as they are presented in appearance, but with complete spontaneity it makes its own order according to ideas, in which it fits the empirical conditions, and according to which it even declares actions to be necessary which still have **not happened** and perhaps will not happen, but presupposing of all of them that reason could have causality in relation to them; for, without that, it would not expect its ideas to have effects in experience.²¹

Despite its reference to ‘ideas’ – which, as I will explain in the following section, has a very specific sense for Kant – the argument here is quite straightforward. It argues simply that to propose, according to reasons, that an action ‘should’ happen is to presuppose that ‘reason could have causality in relation to’ the action, and thus that, although ‘natural conditions’ might constrain the available options and consequences, antecedent conditions do not causally determine the ‘choice’. In other words, it is to presuppose that the choice addressed is not determined by antecedent causes, because it is determined by reasons. On these grounds, then, Kant claims that the ‘freedom’ of choice in the ‘practical’ sense requires that choice is also ‘free’ in the sense that it is not determined by antecedent causes. As he expresses it, ‘the practical concept [of freedom] is grounded on [...the] **transcendental** idea of **freedom**’.²²

This supposed requirement of ‘transcendental’ freedom presents ‘the true moment of the difficulties’ of the ‘practical’ freedom of choice, of course, when it is confronted with a commitment to the principle of causal determinism, and Kant attempts to resolve these difficulties by appeal to the distinction between ‘intelligible’ and ‘empirical’ choice. Thus he writes, for instance, that, as

²¹ *KrV* A547-8/B575-6. See also *KrV* A448/B476, A534/B562, and A547/B575.

²² *KrV* A533/B561.

‘intelligible’ choice, ‘reason in its causality is subject to no conditions of appearance and its temporal course’.²³ He endorses the argument again in his immediately following *Prolegomena to Any Future Metaphysics*, his review of Johann Schultz’s *Versuch einer Anleitung zur Sittenlehre für alle Menschen*, and the latter pages of the third section of the *Groundwork*, and, indeed, he did not modify the relevant chapter of the Transcendental Dialectic when he prepared a second edition of the *Critique*.²⁴

This argument is also popular among contemporary commentators, who often extend it to Kant’s other texts and, in pursuit of metaphysical modesty, combine it with a non-ontological interpretation of the distinction between ‘intelligible’ and ‘empirical’ choice. Henry Allison provides a particularly extensive defence of such a position, according to which Kant’s distinction as a distinction between two epistemological ‘aspects’ of choice and a commitment to, and identification with, motivations is an ineliminable moment of spontaneity in rational choice, which can be admitted only under choice’s ‘intelligible’ aspect. The approach has also been taken up by others, notably Korsgaard, as a commitment to two irreducible ‘standpoints’ on choice.²⁵

However, it is far from clear that rational choice presupposes ‘transcendental’ freedom, as Kant’s argument requires. Kant reasonably denies that any mere ‘experience’ of rational choice, or of practical judgement, could establish an agent’s ‘transcendental’ freedom.²⁶ Nor would a mere conviction of

²³ *KrV* A556/B584. See also *KrV* A548-57/B576-85.

²⁴ See *P* 344-7, *RS* 13-4, *G* 455-7, and also, perhaps, *MS* 381-2 and 384-5.

²⁵ For references to Allison, Korsgaard, and other examples of this approach, see n.10 above.

²⁶ Kant thus anticipates Ameriks’s objection to Allison, in his ‘Kant and Hegel on Freedom’, pp.226-8.

such freedom be adequate.²⁷ Kant also reasonably admits that choice can be conceived as rational without also conceiving of it as spontaneous: in his discussion in the Transcendental Dialectic, for instance, he writes that an agent's 'empirical' choice 'displays a rule, according to which one can derive the rational grounds and the actions themselves according to their kind and their degree, and judge the subjective principles of his choice'.²⁸ Given this, it is not clear why Kant, or anyone else, should insist on a further moment of spontaneity in rational choice.

²⁷ For discussion of this point, see Guevara, 'Two Standpoints on the Will', pp.88-91, and Ameriks, *Kant and the Fate of Autonomy*, p.73.

²⁸ *KrV* A549/B577. Kant thus anticipates compatibilist objections, such as that of Brink in his 'Kantian Rationalism', pp.266ff.

As I mentioned in the preceding section, Kant republished his argument from rational choice in the second edition of the *Critique of Pure Reason* and meanwhile he endorsed it in, among other texts, the third section of the *Groundwork*. However, the latter text also provides a different, and more extensive, argument, which, furthermore, Kant dismissed and replaced in the *Critique of Practical Reason*. In the following two subsections, I consider this argument and its replacement in turn, and argue, in particular, that both are hamstrung by a puzzling confusion.

I. The ambitious argument of the Groundwork, section III

I noted in the previous chapter that Kant's concern in the third section of the *Groundwork* is to explain the possible justification of a moral judgement. In particular, he is concerned to explain, independently of 'common moral rational cognition', how a will could be judged to have a morally good reason to do or refrain from an action, given that such a reason is neither conceptually included in having a will as such, nor derivable from any other reasons which a will might have, since the latter reasons would lack the requisite unconditional character. This is a comprehensible concern. However, the particular doubt about moral judgement that Kant identifies at the beginning of the section, and the argument with which he then attempts to dispel it, are profoundly unconvincing. Indeed, I

will argue that this is so despite the fact that Kant's argument is often dismissed on a variety of other, inappropriate grounds.

Kant begins the section by arguing that if a will is 'free' in a 'negative' sense, then, 'through mere analysis of its concept', it must also have 'autonomy', the property which, as the first and second sections demonstrate, a moral judgement must attribute to a will. He writes here that a 'will is a kind of causality of living beings insofar as they are rational', and that a will which is 'free' in the 'negative' sense is a will which, unlike 'natural necessity', has the 'property [...] that it can be effective independently of alien causes *determining* it'. That is, for Kant, to have a will is to be able to act according to reasons, and to have a negatively free will is to be able to so act undetermined by antecedent causes, and thus undetermined by something other than this ability. The 'negative' sense of freedom, then, includes the 'transcendental' sense with which Kant is occupied in the *Critique of Pure Reason* and elsewhere. Here he argues that negative freedom entails autonomy – which he here refers to as 'the will's property of being a law to itself' and as a will's 'freedom' in a 'positive' sense – because willing must be determined – in Kant's terms, its causality cannot be 'lawless'. That is, he argues that if willing must be determined but is not determined by antecedent causes and thus by something other than will itself, then, by elimination, it must be determined by will itself.²⁹ Thus Kant concludes that, by analysis, a will's negative freedom entails its autonomy.

However, Kant's doubt about moral judgement rests on the opposite entailment, namely, that, by analysis, a will's autonomy entails its negative

²⁹ *G* 446, 447. See also *KpV* 29, 31, and 93-4, and *MS* 222, 223, and 239. Allison provides a similar interpretation of Kant's argument here, although he treat it in terms of the formula of universal law, rather than the formula of autonomy to which Kant refers, and holds that the argument is made clear only at *KpV* 29. See Allison, *Kant's Theory of Freedom*, pp.204-13, and 'Autonomy and Spontaneity in Kant's Conception of the Self', pp.134-42.

freedom. He affirms both entailments when he writes, slightly later in the section, that autonomy and negative freedom 'are both autonomy, [and] hence reciprocal concepts'. But he provides an argument from autonomy to negative freedom only in the *Critique of Practical Reason*. There he argues that the 'determining ground' of a will which has autonomy 'can be represented merely by reason', and is therefore 'distinguished from every determining ground of events in nature according to the law of causality, because in these the determining grounds must themselves be appearances'. From this Kant concludes that a will which has autonomy 'must be thought as completely independent of the natural law of appearances in their relations to one another, namely[,] the law of causality'.³⁰ That is, Kant argues that since a will which has autonomy acts according to a reason which is 'given' by will itself, it cannot be determined to act by something other than itself, as the possible objects of a cognitive judgement are in their subjection to the principle of causal determinism, and it is therefore negatively free.

These arguments for the two 'reciprocal' entailments are deeply puzzling. For autonomy is a quality of the reasons according to which an agent acts, while negative freedom regards the causal determination of her choice itself. Neither is pertinent to the other. Thus Kant's arguments apparently confuse the 'self-determination' of an agent's acting according to a reason concerning mere will, rather than inclination, with the 'self-determination' of an agent's choice being self-caused, rather than caused by antecedent causes.

³⁰ G 450, *KpV* 28-9, 29. Note that although at *KpV* 28-9 Kant expresses moral goodness as 'the mere lawgiving form of maxims alone [being] the sufficient determining ground of a will', and refers only to the 'transcendental' sense of freedom, at *KpV* 29 and 33 he equates the former with 'autonomy', or the 'positive' sense of freedom, and the latter with the 'negative' sense of freedom. For other statements of this argument, see *KpV* 33-4, 38-9, 42, 44, 50, 55, 93-4, and 105, *RU* 455-7, and *MS* 223 and 225.

Furthermore, if this argument is confused, then the doubt that Kant identifies on the basis of it is misplaced. He first refers to the synthetic character of a moral judgement, by stating that ‘such a synthetic proposition is possible only by this, that the two cognitions [of which it consists, namely, a will and its having a morally good reason to do or refrain from an action] are bound together through their connection with a third, in which they are both to be found’.³¹ He then writes, ‘The *positive* concept of freedom creates this third [cognition], which cannot be[,] as in physical causes[,] the nature of the sensible world (in the concept of which the concepts of something as cause in relation to *something other* as effect come together)’. Thus, Kant claims, the ‘third’ cognition to which a moral judgement must refer is a cognition ‘to which freedom points us and of which we have an idea *a priori*’.³² That is, Kant maintains that since a moral judgement attributes autonomy (‘positive’ freedom) to a will, it also attributes to it a freedom which is not compatible with the principle of causal determinism, namely, negative freedom. Given Kant’s claim that a possible object of a cognitive judgement is subject to the principle of causal determinism, a moral judgement thus ‘points us’ towards regarding willing as not such a possible object, or not part of ‘the sensible world’. The doubt with which Kant is occupied in the third section, then, rests upon his puzzling affirmation of autonomy’s entailment of negative freedom: he is concerned that the autonomy which a moral judgement attributes to a will cannot justifiably be attributed to it, because the willing such a judgement requires could

³¹ *G* 447. Note, however, that here Kant presents his concern in the third section as that of defending the possibility of the ‘principle’ of morality itself, which, he claims, cannot be established ‘through analysis of the concept of an absolutely good will’, and is therefore ‘a synthetic proposition’. This claim confuses the derivation of the ‘principle’ of morality from the concept of a good will with the character of the judgement that a will has a morally good reason to do or refrain from an action, in the sense identified by the ‘principle’ of morality. As he indicates at *G* 392, 420, 420n, 428-9, 429n, 431, 440, 444-5, and 454, Kant’s considered position is that the former is analytic, *a priori*, and provided in the first and second sections, while the latter is synthetic, *a priori*, and considered in the third section.

³² *G* 447.

not be a possible object of a cognitive judgement, subject to the principle of causal determinism.³³

In attempting to dispel this doubt in the rest of the section, however, Kant appeals to the equally puzzling opposite entailment for which he argues at the beginning of the section, namely, that a will's negative freedom entails its autonomy. That is, he undertakes to dispel his doubt about autonomy by arguing that, despite its incompatibility with the principle of causal determinism, negative freedom can and must be attributed to a will, and that therefore, by analysis, autonomy can and must also be attributed to it. Indeed, as I mentioned in the preceding chapter and as Kant emphasises later in the section, such an argument would, if successful, also vindicate the account of moral goodness provided in the first and second sections, by providing grounds for it other than those of 'common moral rational cognition'. In the terms of the 'method' he outlines in the preface, such an argument would identify 'sources' of this account other than 'common moral rational cognition'.

As 'preparation' for this argument, Kant first identifies three significant constraints which the doubt places upon any such attempt to dispel it. In short, these constraints are the following: firstly, the argument should demonstrate that negative freedom must be attributed to a will as such, rather than to a will only on further, contingent grounds, since a moral judgement attributes autonomy to a will on equally unconditional grounds; secondly, since a will's negative freedom is not a possible object of a cognitive judgement, the argument cannot demonstrate this freedom as such an object, but rather only as a necessary presupposition regarding a will; and, thirdly, the argument must not rest on grounds which might presuppose

³³ Later in the section, at *G* 456, Kant refers to this doubt as that on which 'the fatalist can build and chase all morality from its supposed property[,] possessed without title'.

the possibility of making a moral judgement, since the argument is intended to demonstrate precisely that possibility.³⁴

Kant identifies the third constraint by considering, as a mere assertion, a claim which would satisfy the first and second constraints. Of this claim he writes the following.

Now I assert: that to every rational being that has a will we must necessarily lend also the idea of freedom, under which alone he acts. For in such a being we think of a reason which is practical, i.e. has causality in view of its objects. Now[,] one cannot possibly think of a reason which[,] with its own consciousness [and] in view of its judgements[,] receives a direction from elsewhere, for then the subject would ascribe the determination of the power of judgement not to his reason, but to an impulse. It must regard itself as originator of its principles, independently of alien influences, [and] consequently as practical reason, or as the will of a rational being[,] it must be regarded of itself as free; i.e. the will of such a being can be its own only under the idea of freedom and must therefore in a practical respect be attributed to every rational being.³⁵

The claim which Kant presents as a mere assertion here is that made in the third, penultimate sentence: that the making of a rational judgement – that is, a judgement according to reasons – is necessarily negatively free. Given this claim as a premise, Kant argues that a will as such must be presupposed to be negatively free, because, as action according to reasons, willing is an instance of rational judging. The asserted claim would thus dispel his doubt about autonomy, and would do so without appealing to grounds which are contingent to a will, or

³⁴ For the first two constraints, see *G* 447-9, and for the third, see *G* 448-50 and 453.

³⁵ *G* 448.

considering a will's negative freedom as an object of a cognitive judgement. Since the asserted claim has this happy implication, however, Kant suspects that it might merely presuppose the possibility of attributing autonomy to a will in moral judgement. Thus he claims that 'a kind of circle shows itself here': 'We take ourselves as free in the order of efficient causes, in order to think ourselves in the order of ends under moral laws, and we afterwards think ourselves as subject to these laws, because we have ascribed to ourselves freedom of will'.³⁶ Kant therefore concludes his 'preparation' by noting that, thus far, he has only extended the analysis of moral goodness which the first and second sections provide, by demonstrating that since moral goodness consists of the autonomy of a will, it also consists of the will's negative freedom.³⁷

In the rest of the third section, Kant attempts to justify the claim which he considers as a mere assertion in his 'preparation', namely, that rational judging is necessarily negatively free. Although not ultimately convincing, his argument is simpler and more plausible than the variety of accounts in the literature would suggest. Its grounds lie in the account of cognitive judgement which he presents in his *Critique of Pure Reason*. In the *Groundwork* he presents this account in a brief and simplified form, emphasising particularly the distinction between 'appearances' and 'what they may be in themselves'. He also mentions here that

³⁶ G 448, 450. See also G 449-50 and 453. Commentators have offered a variety of other interpretations of the 'circle', however. For instance, in his *Kant's Theory of Mind*, pp.203-6, Ameriks maintains that the 'circle' refers to the inadequacy of claim that rational judging is negatively free for justifying the attribution of autonomy, rather than only negative freedom, to a will. This is unconvincing, given that Kant begins the third section by arguing precisely that, by analysis, a negatively free will also has autonomy. Equally unconvincing is Korsgaard's claim, in her 'Morality as Freedom', p.167, that the 'circle' refers to Kant's failure to yet demonstrate our motivational 'interest' in moral goodness. For Kant attempts to demonstrate what she understands as our motivational 'interest' in moral goodness only after having claimed to remove the suspicion of a 'circle', and then only as a consequence of his particular solution to it. Also unconvincing, if more understandable, is Allison's claim, in his *Kant's Theory of Freedom*, pp.217-21, that the 'circle' refers to Kant's suspicion of the claim that a rational being has a will at all. Although some of Kant's remarks in the third section might suggest such a suspicion, he also regularly refers to the having of a will as a basic premise of his argument, and nowhere attempts to demonstrate it.

³⁷ See G 449-50.

the bringing of an object under a concept in a cognitive judgement manifests the judging being's 'self-activity', in contrast to the passivity of its being affected by sensible intuitions.³⁸ This presumably refers not only to the conceptual presuppositions regarding an 'object' in general which, Kant holds, are required to bring an object under a concept in a cognitive judgement, but also to another claim made in the *Critique of Pure Reason*, that this bringing of an object under a concept requires a certain spontaneity of 'apperception'.³⁹

However, in attempting to justify the claim that rational judging is necessarily negatively free, Kant emphasises a further, 'pure self-activity'. This particular 'self-activity', he claims, is manifested by the very distinguishing of the possible objects of a cognitive judgement from 'what they may be in themselves', and also by what he calls 'ideas'. For Kant, as he makes clear at the beginning of the Transcendental Dialectic of the *Critique of Pure Reason*, an 'idea' identifies an unconditioned totality of conditions, and there is thus an 'idea' corresponding to each kind of condition under which the possible objects of a cognitive judgement stand. Since an 'idea' proceeds beyond these conditions to their unconditioned totality, Kant holds that it identifies an object which cannot be an object of a cognitive judgement and, as he emphasises in the *Groundwork*, that it thus manifests a 'self-activity' which is negatively free. Thus in the *Groundwork* he writes of this 'self-activity', along with that of distinguishing of 'appearances' from 'things in themselves', 'reason shows under the name of ideas a spontaneity so pure that through this it goes far beyond everything that sensibility can supply it and proves its most noble occupation in distinguishing the world of sense and the world of understanding from each other, [...and] thereby marking out for the

³⁸ G 452.

³⁹ For the latter claim, see *KrV* A103-10 and B131-6.

understanding itself its limits'. On these grounds, Kant claims that a human rational being must regard itself not only as a possible object of a cognitive judgement, and so subject to the principle of causal determinism, but also as a 'thing in itself' and, crucially, as negatively free in this regard. In Kant's terms, a human rational being not only 'must count himself in the *world of sense*', but also 'in view of what may be in him of pure activity [...] must count himself in the *intellectual world*, which however he does not cognise further'.⁴⁰

By appealing to the spontaneity manifested in 'ideas' and in the distinction between 'appearances' and 'things in themselves', then, Kant claims to justify the claim that rational judging is necessarily negatively free. He intends this justification to satisfy his third constraint by not presupposing the possibility of making a moral judgement, and he argues from the claim in the manner which he proposes in his 'preparation', a manner which satisfies his first and second constraints. That is, given the claim that rational judging is necessarily negatively free, he argues that autonomy must be attributed to a will, because willing is an instance of rational judging and, and, by analysis, a will's negative freedom entails its autonomy. Given the nature of his justification of the claim regarding rational judging, of course, he restricts this conclusion to a will insofar as it is not regarded as a possible object of a cognitive judgement. Thus he concludes that a human rational being 'has two standpoints from which he can regard himself and cognise laws for the use of his powers, [and] consequently for all his actions, *first*, insofar as he belongs to the world of sense, under laws of nature (heteronomy), [and] *second*, as belonging to the intelligible world, under laws which, independent of nature, are not empirical, but grounded merely in reason [that is, as having

⁴⁰ G 452, 451. On the nature of 'ideas', see also G 455 and 463, and KrV A310-38/B366-96, P 327-65, KU 341-4 and 351, and MS 371.

autonomy]’.⁴¹ Furthermore, Kant claims, the latter standpoint constitutes a human rational being’s ‘true self’, from which she must be regarded in a practical judgement, because ‘*the world of understanding contains the ground of the world of sense, hence also of its laws*’ – that is, apparently, because even from the former standpoint, a certain ‘self-activity’ is presupposed.⁴²

Kant claims to thus successfully defend the possibility of a moral judgement against his doubt that it is not possible to attribute to a will the autonomy which such a judgement must attribute to it. As he puts it here, the ‘*categorical ought* [of a moral judgement] represents a synthetic proposition *a priori*, through which to my will affected by sensible desires is added the idea of the same will, but belonging to the world of the understanding[,] pure, [and] for itself practical, [the idea] which contains the supreme condition of the former according to reason’.⁴³

Kant’s argument in these pages of the third section of the *Groundwork* is often unfairly dismissed, even by sympathetic commentators. There is a broad consensus that the argument is intended to demonstrate the negative freedom of a will by appealing to the spontaneity which, Kant claims, is manifested by ‘ideas’ and by the distinction between ‘appearances’ and ‘things in themselves’.⁴⁴ But, beyond this, commentators disagree widely over the nature of the argument’s obscurity or failure. Karl Ameriks, for instance, dismisses Kant’s argument as profoundly obscure, on the grounds that Kant simply assumes that negative freedom must be attributed to a will; that only later in the section does he

⁴¹ G 452. See also G 453, and KrV A547/B575.

⁴² G 457, 453. See also G 454, 457-8, and 461.

⁴³ G 454.

⁴⁴ See, for instance, Henrich, ‘Die Deduktion des Sittengesetzes’, pp.(317-38), Ameriks, *Kant’s Theory of Mind*, pp.203-9 and 211-6, Allison, *Kant’s Theory of Freedom*, pp.217, 221-8, 237, and 242, and Korsgaard, ‘Morality as Freedom’, pp.161 and 170, and ‘Creating the Kingdom of Ends’, p.201 and n.21.

emphasise the element of his argument which, according to Ameriks, is intended to demonstrate a will's autonomy, namely, our membership of the 'intelligible world'; and that it is unclear precisely which kind of spontaneity Kant's argument refers to.⁴⁵ Allison takes another approach, criticising Kant's argument for demonstrating only that a rational being is something beyond the conditions of sensibility, and can therefore have a will, but not that it has a will which is autonomous or governed by moral laws.⁴⁶ Such criticisms profoundly misrepresent Kant's concerns and argument – most unconvincingly, perhaps, by failing to appreciate that if Kant had assumed or demonstrated a will's negative freedom, then he would also have demonstrated its autonomy, since he maintains that one analytically entails the other.⁴⁷ Other commentators, however, simply marginalise Kant's account of cognitive judgement and 'ideas', with the intention of thus revealing a more defensible line of argument. Michael McCarthy, for instance, claims that Kant's primary argument proceeds from his bare assertion of the necessary negative freedom of rational judging to the claim that a will must therefore be regarded as subject to the laws of the 'intelligible world'.⁴⁸ Thomas Hill similarly focuses on a supposed argument from a simple assertion that the will of a rational being is negatively free, or, at most, from what he calls 'a thought experiment' regarding willing. However, as Hill, if not McCarthy, rightly notes, Kant's third constraint implies that he would consider such a premise to risk merely reflecting a contingency of human nature.⁴⁹

⁴⁵ See Ameriks, *Kant's Theory of Mind*, pp.203-9 and 211-6. See also his later account, in *Kant and the Fate of Autonomy*, pp.70-1 and 74, and Sullivan, *Immanuel Kant's Moral Theory*, pp.84-8.

⁴⁶ See Allison, *Kant's Theory of Freedom*, pp.217, 221-8, 237, and 242.

⁴⁷ See also n.36 above, on Ameriks's and Allison's misinterpretations of the 'circle' referred to at G 450.

⁴⁸ See McCarthy, 'Kant's Rejection of the Argument of Groundwork III', pp.175-80.

⁴⁹ See Hill, 'Kant's Argument for the Rationality of Moral Conduct', pp.263-7, and n.6 to p.252. Wood provides a similar interpretation, if somewhat hesitantly, in his *Kant's Ethical Thought*, pp.171-82, and nn.28, 29, and 30.

Nonetheless, even rightly understood, Kant's argument is flawed. There are three flaws in particular. Firstly, as Ameriks observes, although the intended object of 'ideas' or other judgements may not be determined by possible objects of a cognitive judgement and their conditions, nothing follows from this regarding how its *intending*, in judgement, is to be explained. In particular, it does not follow that its intending is not so determined.⁵⁰ Secondly, as Kant himself notes in the preface, he does not demonstrate 'the unity [...] in a common principle' of a 'critique of pure practical reason' with his critique of 'speculative', or theoretical, reason. This unity is nonetheless presupposed by his extension to rational judgement in general, and to willing in particular, of the spontaneity supposedly revealed by 'ideas' and by the distinguishing of 'appearances' from 'things in themselves'.⁵¹ Finally, in arguing from a will's negative freedom to its autonomy, Kant again invokes the puzzling entailment that he affirms at the beginning of the section.

Having completed his supposed defence of the possible justification of a moral judgement, Kant concludes the section by clarifying its nature and implications.⁵² In particular, he emphasises, firstly, that this defence belongs to 'speculative philosophy', rather than 'practical philosophy'; secondly, that it not only dispels the doubt about autonomy, but also provides a 'speculative' vindication of his analysis of moral goodness in the preceding sections; and, thirdly, that 'speculative philosophy' can do no more for 'practical philosophy' than provide this defence and vindication, and, in particular, cannot 'explain' a will's autonomy in terms of antecedent causes.⁵³ Kant does not, however, attend to

⁵⁰ See Ameriks, *Kant's Theory of Mind*, pp.207 and 213-6.

⁵¹ *G* 391. This seems to be Korsgaard's objection, in her 'Morality and Freedom', p.170, and 'Creating the Kingdom of Ends', n.21.

⁵² In doing so, he also provides brief restatements of his argument. See *G* 454-5, 455-7, 458, 459, and 461, and also the earlier restatement of it at *G* 453.

⁵³ See, respectively, *G* 455-7, 457-8, and 458-63.

the confusion which underpins not only his doubt about autonomy, but also his attempt to dispel it, and, indeed, the latter three clarifications. That is, he does not explain the pertinence of a will's negative freedom, a quality of its causal determination, to its autonomy, the quality of its reasons which a moral judgement attributes to it.

II. *The modest argument of the Critique of Practical Reason*

In the *Critique of Practical Reason*, Kant denies that a 'speculative' argument for the necessity of attributing negative freedom to a will can be provided, because such an argument would require a non-sensible intuition of this freedom, and a human rational being does not have such intuitions. In one passage, he also admits that he is not yet able to demonstrate the required 'unity' of speculative and practical reason. Kant insists that 'speculative philosophy' can demonstrate only the possibility of negative freedom, by distinguishing the possible objects of a cognitive judgement, subject to the principle of causal determinism, from what they are 'in themselves'.⁵⁴ Indeed, he considers the demonstration of this possibility to be sufficient to dispel the doubt about autonomy.⁵⁵ In the *Critique*, then, Kant's response to the doubt is decidedly less ambitious than in the *Groundwork*.

Furthermore, given that the doubt is thus dispelled, Kant claims that the common presupposition that a judgement according to (i) and (ii) can be made, a

⁵⁴ See *KpV* 5-7, 6n, 29, 31, 42-56, 93-9, and 103-6, and also *KrV* Bxxiv-xxx and Bxxvi n. For Kant's admission regarding the 'unity' of speculative and practical reason, see *KpV* 90-1.

⁵⁵ Kant refers to the doubt at *KpV* 7-8, 15-6, 30, 44-5, 54-6, 89-90, and 94-103, and *KrV* Bxxviii-xxx.

presupposition which he here calls a ‘fact of reason’, can itself serve to demonstrate the necessity of attributing negative freedom, along with autonomy, to a will ‘in itself’. He makes this claim simply on the grounds that, by analysis, a will’s autonomy entails its negative freedom, and that, in attributing autonomy to a will, a moral judgement therefore also attributes negative freedom to it. In the *Groundwork*, this entailment gives rise to the doubt about autonomy, in response to which a ‘speculative’ demonstration of the possibility and necessity of attributing negative freedom to a will is provided. In the *Critique*, however, the doubt is supposed to be dispelled merely by the ‘speculative’ demonstration of the possibility of such freedom, and the entailment instead serves a ‘practical’ demonstration of the necessity of attributing negative freedom, along with autonomy, to a will ‘in itself’.⁵⁶ Kant also upholds this more modest position in his later texts.⁵⁷ With this retreat, then, Kant no longer offers a ‘speculative’ demonstration of a will’s negative freedom and autonomy, and consequently also no longer offers a ‘speculative’ vindication of the account of moral goodness which, in the *Critique of Practical Reason* as in the *Groundwork*, he attributes to ‘common moral rational cognition’. In effect, he abandons the pursuit of independent vindication for this account.⁵⁸ Notably, this is so despite the fact that

⁵⁶ For this argument, see *KpV* 3-5, 4n, 6, 15-6, 29-32, 35, 42-50, 54-7, 89-94, 98-100, and 103-6. Unsurprisingly, Kant continues to insist on the impossibility of explaining a will’s autonomy in terms of natural causes. See *KpV* 7, 30, 43, 46, 49, 94-5, and 99, and also *TP* 285 and 285n, *RU* 453-4 and 459-60, *MS* 221, 226, 320n, 378, 380n, and 483, and *SF* 83-4. Note also that, for Kant, the ‘practical’ argument for a will’s negative freedom is significant for the ‘system’ of reason, in that it demonstrates what ‘speculative philosophy’ not only shows is possible, but also shows is necessary to avoid the ‘antinomy’ raised by objects’ subjection to natural causality, and provides ‘practical’ grounds for the other ‘ideas’ of God and immortality. See *KpV* 3-7, 10, 12, 30, 48-9, and 103-6, and *KrV* Bxxvii-xxx.

⁵⁷ See *KrV* Bxxi-xxii, Bxxiv-xxx, and Bxxvi n, *RU* 453-4 and 458-60, *KU* 343, 403, and 468, *R* 26n, 41, 49-50, 50n, 62, and 138, *TP* 284-5, 285n, and 287-8, *MS* 221, 225, 226, 239, 252, 273, 280n, 380, 380n, 383, 396, 404-5, and 418, *ApH* 324, and *VE* 27: 506-7.

⁵⁸ The differences between the arguments in *G* and *KpV* presumably explain why in *G* Kant insists that he provides a ‘critique of pure practical reason’ – that is, a consideration of the ultimate justification of a moral judgement – and in *KpV* he insists only on a ‘critique of practical reason’ – that is, a consideration of the ultimate justification of any practical judgement. See *G* 391-2, 447.

he offers two other arguments for attributing spontaneity to choice, arguments which he does not withdraw and which, if cogent and given spontaneity's entailment of autonomy, would provide such vindication.⁵⁹

The broad differences between the *Groundwork* and the *Critique of Practical Reason* in this regard are generally well recognised, although they are sometimes disputed.⁶⁰ Commentators tend to explain Kant's change of strategy as reflecting his belated acknowledgement of the particular weakness which they identify in the *Groundwork* argument. This has some plausibility in the case of Ameriks's persuasive criticism of the latter argument, given Kant's reference in the *Critique of Practical Reason* to the unavailability of a non-sensible intuition of negative freedom.⁶¹ However, although Kant offers a more modest solution to his doubt regarding autonomy, with both doubt and solution he continues to insist on the puzzling entailment of a will's negative freedom by its autonomy. In this, basic respect, the argument of the *Critique of Practical Reason* fails to improve on that of the third section of the *Groundwork*.

454, and 463, and *KpV* 3 and 15-6, and, of course, the titles of the third section of *G* and of *KpV* themselves.

⁵⁹ Kant even notes this at *KpV* 93-4, immediately before presenting one of these two arguments.

⁶⁰ For the consensus, see, for example, Ameriks, *Kant's Theory of Mind*, pp.191-2, 209-11, and 226-7, and *Kant and the Fate of Autonomy*, pp.62-75, Korsgaard, 'Morality as Freedom', pp.161 and 170, and Allison, *Kant's Theory of Freedom*, pp.201 and 230-43, and 'Autonomy and Spontaneity in Kant's Conception of the Self', pp.138-42. Among the dissenters, Beck, for instance, argues that Kant provides a 'speculative' demonstration of a will's negative freedom in both texts, while Paton, Henrich, and, more recently, Voeller claim that no such demonstration is provided in either text. See Paton, *The Categorical Imperative*, pp.199-278, Beck, *A Commentary on Kant's Critique of Practical Reason*, pp.166-75, Henrich, 'Die Deduktion des Sittengesetzes', pp.(322-335), and Voeller, *The Metaphysics of the Moral Law*, esp. chs.2 and 3. See also McCarthy, 'Kant's Rejection of the Argument of Groundwork III', pp.180-90, Sullivan, *Immanuel Kant's Moral Theory*, pp.88-90 and n.15, and O'Neill, 'Reason and Autonomy in *Grundlegung* III', pp.64-5.

⁶¹ See Ameriks, *Kant's Theory of Mind*, pp.211-20, and, for a less plausible example of this tendency, Allison, *Kant's Theory of Freedom*, pp.227-9, 237, and 242.

Later in the *Critique of Practical Reason*, in the ‘Critical Elucidation’ of its Analytic, Kant offers a fourth argument for attributing spontaneity to ‘intelligible’ choice. This argument is a standard incompatibilist argument from responsibility, premised on the claim that an agent is responsible only for what is ‘within her control’ in the sense that she could do otherwise. Given this premise, Kant’s argument is a persuasive one. However, the premise begs the question of the falsity of a standard alternative sense of responsibility, which Kant dismisses in the course of the argument.

Kant’s argument is, in short, that since the principle of causal determinism entails that any action is sufficiently determined by antecedent causes, and these causes are outside an agent’s control at the moment that she acts, the principle precludes responsibility. To reply that these antecedent causes themselves might have been within an agent’s control, Kant implies, is merely to precipitate a regress, for the further causes of those causes must, given the principle of determinism, have been sufficiently determined by antecedent causes in turn. Since the principle also excludes a first, undetermined cause, the agent’s action or its causes must ultimately be sufficiently determined by causes outside her control. Thus Kant writes, ‘at every point of time I still stand under the necessity of being determined to action by *what is not within my control*, and the *a parte priori* endless series of events which I always only continue according to an already

predetermined order would nowhere begin of itself, [...and] therefore my causality would never be freedom'.⁶²

Kant proceeds to dismiss the standard compatibilist reply. This reply affirms an alternative sense of responsibility, according to which an agent is responsible only for what is 'within her control' in the sense of not circumstantially or psychologically constrained. As Kant expresses it, this reply holds that an agent is responsible for an action only if the action is 'not driven from outside', but rather 'necessary through determining grounds which lie *in* the subject' in the sense that they are 'thought with reason' or determine action 'through representations'.⁶³ This is a distinct sense of responsibility, although it might be expressed in terms of what is 'within the control' of the agent, and also might, psychologically-speaking, encourage belief in responsibility in Kant's sense. However, Kant dismisses the compatibilist alternative out of hand, as 'a wretched substitute', 'a little quibbling about words'.⁶⁴ Simply restating his argument from his own sense of responsibility, he insists that, even if an agent's action is determined in the manner required by the compatibilist sense of responsibility, it is still ultimately and sufficiently determined by antecedent causes which are not in her control. In Kant's well-known words, it is still 'nothing better than the freedom of a turnspit'. Against the compatibilist conclusion that the principle of causal determinism is compatible with responsibility, then, Kant merely opposes his own: that without '*transcendental freedom*', 'no moral law, no imputation according to it is possible'.⁶⁵ To accommodate such 'freedom', of

⁶² *KpV* 95. See also *KpV* 94.

⁶³ *KpV* 96.

⁶⁴ *KpV* 96.

⁶⁵ *KpV* 97. See also *KpV* 96.

course, he once again invokes his distinction between ‘intelligible’ and ‘empirical’ choice.⁶⁶

Kant’s argument from responsibility therefore begs the question of the falsity of the compatibilist’s alternative sense of responsibility. I argued in the first part of the thesis that, on the evidence of his criticism of the common sense conception of agency, Nietzsche does not deny responsibility as such, and in the third part, I will argue that he endorses the compatibilist’s sense of responsibility, and distinguishes it from the sense that Kant endorses. Given this, Kant’s argument ought not to persuade Nietzsche, and Nietzsche’s own concern for restricting responsibility to what is ‘within the control’ of the agent ought to be understood in compatibilist, rather than incompatibilist, terms.

With regard to Kant’s argument from responsibility, there is also some warrant to Nietzsche’s allegation that Kant exploits ‘free will’ to promote what Nietzsche considers an over-inflated sense of responsibility. This is unsurprising, since the argument is that responsibility requires not only causally undetermined choice, but also choice which is ultimately causally sufficient for action. Thus, in presenting his argument from rational choice, as I mentioned, Kant emphasises that ‘natural conditions’ constrain possible actions, and in attempting to dispel his doubt regarding autonomy in the *Groundwork* and the *Critique of Practical Reason*, he also provides no indication that he denies the pertinence of such constraints. Immediately after presenting his argument from responsibility in the *Critique of Practical Reason*, however, Kant claims that the ‘intelligible’ choice required is such that no such constraints can ultimately be pertinent. Most clearly, perhaps, in considering the judgement of a bad action, he represents ‘intelligible’

⁶⁶ See *KpV* 95, 97-100. This argument is also expressed in *RU* 455-60, and, perhaps, at *R* 21, and *MS* 223. In *KpV* Kant proceeds to consider a corresponding problem regarding the divine determination of ‘intelligible’ choice. See *KpV* 100-3.

choice as the 'prosecutor' and 'empirical' choice as the 'advocate', and insists that 'the advocate [...] can in no way silence the prosecutor, if only [...the agent] is aware that at the time he committed this wrong that he was in his senses, i.e. had the use of his freedom'.⁶⁷ This is so, Kant argues, because responsibility requires a choice which is ultimately causally sufficient for action, and thus has no place for constraints.⁶⁸

Nietzsche need not accept this conclusion, since he does not endorse the sense of responsibility on which it is premised. On the other hand, given Kant's sense of responsibility, his conclusion is a persuasive one, and he could dismiss Nietzsche's allegation of its over-inflation for its own misconception of responsibility. In this respect, therefore, Nietzsche's allegation is ultimately inconclusive.

⁶⁷ *KpV* 98.

⁶⁸ See *KpV* 98-100, and also *KrV* A554-5/B582-3.

With the exception of Kant's endorsement, through his argument from responsibility, of what Nietzsche would consider an over-inflated sense of responsibility, Kant's four arguments for attributing spontaneity to choice lie at some distance from Nietzsche's critical concerns. However, in the process of providing these arguments, Kant occasionally makes the claims regarding choice and reasons that Nietzsche criticises. I argued in the previous chapter that Kant's conception of moral judgement can accommodate Nietzsche's criticisms of these claims, and here I will argue that, however unconvincing Kant's four arguments for the spontaneity of choice, Nietzsche's criticisms can be similarly accommodated by the first three arguments and are not pertinent to the fourth.

Firstly, Kant's argument from rational choice is presented in terms of the 'practical' freedom of acting according to reasons, and therefore being 'affected', but not 'necessitated', by 'impulses'. He insists that this freedom 'can be proved by experience' of what 'determines human choice' – that is, he insists that we have 'experience' of rational choice as a sufficient cause of an action. This corresponds closely to the common sense conception of agency to which Nietzsche objects. Of this conception, to recall, Nietzsche suspects that better causal 'descriptions' would identify salient causes of actions other than choice, and causes of choice other than reasons. For the purposes of his argument from rational choice, however, Kant might admit such suspicions. For, firstly, this argument is supposed to demonstrate the necessity of presupposing spontaneity of choice, and neither requires nor entails that choice is also a sufficient cause of an action. I suggested in

the previous chapter that, although Kant defines 'choice' as the sufficient cause of an action, his conception of moral judgement could accommodate a weaker claim. Secondly, Kant's argument requires only that some human choices are made according to reasons, and not that every human choice is of this kind. Indeed, with his affirmation of the 'experience' of rational choice he does not claim that every choice is made according to reasons, but insists merely that 'we have an ability' to act by rational choice. This corresponds with the definition of 'choice' that I considered in the previous chapter, and is consistent with the accommodation of Nietzsche's criticism of practical reason that I proposed there.

Unlike the argument from rational choice, Kant's attempts to dispel his doubt about autonomy in the *Groundwork* and the *Critique of Practical Reason* are presented without affirmations of the causal sufficiency of choice, although these arguments are, of course, concerned with establishing choice's causally undetermined character. Still, like the argument from rational choice, Kant's ambitious and modest attempts to dispel his doubt about autonomy are occupied with rational choice. As he writes with regard to 'every rational being that has a will' in the 'preparation' for his argument in the third section of the *Groundwork*, 'in such a being we think of a reason which is practical, i.e. has causality in view of its objects'. It is this claim, along with the presupposed 'unity' of 'practical' and 'speculative' reason, which licenses Kant to extend the negative freedom of 'speculative' reason to 'practical' reason. The 'practical' demonstration of spontaneity in the *Critique of Practical Reason*, on the other hand, concerns the implications of a moral judgement's attribution of autonomy to rational choice. However, although both arguments are thus occupied with rational choice, neither argument requires that choice be identified with rational choice, or, indeed, be

considered as causally sufficient for action. Nietzsche's criticisms therefore could be accommodated by these arguments, as they could be accommodated by the argument from rational choice.

Kant's attempts to dispel his doubt about autonomy touch upon Nietzsche's critical concerns only insofar as Kant expresses 'negative' freedom as self-causation at the beginning of the third section of the *Groundwork*, an expression which Nietzsche would, given his objection to 'free will', consider incoherent. I suggested that Kant's puzzling affirmation of the mutual entailment of a will's autonomy and its 'negative' freedom might derive from a confusion of the 'self-determination' of self-causation with that of autonomy. Still, Kant does not persist in this expression, and elsewhere in these arguments he conforms with his general practice of conceiving of the 'freedom' that he wishes to establish as simply the absence of determination by antecedent causes. To this conception, I suggested in the first section, Nietzsche offers no further objection.

Finally, besides warranting Nietzsche's allegation that Kant endorses an over-inflated sense of responsibility, Kant's argument from responsibility provides grounds for considering 'intelligible' choice to be causally sufficient for action. That is, it purports to show that choice is responsible only if it is an undetermined, ultimate, and sufficient cause of action. However, while this is clearly an instance of the claim regarding choice which Nietzsche criticises, his criticism with reference to other salient causes revealed by better causal 'descriptions' is not pertinent. For the argument is occupied with the antecedent determination of choice, rather than its causal sufficiency, and Kant's conclusion is not extended from 'intelligible' choice to 'empirical' choice, or choice as it would be subject to causal 'descriptions'. Nonetheless, since the argument is ultimately inconclusive

on its own terms, it need not obstruct the accommodation of Nietzsche's criticism that I proposed in the preceding chapter – that is, the admittance of the causal insufficiency of choice for action.

I have argued for two broad conclusions in this chapter. In the first section, I argued that Nietzsche's criticisms of 'free will' and scepticism are misplaced as criticisms of Kant's distinction between 'intelligible' and 'empirical' choice, his particular attempt to show how choice might be coherently considered as undetermined by antecedent causes. However, in the three following sections, I argued that Kant's four arguments for his claim that choice must be considered as undetermined by antecedent causes are also unconvincing, for reasons independent of Nietzsche's critical concerns and the tenability of Kant's distinction between 'intelligible' and 'empirical' choice. Considering the arguments in the light of Nietzsche's critical concerns, I noted that on the basis of his fourth argument Kant endorses a sense of responsibility that Nietzsche would dismiss as over-inflated, and that in presenting his arguments Kant occasionally makes claims to which Nietzsche objects. However, I argued that the arguments themselves nonetheless do not provide grounds for withdrawing the accommodations that I proposed on Kant's behalf in the previous chapter.

Moral anthropology

This short chapter is occupied with Kant's 'moral anthropology', his project of identifying the obstacles and assistances which 'human nature' offers to the fulfilment of moral requirements. In particular, this chapter undertakes, firstly, to explain the intended theoretical status of Kant's affirmations of human progress, and, secondly, to present the basic objects and means of this progress as he conceives it. The chapter thus argues that, in spite of Nietzsche's dismissive remark in *The Antichrist* regarding Kant's opinion of the French Revolution, Kant neither claims that there is evidence of human progress nor conceives of human progress in moral terms.

I. *Speculative, not explanatory*

Kant's 'moral anthropology' is intended to identify the obstacles and assistances which 'human nature' offers to the fulfilment of moral requirements.¹ In its basic elements, it provides a familiar modern account of humanity's moral failings – that is, it accounts for these failings in terms of the general, and empirically observable, lack of coincidence between moral requirements and the requirements of human inclinations, combined with the general imperfection of human rationality. Regarding human inclinations, Kant conceives of human beings' social life as, at least in part, constituted by individuals' natural inclinations to engage in self-interested competition over scarce or essentially relative goods, such as possessions, power, and honour. Kant refers to these inclinations as humanity's 'unsociable sociability', and, unsurprisingly, identifies them as a source of humanity's moral failings.² However, he could not hold that immoral inclinations alone suffice to explain any of humanity's moral failings, because, according to his conception of moral goodness, a rational agent always has better reason to do what is morally required. Thus Kant consistently insists that human rationality is imperfect in the sense that, even when acting according to reasons, human beings do not necessarily do what they have best reason to do, or

¹ However, although Kant pursues this project consistently, his earlier texts describe 'moral anthropology', or 'practical anthropology', as the application of supreme moral requirements to human conditions, while his later texts merely consider this application to refer to 'anthropology', and describe 'moral anthropology', or 'pragmatic anthropology', as the identification of the obstacles and assistances of 'human nature' to the fulfilment of moral requirements. Compare *G* 387-8 with *MS* 216-7 and *ApH* 119-20, for example.

² Kant's most extensive account of humanity's 'unsociable sociability' is at *I* 19-21, but he also refers to it at *I* 22, *TP* 310, *EF* 345, 360-1, and 365-6, and *MS* 471-2. He explicitly refers to it as a source of humanity's moral failings at *MAM* 119 and *EF* 355.

at least do not necessarily do it for the best reason.³ As I mentioned in the sixth chapter, he also accepts that human rationality can be momentarily suspended or obstructed by passing feelings such as anger, fear, or hope, and that certain persistent inclinations, such as hatred or lust for power, can eclipse other inclinations and concerns in rational reflection.

However, Kant's conception of these two familiar sources of humanity's moral failings is also informed and constrained by his teleological conception of human history, and in this respect, his anthropology is more distinctive. That is, while other modern philosophers tend to treat our immoral inclinations and imperfect rationality simply as contingent, unfortunate constraints to be acknowledged in determining moral requirements, Kant holds that there are also grounds for treating them historically and teleologically, as if they were intended to promote a certain end. As he indicates in his lengthy presentation of them in the second part of his *Critique of Judgement*, these grounds lie ultimately in his claim that, beyond a certain point, the theoretical investigation of certain natural beings – such as trees, for example – can proceed only on the principle that their parts are intentionally organised according to certain internal ends – such as growth, for instance. One must thus consider organisms as if their parts were intentionally organised by a rational agent, according to certain reasons, rather than as simply the effects of preceding causes, Kant claims. In turn, he argues, this requires that organisms also be considered as collectively organised in an equally intentional manner. Undertaking to determine the end, or reason, according to which organisms should be considered as collectively organised, he firstly denies that it could be one of these organisms' internal ends, because the relations between these

³ See, in particular, *G* 412-4, *KpV* 20 and 32, and *MS* 213-4, 221-2, and 379-80, and also *I* 23, and *G* 389, 400n, 421n, and 449. For some discussion, see Hill, 'The Hypothetical Imperative', pp.18-9 and 26-7, and 'Kant's Theory of Practical Reason', pp.124 and 127-8.

organisms and ends are not sufficiently organised for this to be plausible. For example, he argues that it is not clear whether plants should be considered as means to the internal ends of herbivores, and herbivores in turn as means to the internal ends of carnivores, or vice versa. Kant then argues that an end other than organisms' internal ones can be given only by human beings, since they are the only organisms which are rational – that is, they themselves act according to ends, or reasons.⁴

Finally, then, Kant considers the two ends, or reasons, which human beings necessarily and ultimately, if imperfectly, act according to. The first is happiness, but Kant excludes this as incoherent, given individuals' varying and insatiable conceptions of it, and as unlikely, given the suffering which nature persistently imposes on human beings. The second is moral goodness. However, according to Kant's account of moral goodness, it (b') can be achieved only by a will. Thus appealing to his analysis of moral goodness in the *Groundwork* – indeed, he here also refers to its grounds in the conception of 'a good will' – Kant denies that the end, or reason, according to which organisms should be considered as collectively organised can be moral goodness as such. Nonetheless, he argues that this end must be the development of rationality, since this serves the end of moral goodness and also satisfies every other condition which his argument identifies. That is, the development of rationality not only is not an internal end of any organism, including human beings, but is also given by human beings as rational agents, and is coherent, likely, and achievable by nature. Kant calls the development of rationality 'culture', and divides it into two developments: firstly, the development of the general ability to achieve ends which he calls 'skill' and, secondly, the

⁴ See KY 359-429.

weakening of immoral inclinations and the cultivation of inclinations which coincide with, rather than oppose, moral requirements.⁵

By means of this argument, then, Kant attempts to demonstrate that our theoretical investigation of natural organisms requires that we consider them as collectively organised according to a certain end – namely, the perfection of human ‘skill’ and the coincidence of human inclinations with moral requirements. In his various attempts to fill out this conception of human progress in the *Critique* and his essays, he presents human progress as driven by the ‘unsociable sociability’ and imperfect rationality which are otherwise simply unfortunate contingencies. These attempts are, therefore, mere speculations, made from the broad and otherwise indeterminate perspective of his conclusion regarding the collective, teleological organisation of organisms, and applied to his broad empirical claim about human beings’ ‘unsociable sociability’ and his insistence on their imperfect rationality. Indeed, the speculative status of these attempts is consistently underlined by Kant. That is, he consistently emphasises that his attempts are not intended to be based upon empirical evidence of human progress, but are nonetheless also not intended as mere fictions, since they are guided by his teleological constraints and his broad claims regarding human ‘unsociable sociability’ and imperfect rationality.⁶

⁵ See *KU* 429-445. Related remarks are made at *TP* 308 and 310, *EF* 360-2, and *ApH* 324 and 327-30, and on the incoherence of happiness, see, for instance, *G* 395-6 and 417-9, and *KpV* 25-6.

⁶ See *I* 29-31, *MAM* 107-10, *TP* 311-2, and *EF* 368, for example.

II. Skills and inclinations

A notable example of Kant's speculations on human progress is provided by his essay, 'Conjectural Beginning of Human History'. There he speculates, by means of an unorthodox interpretation of the biblical story of the Fall, on the initial consequences of humanity's gaining rationality. His basic claim is that, through experiment, rationality extends humanity's inclinations beyond, and even against, its instincts, and thus gives humanity the freedom to choose while simultaneously denying it the guidance of instinct. In the light of this basic claim, Kant then offers two speculations. Firstly, he proposes that the effect of rationality on the sexual instinct was profound, since rationality allows this instinct to be creatively prolonged and complicated, particularly by the withdrawal of its object and thus the postponement of its satisfaction. This suggests a broader speculation, Kant claims: namely, that such creativity with inclinations, particularly by concealment, is the source of social customs, and thus the source of humanity's sociability and the 'unsociable', competitive nature of this sociability.⁷ Importantly, this reflects his persistent view of rationality as developing socially, by experiment, practice, and instruction across individuals and generations, and thus, ultimately, in the species as a whole, rather than in individuals.⁸ Similarly implicit, if not elaborated upon here or elsewhere, is his view of the development of inclinations as, for the most part, equally social.⁹

Secondly, Kant speculates that, by denying free choice the guidance of instinct and also by enabling human beings to anticipate the future, rationality introduced enormous anxiety into human life – in particular, anxieties about how

⁷ See *MAM* 112-3.

⁸ See *I* 18-9 and 22, and *MAM* 114-5.

⁹ Kant emphasises the social and species level of development generally at *TP* 310 and 312.

to choose, about the burdens and uncertainty of the future, and about death. Kant suggests that such anxieties of rationality, along with inevitable dissatisfaction with society's created pleasures, drive human beings to pursue an imagined 'paradise' of rest and peace in which the anxious exercise of rationality would not be much required. The tragedy of this pursuit, however, is that it is precisely their rationality which denies them such a 'paradise' and which they must exercise even in trying to avoid doing so.¹⁰

Thus Kant speculates that humanity's 'unsociable sociability', and also certain other driving anxieties, can be traced to its rationality, a possession which, once gained, cannot be lost. Unsurprisingly, he claims that rationality also introduces moral requirements, but that the rationality which humanity gains is imperfect and that the inclinations which it gains do not all coincide with moral requirements. Thus he concludes that, besides introducing many new pains, sociability and the anxieties of rationality also introduce many moral failings.¹¹ However, the distinctive, teleological character of Kant's speculations is reflected by his further claim, made in 'Conjectural Beginning' as elsewhere, that human beings' imperfectly rational pursuit of their inclinations – and particularly their immoral ones – serves to develop the rationality of the species as a whole. In particular, regarding human 'skill', he speculates that the 'unsociable', competitive nature of humanity's sociability stimulates human beings to develop skills by experiment, practice, and instruction.¹² Regarding inclinations, he suggests that conformity to social customs can coincide with moral requirements and weaken immoral inclinations, and is supported by the modern pursuit of arts and sciences, despite the evils which, for Kant as for Rousseau, these pursuits also presuppose

¹⁰ See *MAM* 110-5 and 122-33. See also *I* 19-20 and *ApH* 324.

¹¹ See, for example, *MAM* 116-8.

¹² See *I* 20-2, *MAM* 114-5, *KU* 433, *TP* 312, and *ApH* 321-5.

and produce.¹³ In general, then, Kant's anthropological speculations confirm his remark in 'What is Enlightenment?', regarding the constraints necessary to develop freedom of thought, that 'a strange, unexpected course is shown in human affairs [...] if one considers it in the large, in which almost everything is paradoxical'.¹⁴

However, Kant's most extensive speculations concern the political realm. He considers this realm to be characterised by human beings' competition over possessions, and, unsurprisingly, he holds that human beings' immoral inclinations and imperfect rationality tend to ensure that this competition collapses into 'war', the exercise or mere threat of violence. However, Kant speculates that this competition and violence also might stimulate the development of human skills in two ways. Firstly, he speculates that this violent competition disperses human beings, and thus forces them to develop the skills necessary to live in less habitable regions of the earth.¹⁵ Secondly, he speculates that this violent competition gives each human being a strong inclination to unite with others in a community which can provide him or her with security, and that this inclination, along with the countervailing inclination to resist constraints for the sake of competitive inclinations, stimulates human beings to pursue the perfect balance between security and freedom from constraint.¹⁶ For Kant, a community can provide security only by the public administration of justice, in however rudimentary a form – that is, only by a civil condition – and only when there is also a certain

¹³ See, in particular, *KU* 432-4, and also *I* 19-21, *MAM* 110-5 and 121-6, and *ApH* 324-5.

¹⁴ *WA* 41.

¹⁵ See *MAM* 118-9 and *EF* 363-4. In these passages, Kant also speculates that human beings' competition over possessions originated in the differing requirements of pastoral and agricultural means of subsistence, when these means had emerged from uncompetitive hunting and gathering.

¹⁶ See *I* 22-6, *MAM* 119, *KU* 432-3, *TP* 310 and 312, *EF* 363-4 and 365-7, and *ApH* 327-8.

‘civil’ relation between states.¹⁷ Furthermore, he considers the pursuit of security in such communities as the development of a skill, and therefore as a pursuit which must proceed by experiment, practice, and instruction.¹⁸ He also considers this skill crucial because its consequences broadly coincide with general political obligations, and because the establishment of a perfect constitution is also a necessary condition for the development of other skills.¹⁹ As he concisely expresses the ‘problem’ in *Toward Perpetual Peace*, ‘The problem of establishing a state [...] is soluble even for a nation of devils (if only they have understanding) and goes like this: “Given a multitude of rational beings all of whom desire universal laws for their preservation but each of whom is inclined covertly to exempt himself from them, so to order this multitude and establish their constitution that, although in their private dispositions they strive against one another, these yet so check one another that in their public conduct the result is the same as if they had no such evil dispositions”’.²⁰ In response to this ‘problem’, here and elsewhere he speculates, in particular, that the economic demands of ‘war’ between states can require them to uphold certain civil freedoms and constrain their belligerence, and that trade between states can weaken inclinations to immoral violence and cultivate inclinations to peace.²¹

In this context, Nietzsche’s misunderstanding of Kant’s claims regarding the French Revolution can be appreciated. To recall, Nietzsche alleges that, according to Kant, the French Revolution ‘proved’, or could only be ‘explained’ by, a ‘moral predisposition’ or “tendency [...] towards the good” on the part of

¹⁷ On the need for a civil condition, see the passages referred to in n.16, and on the need for a ‘civil’ relation between states, see, in particular, *I* 24-8, *KU* 432-3, and *TP* 310-1.

¹⁸ See, for instance, *I* 22-3, and *ApH* 328.

¹⁹ See *I* 21-9, *MAM* 121-6, *KU* 432-3, *TP* 310-12, and *EF* 363-8.

²⁰ *EF* 366. See also *I* 22-6, *KU* 432-3, and *TP* 312.

²¹ See *I* 27-8, *MAM* 119-20, *TP* 311, and *EF* 364 and 368.

humanity. This allegation is presumably directed at claims which Kant makes in the second part of his late work, *The Conflict of the Faculties*, which Nietzsche apparently read between May and June 1887. In particular, this part of the work refers to a public ‘attitude’ of ‘sympathy’, and even ‘passion’ or ‘enthusiasm’, for the constitutional changes of the French Revolution, on the part of spectators of it. Kant insists that this ‘attitude’ not only had rightful concerns, but that it was also both universal and unselfish. He claims, furthermore, that its universality justifies its attribution to human beings as such, and that its object is ‘too momentous, too intimately interwoven with the interests of humanity, and too widespread in its influence’ for the ‘attitude’ not to have a significant influence on later humanity.²² However, despite the apparently empirical nature of these claims, Kant prefaces them by again denying precisely what Nietzsche alleges, namely, that he affirms human progress on the grounds of empirical evidence.²³ Indeed, his following remarks demonstrate that both the speculative status and the objects of these claims correspond with those of Kant’s earlier attempts to fill out his teleological perspective on human history. In particular, he presents the ‘attitude’ towards the French Revolution as admitting the speculation that humanity’s ‘unsociable sociability’ might prompt the development of human skill in political constitutions, but not progress in moral goodness.²⁴ Thus, although Nietzsche rightly points out that Kant describes this ‘attitude’ as ‘moral’, he fails to note that Kant describes it as ‘moral’ only in the ‘unselfish’ sense, rather than in his own sense.²⁵

It might be objected that elsewhere, if only speculatively, Kant affirms human progress in moral goodness in his own sense. In particular, in his earlier

²² *SF* 85, 86, 88. See also *SF* 87.

²³ See *SF* 79 and 83-4.

²⁴ See *SF* 91-3.

²⁵ See *SF* 85-6. It follows, I would suggest, that the ‘attitude’ also does not express a moral commitment or demand, as Krasnoff maintains in his ‘The Fact of Politics’, pp.29-34.

essays he makes a few brief remarks which might suggest that he considers human progress to extend to moral goodness – that is, to rationality not in the sense of ‘culture’, but in the sense of action according to morally good reasons. Most suggestive, perhaps, are Kant’s remarks in ‘Conjectural Beginning’ that ‘the first hint of man’s development as a moral being’ comes with the ‘*sense of decency*’ which informs social customs, and that an ‘obscure’ appreciation of, or ‘a distant preparation’ for, the status of each rational being as an ‘*end in himself*’ also comes with the development of reason.²⁶ These are hardly conclusive contraventions of the *Critique*’s considered denial of speculation in human moral progress, however. For the appreciation, or ‘preparation’, referred to is ‘obscure’, or ‘distant’, and consists simply of the treatment of animals as mere ‘means’ which allegedly follows on the gaining of rationality, while ‘decency’ is supposed to provide merely a ‘first incentive’ to ‘concealing all that might invite contempt’, rather than progress in acting according to morally good reasons.²⁷ In other words, these remarks continue Kant’s speculations about the development of ‘skill’ and morally coincident inclinations, rather than extend these speculations to moral goodness itself. Despite Nietzsche’s – and, indeed, many other commentators’ – claim to the contrary, therefore, Kant does not extend his affirmations of human progress to moral goodness, in any respect.²⁸

In conclusion, however, a final element of Kant’s treatment of human progress should be noted. That is, in ‘On the Common Saying: “That May be

²⁶ *MAM* 113, 114. See also the remarks on Rousseau at *MAM* 116. Also suggestive, perhaps, is Kant’s reference to the development of ‘a *moral whole*’ of human society at *I* 21, but the context, and his following ‘propositions’, indicate that he is thus referring to the development of a civil condition.

²⁷ *MAM* 114, 113.

²⁸ For recent examples of the contrary claim that Kant affirms human moral progress, at least in certain respects, see Kleingeld, ‘Kant, History, and the Idea of Moral Development’, pp.62-4, and Loudon, *Kant’s Impure Ethics*, pp.142-143, and 153-157. Equally implausible, I would suggest, is Wood’s claim that Kant’s teleological human history concerns no determinate and final end at all, but only ‘freedom’ or ‘reason’. See Wood, ‘Kant’s Historical Materialism’, pp.20-23.

Correct in Theory, but it is of no Use in Practice’, Kant apparently claims that one must nonetheless ‘hope’ for historical progress in human moral goodness. He argues that this ‘hope’ is admissible as long as such development cannot be ‘proved’ impossible, and that it is required to fulfil a duty to make some improvement to human ‘posterity’. Furthermore, he claims that this ‘hope’ must be addressed to how ‘human *nature*’ must be considered to promote ‘culture’, rather than to morally better actions as such.²⁹ It is possible that the object of Kant’s ‘hope’ here is merely the development of a civil condition within and between states, since he discusses only this development in the remainder of this part of the essay.³⁰ However, even if its object is human moral progress, Kant’s ‘hope’ does not express a particular theoretical explanation of human history or even a teleological speculation of the kind that he offers, here and elsewhere, regarding human ‘culture’. Essentially, this ‘hope’ simply affirms that historical development in human moral goodness is possible, and that a duty to pursue it is therefore fulfilable, and thus possible. Beyond this, Kant simply suggests that, without such ‘hope’, ‘an earnest desire to do something profitable for the general well-being would never have warmed the human heart’.³¹

²⁹ *TP* 309, 310. See also *TP* 308.

³⁰ See *TP* 310-3. Such an interpretation should be given to the similar remark at *EF* 365, I suggest.

³¹ *TP* 309.

Nietzsche's dismissive claim in *The Antichrist* – that, according to Kant, the French Revolution 'proved', or could only be 'explained' by, a 'moral predisposition' or "'tendency [...] towards the good'" on the part of humanity – therefore misrepresents the intended status and the objects of Kant's affirmations of human progress, in general and in *The Conflict of the Faculties* in particular. Firstly, although Kant's broad teleological perspective on human history has theoretical grounds, his claims about particular instances and means of human progress are not intended as claims based upon empirical evidence, but as speculations guided by his teleological constraints, his broad empirical claim regarding human beings' 'unsociable sociability', and his insistence on their imperfect rationality. Secondly, on the grounds that (b') only a will can achieve moral goodness, these speculations are restricted to progress in human 'skill' and in inclinations coincident with moral requirements, ends which, Kant insists, serve the end of moral goodness, but are distinct from it. Notably, these two basic aspects of Kant's 'moral anthropology' are not only reflected in the second part of *The Conflict of the Faculties*, at which Nietzsche presumably directs his remark, but are also presented at some length in Kant's *Critique of Judgement*, a text with which Nietzsche was apparently better acquainted.

Conclusions of part II

In this part of the thesis, I undertook to defend Kant's conception of the moral agent against Nietzsche's explicit criticisms and certain other criticisms commonly made in Nietzsche's name.

In the fifth chapter, I offered an interpretation of Kant's argument in the first and second sections of the *Groundwork*. This revealed that Kant derives his formulas as expressions of two distinctive features of moral goodness which emphasise different aspects of one or both of these features, and that he identifies these features by analysing a conception of the goodness of 'a good will' which he attributes to moral common sense at the beginning of the first section. His analysis can be summarised as follows. Its object, the conception of the goodness of 'a good will', consists of two claims: a good will is (a) the only good which is good in all contexts, and (b) a condition of the goodness of every other good. From (a), Kant draws the conclusion that (a') a good will does what is morally good because it is morally good. By distinguishing between action 'from duty' and action 'from inclination, he argues that this conclusion follows because, to be (a) good in all contexts, a good will's motivation to do what is morally good must be necessarily concerned with what is morally good, rather than contingently and therefore, at most, coincidentally so concerned. From (b), on the other hand, he draws the conclusion that (b') moral goodness is a goodness which only a will can achieve. This conclusion follows, he argues, because the goodness of a good will, which he equates with moral goodness, could not be (b) a condition of the goodness of every other good if this condition could be fulfilled by other causes. From conclusions

(a') and (b'), Kant proceeds to derive two distinctive features of moral goodness, (i) and (ii). He derives the first such feature from (b'), by arguing that, if (b') moral goodness is a goodness which only a will can achieve, then it must be (i) a goodness of a will's reason for action alone. He then argues that moral goodness must also be (ii) good for a will as such, on the grounds that (a') a good will does what is morally good because it is morally goodness and (ii) this goodness must be a goodness of its reason for action alone.

In the sixth chapter, I proceeded to consider the interpretation of the formulas and their place in Kant's conception of moral judgement. My broadest concern here was to demonstrate that Kant conceives of moral judgement as judgement according to a system of duties, to be derived by considering contingent conditions of varying generality under the distinctive features of moral goodness (i) and (ii), as expressed, emphasised, and clarified by the formulas. I emphasised that this conception reveals a number of common interpretations of the formulas to be misleading, if not simply mistaken. It also reveals Nietzsche's criticisms of the formula of universal law to be misplaced, since they misrepresent this formula either as a 'test' of reasons for action in specific circumstances or as precluding 'pleasure' in the achievement or determination of moral goodness. Nor, I mentioned, does this conception suggest that Kant shares Nietzsche's concern for individual 'autonomy'. However, I also argued that Nietzsche's criticisms of the common sense conception of agency, while not offering conclusive refutations of Kant's conception of moral judgement, nonetheless require certain accommodations from it. In particular, Kant should admit that, considered at a higher level of particularity than prescriptions according to reasons can capture, actions might be considered as not motivated by reasons, and that better causal

'descriptions' might reveal that choice is not causally sufficient for action. Finally, I argued that Nietzsche's demand for independent criticism of moral common sense would be regarded as extravagant by Kant, given his ultimate methodological commitment to common sense.

In the following chapters, I considered two further branches of Kant's moral philosophy to which Nietzsche's explicit criticisms refer. In the seventh chapter, I considered Kant's attempt to accommodate his claim that choice must be considered as undetermined by antecedent causes, and the four arguments that he provides for the claim itself. I argued that, while his attempt to accommodate this claim is not vulnerable to Nietzsche's criticisms, Kant's four arguments for the claim are independently unconvincing. In the eighth chapter, I demonstrated that, despite a dismissive remark of Nietzsche's, Kant neither claims that there is evidence of human progress nor conceives of human progress in moral terms.

Part III

The Animal That May Promise: Nietzsche's Rival Analysis of the Moral Agent

Nietzsche's rival analysis

In this chapter, I defend the third broad conclusion of the thesis, that neglected elements of Nietzsche's moral philosophy provide for a critical engagement with Kant's conception of the moral agent that is more sophisticated, telling, and original than that articulated by Nietzsche's explicit criticisms of Kant. In particular, I argue that, with his account of the 'animal that may promise' in the *Genealogy*, Nietzsche presents an alternative, but nonetheless Kantian, conception of the moral agent which reveals that Kant makes two significant, yet undefended, assumptions in deriving the basis features of his own conception. While one assumption merely regards their ultimately inconclusive differences over motivation, the other presents a substantial challenge to Kant's interpretation of one of his distinguishing features of moral goodness. In response, I propose a resolution that owes something to both Kant and Nietzsche. Furthermore, I argue that Nietzsche's emphasis on the historical origins and current condition of his conception of the moral agent represents an attempt to vindicate that conception, and that Kant could accept this manner of vindication.

The chapter is therefore divided into three sections, each based upon an aspect of Nietzsche's account of the 'animal that may promise' at the beginning of the second essay in the *Genealogy*. The first section identifies the conception of choice presented by this account, and notes how this conception avoids the

common sense claims regarding choice and reasons, and the related claims regarding responsibility, to which Nietzsche objects elsewhere. The second, lengthier section then considers the conception of goodness which accompanies this account of the 'animal that may promise', and its implications for Kant's conception of the moral agent. Finally, the third section turns briefly to Nietzsche's remarks about the historical condition of the 'animal that may promise', and the kind of vindication which they offer for his position.

Nietzsche presents his positive conception of choice at the beginning of the second essay of *The Genealogy*, as an account of what he calls the 'animal that may promise'. Here he conceives of choice as an ability (he refers to it as 'ein Vermögen' and with 'können' and 'dürfen') which itself comprises two abilities and one inability, each of which he articulates in terms of memory. He maintains that these two abilities and one inability oppose each other, but that in choice they are hierarchically arranged such that one ability overcomes the other, which, in turn, overcomes the inability. Thus he describes the ability to forget as first overcoming the inability to forget – that is, the inability to “cope [literally, be finished, *fertig werden*]” with’, or ‘digest’, experiences and desires, and the processes of ‘inanimation [*Einverseelung*]’ through which they are ‘digested’. Forgetting, then, limits an agent’s consciousness – it ‘digests’, ensures the ‘inhibition’ of, or ‘temporarily close[s] the doors and windows of consciousness’ to, her experiences, desires, and inanimistic processes. It thus gives the agent a ‘present’, Nietzsche claims, ‘a little stillness, a little *tabula rasa* of consciousness, so that there is again space for new things, above all for the nobler functions and functionaries, for ruling, foreseeing, predetermining (for our organism is arranged oligarchically)’. He proceeds to claim that this noble ability to forget also makes ‘space’ for a yet ‘nobler’ ability, that of making and keeping a promise, which involves ‘an opposing ability, a memory’, an ability which in choice opposes and overcomes the ability to forget. Thus he writes that promising involves ‘an active *willing-not-to-be-rid-of*, a continuous willing of something once willed, a real

memory of the will'. In promising, then, a choice is actively not forgotten, such that in choosing the agent is able 'to stand security for himself *as future*'.¹ Finally, Nietzsche also maintains that this '*memory of the will*' requires a number of subsidiary abilities,

so that between the original 'I will', 'I will do' and the actual discharge of the will, its *act*, a world of strange new things, circumstances, even acts of will may be safely put in between without this long chain of will breaking. But how much this presupposes! In order to dispose of the future in advance in this way, man must first have learned to distinguish necessary from chance events, to think causally, to see and anticipate the distant as if it were present, to fix with certainty what the purpose is and what the means to it, in general to be able to calculate, compute[.]²

This account of the 'animal that may promise' expresses a conception of conscious choice as causally significant for the chosen action, and as made for reasons. It does not, however, make the two claims which comprise the common sense conception of agency to which Nietzsche objects elsewhere. Firstly, it conceives of choice as causally significant for the chosen action, despite the 'strange new things' that might intervene between choice and action. But, by insisting that choice is an ability that presupposes other abilities and a certain hierarchical arrangement among them, this conception does not consider choice as causally sufficient for action. Indeed, in this as in its hierarchical structure, this conception is suggestive of Nietzsche's phenomenology of choice in *Beyond Good and Evil*, with which, I argued, he intends to raise doubt about the claim

¹ *GM* II 1. For other remarks which associate the ability to will with the ability to forget, see *JGB* 230 and *GD* VIII 6.

² *GM* II 1.

that choice is causally sufficient for action. This conception also does not claim that choice is undetermined by antecedent causes, or endorse a sense of responsibility which Nietzsche would dismiss as underwritten only by such ‘free will’ and as over-inflated. With regard to responsibility in particular, he writes that choice as he conceives it here provides the agent with ‘the extraordinary privilege of *responsibility*’, but he does not suggest that an agent’s responsibility might extend to her ‘strength’ or her ‘being this or that’, as he alleges that proponents of ‘free will’ unjustifiably extend it.³

Notably, here Nietzsche also explicitly distinguishes his sense of responsibility from the sense to which Kant appeals in arguing from responsibility to the spontaneity of choice. Of the notion that punishment might rest upon a ‘presupposition about freedom or unfreedom of the will’, Nietzsche writes, ‘[t]hat thought, [...] which has even had to serve for an explanation of how the feeling of justice came into being on earth at all, “the criminal earned punishment *because* he could have acted otherwise”, is in fact an extremely late attained, even sophisticated form of human judging and inferring’. He distinguishes this particular sense of responsibility from ‘those much more primitive distinctions “intentional”, “negligent”, “accidental”, “accountable”, and their opposites’ with which he and his conception of choice are concerned.⁴ Anticipating this in my discussion of Kant’s argument from responsibility in the seventh chapter, I proposed that, while Kant’s premise is that an agent is responsible for what is ‘within her control’ in the sense that she could do otherwise, Nietzsche considers an agent responsible simply for what is ‘within her control’ in the sense of not being circumstantially or psychologically constrained. By thus endorsing an

³ *GM* II 2.

⁴ *GM* II 4.

alternative, compatibilist sense of responsibility, I argued, Nietzsche might avoid the incompatibilist conclusion of Kant's otherwise persuasive argument from responsibility.

Secondly, although Nietzsche conceives of choice as made for reasons, he does not claim that it is motivated by them. He insists that to choose requires the agent to be able 'to fix with certainty what the purpose is and what the means to it, in general to be able to calculate, compute'. He also claims that this reasoning requires that the agent be able to 'forget' her passing desires and experiences. In this, his conception of choice bears some similarity to Kant's definition of choice, and particularly Kant's claim that to act according to reasons is to be 'free' in the sense of 'affected, but not determined, by impulses'.⁵ However, Nietzsche does not claim that choice is motivated by the cognitive acknowledgement of reasons. Indeed, he would be unlikely to do so, given that, as I have shown in the first part of the thesis, elsewhere he rejects the corresponding common sense claim regarding reasons, and consistently considers choice as instead motivated by noncognitive states informed by reasons. In the following section, I will show that he also confirms this position with regard to his positive conception of choice, as presented by his account of the 'animal that may promise'.

⁵ This aspect of Nietzsche's conception is also emphasised in late notes. See, for instance, *KGW* VIII:2 9 [139, 169, and 178] (Autumn 1887) and 11 [353] (November 1887-March 1888). I emphasise its similarity to Kant's claim about the 'freedom' of acting according to reasons in my earlier account of the 'Kantian' form of Nietzsche's conception, in 'Nietzsche's Kantian Ethics', pp.10-11.

I.

Nietzsche's account of the 'animal that may promise' at the beginning of the second essay of the *Genealogy* is accompanied by an associated conception of goodness.⁶ In particular, he proceeds to write the following of this 'animal', which he now refers to as 'free' and as 'sovereign'.

The 'free' human being, the possessor of a long unbreakable will, also has in this possession his *measure of value*: looking out from himself upon others, he honours or he despises; and just as necessarily as he honours his equals, the strong and reliable (those who *may* promise), – therefore, everyone who promises like a sovereign, weightily, seldom, slowly, who is stingy with his trust, who *distinguishes* when he trusts, who gives his word as something that can be relied upon because he knows himself strong enough to uphold it against accidents, even 'against fate' –: just as necessarily he will hold his kick ready for the feeble windbags who promise but may not, and his switch for the liar who breaks his word at the very moment he has it in his mouth. The proud knowledge of the extraordinary privilege of *responsibility*, the consciousness of this rare freedom, this power over oneself and fate, has in him sunk down to his lowest depths and become instinct, the dominating instinct [...] this sovereign human being calls it his *conscience*...[.]⁷

⁶ I take the following to discredit the common denial that Nietzsche presents the 'animal that may promise' as an ideal. See, for instance, Hatab, *A Nietzschean Defense of Democracy*, pp.37-8. My account expands upon that given in my 'Nietzsche's Kantian Ethics', pp.14-21.

⁷ *GM* II 2.

Nietzsche, then, holds that the ability to choose provides its possessor with 'his *measure of value*'. Furthermore, Nietzsche's account of this 'measure' indicates that he considers it to identify a particular kind of reasons of 'good and bad', the reasons whose general features he presents in the first essay of the *Genealogy*. In interpreting these general features in the second chapter of the thesis, I showed that Nietzsche presents reasons of 'good and bad' as ultimately justified only by appeal to a distinguishing characteristic of exemplary 'good', or 'bad', agents, a characteristic which is supposed to bestow 'goodness', or 'badness', on their actions. As I expressed it, such a reason requires an agent to do, or not do, *x* ultimately because agents with a distinguishing characteristic *y* do, or do not do, *x*. I also demonstrated that Nietzsche insists that obedience to such reasons is motivated by an affirmation of one's 'self' as possessing the appropriate 'goodness'-bestowing characteristics. Since I was concerned with sketching Nietzsche's 'revaluation' of the ostensibly identical reasons of 'good and evil', however, I did not mention how he understands the requirements of reasons of 'good and bad' to be determined. Unsurprisingly, he presents these reasons as requiring what would prove an agent's possession of the appropriate 'goodness'-bestowing characteristics, rather than 'badness'-bestowing ones. However, besides reflecting the nature of the particular characteristics concerned, Nietzsche insists that these reasons' requirements are determined according to a constant and creative mutual measurement among those credited with sharing a 'goodness'-bestowing characteristic. He intimates this in the first essay of the *Genealogy* when he writes of how the "“good ones”" – that is, those credited with sharing a 'goodness'-bestowing characteristic – 'are so strictly held within limits *inter pares* by custom, respect, usage, gratitude, still more by mutual suspicion, by jealousy,

[...and] in relations to each other prove themselves so inventive in consideration, self-control, delicacy, loyalty, pride, and friendship'.⁸ However, he expands upon this remark only in sections of the chapter, 'What is noble?', in *Beyond Good and Evil*. The following passage provides a relatively extensive account.

egoism belongs to the essence of the noble soul, I mean the immovable faith that to a being such as 'we are' other beings must be subordinate by their nature, and sacrifice themselves to us. [...] Under circumstances which make it hesitate at first, it admits that there are equals-in-rights with it; as soon as it is clear as to this question of rank, it moves among these equals and equals-in-rights with the same certainty in modesty and tender reverence as it has in intercourse with itself [...] it honours *itself* in them and in the rights it concedes them, it does not doubt that the exchange of honours and rights, as the *essence* of intercourse, likewise belongs to the natural condition of things. The noble gives as it takes, out of the passionate and sensitive instinct of requital which lies in its ground.⁹

Nietzsche mentions a variety of 'goodness'-bestowing characteristics in terms of which such 'requital' among 'equals' might be practiced, and thus the requirements of reasons of 'good and bad' determined. In the first essay of the *Genealogy* he refers to being 'blond-headed', a 'warrior', or 'truthful', for instance. However, when he presents the ability to choose as a 'measure of value' at the beginning of the second essay, he identifies this ability as such a 'goodness'-bestowing characteristic, and therefore as that according to which, by 'requital' among 'equals', requirements of reasons of 'good and bad' are to be determined. In particular, Nietzsche there describes how, by 'looking out from himself upon

⁸ *GM I* 11. Nietzsche remarks similarly on enmity among the 'good ones' at *GM I* 10.

⁹ *JGB* 265. See also *JGB* 259, 262, 263, 272, and 287, and the similar remarks at *FW* 3.

others', one who is able to choose measures others according to this ability, and 'honours' those equally able to choose. Immediately after the quoted passage, he also emphasises that being able to choose means 'that one *may affirm oneself*'.¹⁰ In the light of his more extensive account of 'honour' among self-affirming 'equals' in *Beyond Good and Evil*, then, Nietzsche can be presumed to consider the ability to choose as a 'measure' which delimits a sphere of those to whom an agent has duties and against whom she has rights, and among whom these duties and rights are to be determined by a constant and creative mutual measurement of this ability.

This presumption finds confirmation and elaboration in Nietzsche's earlier texts. Most notably, in a section of the preface to the *Genealogy* Nietzsche refers to ten sections of his earlier texts which, he insists, prefigure claims made in the *Genealogy*. Six of these ten sections present his notion of 'requital' in some detail, and one of the six sections is one of two successive sections of *Daybreak* which present lengthy analyses of 'requital' precisely in terms of the ability to choose.¹¹ The section of *Daybreak* begins by presenting 'requital' in terms of 'power', as follows.

Our duties – are the rights of others over us. How have they acquired these rights? By taking us to be capable of contracting and of requiting, positing us as similar and equal to them, and consequently entrusting us with something, training, reproving, supporting us. We fulfil our duty – that is: we justify that representation of our power according to which everything was shown to us, we give back in the measure in which one gave to us. It is therefore our pride which bids us do our duty, – when we do something for others in return for something they have done for us, we will the

¹⁰ *GM II 3*.

¹¹ Nietzsche refers to ten earlier passages at *GM V 4*. The other five that concern 'requital' are *MA 45* and *92*, and *WS 22*, *26*, and *33*.

restoration of our autocracy, – for they have intervened in our sphere of power and would have continued to have their hand in it if we did not practise, with ‘duty’, a requital, that is, intervene in their power.¹²

Crucially, here Nietzsche proceeds to equate one’s ‘power’ with one’s ability to choose, and also to emphasise that it is one’s own and others’ beliefs about one’s ‘power’ which determines one’s duties and rights. Thus he writes, ‘the feeling of duty depends upon our having the same *belief* in regard to the extent of our power as others have: that is, that we *are able* to promise certain things and bind ourselves to perform them (“freedom of will”)’.¹³ For Nietzsche, then, one’s duties and rights correspond precisely with the degree of one’s ability to choose, as this is posited by oneself and by others, and determining these duties and rights therefore ‘constantly needs the refined tact of a balance’, to weigh the shifting degrees of one’s own and others’ posited ability to choose.¹⁴ Notably, he includes benevolence among these duties and rights, as a concern for maintaining a ‘sphere of power’ extended to include the spheres of subordinate, unequal others, but he also insists that duties and rights extend no further beyond one’s equals than this.¹⁵

Nietzsche restates this particular account of ‘requital’ in the following section of *Daybreak*, and in sections of *Human, All Too Human* and *The Gay Science*.¹⁶ As he indicates when referring to it in the preface to the *Genealogy*, he also considers it particularly significant for the explanation of political duties and rights, or ‘justice’, as ‘a balance between those of approximately equal power’. It is unsurprising, then, that after presenting his conception of the ability to choose as

¹² *M* 112.

¹³ *M* 112. Nietzsche also emphasises ‘believed’ power at *M* 113, 146, and 326.

¹⁴ *M* 112.

¹⁵ See *M* 112.

¹⁶ See *MA* 44, *M* 113, and *FW* 13. Notably, Nietzsche also refers to *M* 113 in *GM* II 6.

a 'measure of value' at the beginning of the second essay, Nietzsche proceeds to present the requirements of 'justice' as determined by the practice of 'comparing, measuring, calculating power against power'. In particular, he defines 'justice' as 'the good will among those of approximately equal power to come to terms with one another, to "understand" each other through a balance – and, regarding those of less power, to *force* them to a balance among themselves'. Like other duties and rights determined according to this 'measure of value', therefore, political duties and rights are to be determined by constant and creative mutual measurement of the ability to choose. Thus determined, Nietzsche proceeds to claim, political duties and rights are duties and rights '*against the whole*' of a 'community' of agents, which may coerce its members to uphold them. He also considers them to take priority over other, similarly determined duties and rights – as he puts it, political duties and rights establish '*exceptional conditions*, as partial restrictions of the true will of life, which is out after power'.¹⁷ As he proceeds to emphasise, with this account he supposes to dispose of Eugen Dühring's particular kind of retributivist theory, which bases justice precisely on '*ressentiment*', and the more general invoking of '[a]n order of right [*Recht*] conceived of as sovereign and universal', according to which the requirements of 'justice' would be determined.¹⁸

¹⁷ *GM* V 4, II 8, 9, 11. At *GM* II 9, Nietzsche adds that these obligations serve to secure 'all the goods and conveniences of communal life', among which he includes being 'sheltered, taken care of, in peace and trust, carefree with regard to certain harms and hostilities to which the human being *outside*, the "outlaw", is exposed'. For his account of 'justice', see also *JGB* 259 and *GM* II 10, and, for similar earlier accounts not expressed in terms of the ability to choose, see *WS* 22, and the two other sections to which *GM* V 4 refers in this regard, *MA* 92 and *WS* 26.

¹⁸ See *GM* II 11. Dühring's theory is presented in, for instance, his *Der Werth des Lebens: Eine philosophische Betrachtung*, published in 1865, and his *Cursus der Philosophie als streng wissenschaftlicher Weltanschauung und Lebensgestaltung*, published in 1875. Nietzsche's library includes both of these texts, along with a number of Dühring's numerous other publications. In Nietzsche's copy, the appendix of *Der Werth des Lebens* on '*Die transcendente Befriedigung der Rache*' is particularly well-marked, and he made an extensive critical summary of the book in

The rest of the *Genealogy* is occupied with Nietzsche's 'revaluations', as I presented them in the first part of the thesis. Considered with regard to these 'revaluations', however, the conception of goodness introduced at the beginning of the second essay does not simply identify a particular kind of the reasons of 'good and bad', the general features of which are presented in the first essay. It also supplements the 'revaluations' in the second and third essays. In the third essay the supplementation involves a further issue, which I will consider, along with this supplementation, in the third section of this chapter. The supplementation in the second essay, on the other hand, is relatively straightforward. In that essay, to recall, Nietzsche argues that a kind of religious reason to deny certain 'instincts' could be obeyed only from instincts of cruelty, turned against the agent herself. Importantly, however, he also calls such motivation 'bad conscience' and, indeed, he calls its motivating obedience to these religious reasons in particular 'conscience disturbed' and 'conscience-vivisection'.¹⁹ Nietzsche thus refers back to the conception of goodness that he expresses at the beginning of the section – a conception which, as he emphasises there, he considers to constitute an agent's 'conscience' – and indicates that he intends not only to identify the noncognitive state that motivates obedience to this particular kind of reasons, but also to evaluate it according to its failure to achieve this particular goodness.²⁰

1875, to be found at *KGW* VI:1 9 [1] (Summer 1875). For a helpful summary of Dühring's theory, see Small, *Nietzsche in Context*, pp.172-4.

¹⁹ The latter two remarks are made at *GM* II 24.

²⁰ One might offer a corresponding interpretation of Nietzsche's treatment of political obligations at *JGB* 199, the only section of *JGB* that refers to 'bad conscience'. Similar remarks are made at *FW* 5.

II.

The conception of goodness that Nietzsche presents at the beginning of the second essay of the *Genealogy*, and expands upon elsewhere, effectively provides an analysis of the conception of the goodness of 'a good will' which Kant analyses in arriving at his own conception of moral goodness. Furthermore, Nietzsche's analysis differs from Kant's in two substantial ways, and thus reveals that Kant imports two extraneous and controversial claims into his analysis.

I demonstrated in the fifth chapter that Kant's analysis of the conception of the goodness of a good will in the *Groundwork* begins as follows. He first identifies two claims that constitute this conception: a good will is conceived of as (a) the only good which is good in all contexts, and (b) a condition of the goodness of every other good. From (a), Kant draws the conclusion that (a') a good will does what is morally good because it is morally good. By distinguishing between action 'from duty' and action 'from' inclination, he argues that this conclusion follows because, to be (a) good in all contexts, a good will's motivation to do what is morally good must be necessarily concerned with what is morally good, rather than contingently and therefore, at most, coincidentally so concerned. From (b), on the other hand, he draws the conclusion that (b') moral goodness is a goodness which only a will can achieve. This conclusion follows, he argues, because the goodness of a good will, which he equates with moral goodness, could not be (b) a condition of the goodness of every other good if this condition could be fulfilled by other causes.

With the conception of goodness that he presents at the beginning of the second essay of the *Genealogy*, Nietzsche effectively endorses these conclusions.

Firstly, he presents this goodness as achieved only by an agent motivated by her affirmation of herself as possessing the appropriate 'goodness'-bestowing characteristic – in this case, as possessing the ability to choose. Since this characteristic is precisely what ultimately bestows 'goodness' on the reasons which she thus obeys, her motivation is necessarily, rather than contingently, concerned with the pertinent 'goodness'. Like Kant's 'good will', then, she does what is good because it is good. Secondly, the reasons of 'good and bad' that a 'goodness'-bestowing characteristic ultimately justifies are requirements to do what would prove the appropriate characteristic, and therefore, by definition, can be fulfilled only by an agent with the appropriate characteristic. In the case of Nietzsche's conception of goodness, the appropriate characteristic is the ability to choose, and therefore this goodness can be achieved only by an agent with the ability to choose. Like the goodness of Kant's 'good will', then, it is a goodness which only a 'will' can achieve.

However, Nietzsche's conception of goodness also throws doubt upon the two distinctive features of moral goodness which Kant proceeds to derive from (a') and (b'). As I demonstrated in the fifth chapter, Kant derives the first such feature from (b'), by arguing that, if (b') moral goodness is a goodness which only a will can achieve, then it must be (i) a goodness of a will's reason for action alone. He then argues that moral goodness must also be (ii) good for a will as such, since (a') a good will does what is morally good because it is morally good and (ii) this goodness must be a goodness of its reason for action alone. Importantly, as he indicates with his emphasis and employment of (ii) by means of the formula of universal law, he equates what is (ii) good for a will as such with what holds equally for any agent in relevantly similar circumstances.

Nietzsche's conception of goodness reveals that (i) and (ii) follow from (a') and (b') only on the grounds of two extraneous and controversial assumptions about the ability to choose, or 'will'. Firstly, although Nietzsche's conception of goodness effectively endorses (b'), it considers the goodness which (b') only a 'will' can achieve to be (Nietzsche's modified version of (i)) a goodness of its noncognitive motivating state alone, rather than (i) a goodness of its reason for action alone. For, as I anticipated in the first section, with his account of the 'animal that may promise' Nietzsche again considers choice to be motivated by noncognitive states informed by reasons, rather than by reasons. He indicates this by referring to the state of self-affirmation that motivates obedience to the pertinent reasons of 'good and bad' as an 'instinct', and, when it motivates, as a 'dominating instinct'. Thus he writes of the 'free' or 'sovereign' agent – that is, the agent who achieves the goodness which (b') only his ability to choose can achieve – that 'the consciousness of this rare freedom, this power over oneself and fate, has in him sunk down to his lowest depths and become instinct, the dominating instinct'.

Notably, that Nietzsche thus endorses a modified version of (i) explains his employment of a conception of 'community' which corresponds to that employed by Kant in deriving other-regarding duties and rights in *The Metaphysics of Morals*. As I demonstrated in the sixth chapter, Kant employs his conception of a 'community' of mutually vulnerable agents in association with formulas that he intends to emphasise (i), including the formula that expresses moral goodness as a will's taking will itself to provide a reason for action. Correspondingly, Nietzsche presents other-regarding duties and rights as determined by 'requit' between agents each of whom extends her own noncognitive 'self-affirmation' of the ability

to choose to others equally able to choose, and each of whom is vulnerable to others' choices 'intervening' in her own ability to choose. With regard to political duties and rights in particular, I suggest, he identifies mutual vulnerabilities to 'ressentiment' and to the misconception of political duties and rights, in contrast with the mutual vulnerabilities regarding material things' 'place' to which Kant appeals.²¹

Despite this correspondence, however, the difference between Kant's and Nietzsche's versions of (i) reveals that Kant's argument from (b') to (i) relies upon the extraneous and controversial assumption that choice is made for reasons, in the sense that it is motivated by the cognitive acknowledgement of reasons. In discussing Nietzsche's criticism of the equivalent common sense claim regarding reasons in the sixth chapter, I argued that Kant could uphold this claim against Nietzsche's criticism, with a certain accommodation. Nonetheless, Kant also offers no positive grounds for the claim other than its supposed place in the conception of moral goodness as (i) that he derives by analysis from the conception of the goodness of 'a good will'. Thus in the *Groundwork* he introduces an account of 'practical reason' in order to express this feature of moral goodness. Comparison with Nietzsche's conception of goodness reveals that even these grounds for the claim are lacking, for the conception of the goodness of 'a good will' in fact includes no commitment to practical reason. Indeed, it includes no commitment regarding the nature of motivation at all, and thus, like Kant's commitment to practical reason, Nietzsche's contrary commitment remains ultimately undefended. That said, Nietzsche's conception of goodness nonetheless performs the valuable function of showing how, if Kant's commitment were surrendered, a conception of

²¹ I argue for and evaluate this correspondence between Kant's and Nietzsche's conceptions of 'community' in more detail, and with respect to political duties and rights in particular, in my 'Vulnerabilities of Agency'.

moral goodness might still be upheld that is 'Kantian' in the sense of Nietzsche's modified version of (i), and its elucidation by (b) and (b').

Nietzsche's conception of goodness reveals that Kant also makes a second extraneous and controversial assumption about the ability to choose, or 'will', when he argues from (a') and (i) to (ii). Nietzsche effectively endorses a modified version of this argument – that is, he effectively holds that, since the agent that achieves goodness (a') does what is good because it is good and (Nietzsche's modified version of (i)) achieves a goodness of her noncognitive motivating state alone, this goodness is (ii) good for a will as such. However, while Kant equates what is (ii) good for a will as such with what holds equally for any agent in relevantly similar circumstances, Nietzsche considers what is (ii) good for a will as such to vary not only by circumstances, but also across agents and over time. This is because, unlike Kant, Nietzsche considers the ability to choose, or 'will', itself to vary across agents and over time. Indeed, this is the sense of his treatment of the ability to choose as a 'goodness'-bestowing characteristic whose requirements are determined according to the degrees and constant mutual measurement of agents' possession of it. In presenting this ability as a 'measure of value' at the beginning of the second essay of the *Genealogy*, for instance, Nietzsche emphasises the distinction between any particular sphere of 'equals', defined by a particular degree of ability to choose, and the 'feeble windbags' who are unable to choose to this degree. For Nietzsche, therefore, a goodness that is (i) good for a will as such varies across agents and over time because it varies according to the degree of the ability to choose, or 'will', and this degree varies across agents and over time. Notably, this position explains Nietzsche's often-voiced suspicion of appeals to 'equality'. In *Twilight of the Idols*, for instance, he writes that '[t]he doctrine of

equality' is a 'poison' which '*appears* to be preached by justice itself, whereas it is the *end* of justice... "Equal for equals, unequal for unequals" – *that* would be the true voice of justice: and what follows from it, "Never make unequals equal"'.²²

Nietzsche's conception of goodness thus reveals that, by equating what is (ii) good for a will as such with what holds equally for any agent in relevantly similar circumstances and so assuming that the ability to choose, or 'will', does not vary across agents or over time, Kant makes an assumption that is not included in the conception of the goodness of 'a good will' from which he intends to derive (ii). Given that, like the first assumption, Kant provides this assumption with no other grounds than its supposed place in the conception of moral goodness implicit in 'a good will', Nietzsche's conception of goodness presents a serious challenge to Kant's conception of the moral agent. For the premise of variation in the ability to choose, or 'will', is a plausible one, and, if so, such variation should be admitted as morally pertinent by a conception of goodness distinguished by (ii). Furthermore, if such variation in the ability to choose is admitted to the basic conception of goodness as (i) and (ii), then, as Nietzsche emphasises, the requirements of this goodness must vary according to the degree of this ability. In particular, agents' other-regarding duties and rights must hold only across their 'equals' in this ability, and extend to other agents only derivatively, if at all. I take it that, despite Nietzsche's enthusiasm for them, these implications of the premise are profoundly unattractive, for Kant and for moral common sense. But they account equally for the conception of the goodness of 'a good will' which Kant

²² *GD* IX 48. See also *MA* 92, *WS* 22, 26, and 33, *Z* II 7, *JGB* 202, 203, 259, and 272, *GD* IX 37, and *A* 43 and 57, and *KGW* VII:3 37 [8] (June-July 1885), VIII:2 9 [173] (Autumn 1887) and 11 [127, 142, 148, and 156] (November 1887-March 1888).

attributes to moral common sense, and from which he derives his conception of moral goodness as (i) and (ii).

It is therefore fortunate that, once again, Kant's conception of moral judgement can accommodate Nietzsche's plausible premise while resisting the conclusions which Nietzsche draws from it. For, as I explained in the sixth chapter, Kant admits the moral pertinence of contingent conditions of wills, of varying generality, insofar as these conditions are considered under (i) and (ii) in the derivation of a system of duties, whose generality therefore varies accordingly. Although he displays no particular appreciation of variation in the ability to will, therefore, he could admit such variation as a contingent condition of wills, and so admit its pertinence to the derivation of duties and rights, while refusing to admit it to the basic conception of moral goodness as (i) and (ii). He could thus uphold his equation of what is (i) good for a will as such with what holds equally for any agent in relevantly similar circumstances, and resist Nietzsche's contrary conclusion, that such goodness varies across agents and over time, and, in particular, that the requirements of such goodness hold only across 'equals' in the ability to choose. Just as his general duty of beneficence is derived by admitting the 'need' and vicinity of human wills without holding on the grounds of such 'need' and vicinity, and his general political obligation is derived by admitting the finitude of material things' 'places' without holding on the grounds of such finitude, then, Kant could derive general and specific duties and rights by appeal to variation in the ability to choose without insisting that they hold on the grounds of such variation. Like the response to Nietzsche's criticism of practical reason, this is to accommodate, rather than defend, Kant's assumption. But it demonstrates

how a defensible Kantian conception of the moral agent must owe something to both Kant and Nietzsche.

In presenting his account of the ‘animal that may promise’ at the beginning of the second essay of the *Genealogy*, Nietzsche also emphasises its historical origins and current condition. In the sections immediately following his account, he attributes the development of the ability to choose, and of the commitment to this ability as a ‘goodness’-bestowing characteristic, to the relationship between creditor and debtor, under which promising was first practised.²³ However, in first presenting his account, he also maintains that to choose, and particularly to reason over choice, requires that the agent ‘must for the moment have become *calculable, regular, necessary*, also in his own representation of himself’, and that a ‘condition and preparation’ for the ‘task’ of developing the ability to choose was therefore ‘the more specific task of first *making* men to a certain degree necessary, uniform, like among like, regular, and consequently calculable’. This preparatory ‘task’ Nietzsche attributes to what he calls ‘the morality of custom’, such that the fulfilment of this task retrospectively provides this ‘morality’ with ‘its meaning, its great justification, however much hardness, tyranny, mindlessness, and idiocy is inherent in it’.²⁴ Crucially, in proceeding from this preparatory ‘task’ to the basic ‘task’ of developing the ability to choose and the commitment to it as a ‘goodness’-bestowing characteristic, he also implies that ‘the morality of custom’ provided the practices of promising with which, under the relationship between creditor and debtor, this further ‘task’ was fulfilled. Implicitly, therefore,

²³ See *GM* II 3-6, 8-11, and 14-15.

²⁴ *GM* II 1, 2.

Nietzsche's claim is that 'the morality of custom' should be credited with both 'tasks'.

As he explains at some length in *Human, All Too Human* and *Daybreak*, Nietzsche considers 'the morality of custom' to constitute a particular, primitive relation between a community, its customs, and natural events.²⁵ For the purposes of identifying the historical origins and current condition of the 'animal that may promise', however, 'the morality of custom' simply represents arbitrary moral standards – or 'hardness, tyranny, mindlessness, and idiocy' – obedience to which provides the practices of promising which develop the ability to choose and the commitment to this ability as a 'goodness'-bestowing characteristic. By emphasising 'the morality of custom', then, Nietzsche echoes Kant's affirmation of human history's contingency and lack of moral goodness. However, while Nietzsche also echoes Kant by insisting that the goodness required can be *achieved* only by the ability to choose, and not by human history, he differs from Kant by nonetheless insisting that human history develops the ability to choose and the commitment to the associated conception of goodness. In particular, he writes that, if 'we place ourselves at the end of the enormous process, [...] where society and its morality of custom finally brings to light that *to which* it was merely a means: then we find as the ripest fruit on its tree the *sovereign individual*, the individual similar only himself, freed again from the morality of custom, autonomous [and] supermoral [...], in short the human being with his own independent long will, who *may promise*'.²⁶ Apparently, then, Nietzsche considers human history not to account for our achievement of goodness as he conceives of it, but to nonetheless

²⁵ See, in particular, *MA* 96-9, and 111, and *M* 9, 10, 18, 23, 31, 33, 40, 130, and 142. I provide a more extensive account of Nietzsche's understanding of this relation, and note some of its origins in his reading, in my "The Animal That May Promise", pp.114-5 and 119-20, and n.7.

²⁶ *GM* II 2.

account for our commitment to this goodness and our ability to achieve it, and for this commitment's being 'freed' from other commitments under which it, along with the ability to choose, developed.

Nietzsche thus, I propose, intends to vindicate his conception of goodness on grounds internal to moral common sense. As I have shown in the first part of the thesis, Nietzsche insists that moral common sense articulates a plurality of moral conceptions and he generally complains that moral philosophers such as Kant have simply stated a selection of these moral conceptions, without criticising them on grounds independent of common sense, and, in particular, without raising his particular concern for 'humanity'. In the sixth chapter, I argued that Kant would be unmoved by such complaints, since, at least according to his considered position, he recognises no such independent grounds. However, with regard to the 'animal that may promise', Nietzsche's position is a little different. Admittedly, he again raises his concern for 'humanity', by writing of the 'sovereign individual' that there is 'in him a proud consciousness, quivering in every muscle, of *what* has at length been achieved and become flesh in him, a consciousness of his own power and freedom, a feeling of the completion of humanity'.²⁷ But here Nietzsche's emphasis is on the agent's own 'feeling', and his remarks about the historical origins and current condition of this 'feeling' suggest that, while he includes it within common sense, he also considers it to take priority over other common sense moral conceptions – as represented by the 'hardness, tyranny, mindlessness, and idiocy' of 'the morality of custom'. His grounds for so considering this particular common sense 'feeling' regard the relations between, on the one hand, its origins and current condition and, on the other, those of other

²⁷ *GM II 2.*

common sense moral conceptions. In particular, he claims that the commitment to his conception of goodness has been 'freed' from other commitments under which it developed. Crucially, although this claim provide grounds for criticising and ranking common sense conceptions, it is also internal to common sense. It therefore, I suggest, exemplifies a means of criticising and ranking common sense conceptions which, given his commitment to the ultimate authority of common sense, Kant could accept, although he does not himself appreciate or employ it.

Incidentally, and in conclusion, this means of criticising common sense conceptions might also explain the atypical nature of Nietzsche's appeals to 'humanity' in the second and third essays. As I noted in the second chapter of the thesis, the first essay offers typical remarks about the consequences for 'humanity' of obedience to the reasons discussed there, and are presented as reflecting Nietzsche's own particular concern. In the second essay, however, his brief remarks about 'humanity' are restricted to his account of political duties and rights, and, in particular, to the consequences for other duties and rights determined by 'requital' of the priority of these distinctively political ones.²⁸ In other words, these remarks express an exclusive concern for the conception of goodness presented at the beginning of the second essay, and are not presented as reflecting Nietzsche's own particular concern. Furthermore, although in evaluating 'ascetic ideals' according to 'humanity' in the third essay Nietzsche again expresses his typical concerns, there he also suggests a concern for this conception of goodness. As I noted when briefly discussing this evaluation in the second chapter, Nietzsche expresses it in terms of what he calls 'the basic fact of the human will, its *horror vacui: it needs a goal*', and he proceeds to treat the 'goal'

²⁸ See *GM* II 11.

required as an answer to the question, “‘what is humanity in general for?’”. I treated this expression as atypical, in the sense that it attributes to ‘the human will’ as such a concern that Nietzsche elsewhere presents as his own particular concern, determined by his own particular noncognitive motivating states. However, this atypical feature is perhaps explained by a remark regarding ‘self-affirmation’ that Nietzsche makes in the last section of the essay: he remarks that only with an answer to the question, “‘what is humanity in general for?’”, is humanity able ‘to justify, to explain, to affirm itself’.²⁹ Such ‘self-affirmation’ could, of course, refer simply to humanity’s having a ‘goal’ of which Nietzsche approves, on the basis of his own particular noncognitive motivating states or, perhaps, some other standard. However, that here he extends his concern for such a ‘goal’ precisely to ‘the human will’ as such – rather than to, say, other human abilities, instincts, characteristics – suggests that the ‘self-affirmation’ concerned might instead correspond to the ‘self-affirmation’ of the ability to choose, or ‘will’, that he articulates at the beginning of the second essay. If so, in the third essay Nietzsche would be proposing to evaluate humanity collectively precisely as in the second essay he proposes to evaluate each agent individually – that is, according to its fulfilment of requirements whose ultimate justification lies in their proving its ability to choose. Furthermore, his failure to present his appeals to ‘humanity’ in the second and third essays as reflecting his own particular concern would be explained by the internal criticism of common sense with which, I have suggested, he explains the priority of his conception of goodness at the beginning of the second essay.

²⁹ *GM* III 28.

In this chapter, I have argued that the first sections of the second essay of the *Genealogy* provide for a critical engagement with Kant's conception of the moral agent that is more telling, and, indeed, also more original and sophisticated, than that articulated by Nietzsche's explicit criticisms of Kant. Besides offering a conception of choice that is consistent with Nietzsche's critical treatment of the common sense conception, these sections present a conception of goodness that, while Kantian, raises a serious challenge to Kant's interpretation of the 'unconditional' character of moral goodness. Furthermore, I proposed that these sections also provide a means of internally criticising and ranking common sense moral conceptions that would develop Kant's own methodological commitment to moral common sense. Although I argued that a certain feature of Kant's understanding of the derivation of duties and rights could be exploited in order to defuse the challenge to his interpretation of the 'unconditional' character of moral goodness, and that Kant could also accept the internal criticism of moral common sense, these are nonetheless substantial 'Kantian' modifications of Kant's own conception of the moral agent.

Bibliography

Kant and Nietzsche

- Alberts, Jörg, ed., *Kant und Nietzsche: Vorspiel einer künftigen Weltauslegung?*, Hofheim: Freien Akademie, 1988.
- Ansell-Pearson, Keith, 'Nietzsche's Overcoming of Kant and Metaphysics: From Tragedy to Nihilism', in *Nietzsche-Studien*, vol.16, 1987, pp.310-39.
- Bailey, Tom, 'Nietzsche's Kantian Ethics', in *International Studies in Philosophy*, v.35, no.3, 2003, pp.5-27,
- 'Anthropology and Moral Judgement: Kant and Nietzsche on Human Evil', forthcoming, 2004,
- 'Vulnerabilities of Agency: An Alternative Conception of Political Community', forthcoming, 2004.
- Bernstein, J.M., 'Autonomy and Solitude', in Ansell-Pearson ed., *Nietzsche and Modern German Thought*, pp.192-215.
- Cartwright, David E., 'Kant, Schopenhauer, and Nietzsche on the Morality of Pity', in *Journal of the History of Ideas*, vol.45, 1984, pp.83-98.
- Conrad, Mark T., 'Nietzsche's Kantianism', in *International Studies in Philosophy*, v.33, no.3, 2001, pp.25-36.
- Davies, Paul, 'Kant's Joke (Or, On Continuing to Use the Word "God")', in Rehberg and Jones eds., *The Matter of Critique*, pp.110-28.
- De Gauthier, Jules, *From Kant to Nietzsche*, trans. Merold G. Spring, New York: Books for Libraries, 1971.

- Dickopp, Heinz, 'Aspekte zum Verhältnis Nietzsche-Kant und ihre Bedeutung für die Interpretation des „Willens zur Macht“', in *Kant-Studien*, v.61, 1970, pp.97-111.
- Du Plessis, S.I.M., 'Nietzsche's Use and Abuse of Kant's Philosophy', in Beck ed., *Proceedings of the Third International Kant Congress*, pp.270-74.
- Fowler, Mark C., 'Nietzsche as Instructor in Autonomy?: Comments on "Nietzsche Contra Kant and the Problem of Autonomy"', in *International Studies in Philosophy*, v.22, no.2, 1990, pp.13-16.
- Green, Michael Steven, *Nietzsche and the Transcendental Tradition*, Chicago: University of Illinois, 2002.
- Hamacher, Werner, 'Das Versprechen der Auslegung. Zum hermeneutischen Imperativ bei Kant und Nietzsche', in his *Entferntes Verstehen*, pp.49–112, trans. Jane O. Newman and John H. Smith as 'The Promise of Interpretation: Reflections on the Hermeneutical Imperative in Kant and Nietzsche', in Rickels ed., *Looking After Nietzsche*, pp.19-47, *Entferntes Verstehen. Studien zu Philosophie und Literatur von Kant bis Celan*, Frankfurt, 1998.
- Heftrich, Urs, 'Nietzsches Auseinandersetzung mit der "Kritik der Ästhetischen Urteilskraft"', in *Nietzsche-Studien*, v.20, 1991, pp.238-66.
- Heinrichs, Johannes, 'Kant und Nietzsche: Der kategoriale and der esistentielle Denker', in Alberts ed., *Kant und Nietzsche*.
- Hill, R. Kevin, *Nietzsche's Critiques: The Kantian Foundations of his Thought*, Oxford: Oxford University Press, 2003.
- Houlgate, Stephen, 'Kant, Nietzsche, and the Thing in Itself', in *Nietzsche-Studien*, vol.22, 1993, pp.115-57.
- Ibáñez-Noé, Javier, 'Nietzsche and Kant's Copernican Revolution: On Nietzsche's Subjectivism', in *New Nietzsche Studies*, v.5, nos.1-2, Spring-Summer 2002, pp.132-49.
- Kerckhove, Lee F., 'Nietzsche's Critique of Kantian Moral Autonomy', in *Eidos: The Canadian Graduate Journal of Philosophy*, v.14, no.2, June 1997, pp.15-31.

- Kittmann, Siegfried, *Kant und Nietzsche: Darstellung und Vergleich ihrer Ethik und Moral*, Frankfurt: Peter Lang, 1984.
- Krell, David Farrell, 'Der Maulwurf: Die philosophische Wühlarbeit bei Kant, Hegel, und Nietzsche / The Mole: Philosophic Burrowings in Kant, Hegel, and Nietzsche', in O'Hara ed., *Why Nietzsche Now?*, pp.155-85.
- Land, Nick, 'Art as Insurrection: The Question of Aesthetics in Kant, Schopenhauer, and Nietzsche', in Ansell-Pearson ed., *Nietzsche and Modern German Thought*, pp.241-56.
- Miller, Elaine P., 'Sublime Time: Nietzsche's Tragic Re-Thinking of Kant's Aesthetic Temporality', in *Eidos: The Canadian Graduate Journal of Philosophy*, v.14, no.2, June 1997, pp.49-68.
- Mosser, Kurt, 'Nietzsche, Kant, and the Thing in Itself', in *International Studies in Philosophy*, vol.25, 1993, no.2, pp.67-77.
- Müller-Lauter, Wolfgang, 'Nietzsches Auf-lösung des Problems der Willensfreiheit', in Bauschinger, Cocalis, and Lennox eds., *Nietzsche heute*, pp.23-73.
- Nancy, Jean-Luc, 'Nietzsche's Thesis on Teleology', trans. Peter Connor, in Rickels ed., *Looking After Nietzsche*, pp. 49-65.
- Newman, Michael, 'Reading the Future of Genealogy: Kant, Nietzsche, Plato', in Ansell-Pearson ed., *Nietzsche and Modern German Thought*, pp.257-81.
- Olafson, Frederick A., 'Nietzsche, Kant, and Existentialism', in Solomon ed., *Nietzsche*, pp.194-201.
- Owen, David, 'The Contest of Enlightenment: An Essay on Critique and Genealogy', in *The Journal of Nietzsche Studies*, v.25, Spring 2003, pp.35-57.
- Pothen, Philip, 'Nietzsche *contra* Kant: Aesthetic Disinterestedness and the Judgment of Taste', in *Eidos: The Canadian Graduate Journal of Philosophy*, v.14, no.2, June 1997, pp.69-84.
- Reboul, Olivier, *Nietzsche Critique de Kant*, Vendôme: Presses Universitaires de France, 1974.

- Rollinger, Robin, 'Two Criticisms of Kant', in *Dialogue*, Oct. 1983, pp.27-31.
- Salem-Wiseman, Jonathan, 'Nature, Deception, and Politics of Art: Divisions of Labour in Kant and Nietzsche', in *International Studies in Philosophy*, v.30, no.1, 1998, pp.107-20.
- Simon, Josef, 'Die Krise des Wahrheitsbegriffs als Krise der Metaphysik: Nietzsches Alethiologie auf dem Hintergrund der Kantischen Kritik', in *Nietzsche-Studien*, v.18, 1989, pp.242-59,
- 'Moral bei Kant und Nietzsche', in *Nietzsche-Studien*, v.29, 2000, pp.178-98.
- Stack, George J., 'Kant and Nietzsche's Analysis of Knowledge', in *Dialogos*, v.22, 1987, pp.7-40,
- 'Kant, Lange, and Nietzsche: Critique of Knowledge', in Ansell-Pearson ed., *Nietzsche and Modern German Thought*, pp.30-58.
- Weiss, Steven D., 'Nietzsche and the Thing in Itself: Surviving Modern Kant Scholarship', in *International Studies in Philosophy*, v.25, no.2, 1993, pp.79-84.
- White, Richard, 'Nietzsche contra Kant and the Problem of Autonomy', in *International Studies in Philosophy*, v.22, no.2, 1990, pp.3-11.
- Williams, Garath, 'Nietzsche's Response to Kant's Morality', in *Philosophical Forum*, v.30, no.3, September 1999, pp.201-16.

Kant

Allison, Henry E., *Kant's Transcendental Idealism: An Interpretation and Defense*, London: Yale University Press, 1983,

Kant's Theory of Freedom, Cambridge: Cambridge University Press, 1990,

'On a Presumed Gap in the Derivation of the Categorical Imperative', in his *Idealism and Freedom*, pp.143-54,

'Autonomy and Spontaneity in Kant's Conception of the Self', in his *Idealism and Freedom*, pp.129-42,

'Kant on Freedom: A Reply to My Critics', in his *Idealism and Freedom*, pp.109-28,

'Transcendental Idealism: A Retrospective', in his *Idealism and Freedom*, pp.3-26,

Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy, Cambridge: Cambridge University Press, 1996.

Ameriks, Karl, 'Kant's Deduction of Freedom and Morality', in *Journal of the History of Philosophy*, v.19, January 1981, pp.53-79,

Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason, Oxford: Clarendon, 1982,

'Kant on the Good Will', in Höffe ed., *Grundlegung zur Metaphysik der Sitten*, pp.45-65,

'Kant and Hegel on Freedom: Two New Interpretations', in *Inquiry*, no.35, 1992, pp.219-32,

Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy, Cambridge: Cambridge University Press, 2000.

Aune, Bruce, *Kant's Theory of Morals*, Princeton: Princeton University Press, 1979.

- Axinn, Sidney, and Jane Kneller, eds., *Autonomy and Community: Readings in Contemporary Kantian Social Theory*, Albany: State University of New York Press, 1998.
- Bailey, Tom, 'Common Sense, Right, and Moral Judgement: Two Recent Additions to the Kant Literature', forthcoming in *Res Publica*, 2004.
- Barbone, Steven, 'Compatibilism in the First Critique', in *Idealistic Studies*, v.24, no.2, Spring 1994, pp.111-22.
- Baron, Marcia W., *Kantian Ethics Almost Without Apology*, London: Cornell University Press, 1995,
- 'Love and Respect in the *Doctrine of Virtue*', in Timmons ed., *Kant's Metaphysics of Morals*, pp.391-407.
- Beck, Lewis White, *A Commentary on Kant's Critique of Practical Reason*, Chicago: University of Chicago Press, 1960,
- 'The Fact of Reason: An Essay on Justification in Ethics', in his *Selected Essays on Kant*, pp.45-56,
- ed., *Proceedings of the Third International Kant Congress*, Dordrecht: Reidel, 1972,
- 'Five Concepts of Freedom in Kant', in Szrednicki ed., *Stephan Körner*, pp.35-51, *Selected Essays on Kant*, ed. by Hoke Robinson, Rochester: University of Rochester, 2002.
- Bennett, Jonathan, *Kant's Dialectic*, Cambridge: Cambridge University Press, 1974,
- 'Kant's Theory of Freedom', in Wood ed., *Self and Nature in Kant's Philosophy*, pp.102-12.
- Brink, David O., 'Kantian Rationalism: Inescapability, Authority, and Supremacy', in Cullity and Gaut eds., *Ethics and Practical Reason*, pp.255-91.
- Cummiskey, David, *Kantian Consequentialism*, Oxford: Oxford University Press, 1996.
- Doore, Gary, 'Contradiction in the Will', in *Kant-Studien*, v.76, 1985, pp.138-51.

- Engstrom, Stephen, and Jennifer Whiting eds., *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, Cambridge: Cambridge University Press, 1996,
- ‘Kant’s Conception of Practical Wisdom’, in *Kant-Studien*, v.88, 1997, pp.16-43.
- Flikschuh, Katrin, *Kant and Modern Political Philosophy*, Cambridge: Cambridge University Press, 2000.
- Förster, Eckart, ed., *Kant’s Transcendental Deductions: The Three Critiques and the Opus Postumum*, Stanford: Stanford University Press, 1989.
- Frierson, Patrick R., *Freedom and Anthropology in Kant’s Moral Philosophy*, Cambridge: Cambridge University Press, 2003.
- Gaut, Berys, ‘The Structure of Practical Reason’, in his and Cullity eds., *Ethics and Practical Reason*, pp.161-88,
- and Samuel Kerstein, ‘The Derivation without the Gap: Rethinking *Groundwork I*’, in *Kantian Review*, v.3, 1999, pp.18-40.
- Guevara, Daniel, ‘The Two Standpoints on the Will’, in *Kantian Review*, v.1, 1997, pp.82-114.
- Guyer, Paul, ed., *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press, 1992,
- ‘Kant’s Morality of Law and Morality of Freedom’, in his *Kant on Freedom, Law, and Happiness*, pp.129-71,
- ‘Duty and Inclination’, in his *Kant and the Experience of Freedom*, pp.335-93,
- Kant and the Experience of Freedom: Essays on Aesthetics and Morality*, Cambridge: Cambridge University Press, 1993,
- ‘The Possibility of the Categorical Imperative’, in his *Kant on Freedom, Law, and Happiness*, pp.172-206,
- ‘The Strategy of Kant’s *Groundwork*’, in his *Kant on Freedom, Law, and Happiness*, pp.207-231,
- ‘The Value of Reason and the Value of Freedom’, in *Ethics*, v.109, no.1, October 1998, pp.22-35,

- ed., *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, Oxford: Rowman and Littlefield, 1998,
- 'Moral Worth, Virtue, and Merit', in his *Kant on Freedom, Law, and Happiness*, pp.287-329,
- Kant on Freedom, Law, and Happiness*, Cambridge: Cambridge University Press, 2000,
- 'Kant's Deductions of the Principles of Right', in Timmons ed., *Kant's Metaphysics of Morals*, pp.23-64,
- 'Ends of Reason and Ends of Nature: The Place of Teleology in Kant's Ethics', in *Journal of Value Enquiry*, v.36, nos.2-3, 2002, pp.161-86,
- 'Kant on Common Sense and Scepticism', in *Kantian Review*, v.7, 2003, pp.1-37.
- Harbison, Warren G., 'The Good Will', in *Kant-Studien*, v.71, 1980, pp.47-59,
- 'Self-Improvement, Beneficence, and the Law of Nature Formula', in *Kant-Studien*, v.91, 2000, pp.17-24.
- Harper, William L., and Ralf Meerbote, eds., *Kant on Causation, Freedom, and Objectivity*, Minneapolis: University of Minnesota Press, 1984.
- Harré, Rom, and Daniel N. Robinson, 'The Demography of the Kingdom of Ends', in *Philosophy*, v.69, no.267, January 1994, pp.5-19.
- Harrison, Jonathan, 'Kant's Examples of the First Formulation of the Categorical Imperative', in Wolff ed., *Kant*, pp.228-45,
- 'The Categorical Imperative', in Wolff ed., *Kant*, pp.259-65.
- Henrich, Dieter, 'Die Deduktion des Sittengesetzes: Über die Gründe der Dunkelheit des letzten Abschnittes von Kants Grundlegung zur Metaphysik der Sitten', in Schwann ed., *Denken im Schatten des Nihilismus*, pp.55-110, excerpt trans. Guyer as '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 Guyer ed., *Kant's Groundwork of the Metaphysics of Morals*, pp.303-41,

‘Das Problem der Grundlegung der Ethik bei Kant und im spekulativen Idealismus’, in his *Selbstverhältnisse*, pp.6-56, trans. Louis Hunt as ‘Ethics of Autonomy’, in Henrich, *The Unity of Reason*, pp.89-121,

Selbstverhältnisse: Gedanken und Auslegungen zu den Grundlängen der klassischen deutschen Philosophie, Stuttgart: Reclam, 1982,

The Unity of Reason: Essays on Kant’s Philosophy, ed. Richard L. Velkley, London: Harvard University Press, 1994.

Henson, Richard G., ‘What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action’, in *Philosophical Review*, v.88, 1979, pp.39-54.

Herman, Barbara, ‘On the Value of Acting from the Motive of Duty’, in her *The Practice of Moral Judgment*, pp.1-22,

‘Mutual Aid and Respect for Persons’, in her *The Practice of Moral Judgment*, pp.45- 72,

‘The Practice of Moral Judgment’, in her *The Practice of Moral Judgment*, pp.73-93,

‘Murder and Mayhem’, in her *The Practice of Moral Judgment*, pp.113-31,

‘What Happens to the Consequences?’, in her *The Practice of Moral Judgment*, pp.94-112,

‘Moral Deliberation and the Derivation of Duties’, in her *The Practice of Moral Judgment*, pp.132-58,

‘Leaving Deontology Behind’, in her *The Practice of Moral Judgment*, pp.208-40,

The Practice of Moral Judgment, London: Harvard University Press, 1993,

Andrews Reath, and Christine Korsgaard, eds., *Reclaiming the History of Ethics: Essays for John Rawls*, Cambridge: Cambridge University Press, 1997.

Hill Jr., Thomas E., ‘The Kingdom of Ends’, in his *Dignity and Practical Reason in Kant’s Moral Theory*, pp.58-66,

‘The Hypothetical Imperative’, in his *Dignity and Practical Reason in Kant’s Moral Theory*, pp.17-37,

- 'Humanity as an End in Itself', in *Ethics*, v.91, October 1980, pp.84-99,
- 'Kant's Argument for the Rationality of Moral Conduct', in Guyer ed., *Kant's Groundwork of the Metaphysics of Morals*, pp.249-72,
- 'The Kantian Conception of Autonomy', in his *Dignity and Practical Reason in Kant's Moral Theory*, pp.76-96,
- 'Kant's Theory of Practical Reason', in his *Dignity and Practical Reason in Kant's Moral Theory*, pp.123-46,
- Dignity and Practical Reason in Kant's Moral Theory*, London: Cornell University Press, 1992,
- 'Beneficence and Self-Love', in his *Human Welfare and Moral Worth*, pp.99-124,
- 'Donagan's Kant', in his *Respect, Pluralism, and Justice*, pp.119-51,
- 'Moral Dilemmas, Gaps, and Residues', in his *Human Welfare and Moral Worth*, pp.362-402,
- 'Is a Good Will Overrated?', in his *Human Welfare and Moral Worth*, pp.37-60,
- Respect, Pluralism, and Justice: Kantian Perspectives*, Oxford: Oxford University Press, 2000,
- 'Kantian Analysis: From Duty to Autonomy', in his *Human Welfare and Moral Worth*, pp.13-36,
- 'Personal Values and Setting Oneself Ends', in his *Human Welfare and Moral Worth*, pp.244-74,
- 'Meeting Needs and Doing Favours', in his *Human Welfare and Moral Worth*, pp.201-43,
- Human Welfare and Moral Worth: Kantian Perspectives*, Oxford: Oxford University Press, 2002.
- Höffe, Otfried, ed., *Grundlegung zur Metaphysik der Sitten: Ein kooperativer Kommentar*, Frankfurt am Main: Vittorio Klostermann, 1989.
- Hudson, Hud, *Kant's Compatibilism*, London: Cornell University Press, 1994.

- Johnson, Robert N., 'Expressing a Good Will: Kant on the Motive of Duty', in *Southern Journal of Philosophy*, v.34, no.2, 1996, pp.147-68.
- Jones, Hardy E., *Kant's Principle of Personality*, London: University of Wisconsin Press, 1971.
- Jones, Rachel, and Andrea Rehberg eds., *The Matter of Critique: Readings in Kant's Philosophy*, Manchester: Clinamen, 2000.
- Kemp, J., 'Kant's Examples of the Categorical Imperative', in Wolff ed., *Kant*, pp.246-58.
- Kerstein, Samuel, 'Korsgaard's Kantian Arguments for the Value of Humanity', in *Canadian Journal of Philosophy*, v.31, no.1, March 2001, pp.23-52.
- Kersting, Wolfgang, *Wohlgeordnete Freiheit*, Frankfurt: Suhrkamp, 1993.
- Kleingeld, Pauline, 'Kant, History, and the Idea of Moral Development', in *History of Philosophy Quarterly*, v.16, no.1, January 1999, pp.59-80.
- Korsgaard, Christine M., 'Kant's Formula of Universal Law', in her *Creating the Kingdom of Ends*, pp.77-105,
- 'Two Distinctions in Goodness', in her *Creating the Kingdom of Ends*, pp.249-74,
- 'Kant's Formula of Humanity', in her *Creating the Kingdom of Ends*, pp.105-132,
- 'Kant's Analysis of Obligation: The Argument of *Groundwork* I', in her *Creating the Kingdom of Ends*, pp.43-76,
- 'Morality as Freedom', in her *Creating the Kingdom of Ends*, pp.159-87,
- 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations', in her *Creating the Kingdom of Ends*, pp.188-221,
- Creating the Kingdom of Ends*, Cambridge: Cambridge University Press, 1996,
- 'The Normativity of Instrumental Reason', in Cullity and Gaut eds., *Ethics and Practical Reason*, pp.215-54.
- Krasnoff, Larry, 'The Fact of Politics: History and Teleology in Kant', in *European Journal of Philosophy*, v.2, no.1, 1994, pp.22-40.
- Kuehn, Manfred, *Scottish Common Sense in Germany, 1768-1800: A Contribution to the History of Critical Philosophy*, Montreal: McGill-Queen's University Press, 1987.

- Latham, Noa, 'Causally Irrelevant Reasons and Action Solely from the Motive of Duty', in *Journal of Philosophy*, v.91, 1994, pp.599-618.
- Louden, Robert B., *Kant's Impure Ethics: From Rational Beings to Human Beings*, Oxford: Oxford University Press, 2000.
- Ludwig, Bernd, *Kants Rechtslehre: Ein Analytischer Kommentar*, Hamburg: Felix Meiner, 1988,
- "The Right of a State" in Immanuel Kant's *Doctrine of Right*, in *Journal of the History of Philosophy*, v.28, 1990, pp.403-15,
- 'Whence Public Right?: The Role of Theoretical and Practical Reasoning in Kant's *Doctrine of Right*', in Timmons ed., *Kant's Metaphysics of Morals*, pp.159-83.
- Mariña, Jacqueline, 'Kant's Derivation of the Formula of the Categorical Imperative: How to Get it Right', in *Kant-Studien*, v.89, 1998, pp.167-78.
- McCarthy, Michael H., 'Kant's Rejection of the Argument of Groundwork III', in *Kant-Studien*, v.73, 1982, pp.169-90.
- McCloskey, Mary A., 'Kant's Kingdom of Ends', in *Philosophy*, v.51, 1976, pp.391-9.
- McNair, Ted, 'Universal Necessity and Contradictions in Conception', in *Kant-Studien*, v.91, 2000, pp.25-43.
- Meerbote, Ralf, 'Kant on the Nondeterminate Character of Human Actions', in his and Harper eds., *Kant on Causation, Freedom, and Objectivity*, pp.138-63.
- Mulholland, Leslie, *Kant's System of Rights*, New York: Columbia University Press, 1991.
- Murphy, Jeffrie, *Kant: The Philosophy of Right*, Macon: Mercer University Press, 1994.
- O'Neill, Onora, (as Onora Nell) *Acting on Principle: An Essay on Kantian Ethics*, London: Columbia University Press, 1975,
- 'Consistency in Action', in her *Constructions of Reason*, pp.81-104,
- 'Between Consenting Adults', in her *Constructions of Reason*, pp.105-25,
- 'Constructivisms in Ethics', in her *Constructions of Reason*, pp.206-18,

'Universal Laws and Ends in Themselves', in her *Constructions of Reason*, pp.126-44,

'Reason and Autonomy in *Grundlegung III*', in her *Constructions of Reason*, pp.51-65,

'Agency and Anthropology in Kant's *Groundwork*', in Yovel ed., *Kant's Practical Philosophy Reconsidered*, pp.63-82,

Constructions of Reason: Explorations of Kant's Practical Philosophy, Cambridge: Cambridge University Press, 1989,

'Kant's Virtues', in Crisp ed., *How Should One Live?*, pp.77-97,

'Instituting Principles: Between Duty and Action', in Timmons ed., *Kant's Metaphysics of Morals*, pp.331-47,

'Autonomy: The Emperor's New Clothes', in *Mind*, 2003, pp.1-21.

Paton, H.J., 'Analysis of the Argument', in his *The Moral Law*, pp.13-49,

The Moral Law: Kant's Groundwork of the Metaphysic of Morals, London: Hutchinson, 1948,

The Categorical Imperative: A Study of Kant's Moral Philosophy, 4th ed., London: Hutchinson, 1963.

Pogge, Thomas W., 'Kant's Theory of Justice', in *Kant-Studien*, v.79, 1988, pp.407-33,

'The Categorical Imperative', in Guyer ed., *Kant's Groundwork of the Metaphysics of Morals*, pp.189-213.

Potter, Nelson, 'The Argument of Kant's *Groundwork*, Chapter 1', in Guyer ed., *Kant's Groundwork of the Metaphysics of Morals*, pp.29-49.

Rawls, John, 'Kantian Constructivism in Moral Theory', in *The Journal of Philosophy*, v.77, 1980, pp.515-72,

'Themes in Kant's Moral Philosophy', in Förster ed., *Kant's Transcendental Deductions*, pp.81-113,

Lectures on the History of Moral Philosophy, ed. by Barbara Herman, London: Harvard University Press, 2000.

- Robinson, Daniel N., and Rom Harré, 'The Demography of the Kingdom of Ends', in *Philosophy*, v.69, no.267, January 1994, pp.5-19.
- Rosen, Allen, *Kant's Theory of Justice*, Ithaca: Cornell University Press, 1993.
- Rotenstreich, Nathan, 'On the Formalism of Kant's Ethics', in Yovel ed., *Kant's Practical Philosophy Reconsidered*, pp.49-62.
- Schönecker, Dieter, 'Gemeine sittliche und philosophische Vernunftkenntnis: Zum ersten Übergang in Kants *Grundlegung*', in *Kant-Studien*, v.88, 1997, pp.311-33.
- Steinberger, Peter J., 'The Standard View of the Categorical Imperative', in *Kant-Studien*, v.90, no.1, 1999, pp.91-9.
- Sullivan, Roger J., *Immanuel Kant's Moral Theory*, Cambridge: Cambridge University Press, 1989.
- Teale, A.E., *Kantian Ethics*, London: Oxford University Press, 1951.
- Thompson, Kevin, 'Kant's Transcendental Deduction of Political Authority', in *Kant-Studien*, v.92, no.1, 2001, pp.62-78.
- Timmons, Mark, ed., *Kant's Metaphysics of Morals: Interpretative Essays*, Oxford: Oxford University Press, 2002.
- Tittel, Gottlob August, *Über Herr Kants Moralreform*, Frankfurt: Pfahler, 1786.
- Voeller, Carol W., *The Metaphysics of the Moral Law: Kant's Deduction of Freedom*, London: Garland, 2001.
- Walker, Ralph C.S., *Kant*, London: Routledge, Kegan, and Paul, 1978.
- Willaschek, Marcus, 'Why the *Doctrine of Right* Does Not Belong in the *Metaphysics of Morals*: On Some Basic Distinctions in Kant's Moral Philosophy', in *Jahrbuch für Recht und Ethik*, v.5, 1997, pp.205-27,
- 'Which Imperatives for Right?: On the Non-Prescriptive Character of Juridical Laws in Kant's *Metaphysics of Morals*', in Timmons ed., *Kant's Metaphysics of Morals*, pp.65-87.
- Wolff, Robert Paul, ed., *Kant: A Collection of Critical Essays*, London: University of Notre Dame Press, 1968,

- 'Remarks on the Relation of the *Critique of Pure Reason* to Kant's Ethical Theory', in Den Ouden and Moen eds., *New Essays on Kant*, pp.139-53.
- Wood, Allen W., 'Kant on False Promises', in Beck ed., *Proceedings of the Third International Kant Congress*, pp.614-9,
- 'Kant's Compatibilism', in his ed., *Self and Nature in Kant's Philosophy*, pp.73-101,
- ed., *Self and Nature in Kant's Philosophy*, London: Cornell University Press, 1984,
- 'Humanity as End in Itself', in Guyer ed., *Kant's Groundwork of the Metaphysics of Morals*, pp.165-87,
- 'Kant's Project for Perpetual Peace', in Cheah and Robbins eds., *Cosmopolitics*, pp.59-76,
- 'Kant's Historical Materialism', in Axinn and Kneller eds., *Autonomy and Community*, pp.15-37,
- Kant's Ethical Thought*, Cambridge: Cambridge University Press, 1999,
- 'The Final Form of Kant's Practical Philosophy', in Timmons ed., *Kant's Metaphysics of Morals*, pp.1-21.
- Yovel, Yirmiyahu, ed., *Kant's Practical Philosophy Reconsidered*, London: Kluwer Academic Publishers, 1989.

Nietzsche

Allison, David, ed., *The New Nietzsche*, New York: Delta Books, 1977.

Ansell-Pearson, Keith, 'Nietzsche and the Problem of the Will in Modernity', in his ed., *Nietzsche and Modern German Thought*, pp.165-91,

ed., *Nietzsche and Modern German Thought*, London: Routledge, 1991,

'Nietzsche on Autonomy and Morality: The Challenge to Political Theory', in *Political Studies*, v.39, 1991, pp.270-86,

'The Significance of Michel Foucault's Reading of Nietzsche: Power, the Subject, and Political Theory', in *Nietzsche-Studien*, v.20, 1991, pp.267-83,

and Howard Caygill eds., *The Fate of the New Nietzsche*, Aldershot: Ashgate, 1993.

Bailey, Tom, 'Two Interpretations of Nietzsche's Epistemology', unpubl., 1998,

'Friedrich Nietzsche and the "Real World": A Critique of Maudemarie Clark's Interpretation of Nietzsche', unpubl., 1998,

"The Animal That May Promise": Nietzsche on the Will, Naturalism, and Duty', in *Pli: The Warwick Journal of Philosophy*, v. 11, 2001, pp. 103-21.

Bauschinger, Sigrid, Susan L. Cocalis, and Sara Lennox eds., *Nietzsche heute: Die Rezeption seines Werkes nach 1968*, Stuttgart: Francke, 1995.

Beardsworth, Richard, 'Nietzsche, Nihilism, and Spirit', in Ansell-Pearson and Morgan eds., *Nihilism Now!*, pp.37-69.

Bergmann, Frithjof, 'Nietzsche's Critique of the Modality of Moral Codes', in *International Studies in Philosophy*, v.17, no.2, 1985, pp.99-116,

'Nietzsche's Critique of Morality', in Solomon and Higgins eds., *Reading Nietzsche*, pp.29-45.

- Berkowitz, Peter, *Nietzsche: The Ethics of an Immoralist*, London: Harvard University Press, 1995.
- Blondel, Éric, *Nietzsche: le corps et la culture: La philosophie comme généalogie philologique*, Paris: Presses Universitaires de France, 1986, trans. Seán Hand as *Nietzsche: The Body and Culture: Philosophy as a Philological Genealogy*, Stanford: Stanford University Press, 1991.
- Brobjer, Thomas H., *Nietzsche's Ethics of Character: A Study of Nietzsche's Ethics and its Place in the History of Moral Thinking*, Uppsala: Uppsala University Press, 1995, 'Nietzsche's Reading and Private Library, 1885-1889', in *Journal of the History of Ideas*, v.58, no.4, October 1997, pp.663-93.
- Clark, Maudemarie, *Nietzsche on Truth and Philosophy*, Cambridge: Cambridge University Press, 1990, 'Nietzsche's Immoralism and the Concept of Morality', in Schacht ed., *Nietzsche, Genealogy, Morality*, pp.15-34, 'On Knowledge, Truth, and Value: Nietzsche's Debt to Schopenhauer and the Development of his Empiricism', in Janaway ed., *Willing and Nothingness*, pp.37-78, 'On the Rejection of Morality: Bernard Williams's Debt to Nietzsche', in Schacht ed., *Postmoralism*, pp.100-22.
- Conway, Daniel W., and Peter S. Groff eds., *Nietzsche: Critical Assessments*, 4 vols., London: Routledge, 1998.
- Danto, Arthur, *Nietzsche as Philosopher*, Guildford: Columbia University Press, 1980.
- Deleuze, Gilles, *Nietzsche et la Philosophie*, Paris: Presses Universitaires de France, 1962, trans. Hugh Tomlinson as *Nietzsche and Philosophy*, London: Athlone, 1983.
- Dudley, Will, *Hegel, Nietzsche, and Philosophy: Thinking Freedom*, Cambridge: Cambridge University Press, 2002.
- Foucault, Michel, 'Nietzsche, la Généalogie, l'Histoire', in his *Dits et Écrits 1954-1988*, v.2, pp.136-56, trans. Donald F. Bouchard and Sherry Simon as 'Nietzsche, Genealogy, History', in Foucault, *Language, Countermemory, Practice*, pp.139-64.

- Gerhardt, Volker, “‘Das Their das versprechen darf’: Mensch, Gesellschaft und Politik bei Friedrich Nietzsche”, in Höffe ed., *Der Mensch: ein politisches Thier?*, pp.134-156,
- ‘Selbstbegründung: Nietzsches Moral der Individualität’, in *Nietzsche-Studien*, v.21, 1992, pp.28-49, trans. Peter Poellner as ‘Self-grounding: Nietzsche’s Morality of Individuality’, in Ansell-Pearson and Caygill eds., *The Fate of the New Nietzsche*, pp.283-301,
- ‘Die Tugend des freien Geistes: Nietzsche auf dem Weg zum individuellen Gesetz der Moral’, in *Synthesis Philosophica*, v.11, no.1, 1996, pp.111-22.
- Geuss, Raymond, ‘Nietzsche and Genealogy’, in *European Journal of Philosophy*, v.2, no.3, December 1994, pp.274-92,
- ‘Nietzsche and Morality’, in *European Journal of Philosophy*, v.5, no.1, April 1997, pp.1-20,
- ‘Genealogy as Critique’, trans. Nicholas Walker, in *European Journal of Philosophy*, v.10, no.2, August 2002, pp.209-15.
- Guay, Robert, ‘Nietzsche on Freedom’, in *European Journal of Philosophy*, v.10, no.3, December 2002, pp.302-327.
- Hatab, Lawrence, *A Nietzschean Defense of Democracy: An Experiment in Postmodern Politics*, Chicago and La Salle: Open Court, 1995.
- Havas, Randall, *Nietzsche’s Genealogy: Nihilism and the Will to Knowledge*, London: Cornell University Press, 1995,
- ‘Nietzsche’s Idealism’, in *Journal of Nietzsche Studies*, no.20, Fall 2000, pp.90-9.
- Heidegger, Martin, *Nietzsche*, 2 vols., Pfullingen: Günther Neske, 1961, partial trans. David Farrell Krell, 3 vols., London: Routledge and Kegan Paul, 1979.
- Higgins, Kathleen Marie, and Robert C. Solomon eds., *Reading Nietzsche*, Oxford: Oxford University Press, 1988,
- and Bernd Magnus eds., *The Cambridge Companion to Nietzsche*, Cambridge: Cambridge University Press, 1996.

- Hoy, David Couzens, 'Nietzsche, Hume, and the Genealogical Method', in Yovel ed., *Nietzsche as Affirmative Thinker*, pp.20-38.
- Hunt, Lester H., *Nietzsche and the Origin of Virtue*, London: Routledge, 1991.
- Janaway, Christopher, ed., *Willing and Nothingness: Schopenhauer as Nietzsche's Educator*, Oxford: Oxford University Press, 1998.
- Jenkins, Scott, 'Morality, Agency, and Freedom in Nietzsche's *Genealogy of Morals*', in *History of Philosophy Quarterly*, v.20, no.1, January 2003, pp.61-80.
- Kaufmann, Walter, *Nietzsche: Philosopher, Psychologist, Antichrist*, 4th ed., New Jersey: Princeton University Press, 1974.
- Lackey, Michael, 'Killing God, Liberating the "Subject": Nietzsche and Post-God Freedom', in *Journal of the History of Ideas*, v.60, no.4, 1999, pp.737-54.
- Leiter, Brian, 'Morality in the Pejorative Sense: On the Logic of Nietzsche's Critique of Morality', in *British Journal of the History of Philosophy*, v.3, no.1, 1995, pp.113-45,
 'The Paradox of Fatalism and Self-Creation in Nietzsche', in Janaway ed., *Willing and Nothingness*, pp.217-57,
 'Nietzsche's Metaethics: Against the Privilege Readings', in *European Journal of Philosophy*, v.8, no.3, 2000, pp.277-97,
Nietzsche on Morality, London: Routledge, 2002.
- Lippitt, John, ed., *Nietzsche's Futures*, London: Macmillan, 1999.
- Magnus, Bernd, 'The Use and Abuse of *The Will to Power*', in Solomon and Higgins eds., *Reading Nietzsche*, pp.218-35.
- May, Simon, *Nietzsche's Ethics and His War on 'Morality'*, Oxford: Oxford University Press, 1999.
- Moles, Alastair, *Nietzsche's Philosophy of Nature and Cosmology*, New York: Lang, 1989.
- Müller-Lauter, Wolfgang, *Nietzsche: Seine Philosophie der Gegensätze und die Gegensätze Seiner Philosophie*, Berlin: De Gruyter, 1971, trans. David J. Parent as

Nietzsche: His Philosophy of Contradictions and the Contradictions of His Philosophy, Chicago: University of Illinois, 1999,

‘Nietzsches “Lehre” vom Willen zur Macht’, in *Nietzsche-Studien*, v.3, 1974, pp.1-60, trans. Parent as ‘Nietzsche’s “Doctrine” of the Will to Power’, in Müller-Lauter, *Nietzsche*, pp.122-60.

Nehamas, Alexander, *Nietzsche: Life as Literature*, London: Harvard University Press, 1985.

O’Hara, Daniel, ed., *Why Nietzsche Now?*, Bloomington: Indiana University Press, 1981.

Oaklander, L. Nathan, ‘Nietzsche on Freedom’, in *Southern Journal of Philosophy*, v.22, no.2, 1984, pp.211-22.

Owen, David, *Maturity and Modernity: Nietzsche, Weber, Foucault, and the Ambivalence of Reason*, London: Routledge, 1994,

Nietzsche, Politics, and Modernity: A Critique of Liberal Reason, London: Sage, 1995,

‘Nietzsche, Enlightenment, and the Problem of Noble Ethics’, in Lippitt ed., *Nietzsche’s Futures*, pp.3-29,

‘Equality, Democracy, and Self-Respect: Reflections on Nietzsche’s Agonal Perfectionism’, in *Journal of Nietzsche Studies*, v.24, 2002, pp.113-31.

Poellner, Peter, *Nietzsche and Metaphysics*, Oxford: Oxford University Press, 1995.

Richardson, John, *Nietzsche’s System*, Oxford: Oxford University Press, 1996.

Rickels, Laurence A., ed., *Looking After Nietzsche*, Albany: State University of New York Press, 1990.

Ridley, Aaron, *Nietzsche’s Conscience: Six Character Studies from the ‘Genealogy’*, London: Cornell University Press, 1998,

‘Ancillary Thoughts on an Ancillary Text’, in *Journal of Nietzsche Studies*, no.20, Fall 2000, pp.100-8.

- Risse, Mathias, 'The Second Treatise in *On the Genealogy of Morality*: Nietzsche on the Origin of Bad Conscience', in *European Journal of Philosophy*, v.9, no.1, 2001, pp.55-81.
- Schacht, Richard, *Nietzsche*, London: Routledge, 1983,
 ed., *Nietzsche, Genealogy, Morality: Essays on Nietzsche's Genealogy of Morals*, London: University of California Press, 1994,
 ed., *Nietzsche's Postmoralism: Essays on Nietzsche's Prelude to Philosophy's Future*, Cambridge: Cambridge University Press, 2001.
- Slote, Michael, 'Nietzsche and Virtue Ethics', in *International Studies in Philosophy*, v.30, no.3, 1998, pp.23-7.
- Small, Robin, *Nietzsche in Context*, Aldershot: Ashgate, 2001.
- Solomon, Robert C., ed., *Nietzsche: A Collection of Critical Essays*, Indiana: University of Notre Dame Press, 1980,
 'A More Severe Morality: Nietzsche's Affirmative Ethics', in *Journal of the British Society for Phenomenology*, v.16, no.3, October 1985, pp.250-67,
 'Nietzsche as Existentialist and as Fatalist: The Practical Paradoxes of Self-Making', in *International Studies in Philosophy*, v.34, no.3, 2002, pp.41-54,
 'Nietzsche on Fatalism and "Free Will"', in *Journal of Nietzsche Studies*, no.23, Spring 2002, pp.63-87.
- Stack, George J., *Lange and Nietzsche*, Berlin: De Gruyter, 1983.
- Strong, Tracy B., *Friedrich Nietzsche and the Politics of Transfiguration*, London: University of California Press, 1975.
- Swanton, Christine, 'Outline of a Nietzschean Virtue Ethics', in *International Studies in Philosophy*, v.30, no.3, 1998, pp.29-38.
- Thatcher, David S., 'Zur Genealogie der Moral: Some Textual Annotations', in *Nietzsche-Studien*, v.18, 1989, pp.587-99.
- Warren, Mark, *Nietzsche and Political Thought*, London: MIT Press, 1988.

White, Richard, *Nietzsche and the Problem of Sovereignty*, Chicago: University of Illinois Press, 1997.

Williams, Bernard, 'Nietzsche's Minimalist Moral Psychology', in Schacht ed., *Nietzsche, Genealogy, Morality*, pp.237-47.

Other literature

- Ansell-Pearson, Keith, and Diane Morgan eds., *Nihilism Now!: Monsters of Energy*, Basingstoke: Macmillan, 2000.
- Augustine, *De libero arbitrio*, 388, 391-5, trans. Thomas Williams as *On Free Choice of the Will*, Indianapolis: Hackett, 1993,
- De spiritu et littera*, 412, trans. P. Holmes as *On the Spirit and the Letter*, in Augustine, *Basic Writings of Saint Augustine*, v.1,
- De civitate Dei*, 413-26, trans. J. O'Meara as *The City of God*, Harmondsworth: Penguin, 1972,
- Basic Writings of Saint Augustine*, ed. W. Oates, New York: Random House, 1948.
- Bagehot, Walter, *Physics and Politics, or Thoughts on the Application of the Principles of 'Natural Selection' and 'Inheritance' to Political Society*, 1872, London: Kegan Paul, Trench, Trübner, and Co., 1896.
- Calvin, John, *Institution de la religion chrestienne*, trans. Henry Beveridge as *Institutes of the Christian Religion*, 2 vols., Grand Rapids, Michigan: Eerdmans, 1957.
- Cheah, Pheng, and Bruce Robbins eds., *Cosmopolitics: Thinking and Feeling Beyond the Nation*, London: University of Minnesota Press, 1998.
- Cullity, Garrett, and Berys Gaut, eds., *Ethics and Practical Reason*, Oxford: Oxford University Press, 1997.
- Dihle, Albrecht, *The Theory of Will in Classical Antiquity*, London: University of California, 1982.
- Fichte, Johann Gottlieb, *Das System der Sittenlehre nach den Principien der Wissenschaftslehre*, 1798, in his *Werke*, v.4, trans. A.E. Kroeger as *The Science of Ethics*, London: Kegan Paul, Trench, Truebner, 1897,
- Werke*, ed. I.H. Fichte, Berlin: De Gruyter, 1971.

- Foucault, Michel, *Language, Countermemory, Practice: Selected Essays and Interviews*, trans. Donald F. Bouchard and Sherry Simon, Ithaca: Cornell University Press 1977,
Dits et Écrits 1954-1988, ed. Daniel Defert and François Ewald, 4 vols., Paris: Éditions Gallimard, 1994.
- Frankena, William K., *Ethics*, 2nd ed., New Jersey: Prentice Hall, 1973.
- Goethe, Johann Wolfgang, *Werke*, ed. Erich Trunz, 14 vols., Hamburg: Christian Wegner, 2nd ed., 1952.
- Hegel, Georg Wilhelm Friedrich, *Über die Wissenschaftlichen Behandlungsarten des Naturrechts, seine Stelle in der praktischen Philosophie, und sein Verhältniss zu den positiven Rechtswissenschaften*, 1802-3, in his *Werke*, v.4, trans. T.M. Knox as *Natural Law: The Scientific Ways of Treating Natural Law, Its Place in Moral Philosophy, and Its Relation to the Positive Sciences of Law*, Philadelphia: University of Pennsylvania Press, 1975,
Phänomenologie des Geistes, 1807, trans. A.V. Miller as *Hegel's Phenomenology of Spirit*, Oxford: Oxford University Press, 1977,
Grundlinien der Philosophie des Rechts, 1821, trans. H.B. Nisbet as *Elements of the Philosophy of Right*, ed. Allen W. Wood, Cambridge: Cambridge University Press, 1991,
Enzyklopädie der philosophischen Wissenschaften, 1817, rev. 1827, 1830, pt.1 trans. T.F. Geraets, W.A. Suchting, and H.S. Harris, Indianapolis: Hackett, 1991,
Vorlesungen über die Geschichte der Philosophie, 1840, trans. E.S. Haldane and Frances H. Simson as *Lectures on the History of Philosophy*, 3 vols., London: University of Nebraska Press, 1995,
Werke, ed. H. Büchner and O. Pöggeler, Hamburg, 1968.
- Hobbes, Thomas, *Leviathan*, 1651, ed. Edwin Curley, Indianapolis: Hackett, 1994.
- Höffe, Otfried, ed., *Der Mensch: ein politisches Thier?: Essays zur politischen Anthropologie*, Stuttgart: Reclam, 1992.

- Hume, David, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 1748 / 1751, ed. L.A. Selby-Bigge, rev. P.H. Nidditch, 3rd ed., Oxford: Oxford University Press, 1975,
- A Treatise of Human Nature*, 1739-40, ed. L.A. Selby-Bigge, rev. by P.H. Nidditch, 2nd ed., Oxford: Oxford University Press, 1978.
- Ihering, Rudolf von, *Der Kampf ums Recht*, 1872, trans. J.J. Lalor as *The Struggle for Law*, Chicago, 1915,
- Der Zweck in Recht*, 1877-83, trans. Isaac Husik as *Law as a Means to an End*, Boston: Boston Book Company, 1913.
- Lecky, William Edward Hartpole, *History of the Rise and Influence of Rationalism in Europe*, 2 vols., London: Longman, Green, Longman, Roberts, and Green, 1865,
- History of European Morals from Augustus to Charlemagne*, 2 vols., 3rd ed., London: Longmans, Green, and Co., 1877.
- Leibniz, Gottfried Wilhelm, 'On Freedom', 1689 (?), trans. Roger Ariew and Daniel Garber as 'On Freedom', in Leibniz, *Philosophical Essays*, pp.94-8,
- Discours de Métaphysique*, 1686, trans. R.S. Woolhouse and Richard Francks as 'Discourse on Metaphysics', in Leibniz, *Philosophical Texts*, pp.54-89,
- Principes de la nature et de la grace fondés en raison*, 1714, trans. Woolhouse and Francks as 'Principles of Nature and Grace, Based on Reason', in Leibniz, *Philosophical Texts*, pp.258-66,
- Principes de la philosophie ou monadologie*, 1714, trans. Woolhouse and Francks as 'Monadology', in Leibniz, *Philosophical Texts*, pp.267-81,
- Philosophical Essays*, trans. and ed. Ariew and Garber, Indianapolis: Hackett, 1989,
- Philosophical Texts*, trans. and ed. Woolhouse and Francks, Oxford: Oxford University Press, 1998.
- Locke, John, *An Essay Concerning Human Understanding*, 1690, abridgement ed. John W. Yolton, London: Dent, 1947.

- Lubbock, John, *The Origin of Civilisation and the Primitive Condition of Man: Mental and Social Condition of Savages*, London: Longmans, Green, and Co., 1870.
- MacIntyre, Alasdair, *After Virtue: A Study in Moral Theory*, 2nd ed., London: Duckworth, 1985.
- Mill, John Stuart, *Utilitarianism*, in his *On Liberty and Other Essays*, pp.129-201, *On Liberty and Other Essays*, Oxford: Oxford University Press, 1991.
- Nietzsche, Friedrich, *Menschliches, Allzumenschliches: Ein Buch Für Freie Geister*, trans. Gary Handwerk as *Human, All Too Human I*, v.3 of *The Complete Works of Friedrich Nietzsche*, ed. by Ernest Behler, Stanford: Stanford University Press, 1997.
- Pippin, Robert B., *Modernism as a Philosophical Problem: On the Dissatisfactions of European High Culture*, 2nd ed., Oxford: Blackwell, 1999.
- Rée, Paul, *Der Ursprung der moralischen Empfindungen*, Chemnitz: Ernst Schmeitzner, 1877, *Die Entstehung des Gewissens*, Berlin: Carl Duncker, 1885.
- Rousseau, Jean-Jacques, *Du Contrat Social*, 1762, trans. Victor Gourevitch as *Of the Social Contract*, in Rousseau, *The Social Contract and Other Later Political Writings*, pp.39-152, *The Social Contract and Other Later Political Writings*, trans. and ed. Gourevitch, Cambridge: Cambridge University Press, 1997.
- Russell, Bertrand, *History of Western Philosophy, and its Connection with Political and Social Circumstances from the Earliest Times to the Present Day*, 2nd ed., London: Allen and Unwin, 1961.
- Schiller, Friedrich, *Xenien*, in Goethe, *Werke*, v.1, pp.208-34.
- Schneewind, Jerome B., *The Invention of Autonomy: A History of Modern Moral Philosophy*, Cambridge: Cambridge University Press, 1998.
- Schopenhauer, Arthur, *Über die vierfache Wurzel des Satzes vom zureichenden Grunde*, 1813, trans. E.F.J. Payne as *On the Fourfold Root of the Principle of Sufficient Reason*, La Salle: Open Court, 1974,

- Über die Freiheit des Menschlichen Willens*, 1840, trans. Konstantin Kolenda as *On the Freedom of the Will*, 2nd ed., Oxford: Blackwell, 1985,
- Über das Fundament der Moral / Über die Grundlage der Moral*, 1840/1, trans. E.F.J. Payne as *On the Basis of Morality*, rev. ed., Oxford: Berghahn Books, 1995,
- Die Welt als Wille und Vorstellung*, 2nd ed., 2 vols., 1844, trans. E.F.J. Payne as *The World as Will and Representation*, 2 vols., Indiana Hills: Falcon's Wing, 1958.
- Schwann, Alexander, ed., *Denken im Schatten des Nihilismus*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1975.
- Singer, Marcus George, *Generalization in Ethics: An Essay in the Logic of Ethics, with the Rudiments of A System of Moral Philosophy*, London: Eyre and Spottiswoode, 1963.
- Srzednicki, J.T.J., ed., *Stephan Körner: Philosophical Analysis and Reconstruction*, Dordrecht: Nijhoff, 1987.
- Williams, Bernard, *Shame and Necessity*, London: University of California Press, 1993,
- 'How Free Does the Will Need to Be?', in his *Making Sense of Humanity and Other Philosophical Papers 1982-1993*, pp.3-21,
- Making Sense of Humanity and Other Philosophical Papers 1982-1993*, Cambridge: Cambridge University Press, 1995.
- Wood, Allen W., *Hegel's Ethical Thought*, 1991.