THE UNIVERSITY *of* EDINBURGH

MORALITY, *ID EST*, WORTHINESS TO BE HAPPY

KANT'S RETRIBUTIVISM, THE 'LAW' OF UNHAPPINESS, AND THE ESCHATOLOGICAL REACH OF KANT'S 'LAW OF PUNISHMENT'

Cameron M. Thomson

Ph.D. Divinity

The University of Edinburgh

2011

DECLARATION

I hereby declare that I am the author of this thesis, that the work contained herein is mine alone, and that this work has never been submitted for any other degree or professional qualification.

Cameron Matthew Thomson

24 July 2012

ABSTRACT

Throughout his work, Kant regularly glosses 'morality' (and cognate expressions) as 'worthiness to be happy' (*Würdigkeit glücklich zu sein*). As a rule, Kant's commentators do not find this remarkable. Correctly understood, however, Kant's gloss on 'morality' is remarkable indeed. This thesis shows why.

In it, I argue that whenever we encounter Kant's gloss, we are faced with an implicit, durable cluster of unjustified commitments; that these commitments both antedate and survive his 'critical period'; that they are fundamentally practical in nature (i.e., that they are unexamined commitments to particular practices); and that these commitments entail a number of problematic theological consequences.

I argue, in particular, that Kant's gloss is a habit that signals, obscurely and implicitly, his antecedent commitments to the practice of capital punishment, on the one hand, and to a particular set of practical attitudes towards the happiness and unhappiness of immoral agents, on the other. I show that this habit has key implications for Kant's thinking about the agent that he calls 'God.'

My point of departure is Kant's claim, in his *Religion*, that the human being's particular deeds are imputable to her 'all the way down,' *only on condition* that the underlying 'disposition' (*Gesinnung*) from which they arise (according to their kind, *qua* moral or immoral) is imputable to her as well—that is, only if her (im)moral character may be regarded as the upshot of, or in some sense identical to, an utterly unassisted, unmotivated, originary deed on her part. I argue that Kant evades the question whether we really are permitted, without further ado, to regard this disposition (and with it an agent's deeds) as so imputable. He simply affirms his commitment to the practice of imputing particular deeds to particular agents and, with this affirmation, affirms that he takes the warrant that it requires (the imputability of '*Gesinnung*') to be secure.

I argue, then, that the theoretical significance of imputation, as expressed in this extraordinary, evasive leap, supervenes on the urgency of the commitments that are expressed in Kant's habitual glossing of 'morality' as 'worthiness to be happy.' The practice for which we would lack a warrant if the human being's character were not imputable to her is the imputation of her deeds under a description (of imputation) that has immediate reference to this same 'one's' *punishment*—specifically and only, however, to the extent that Kant takes punishments to be justifiable in none but strictly *retributivist* terms. These stakes and the constraining role of Kant's habitual gloss are clearest, I argue, in his thinking about the practice of putting murderers to death—a practice, I argue, that has both a political and an eschatological significance for him.

…it is from the necessity of punishment that the inference to a future life is drawn.

— Kant, *The Metaphysics of Morals*

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to recognize a few of the many people who have supported me as I worked towards the completion of this thesis. Foremost among these is my spouse, my best friend, Jennifer. I would also like to express my deep gratitude towards my daughters, Dellis, Abigail, and Violet, who accompanied me on this often-difficult journey. Completion of this project would have been a far more difficult task than it was without their forbearance and companionship.

I also thank Nicholas Adams, my primary supervisor and teacher at the University of Edinburgh. This thesis owes much to his wise guidance and good sense.

I thank Marsha Hewitt, too, a key influence in the early stages of my graduate studies, who first got me worrying about Kant. Her kindness and compassionate understanding of the challenges facing a married graduate student with three daughters have been a great comfort to me during the process leading to this thesis' completion.

Other teachers whose beneficent impact on my life and thinking bears acknowledgement include Phillip Wiebe, Ian Angus, David Mirhady, Ray Jennings, Christopher Morrissey, Otfried Höffe, and Christoph Schwöbel.

The final year of work on this project was fraught with personal challenges that would have been unnavigable without the loving support (and extended presence in the winter of 2011) of my mother and father. Many thanks are due, too, to my parents-in-law, Harry and Terry. I am grateful, in particular, for my elders' acceptance of my weaknesses and their willingness to help. I am particularly grateful to my sister Carmen, as well, for her unconditional love and supportive friendship; she too played a role in bringing this thesis to completion.

Others offered integral support too, in various capacities, without which it is unlikely that I would ever have completed this work. I am thinking, especially, of Wade Larson (friend of my youth, faithful friend and supporter ever after), Altuğ Hasözbek (a refuge in Izmir, friendship in Tübingen), Mark Wallin (host and dear companion of the final, horrible push), the good people of Scarboro Missions (quiet, peace, prayer), Marilyn Elphick (loving friend, gifted listener), Valerie Ha (the best 'real' doctor ever), and also Phil and Michelle Wiebe, Steven and Christine Berg-Nederveen, Maria Calderone, Shawn and Magda Kazubowski-Houston, Fr.'s Ed

Hone and Michael Henesy, Sarah Haley, Vance Tschritter, Jeremy Roberts, and Nick Olkovich (friends and dear encouragers all).

A number of other friends and colleagues specific to my time in Edinburgh and Tübingen have left their mark on my life and work during these last few years. In particular, I would like to acknowledge the influence and loving support of three friends and fellow students at the University of Edinburgh, Blair Wilgus, Steve Martz, and (above all, dear one, for so many deeply formative conversations) Rob Burns; and a number of friends from Tübingen, especially Tina and Marco, and Edison and Maria.

I would also like to thank my examiners, Professors David Fergusson and Paul Janz, for the priviledge of their attention to my work, and for their very stimulating and helpful comments upon it.

Finally, thanks are due to the Social Sciences and Humanities Research Council of Canada, the University of Edinburgh's College of Humanities and Social Science, Universities UK, New College (University of Edinburgh), and the Deutscher Akademischer Austausch Dienst for their substantial financial support of my work.

ABBREVIATIONS AND NOTE ON PRIMARY SOURCES

All references to Kant use the abbreviations listed below and are, first, to the volume and page number of *Kants Gesammelte Schriften* (hereafter *KGS*), herausgegeben von der Deutschen (formerly Königlichen Preussischen) Akademie der Wissenschaften, 29 volumes (Berlin: Walter de Gruyter [et al], 1902 -) and second, in brackets, to the English translations listed below. Where I indicate no such pages (in brackets) the translation, if there is one, is mine. Note, too, that whenever an English translation renders Kant's emphasis in bold font, I have changed this to italic.

KrV A/B *Kritik der reinen Vernunft* (*KGS* 3 and 4)
Critique of Pure Reason. Translated by P. Guyer and Allen W. Wood The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1998.

Anthro *Anthropologie in pragmatischer Hinsicht* (*KGS* 7)
Anthropology from a Pragmatic Point of View. Translated by Robert B. Louden Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2006.

Anthro-C *Anthropologie Collins* (*KGS* 25)

Anthro-M *Anthropologie Mrongovius* (*KGS* 25)

Anthro-P *Anthropologie Parow* (*KGS* 25)

Bem *Bemerkungen zu den Beobachtungen über das Gefühl des Schönen und Erhabenen* (*KGS* 20)

DWL *Logik Dohna-Wundlacken* (*KGS* 24)
'The Dohna-Wundlacken Logic.' In *Lectures on Logic*, edited by J. Michael Young, 431-520. The Cambridge Edition of the

Works of Immanuel Kant. New York: Cambridge University Press, 1992.

Ende *Das Ende aller Dinge* (*KGS* 8)

'The End of All Things.' In *Religion and Rational Theology*, edited by Alan W. Wood and George Di Giovanni, 217-32. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

Fort *Preisschrift über die Fortschritte der Metaphysik* (*KGS* 20)

'What Real Progress has Metaphysics Made in Germany since the Time of Leibniz and Wolff?' In *Theoretical Philosophy after 1781*, edited by Henry Allison and Peter Heath, 337-424. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

Gr *Grundlegung zur Metaphysik der Sitten* (*KGS* 4)

Groundwork of the Metaphysics of Morals. Translated by Mary Gregor Cambridge Texts in the History of Philosophy. New York: Cambridge University Press, 1997.

Idee *Idee zu einer allgemeinen Geschichte in weltbürgerliche Absicht* (*KGS* 8)

'Idea for a Universal History with a Cosmopolitan Aim.' In *Anthropology, History, and Education*, edited by Günter Zöller and Robert Louden, 107-20. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2007.

KdU *Kritik der Urtheilskraft* (*KGS* 5)

Critique of the Power of Judgment. Translated by Paul Guyer and Eric Matthews The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2000.

KpV *Kritik der praktischen Vernunft* (*KGS* 5)

Critique of Practical Reason. Translated by Mary Gregor Cambridge Texts in the History of Philosophy. New York: Cambridge University Press, 1997.

LEC *Moralphilosophie Collins* (*KGS* 27)

'Moral Philosophy: Collin's lecture notes.' In *Lectures on Ethics*, edited by Peter Heath and J. B. Schneewind, 37-222. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LEM₂ *Moralphilosophie Mrongovius II* (*KGS* 29)

'Morality according to Prof. Kant: Mrongovius's second set of lecture notes.' In *Lectures on Ethics*, edited by Peter Heath and J. B. Schneewind, 223-48. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LEV *Moralphilosophie Vigilantius* (*KGS* 27)

'Kant on the metaphysics of morals: Vigilantius's lecture notes.' In *Lectures on Ethics*, edited by Peter Heath and J. B. Schneewind, 249-452. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LMD *Metaphysik Dohna* (*KGS* 28)

'Metaphysic Dohna.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 357-94. The Cambridge Edition

of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LMH *Metaphysik Herder* (*KGS* 28)

'Metaphysic Herder.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 3-18. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LMK$_2$ *Metaphysik K$_2$* (*KGS* 28)

'Metaphysic K_2.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 395-416. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LML$_1$ *Metaphysik L$_1$* (*KGS* 28)

'Metaphysic L_1.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 19-108. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LMM *Metaphysik Mrongovius* (*KGS* 29)

'Metaphysic Mrongovius.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 109-288. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

LMV *Metaphysik Volckmann* (*KGS* 28)

'Metaphysic Volckmann.' In *Lectures on Metaphysics*, edited by Karl Ameriks and Steve Naragon, 289-98. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1997.

MdS *Die Metaphysik der Sitten* (*KGS* 6)

The Metaphysics of Morals. Translated by Mary J. Gregor Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 1996.

ND *Principiorum Primorum Cognitionis Nova Dilucidatio* (*KGS* 1)

'A New Elucidation of the First Principles of Metaphysical Cognition.' In *Theoretical Philosophy, 1755-1770*, edited by David Walford and Ralf Meerbote, 1-46. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1992.

OP *Opus Postumum* (*KGS* 21 and 22)

Opus Postumum. Translated by Eckart Forster and Michael Rosen The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1993.

Orient *Was heißt: Sich im Denken orientieren* (*KGS* 8)

'What Does it Mean to Orient Oneself in Thinking.' In *Religion and Rational Theology*, edited by Alan W. Wood and George Di Giovanni, 1-18. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

R *Reflexionen* (*KGS* 17-19)

Notes and Fragments: Logic, Metaphysics, Moral Philosophy, Aesthetics. Translated by P. Guyer The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2005.

Rel *Die Religion innerhalb der Grenzen der blossen Vernunft* (*KGS* 6)

Religion within the Limits of Reason Alone. In *Religion and Rational Theology*. Translated by Alan W. Wood and George Di Giovanni The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2001.

Streit *Der Streit der Fakultäten* (*KGS* 7)

'The Conflict of the Faculties.' In *Religion and Rational Theology*, edited by Alan W. Wood and George Di Giovanni, 233-328. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

ÜdG *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis* (*KGS* 8)

'On the Common Saying: That May be True in Theory, but it is of No Use in Practice.' In *Practical Philosophy*, edited by Mary J. Gregor and Alan W. Wood, 273-310. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

ÜdM *Über das Mißlingen aller philosophischen Versuche in der Theodicee* (*KGS* 8)

'On the Miscarriage of all Philosophical Trials in Theodicy.' In *Religion and Rational Theology*, edited by Alan W. Wood and George Di Giovanni, 19-38. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2001.

Versuch *Versuch den Begriff der negativen Größen in die Weltweisheit einzuführen* (*KGS* 2)

'Attempt to Introduce the Concept of Negative Magnitudes into Philosophy.' In *Theoretical Philosophy, 1755-1770*, edited by David Walford and Ralf Meerbote, 203-42. The Cambridge

Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1992.

Vorarbeiten-MdS *Vorarbeiten zu MdS* (*KGS* 23)

Vorarbeiten-ÜdG *Vorarbeiten zu ÜdG* (*KGS* 23)

Vorlesungen-Religionslehre *Vorlesungen über die philosophische Religionslehre* (*KGS* 28)

'Lectures on the Philosophical Doctrine of Religion.' In *Religion and Rational Theology*, edited by Alan W. Wood and George Di Giovanni, 335-452. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 2001.

ZeF *Zum ewigen Frieden* (*KGS* 8)

'Toward Perpetual Peace.' In *Practical Philosophy*, edited by Mary J. Gregor and Alan W. Wood, 311-52. The Cambridge Edition of the Works of Immanuel Kant. New York: Cambridge University Press, 1996.

INTRODUCTION

Throughout his work, Kant regularly glosses 'morality' (and cognate expressions) as 'worthiness to be happy' (*Würdigkeit glücklich zu sein*).[1] As a rule, Kant's commentators do not find this remarkable.[2] Correctly understood, however, Kant's gloss on 'morality' is remarkable indeed. This thesis shows why. In it, I argue that whenever we encounter Kant's gloss, we are faced with an implicit, durable cluster of unjustified commitments on his part. I contend that these commitments both antedate and survive his 'critical period.' I show that they are fundamentally practical in nature—i.e., that they are unexamined commitments to particular practices. And I demonstrate that these commitments entail problematic theological consequences.

The main motivation for the descriptive and analytical task encapsulated here—the unearthing and detailed unpacking of Kant's gloss on 'morality'—is one whose full articulation would take us far beyond the limits of the present project. The research embodied in this thesis constitutes the first part of a far more extensive account of the developments in Kant's thinking about freedom and moral accountability that terminate, ultimately, in his theory of 'radical evil' (*radikales Böse*). In short, I hypothesize that the commitments that find expression in Kant's gloss on 'morality' condition the trajectory by which his 'critical period' thinking about freedom and moral accountability passes from its initial state in (roughly) the 'Third Antinomy' section of the first *Critique* (1781) to its final state in the *Religion*'s (1793) theory of the imputable '*Gesinnung*' (see below). The limitations of this thesis do not allow me to identify and account for the various phases of this trajectory.[3]

---

[1] In the main, I will refer to this practice as 'Kant's gloss,' or 'Kant's glossing "morality" as "worthiness to be happy."' Other locutions, however, will convey the same basic idea. Sometimes I will speak, for example, of Kant's deploying the notion or concept of 'worthiness to be happy' and, at others, of his use of the '"worthiness to be happy" idiom.'

[2] I discuss exceptions to this rule—and distinguish my position from theirs—in chapter 1.

[3] In my conclusion, however, I sketch the contours of this line of development and demonstrate in a rudimentary fashion that Kant's thinking may be fruitfully read in these terms.

I leave these matters aside for the present in order to pursue a more preliminary, fundamental aim.

In what follows I execute a task akin to the one a cosmologist takes up, when, having made a close study of the puzzling movements of some heavenly body, he proposes that the presence of a second massive, but invisible body explains the current, or last observed, position of the first. This thesis does not deal directly with these puzzling movements (the developments leading to Kant's theory of radical evil); and it deals only briefly (i.e., here in this introduction) with the ultimate disposition of Kant's explicit claims about freedom and moral accountability. My immediate goal rather, which I pursue in the main body of this work, is to identify and describe the 'invisible body' in question.

I contend, then, that Kant's theory of radical evil lies at the further end of a series of theoretical developments whose trajectory is constrained, in part, by some of his deepest practical commitments. I argue that these commitments are unexamined ones. And I demonstrate that these find their earliest, subsequently most pervasive, and yet least noticed mode of expression in Kant's regular glossing of 'morality' as 'worthiness to be happy.' In particular, as I will show, these commitments are Kant's deeply antecedent allegiance to the practice of capital punishment, on the one hand, and his commitment to a particular set of practical attitudes towards the happiness and unhappiness of immoral agents—with key implications for Kant's thinking about God—on the other. I will qualify and temper this rather strong and still somewhat vague proposal as we go along. In order to do this, it is necessary to begin where Kant's thinking about moral accountability ends: with his theory of radical evil.

The latter is a highly complex machine.[4] Fortunately, for the purposes of this introduction I need focus on no more than one very circumscribed (but integral) aspect of it: Kant's claims regarding the imputability (to her) of the human being's deepest,

[4] Beyond what I have to say in this introduction, I will not go very deeply into this topic. For helpful readings in this area see Henry E. Allison, 'On the Very Idea of a Propensity to Evil,' *Journal of Value Inquiry* 36, no. 2-3 (2002); Richard J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge: Polity Press, 2002); Peter Dews, *The Idea of Evil* (Malden, MA: Blackwell Pub., 2007); P. Formosa, 'Is Radical Evil Banal? Is Banal Evil Radical?,' *Philosophy & Social Criticism* 33, no. 6 (2007); Robert Gressis, 'How to Be Evil: Kant's Moral Psychology of Immorality,' in *Rethinking Kant*, ed. P. Muchnik (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008); Paul Guyer, *Kant*, Routledge Philosophers (London: Routledge, 2006), 226-30; Gordon E. Michalson, *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge: Cambridge University Press, 1990).

radical, most fundamental character, that is, the moral or immoral 'disposition' (*Gesinnung*) from which her deeds spring and relative to which (alone) each and all of her particular deeds may be deemed determinately and 'rigoristically'[5] either moral or immoral.

In his *Religion*, Kant argues that an agent's particular deeds are imputable to her only on condition that the underlying disposition that governs and determines these deeds (according to their kind, *qua* moral or immoral) is itself imputable. First, after describing what he calls 'the subjective ground…of the exercise of the human being's freedom *in general*' and after pointing out that as 'ground' and as 'general' this must be '*antecedent to every deed* that falls within the scope of the senses,' Kant argues that this 'ground' must *itself* arise from freedom—and be itself regarded as a deed[6]—because '*otherwise* [*denn sonst*] the use or abuse of the human being's power of choice with respect to the moral law *could not be imputed to him*.'[7] Then later, describing this 'disposition' as the unified, 'first subjective ground of the adoption of…maxims,' a principle that 'applies to [the individual agent's] entire use of freedom universally,' Kant insists that '[t]his disposition…must be adopted through the free power of choice, *for otherwise [*denn sonst*] it could not be imputed*.'[8] And again, in another instance, Kant argues that '[t]he human being must make or have made *himself* into whatever he is or should become in a moral sense, good or evil.' He continues in the now familiar vein: '[t]hese two [characters] must be an effect of his free power of choice, *for otherwise [*denn sonst*] they could not be imputed to him* and, consequently, he could be neither *morally* good nor evil.'[9]

Taken together, these three passages set forth two distinct, but closely related conditionals. The first conditional states that:

(i) *If* an agent's '*Gesinnung*,' or, more precisely, her will's having the *kind* of '*Gesinnung*' (moral or immoral) that it has, is not in some (sufficiently robust) sense the upshot of 'an intelligible deed'[10] (which means, for

[5] Cf. *Rel* 6: 22-5 (71-3).
[6] For more on Kant's unusual use of 'deed' (*Tat*) in this context see *Rel* 6: 31 (78-9).
[7] *Rel* 6: 21 (70) (my emphasis).
[8] *Rel* 6: 25 (74) (my emphasis).
[9] *Rel* 6: 44 (89) (the first italicization is mine).
[10] *Rel* 6: 31 (79).

Kant, among other things, that she is its author, wholly and simply), *then* her '*Gesinnung*' cannot be imputed to her.

The second conditional states that:

(ii) *If* an agent's '*Gesinnung*' cannot be imputed to her (in the sense that follows, for Kant, from her '*Gesinnung's*' being, or in some sense being grounded in, 'an intelligible deed'), *then* her individual empirical deeds—which, for Kant, must be (in principle) qualifiable, with strict reference to the fundamental quality of the '*Gesinnung*' that governs their production, 'rigoristically' (that is, without any ambiguity), as either wholly moral or wholly immoral—cannot be imputed to her (i.e., in Kantian terms, *both* ascribed to her, as their dynamic source or cause, *and* qualified in moral terms as instances of the kinds 'moral' or 'immoral').

Note the relationship between these two conditionals: (i)'s antecedent ('an agent's "*Gesinnung*" is in no sense 'a deed of freedom') directs us to a consequent that figures, in turn, as (ii)'s antecedent (namely, the claim that 'this same agent's "*Gesinnung*" cannot be imputed to her'). If we schematize (i) and (ii) as $\sim P \rightarrow \sim Q$and $\sim Q \rightarrow \sim R$, respectively, where '$\sim R$' refers to the claim that 'individual deeds cannot be imputed to this agent,' then, by transitive inference, we also have $\sim P \rightarrow \sim R$, or the claim that:

(iii) *If* an agent's '*Gesinnung*' is not 'a deed of freedom' on her part, *then* her individual deeds cannot be imputed to her.

Or again, we are presented with a series of conditionals such that:

(i.a) ~('*Gesinnung*' is 'a deed of freedom') → ~('*Gesinnung*' is imputable)

(ii.a) ~('*Gesinnung*' is imputable) → ~(individual deeds, d_1, d_2, $d_3 \ldots d_n$, are imputable)

and

(iii.a) ~('*Gesinnung*' is 'a deed of freedom') → ~(individual deeds, d_1, d_2, $d_3 \ldots d_n$, are imputable)

But by contraposition we may deduce from (iii.a) that:

(iv.a) ~~(individual deeds, d_1, d_2, $d_3 \ldots d_n$, are imputable) → ~~('*Gesinnung*' is 'a deed of freedom')

which is to say (by the definition of double negation) that:

(v.a) individual deeds, d_1, d_2, $d_3 \ldots d_n$, are imputable → '*Gesinnung*' is 'a deed of freedom'

Of course, in order to infer from (v.a) that an agent's '*Gesinnung*' really *is* 'a deed of freedom,' we (or Kant) would have to show, first, that the antecedent in this case (the claim that individual deeds, d_1, d_2, $d_3 \ldots d_n$, are imputable) really is true. It is not possible to *demonstrate* that this is true so, however, and Kant does not attempt to do so. The claim that an agent's particular deeds really are imputable to *her*, wholly and simply, so that she is identified as their unique author, can only be *presupposed*. And Kant, I suggest, is always already committed to the *practice* that consists in the imputation of particular deeds to their agents, a practice in the course of whose enactment it does seem simply to be *taken for granted* that particular deeds really are imputable, that is, a practice by which the individual agent really is identified, *wholly and simply*, as the author of her particular deeds. This emphatic qualification, 'wholly and simply,' gives expression, furthermore, to Kant's equally antecedent commitment to the practice of punishment, but under a description of this practice (which remains, perhaps, implicit in the unthematized self-understanding of its practitioners) that marks it as an undertaking that is addressed, also 'wholly and simply,' to *individuals* and not—as for example in the practice of 'restorative justice'—to the community (under which alternative conception it might be regarded, instead, as a strategy for navigating a *shared* predicament—shared in a way that guilt, as Kant conceives of the latter, cannot be).

Now, it might be objected that I ought really to focus directly on the question how the individual human being's '*Gesinnung*' can qualify, after all, as a peculiar kind of 'intelligible deed,' an authentic deed of freedom, as Kant claims that it does. Someone might object that it would be better to simply move straightaway to a disputation of Kant's claim that this is so. My reply to this objection would be that to take Kant to task in this regard, at once—to deny that '*Gesinnung*' can be regarded as a 'deed of freedom' and to argue that Kant shows no more than that a particular pair of practices (the imputation of particular deeds, the punishment of guilty parties) cannot be warranted without this sheer *presupposition*—would be to say too much, given the main purpose of this thesis. The main aim of this thesis is to show, rather,

that Kant is simply committed to these practices and that this suffices to explain why he makes the claims concerning '*Gesinnung*' that he makes in the *Religion*. It is possible that those claims are overdetermined—that his making them can be explained *both* without reference to any argumentation on his part for the claim that an agent's '*Gesinnung*' really is an 'intelligible deed' *and*, at the same time, that Kant also seeks to *justify* this claim (he does not), and does not simply presuppose it in order to stabilize and perpetuate of those practices.

But to argue directly with Kant about the freedom (and so, too, the imputability) of '*Gesinnung*' would take us too far afield. My foray into the *form* (the logic) of Kant's reasoning, which I pursued above, serves a point that does not require that I argue for more substantive claims about the *content* of Kant's argumentation in the *Religion*. This point is actually a central one for this thesis, which is to show what is at stake here for Kant—namely, the *practice* that consists in identifying individual agents as the authors of particular deeds and in qualifying those deeds as either moral or immoral—to show, that is, why it is that '*Gesinnung's*' being a 'deed of freedom' matters to Kant *at all*, in the first place, from the outset—before he ever thematizes this commitment. For the moment, in other words, I am not going to argue with Kant about whether the underlying character of the individual human agent's will really *is* like that, or even about whether we are *warranted* in so regarding it. My claim here is simply that, whatever its status (true or false, warranted or unwarranted), and whatever the arguments that Kant adduces for it, he is *constrained* (in advance of, and in spite of, any argumentation) to a theoretical description of the human agent as a free being the character of whose will is imputable to her, without remainder; and that he is constrained to this description of the human agent to the extent that he is deeply and durably committed to the practice of imputing deeds to agents—but again more particularly, that is, in a manner that opens onto the practice of punishment under a retributivist (hence entirely individualistic) description of that practice.

The main point in *this* context, then, is not that Kant claims (problematically) that '*Gesinnung*' is a deed of freedom (something for which he should be taken to task, certainly), but that there are practical reasons (again, I bracket out the question of whether Kant adduces any theoretical reasons, ultimately, in support of this move)

for taking it for granted that the claim that particular deeds are *not* imputable is indeed a *counterfactual* one.

One of this thesis' main contributions, then, is to prepare the ground for an account of how Kant ends up claiming that '*Gesinnung*' is a deed of freedom—which is distinct from an account that follows his reasoning to that claim (which I simply deny is present, in any case, but without arguing the point in this context). And I prepare the ground, here, by way of an analysis of his habitual deployment of the 'worthiness to be happy' idiom, offering an account of this habit of thought and expression that finds in Kant's deployment of this idiom a sign of his commitment to the practice of imputation *under a particular description*, namely, again, one that has direct reference to the practice of punishment—with capital punishment as its paradigmatic instance.

To sum up: in each of the three passages that I referred to above Kant's 'for otherwise' (*denn sonst*) expresses an implicit claim in conditional form. In the first place, Kant holds that if we cannot impute an agent's underlying disposition to her, then we cannot impute her particular deeds to her (in any sense that is relevant for 'morality'). But then, too, he holds that if her disposition is a not a product of (her) freedom and so in some sense a deed (of hers) like her other deeds, then we cannot impute *it*. There is a palpable urgency here: if we cannot do *this*, then we cannot do *that*. And *that* matters. The resolution of this urgent problem consists in Kant's taking the consequent in each case (we cannot impute her particular deeds to her; we cannot impute her character to her) to be a counterfactual claim. Implicitly, in each case, he denies the consequent and thus, by contraposition, is entitled to deny the antecedent as well. He presupposes that we *can* impute the human being's actions to her—we do and we may. Thus we may impute to her the underlying disposition from which these spring as well. Indeed, when we impute her actions to her, this shows that we already regard her disposition to act in the '*way*' that she does as something of which she is the author too. And if we are entitled to regard her as the author of *that*, then we must already regard her underlying character, at least implicitly, as a product of her freedom.

In a sense, this is simply to describe what we *do* and to assert of what we do that it shows that we are already thinking in a certain way, implicitly, just by doing it.

Kant does not actually claim that the human being's particular deeds, along with her underlying character, really *are* imputable to her. He claims only that if we are not entitled to affirm that her underlying character is imputable to her, then a particular practice—one in which we are always already engaged in any case—would be called into question in some way. However, Kant does *not* take the practice in question to be problematic. To the contrary, he takes it to be unproblematic; and he takes it therefore, too, that any condition whose absence would inevitably render it problematic must then be present. We *must* impute the agent's character to her, as something that she has done, '*for otherwise*' the practice of imputing to her 'the use or abuse of [her] power of choice with respect to the moral law' (i.e., the 'use or abuse' of freedom on this or that occasion, as manifest in particular deeds) would be without foundation and, in some sense, impermissible.

Admittedly, this reasoning involves a crass leap and I hesitate to attribute it to Kant. I do not attribute it to him, however, as a piece of reasoning. I attribute it to him as an antecedent, unexamined commitment—a kind of bias—whose inevitability and goodness he simply takes for granted. Thus it is not simply a position, or thesis, or intellectual point of view that is at stake here. Rather, it is a particular practice: the practice of imputing deeds to agents. This gets near to the heart of the matter. And Kant proceeds as though we are called to this practice inexorably, as though we were barred from the outset from saying, 'It is going too far to impute an agent's character to her in this way. If by "imputation" we mean something that cannot be done at all, save on condition that we do *that*, then let us simply give up imputing her deeds to her. Let us do something else, instead.' Or, alternatively, 'Let us think about whether or not we might not give a different account of what we are doing when we identify someone as "the one that did it" and then see where that gets us.' For Kant, the practice of imputing agents' deeds to them—in a mode that expresses the idea that they are these deeds' authors *all the way down*[11]—is so important and so untouchable that Kant does not really investigate it.

[11] For a cognate use of this idiom, which I will deploy from time to time throughout this thesis, see for example Pablo Muchnik, *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Aprioricity of History* (Lanham: Lexington Books, 2009), 52 and cf. J. B. Schneewind, 'Kant and Stoic Ethics,' in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, ed. S. P. Engstrom and J. Whiting (Cambridge: Cambridge, 1994), 290.

In order to see why this is so, it is important to recognize, first, that what Kant holds too dear to investigate is not just imputation, but 'imputation' precisely in a sense that demands that, in addition to all of her particular deeds, an agent's '*Gesinnung*' be imputable to her as well. Not every description of that practice—the practice that consists, say (roughly), in identifying, relative to some event 'that falls within the scope of the senses,' 'the one that did it' and qualifying this 'it' as either good or bad in some sense—not every further specification of this practice picks out a practice that we may legitimately undertake *only on condition* that the underlying disposition from which an agent acts is itself imputable to her as though it too were one of her deeds. For Kant, however, 'imputation' does refer to such a practice—and, again, it is so important that it is immune to investigation.

I submit that, for Kant, the practice for which we would (counterfactually, he assumes) lack a warrant if '*Gesinnung*' does not arise from, and express, an originary free choice of 'the one that did it,' is the imputation of an agent's deeds under a description (of imputation) that has immediate reference to this same 'one's' *punishment*, specifically and only, moreover, to the extent that Kant takes each instance of punishment to be justifiable (if at all) in none but strictly *retributivist* terms.[12] We might refer to this condition as the originary and terminal unity of the acting and suffering subject: 'the one that did it' and 'the one that ought to be punished' have to be unambiguously one and the same, an utterly consolidated identity. The reason that Kant is interested in the imputability of the agent's disposition in the first place, then, is that the imputation of deeds—precisely in this retribution-sensitive sense—is a practice to which he is antecedently committed. And his commitment to the practice of imputation (in this special sense) is a function of his antecedent commitment to the practice of punishment, so construed. This conception of punishment, as we shall see, is clearest in Kant's thinking about the practice of putting murderers to death, but it has a much wider scope of application. In fact, as I will argue, in its 'capital' mode, punishment has both a political and an eschatological character for Kant: the

[12] My claim that this antecedent commitment is to be taken in precisely *retributivist* terms sounds particularly odd. It is as though I am claiming that Kant was a retributivist before he ever reflected on the problem of justifying punishment. In a sense this *is* what I claim. Kant's retributivism, as I show in chapter 3, translates the extreme immediacy of his commitment to the practice of capital punishment into the language of a justifying account of the latter. The answer to the question: 'Why put murderers to death?' is always, 'Because it is absolutely necessary that they die.'

murderer, an agent that deserves to *die* for what she has done, deserves to die in light of a deed which, even in the empirical context of the *polis*, exhibits the same 'inner wickedness' that, in the case of immoral agents in general, demands their eschatological unhappiness. Unique among criminals, the murderer deserves to die, not merely for what she has *done*, but for who she is, as it were, 'deep inside.' Kant is committed, not merely to the political practice of punishment, conceived as retribution, but to a practice that shades off, as it were, into (Kant's) God's practice of *infernal* retribution.

As Kant's brief foray into eschatology at the end of *The Metaphysics of Morals* shows, he takes it that the ultimate punishment of immoral agents is an absolute 'necessity.' Indeed, as Kant puts it there, '*it is from the necessity of punishment that the inference to a future life is drawn*.'[13] And he means that eschatologically rendered punishment is necessary, I suggest, in the same sense and for the same reason as the temporal sovereign's punishment of murder by way of death on the scaffold. This necessity and the assimilation of punishing action to action that makes-unhappy—that denies happiness to agents who, by failing to be moral, are unworthy to be happy—allows me to generalize from Kant's 'law of punishment,' which he explicitly asserts is a categorical imperative,[14] to what I will be calling Kant's 'law of unhappiness.'

I will elaborate on Kant's 'law of punishment' and the latter's relation to this so-called 'law of unhappiness' in due course. First, however, by way of preventing a serious misunderstanding, which threatens to arise in connection with this expression—that is, my use specifically of 'law' in the expression 'law of unhappiness'—I must offer an initial clarification of my intentions in this regard. In fact, Kant's use of 'law' in his reference to a 'law of punishment' poses a similar problem. But since Kant asserts the existence of the latter and clearly expresses his commitment to what it prescribes, while (as I will argue) only *implying* that he is committed to the 'law of unhappiness' (or to what it prescribes), it is particularly pressing that I address the ostensible 'lawfulness' (the universal bindingness or practical necessity) of the Kantian principle that immoral agents ought to be *unhappy*. In addressing this matter, I

[13] *MdS* 6: 490 n. (232 n.) (my emphasis). I discuss this remarkable claim, which I have made the epigraph of this thesis, in chapter 3.
[14] *MdS* 6: 331 (105).

will be obliged to refer in a rudimentary manner to a number of key points whose full elaboration must be deferred to the subsequent chapters of this thesis.

In particular, I must acknowledge the appearance of a problem here, which would bear on my understanding of the relationship that Kant takes, at least some of the time, to hold between immorality, on the one hand, and unhappiness, on the other. I need to be as clear as possible about my treatment of Kant's notion of 'practical law,' specifically that is, in regards to what I will characterize as the eschatological 'reach' of the retributivism expressed by Kant's key claim that 'the law of *punishment* is a categorical imperative.'[15]

Let me be very clear: I will not be claiming that the 'matter' of this or that empirical notion of 'unhappiness' can, just as such, give content to a genuinely Kantian 'practical *law*,' or categorical imperative—a law that commands as its primary end that the empirical state of affairs consisting in this or that agent's being *unhappy* be realized.

In the following, as too in Kant's thinking on this topic, 'unhappiness' refers to a particular state of this or that agent's practical affairs, namely, the state of affairs that obtains when things do not go her way, or do not conform to the empirical desires that she happens to have (whether this is taken to be so at some particular time, or on the whole). I shall regard an agent as unhappy, in other words, when her experience of the world fails to include the realization of the particular ends that she has just in virtue of her empirically (and so contingently) given inclinations. One way in which the thwarting of her desires might come about is through such an agent's being punished. This is so to the extent, particularly, that punishment (which is always, for Kant, an *unwanted* imposition,[16] that is, 'unilateral,' but nevertheless 'rightful' coercion[17]) consists fundamentally in 'physical harm,'[18] or the infliction of 'pain,'[19] and so (according to what I regard as a centrally important definition) in the punished agent's forfeiture of (some measure of) happiness.[20]

---

[15] *MdS* 6: 331 (105).
[16] *MdS* 6: 331 (104).
[17] *MdS* 6: 363 n. (130 n.).
[18] See *KpV* 5: 37 (34).
[19] *MdS* 6: 331 (104).
[20] *KpV* 5: 37 (34). For more detailed discussion of Kant's definition of punishment see chapter 3.

When I refer to Kant's 'law of unhappiness' and indicate that I take the latter to be an implicit principle in Kant's thinking which, when thematized (as Kant does not do), turns out to have the dignity *for Kant* of a '*law'* (always with the qualification that, unlike his 'law of punishment,' of which the 'law of unhappiness' is a kind of eschatological analogue, Kant's taking this to be a 'law,' along with his endorsement of it and of what it commands, is almost entirely *implicit* in his thinking) I do not at all mean to say that 'the law of unhappiness' is an imperative that commands its *subject* to *be unhappy*. Nor do I mean to say that this law is an imperative that commands an end that is determined by reflection on the empirical 'matter' of unhappiness.

This would be a truly problematic claim. In general, on Kant's understanding of them, the ends that are practically necessitated by categorical imperatives cannot be ends, after all, whose content is given by this or that finite subject's inclinations or, as in the case of unhappiness, her disinclinations. The ends that categorical imperatives prescribe cannot be ones that are given, in other words, by any finite agent's pragmatic insight into, or 'wisdom' concerning, what would actually render this or that particular human being happy or unhappy. Reflection on the latter, as Kant argues, can only offer up *hypothetical* imperatives (e.g., 'If you would be happy, do *X*'; 'If you would render so-and-so unhappy, do *Y*').

If this were what I meant by the claim that, for Kant, 'the law of *punishment*,' which he does claim to be a *true* practical law (i.e., an authentically *categorical* imperative), finds an eschatological 'extension' in his implicit 'law of unhappiness,' then a critic would be right to assert that there can no more be a 'law of unhappiness' than there can be a 'law of happiness.' Such a critic would also be correct to point to a serious confusion, on my part, concerning the distinction between what Kant means by 'punishment' (when he explicitly makes the latter the aim of a categorical practical principle, that is, *The Metaphysics of Morals*' 'law of punishment') and what Kant means by 'unhappiness.' But without ever suggesting that such a thing is possible, I claim no more than to find in Kant's regular use of his 'worthiness to be happy' idiom (that is, his habit of glossing 'morality' as 'worthiness to be happy') a sign of his commitment to an implicit eschatological/ethical analogue of his 'law of punishment.' And, regarding this eschatological/ethical analogue as a principle that in-

herits the categorical nature that, *for Kant*, qualifies his 'law of punishment,' I refer to this as 'the law of unhappiness.'

Again too, as I have intimated, this might also be expressed in terms of what I claim is an eschatological and so also a theological 'extension' of the 'reach' of Kant's *politically* situated 'law of punishment.' This means that Kant's *undefended*, but nevertheless clear, explicit assertion that 'the law of punishment is a categorical imperative,' gives expression to a certain thought about the notional relationship between an essential property of the murderous will (what Kant refers to as the murderer's 'inner wickedness') and the putting to death of that will's 'bearer,' which thought 'shades off,' as I put it, into a related conviction bearing on the relationship between immorality (again, too, construed as 'inner wickedness') and unhappiness *more generally* (i.e, beyond the definitive forfeiture of happiness that occurs in capital punishment).

For Kant, as we shall see, the law of *punishment* (at least where the crime in question is murder) commands, unconditionally, that the supposedly *a priori* (which aprioricity is implied by the claim that the injunction is a categorical command), notional connection between the concepts of 'inner wickedness' and 'death-for-murder' be actualized, in practice—by the sovereign or the latter's representative in the experience of the criminal—*and* it calls upon our *approbation*, which affective state signifies (I argue) our impotent, or 'virtual' and 'deferred,' *intention* or *will* that this notional connection be actualized. The law of punishment commands, indeed, not only that the sovereign make the occlusion of the murderer's access to happiness (in short, I suggest, her unhappiness) his or her end, but also forbids the sovereign's particular subjects to creep 'through the windings of eudaemonism'[21] in order to find a way out of the endorsement that is demanded of them.

However, Kant's reference to the 'law of punishment,' in the literary context of his *Metaphysics of Morals*, bears on something that he thinks *ought* (unconditionally, not simply in service of some other end, however laudable) to happen in the spatio-temporal milieu of the human *polis*: the putting-to-death of the members of a particular class of criminals. By contrast, the 'law of unhappiness' commands that a particular subject, that is, (one version of) Kant's '*God*,' insure (ultimately) that immoral

[21] *MdS* 6: 331 (105).

agents are unhappy—that this subject ('God') make their unhappiness an end. Of course, this eschatological unhappiness may be regarded as *punishment* (and Kant does so regard it), but it is punishment in a sense that is distinct from (if also analogous to) punishment in the sense that Kant intends when he refers to the *political* 'law of punishment.' What both 'laws' have in common is the retributivism of the procedure that asserts (but does not show) that their aim is warranted. Again, this aim is not punishment, or the 'forfeiture' of happiness (hence unhappiness), '*per se*,' but rather the actualization of the notional relationship that is supposed to hold between the concepts of crime and punishment (or punishing), on the one hand, and immorality and unhappiness (or making-unhappy), on the other.

In order to avoid misunderstanding here, too, it is important to recognize that I do not see unhappiness, in Kant's thinking, as a mere *effect* of punishment. If this were my view, then a critic would be right to claim that my use of the expression 'law of unhappiness' refers to a law that, unlike Kant's 'law of punishment,' commands something *other than*, or *in addition to*, punishment—whether the latter be conceived of in limited political, or more broadly eschatological-ethical terms. Someone might protest that 'unhappiness' refers to a sensible state of affairs while 'punishment' refers to a particular concept—to the extent, that is, that for Kant 'punishment' (*Straf*) names a kind of logical equivalence between two modes of unilateral limitation of 'external' freedom, one unjustified (crime), the other justified (the sovereign's punishing action).

This latter claim stands, certainly: punishment can be regarded in these strictly formal terms, prescinding from its content (the specific making-unhappy of particular human beings). But nevertheless, according to Kant's own definition, the punishment of a human being may also be regarded as identical to her forfeiture of happiness (in whatever degree, up to and including the total occlusion of an agent's happiness by way of her being put to death). Or, better, punishment consists, from the point of view of its *agent*, in action that aims at occluding the punished subject's happiness (or the opportunity to be happy, or to pursue happiness) while, from the point of view of its object (the one that suffers it), punishment is constituted by this forfeiture. There is a certain formalism in the background, here, in that Kant thinks that the calculus by which we judge what is deserved by law-breakers is as *a priori*

as that which determines 'right' in general—the right to property, to the truth, to freedom from unwarranted external constraint. That Kant regards punishment as a forfeiture of happiness shows, however, that he has (among his several other modes of thinking about this matter) an implicit account of happiness according to which the latter is constituted by action whose execution depends upon the absence of such constraints.

I might have designated the strange imperative that prescribes the relevant action here (or, more radically, the taking-as-an-*end* of the goal of such action), which is binding on 'God' and, too, on us in the 'virtual' and 'deferred' format that I will describe (as approbation and as the will that-it-be-so), 'the law of the-making-unhappy-of-immoral-agents,' where the action that 'makes unhappy' is simply what is meant by 'punishment' in an eschatological context. But this strikes me as awkward. Instead, I call this imperative 'the law of unhappiness,' but mean by 'unhappiness' something that *subsumes* the achievement that is secured, too, by punishment (although it might just *happen to be the case*, in the course of things, that an agent is unhappy—without any agent's having to intervene to bring this about). Obviously, as I said earlier (but this key point needs to be emphasized), this 'law' does not command its *subject* (primarily 'God,' secondarily us) to *be unhappy*, nor simply ('*per se*') to make this or that other agent unhappy, but rather to bring it about that a certain, supposedly *a priori*, notional connection (the one that Kant's use of the 'worthiness to be happiness idiom' shows, obliquely, that he thinks holds between immorality and happiness) is actualized. The connection's actualization is, of course, an empirically contingent matter—it is an extrinsic connection, rather than one that just happens as a matter of course in the unfolding of nature, or one that holds by definition when an agent is immoral. But the command that this connection, already present in reason, be 'forged,' in fact, by 'God,' as I will put it, and, at the same time, the command that we will and approve of this forging prospectively, is *practically necessary*, for reasons that I will explain later.

For now, let us note that there are two distinct issues here: the question of *what* this ostensible 'law' commands and the question *whether* this principle really is authentically law-like in Kant's sense, that is, categorical. The claim that the law is categorical, taken together with the fact that it is *framed* by a 'hypothetical' reference to

some established fact about its object (*If A* is guilty of *X, then* she absolutely *must* suffer *Y*), implies the claim that the notions of being-guilty-of-*X* and suffering-*Y* are bound together *a priori*. This belonging-together, however, cannot be conceived in physical or analytic terms. Instead the connection, as expressed here by the inexorable 'must,' is a normative relationship, one that *ought to be* actualized, even if it is not—which is what I mean by saying, as I will do later, that 'deserves to suffer *Y*' subsumes 'ought to suffer *Y*' and also 'is such that someone-or-other ought to bring it about that *A* suffers *Y*.'

Certainly, a will that aims at some agent's unhappiness, *just as such*, cannot be a will whose willing has the formal property of being a will constrained to what it wills by a categorical law—anymore than a will that aims at happiness, just as such, would be. In both cases—whether with a view to unhappiness or happiness—the 'matter' of the end, which would be contingent, particular, or empirical would be determinative and not, after all, the sheer universality and necessity of the form of willing in question. Each agent's happiness is a distinct, contingent, particular matter; so, too, each one's unhappiness. But if the action of *making-unhappy* aims at an end whose universal *form* is in the foreground, then the maxim of such action may at the same time be regarded as a law. I am *not* saying that the (divine) action that (*ex hypothesi*) consists in the 'making-unhappy' of human agents really *does* qualify as action that could conform to a categorical imperative (given that the objects of such action were immoral). In fact, I do not think that the eschatological 'making-unhappy' of immoral agents can or would conform to this requirement. But neither, I suggest, does the sovereign's punishing action in its political context do so

I propose only that Kant *holds*, implicitly, that the same qualification applies in respect of the notional relationship between the concepts of immorality and unhappiness as holds—*just in case* (*only* in case) the 'law of punishment' really *is* a categorical imperative' (which Kant claims, but not does not show)—between the concepts of the murderous will and the concept of death-for-murder. The end that the law of unhappiness practically necessitates is that the happiness of immoral agents be occluded, eschatologically, *whatever each such agent's unhappiness actually happens to consist in*. The law of unhappiness does not look to the agent's parochial idea of her own unhappiness in the determination of its aim, but rather to her will and its

misuse of freedom—and to the sheer, supposedly *a priori* norm that precludes such a will's being allowed to have its (contingent, parochial) 'way.'

The objection that I am anticipating and addressing here would be based, then, on the idea that the law of unhappiness—if there were such a thing—would have *unhappiness* precisely '*per se*' in its sights, rather than the effecting, in some agent's experience, of the *notional* relationship that is supposed to hold *a priori* between the concepts that her immoral deeds and her actual unhappiness instantiate (or would instantiate). But I am not suggesting that this is Kant's view. Instead, I will claim that this 'law' is *blind* to what it is exactly that the unhappiness of this or that agent actually consists in—and that this is because the latter's constitution is empirical—and looks only to the will's forfeiture of *whatever* happiness is possible or desirable to it, given the agent's empirical constitution, that is, where happiness is given a most general, ideal definition as what the agent *wills* when she wills her ends independently of any consideration of what morality demands that she will. Happiness and unhappiness, here, are *ideals*. The law of unhappiness commands the occlusion of the actualization of the first ideal (happiness) and it commands that the second (unhappiness) be actualized instead.

That the ideal, 'unhappiness,' *ought* unconditionally to be actualized in the experience of immoral agents—so that even 'God' is bound to forego mercy just in case an agent persists in a state of immorality (in a sense of 'immorality' that is untouched by any possible transformation of her will, even its being rendered a good one, that is not effected by *her*)—Kant does not demonstrate. Nor, in my view, can he demonstrate this. But neither does he, nor can he, show that the 'law of *punishment*' really is a categorical imperative rather than a hypothetical one (as I show in chapter 3, Kant sometimes entertains the latter possibility). He is simply aware—and rightly so—that *if* the law of punishment were not a categorical imperative, *then* the practice of putting murderers to death would have to be justified instrumentally.

And this, as Kant recognizes, would render capital punishment a hideous thing. Implicitly, however, Kant sees eternal punishment, or the eschatological making-unhappy of immoral agents (which 'making-unhappy' is regarded as their punishment for having preferred immorality to morality), in the same terms. In Kant's terms, the rationale for claiming that the notional connection between immorality and

unhappiness is no merely 'reactive attitude' (resentment, or a desire for revenge, say), but rather anticipates a form of action and a result that would be absolutely good, has to take a retributivist form. But this is because Kant conceives of eternal punishment on the model of capital punishment. Just as the latter is a *permanent* ejection of its object from the human community, a placing of the executed person outside the context in which she might be drawn back into and reconciled with her community, so too, eternal punishment places its object beyond all help, all assistance, all drawing-back-in.

In other words, Kant rightly sees capital punishment as an undertaking that can *never* be regarded as an invitation to *return* to the community, let alone a mode of rehabilitative assistance or support. Kant is not obviously warranted in taking eschatological 'punishment' in the same terms. But he shows that he does so precisely by giving the distorted, highly individualistic account of agency (and of grace and original sin) that he gives in the *Religion*. He would not *need* to give such an account, I claim, save for the fact that (i) he regards the connection between the notions of unhappiness and immorality by way of an analogy with the connection that he takes to hold between the notions of murder and death-for-murder (in short, he takes action that would aim at the occlusion of an immoral agent's happiness to be an eschatologically inflection of the action by which actual sovereigns occlude the happiness of actual murderers, on the scaffold); and (ii) he (rightly) regards capital punishment as a practice the objective goodness of whose end (the murderer's death) can *only* be asserted in retributivist terms.

In other words, the *Religion*'s treatment of '*Gesinnung*' expresses Kant's awareness that the practice of (retributive) punishment, the approbation of its (justified) instances, and the (prospective) approbation of the unhappiness of immoral agents (also regarded as retribution), is called into question by the counter-claim that, while we may in some sense ascribe her deeds to her as their source, we are not entitled to impute the punishable agent's deeds to her *all the way down*, to the point where our imputation subsumes the very character relative to which her deeds are deemed good or bad ones in the first place. Kant does not demonstrate that we are not entitled to this counter-claim; he simply forestalls it. He does not demonstrate that the practice of (retributive) punishment to which he is antecedently committed is unproblematic;

he forestalls objections to that too. He does this with great perspicacity, offering in the *Religion* a description of human agency that includes, from the outset, an account of the human being's susceptibility to the imputation of deeds that is retro-fitted, as it were, to the ambiguously political-eschatological practice of punishment *qua* retribution.

For Kant, the practice of punishment—in a mode that expresses the idea that the punishable agent is the author of her punishable deed(s) *all the way down*—is too important to give up. When it comes to the 'radical' imputation of deeds, we can answer the question why this is so important for Kant: this imputation's importance is a function of the importance that Kant places on punishment. With respect to the latter, however, this thesis cannot answer the question why punishment is so important to Kant. I only establish that it is, show how Kant's commitment to the practice of punishment is expressed, and argue that this has significant consequences.

In this thesis, then, I argue that Kant's regular glossing of 'morality' as 'worthiness to be happy' is the place in his thinking where these antecedent commitments to (retributive) punishment and (retribution-sensitive) imputation are sheltered and passed along throughout his work. The theory of radical evil is the place where the practical-theoretical denizens of that *topos* come most nearly into the light. Kant's gloss signifies a blind spot in his thinking. The *Religion* discloses that blind spot's contents. Neither that work nor any other, however, thematizes this obscurity and come to grips with it.

I do not claim, of course, that the practical commitments that I have described are decisive for all of Kant's thinking about freedom and moral accountability. The *Religion*'s description of human agency is foreshadowed in various ways throughout the whole body of his writings about morality, but the interests that motivate it do not determine everything that one finds there. Kant's thinking about morality exhibits other tendencies as well and proceeds, accordingly, along other avenues that, just as such, would never have encountered the theory of radical evil. This thesis is no exercise in unlimited suspicion. And yet Nietzsche is onto something when he claims to detect a 'certain odor of blood and torture' in Kant's moral theory.[22] This 'odor' is

[22] Friedrich Nietzsche, *On the Genealogy of Morality: A Polemic*, trans., Alan J. Swensen and Maudemarie Clark (Indianapolis, IN: Hackett Publishing Co., 1998), 41. For a similar, more recent

by no means omnipresent, but it is there—and it signals a problem. The commitments with which this thesis is concerned—constraints that do suggest a kind of 'bloodiness'—entail a particularly strong tendency in Kant's thinking and this explains a particular progression of theoretical moves. This introduction adverts to the last of these: the *Religion*'s treatment of '*Gesinnung*' in the context of that work's theory of radical evil. Again, however, the full story of the operation of these constraints and a detailed account of the moves by which Kant's thinking progresses towards that theory, lie outside the scope of this project.

For now, I claim only that the tendency of Kant's thinking, to the extent that these antecedent commitments condition it, is always already a leaning that inclines—or will have inclined—towards something like the theory of radical evil. The practice of punishment—and of capital punishment in particular—plays the role of a 'strange attractor,' so to speak, which, given the initial state of Kant's critical period thinking about freedom and moral accountability, explains some aspects of the latter's development and ultimate disposition. I do not claim that Kant's thinking about the imputable '*Gesinnung*,' or any of the other moves along the trajectory leading to it, can be straightforwardly deduced from this initial state, or from any of these other moves, in advance. Nor do I claim that Kant's commitments somehow *cause* him inevitably to end up where he does. I propose, rather, that Kant is attracted to what will turn out to have been his account of an imputable '*Gesinnung*,' because assent to something like this notion is always already implicit in the practices to which he is committed, the practices to which his glossing of 'morality' as 'worthiness to be happy' draw attention. In a sense, Kant comes to his notion of the human being's radically imputable, originary disposition towards good or evil and brings it to expression because, given certain of his practical commitments and intellectual discomfitures, something along these lines had been where he was heading, all along.[23]

Now, while Kant's regular glossing of 'morality' as 'worthiness to be happy' signifies the practical commitments that I have described, it does not do so directly.

---

critique see Annette C. Baier, 'Moralism and Cruelty: Reflections on Hume and Kant,' in *Moral Prejudices* (Cambridge, MA: Harvard University Press, 1994).

[23] In broad strokes at least, the distinction that I make here, between a causal account, on the one hand, and an account that demonstrates the attractiveness of a particular idea (to someone in particular, given his or her practical commitments and a certain strain of intellectual uneasiness), even as that idea is still in the process of formation, is inspired by Charles Taylor (see Charles Taylor, *Sources of the Self: The Making of the Modern Identity* (Cambridge: Cambridge University Press, 1989), 202-3).

Rather, *prima facie*, it expresses a particular thought about the relationship between morality and happiness. Kant's gloss expresses the thought that morality and happiness are states of the human agent's practical and empirical affairs that are related to one another; and, albeit more obscurely, it indicates the nature of this relation. In this thesis, I will remedy this obscurity. I will show that Kant's gloss expresses the complex thought that *unhappiness* and *immorality*, in particular, are extrinsically and normatively related,[24] that this normativity is a matter of universal practical necessity,[25] and that the claim that they are related in this way (i.e., as concepts whose *a priori* combination *ought* to be realized in fact) is an instance (along with all categorical imperatives, including 'the law of punishment' to which, in its eschatological extension, the actualization of this notional combination would conform) of what Kant will ultimately characterize as 'a priori synthetic practical proposition[s].'[26] The practices to which Kant's gloss shows him to be committed, the practice of punishment (understood in an implicitly retributivist manner) and the practice of imputation (understood with reference to punishment, so construed), are practices that put into effect what this norm prescribes.

The somewhat convoluted thought that Kant's gloss expresses, however, is not a conclusion that he reaches in the course of his many decades of thinking about the *primary* question of ethics, 'What ought I to do?'.[27] Rather, the thought expressed here is both anterior to Kant's 'critical' moral theory, in the sense that it represents his settled view well in advance of the latter's development, and exterior to it, in the sense that none of that theory's claims are deduced or even deducible from it. I ad-

[24] By saying that these notions are 'normatively related' I am adverting to Kant's conviction that immoral agents *ought*, actually, to be unhappy and that their unhappiness is not simply assured just given that they are immoral (as with the happiness of moral agents, the unhappiness of immoral ones would have to be connected with their being-immoral by way of an 'extra step' [cf. Paul Guyer, *Kant on Freedom, Law, and Happiness* (Cambridge: Cambridge University Press, 2000), 118]).

[25] This is the necessity that Kant ascribes to the eschatologically situated practice of divine punishment when, in *The Metaphysics of Morals*, he avers that 'it is from the necessity of punishment that the inference to a future life is drawn' (*MdS* 6: 490 n. [232 n.]). I discuss this remarkable claim, which I have set as the epigraph to this thesis, in chapter 3.

[26] *Gr* 4: 420 (30). Kant is speaking here, of course, of the 'categorical imperative or law of morality.' See also ibid., 4: 440, 454.

[27] In what follows, I will refer to the 'science,' 'doctrine,' 'inquiry,' 'interest,' etc. that pertains to this and cognate questions as 'morals' or 'morality' in Kant's 'primary,' or 'forward-looking' sense. I will refer to the 'science,' 'doctrine,' etc. that pertains to the question 'How are morality and happiness related?' and other, cognate questions, as 'morals' or 'morality' in Kant's 'secondary,' or 'backward-looking' sense. I take this distinction up again, briefly, in this thesis' conclusion. Sustained discussion of it, however, lies without the limits of this project.

vert to this anteriority and exteriority by characterizing Kant's regular glossing of 'morality' as 'worthiness to be happy' as a *habit*—as both a habitual mode of expression and as a habitual mode of thinking. It is a habitual mode of thinking, moreover, that *remains* simply habitual, for Kant. Contrast this situation with Kant's 'transcendental deduction,' or *a priori* justification, of our 'habit' (so construed by Hume) of subsuming spatio-temporal events, *qua* effects, under the concept of causality (as, too, for the whole class of cognate cognitive 'habits'). Kant's habitual thought that immorality and unhappiness really *are* related to one another in the way that he (or whoever) *takes* them to be, remains a mere habit, without a 'deduction' of its own—indeed, without ever being thematized like this at all. The commitments that the habit expresses are simply too deeply rooted and too important and too questionable—and, one surmises, too much under threat—to be brought fully into view.

Here, then, is why Kant's habit of glossing 'morality' as 'worthiness to be happy' is so remarkable: because of it, Kant's writing is strewn with oblique references to fundamental thoughts and commitments that this same writing fails ever to really address. Given the significance that I claim for it, it is also remarkable that this habit has not attracted the direct attention of more readers. In the rare instances where it is taken up for discussion, the attention that it receives does not, in any case, penetrate to the bedrock that this thesis uncovers.[28]

My argument unfolds in the course of four chapters, as follows.

Chapter 1 has four main aims. First I establish the originary presence, pervasiveness, and longevity of Kant's habit of glossing 'morality' as 'worthiness to be happy.' I show that the habit is in effect early on, that it is pervasive throughout Kant's critical period, and that it survives the latter intact.

Second, I discuss the expression's logical structure. I characterize 'worthiness to be happy' as a particular kind of predicate: a three-place relation. In this connection, I point out that the expression's main effect is not to *identify* morality with worthiness to be happy (as it sometimes seems to do), but rather to represent the moral agent's morality, on the one hand, and her (mainly prospective) happiness, on the

[28] Other commentator's ranging from Nietzsche to Bernard Williams, have noticed, certainly, that Kantian 'morality' is geared towards 'backward-looking' questions of blame and praise, punishment and reward, but they do not disentangle the two main threads in Kant's thinking as I do here, or isolate Kant's habit of glossing 'morality' as 'worthiness to be happy' for special treatment.

other, as distinct states of her affairs that stand in a particular *relationship* to one another. I characterize this as a 'conditioned-by' relation in which morality relates to happiness as one of its necessary conditions. I argue that the idiom's typical form tends to draw Kant's readers' attention away from the fact that the gloss *adds something* to 'morality' and that it does so without Kant's justifying the implication that what is added is somehow internal to the latter, or an immediately obvious corollary of it.

Third, I distinguish the *relata* whose connection the idiom represents (the agent, her happiness, her morality) and show that the variety of ways in which Kant represents 'morality,' in the latter's primary, 'forward-looking' sense,[29] are independent of the idiom's representation of the relationship between that (i.e., morality, irrespective of what *that* turns out to be) and happiness. In other words, I argue, while Kant's thinking about morality varies and changes, his thinking about the relationship between morality and happiness, *to the extent* that this is encapsulated in his regular glossing of 'morality' as 'worthiness to be happy,' remains the same. In other words, I suggest, the expression, 'worthiness to be happy,' always has more or less the same sense.

Fourth, I point to the paucity of commentary on Kant's use of the 'worthiness to be happy' idiom, problematize this lack, and discuss its significance. I also acknowledge that there are, nevertheless, a handful of readers who have addressed this topic directly. I discuss their explanations of Kant's use of the concept in question and distinguish my approach to it in from theirs.

In chapter 2, I present what I take to be Kant's answer to the question, 'How are morality and happiness related?' as this answer is encapsulated, specifically, in his gloss. This, I argue, is the gloss's most proximate, theoretical (as opposed to practical) sense. The main upshot of this chapter's discussion is my claim that, for Kant, immoral agents are just as capable of happiness as moral ones, but that they *ought not to* be happy—that, to the contrary, there is a special sense in which Kant thinks that immoral agents *ought to be* unhappy.

This chapter executes three main tasks. First, I show that Kant's gloss represents morality and happiness as states of an agent's affairs that are *extrinsically* related.

[29] See note 27 above.

Next, I specify this claim more closely by arguing that Kant's gloss represents happiness and morality as states of affairs that are not only extrinsically, but also *necessarily* and *normatively* related. I concede that some of Kant's thinking about happiness, in particular, entails that the latter is intrinsically related to morality, or that happiness is constituted in such a way that morality is internal to it, so that the happiness of immoral agents is precluded from the outset. I also concede that, particularly in the second *Critique*, Kant expresses the view that, given certain of our aims as practical-rational beings, morality may be regarded as a *cause* of happiness—or as able, at least under some ideal set of circumstances, to bring happiness into existence. Certainly, on Kant's view of causality, this way of relating morality to happiness entails that the connection between them is a necessary one.

I argue, however, that while Kant certainly entertains these views, neither of these represents the relationship between morality and happiness in the way that his gloss does. Neither of these is the representation to which his gloss refers. Indeed, admittedly, Kant associates a variety of representations with his concept of worthiness to be happy—but not all of them fit there. The causal account comes closest, I argue, relating the two elements, happiness and morality, extrinsically and necessarily. But it does not capture the normativity that is implicit in the notion of 'worthiness to be happy.' And this is of the essence. The necessity in Kant's causal account of the relationship between morality and happiness makes the ultimately perfect disposition of these elements physically *inexorable* (i.e., given the posit of God's authorship of the laws of nature).[30] But Kant's gloss directs us to a way of conceiving of the relationship between morality and happiness where the relevant necessity is not physical (not even in this peculiar, theologically inflected sense), but rather *normative*, and where the ultimately perfect disposition would have to be *forged* by being put, as it were, into practice—by a third party for whom this task was merely *practically* necessary. In short, Kant's gloss expresses the idea that immoral agents are perfectly capable of happiness (my extrinsicness thesis), but that they *ought to be*

[30] By using the phrase 'physically inexorable,' here, I am signaling that Kant takes everything that happens in nature to be governed by universally binding laws. If there is a natural law to the effect that $P \rightarrow Q$, then given P, Q. This is all that I mean when I say that, for Kant, the connection between P and Q, in this case, would be 'physically' inexorable. That is, given the law that (physically) necessitates Q's obtaining, Q cannot fail to obtain. I call this inexorability 'physical' in order to distinguish it from both practical (hypothetical or categorical) and analytic modes of necessitation.

unhappy (my normativity claim) and, indeed, that they unconditionally *must* be unhappy (my necessity thesis).

Third, I point out that the *necessity* to which Kant's gloss adverts does not pertain to the relationship between morality and happiness. Certainly, as per his gloss, Kant takes morality and happiness to be normatively related, but he does not take it to be the case that morality *necessitates* happiness. Kant does not hold that moral agents positively ought to be happy.

Obviously, I am not saying that Kant holds that moral agents *ought not* to be happy. In order to clarify what I mean and, at the same time, to clear up a particularly widespread bit of confusion in the commentary, I treat the practical necessity implied by Kant's gloss as a particular kind of deservingness. Indeed, Kant's commentators regularly take '*is worthy to be happy*' and '*deserves to be happy*' (or equivalent constructions) to be synonymous expressions. To a certain extent, I argue, they are onto something: Kant's gloss does have reference to a particular notion of desert. I argue, however, that in spite of some rare instances in which Kant appears to proceed otherwise, whenever he deploys his notion of 'worthiness to be happy,' he refers to an extrinsic, normative, necessary relation—in short, an 'ought'—that holds, not between morality and happiness, but between *unhappiness* and *immorality*.

In chapter 3 I explicate this '*ought*' more fully and demonstrate its retributivist connections. In other words, I demonstrate that a form of retributivism is present (mostly in abeyance) whenever we encounter Kant's gloss. To this end, chapter 3 has three main aims. First, I explore Kant's thinking about punishment and affirm that when it comes to the latter's specification and justification Kant is a kind of retributivist. My defense of this claim, here, follows majority consensus, but is qualified in a number of ways.

Second, I show that, especially in his late treatment of the topic in *The Metaphysics of Morals*, Kant's conceptions of the legal and the ethical encounter one another in the practice and justification of capital punishment. Indeed, I characterize Kant's 'scaffold' as a liminal *topos* in which his thinking about law and politics punctures and extends deep into his thinking about ethics and eschatology. I argue that the unconditional, immediate necessity that Kant ascribes to capital punishment

in cases of murder is key to understanding the retributivist tendencies of his thinking about politically situated punishment and eschatological unhappiness more generally.

Third, I argue that the 'ought' that arises from Kant's implicit conviction that immoral agents deserve to be unhappy (the 'ought' implicit in his gloss) may be expressed in the form of a categorical law: 'It is practically necessary that all immoral agents be unhappy.' I refer to this as Kant's 'law of unhappiness' and argue that it is the ethical and ultimately eschatological expression of his political 'law of punishment.'[31] Kant takes 'the law of punishment,' 'Punish all (and only) criminals (in proportion to their crimes)!' to be a categorical imperative. I argue that 'the law of unhappiness' ('Let all immoral agents be unhappy!) is a categorical imperative as well. By so characterizing it, however, I suggest that this law and so, too, Kant's gloss have 'action-guiding' significance.[32]

In chapter 4 I discuss the 'action-guiding' significance of the law of unhappiness. This chapter has four main tasks. First, I open with a brief discussion of the practical significance of Kant's law of punishment. I point out that Kant's commitment to the thesis that all murderers ought to die (just given that they are murderers) is a fundamentally practical one. My discussion sets up a framework for the chapter's main inquiry.

Second, I demonstrate that Kant's commitment to the thesis that immoral agents ought to be unhappy is also fundamentally practical. I argue that, by implying that there is such a law (the categorical 'law of unhappiness'), Kant's gloss suggests indirectly, too, that someone or other is bound by it and that there is (or are) a practice (or practices) that would count as enactments of that law. In other words, I argue, the 'action-guiding' significance of Kant's gloss is, at the same time, the action-guiding significance of the law of unhappiness to which it adverts. And whenever Kant

[31] Note, however, that the law of punishment states that 'all *and only* malefactors ought to be punished.' The punishment of law-abiding citizens is problematic from the point of view of the order within which the law of punishment is binding. Their punishment would be positively unjust (not 'right') and hence illegal. The unhappiness of moral agents is not problematic, however, from the point of view of the order within which the law of unhappiness is binding; their unhappiness offends *kindness*, not justice, but so too does the unhappiness of immoral ones. Kindness may be present, here, but its deliverances do not give expression to any law (see chapter 4).

[32] I borrow the epithet from Smith who applies, it in a related vein, to Kant's use of the 'worthiness to be happy idiom' (see Steven G. Smith, 'Worthiness to Be Happy and Kant's Concept of the Highest Good,' *Kant-Studien* 75, no. 2 (1984)). See also Thomas E. Hill, Jr., 'Wrongdoing, Desert, and Punishment,' in *Human Welfare and Moral Worth: Kantian Perspectives*, ed. Thomas E. Hill, Jr. (Oxford: Clarendon Press, 2002), 324.

glosses 'morality' as 'worthiness to be happy,' I argue, he signals his commitment to a practice (or practices) that execute(s) and conform(s) to this law.

Third, I answer the questions, 'Who is subject to Kant's law of unhappiness?' and 'How is it enacted?' I argue that, for Kant, there are two kinds of agent that this law binds and two contexts in which it does so. These distinct contexts correspond to two distinct ways in which the law of unhappiness is put into practice. Kant's God is subject to the law of unhappiness. Human beings are too. The first context in which the law of unhappiness is enacted is the mundane context in which human beings are called to live in accordance with the demands of morality in Kant's primary, forward-looking sense and the demands of the political law of punishment. This mundane context has both an external aspect, in so far as it the spatio-temporal context in which the deeds of moral or immoral agents are enacted, and an internal one, which is the individual human being's conscience.

With respect to the second context in which the law of unhappiness is enacted my task is to make explicit a point of view that is generally implicit in Kant's thinking. This second context is the eschatological scenario in which God is called upon to omit the mercy to which he is universally inclined (as benevolent) and to act in strict accordance with the law of unhappiness. Far from being a law of God's very nature (as Kant takes the primary, forward-looking moral law to be), the law of unhappiness is an imperative that puts pressure[33] on that nature, to the extent that Kant's God is a kind being. In respect of this law, I construe Kant's God not as 'holy' (as Kant takes him to be vis-à-vis the primary moral law) but 'virtuous.' In order to 'obey' the law here, as I make explicit, he must overcome his powerful, ineliminable desire to make and see his creatures happy. Relative to the happiness of immoral agents, God's desire to show mercy has the form of a constant temptation to disobey.

Fourth, I argue that the mundane and the eschatological contexts coincide to the extent that, under the law of punishment (but not under the primary moral law), we, together with God, are members in common of a single community, a community for

[33] I owe this very evocative mode of expression to Ameriks, who uses it in a distinct, but related connection. See Karl Ameriks, *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy*, Modern European Philosophy (Cambridge: Cambridge University Press, 2000), 137.

which the law of unhappiness is uniquely constitutive. The upshot of this constitution, I argue, is that, like the merely human *polis*, this larger community is a mercy-free zone—here, however, with a view to the broad scope of immorality (regarded, in a sense, as crime against morality), as such, rather than the narrow scope, merely, of external crimes against 'right,' or 'public law.' I characterize this coinciding in terms of a certain porousness of the ethical, by which the latter is susceptible to incursions of an antecedently, politically constituted notion of the Good, or a conflation of good and right.

CHAPTER ONE

Morality, id est, Worthiness to be Happy

> [M]orals is the doctrine…of how we are to become worthy of happiness.[34]

Introduction

In this chapter I execute four main tasks. First, I establish the originary presence, pervasiveness, and longevity of Kant's habit of glossing 'morality' as 'worthiness to be happy.' Second, I discuss the expression's logical structure. I characterize 'worthiness to be happy' as a particular kind of predicate: a three-place relation. Third, I distinguish the *relata* whose connection the idiom represents and argue that the variety of ways in which Kant represents 'morality,' in the latter's primary, 'forward-looking' sense are independent of the idiom's representation of the relationship between that (i.e., morality, irrespective of what *that* turns out to be) and happiness. Fourth, I observe that there is a surprising lack of commentary in this area. I acknowledge, however, that a handful of Kant's readers have addressed this topic directly. I discuss their explanations of his use of the concept of worthiness to be happy and distinguish my approach to it in from theirs.

[34] *KpV* 5: 130 (108).

Kant's habit

In this section I show that Kant's habit of glossing 'morality' as 'worthiness to be happy' is in effect early on, that it is pervasive throughout his critical period, and that it survives the latter intact.

In a wide variety of published and unpublished contexts, Kant deploys variants of the idiomatic formula, '*Würdigkeit glücklich zu sein*,' on over three hundred occasions.[35] It first appears in some of his earliest handwritten notes to Baumgarten's *Initia Philosophiae Practicae*[36] (hereafter *IPP*) and shows up frequently in other *Reflexionen* and marginalia, up to and including instances in the incomplete *Opus Postumum* (1804). It is also pervasive throughout Kant's published authorship. In order to demonstrate my claim that the idiom appears early in Kant's work and then frequently, again, throughout his career, I will briefly review a number of typical cases. After that, I will offer a more sustained discussion of some key instances that support my practice of characterizing Kant's habitual usage as a '*gloss*' on 'morality.'

In an early note to *IPP*, likely from the mid to late 1760s, Kant equates 'worthiness to be happy' with 'morality [*Sittlichkeit*],' asserting that the latter 'lies in conduct,' and affirms that '[a]ll worthiness lies in the use of freedom.'[37] An entry in Kant's *Stammbuchblätter* for 1772-7 shows that in the fall of 1772 he instructed a young Ernst Theodor Langer to defend the thesis that '[t]he human being's first concern [*Sorge*] is: not how he becomes happy, but rather how he becomes worthy of happiness.'[38] The first *Critique* (1781) provides a well-know example when Kant

---

[35] I will not cite all of these here; I cite many of them, however, in the following. In any case, Kant's 'use of the language of desert to characterize the relation between virtue and Glückseligkeit' is hardly 'occasional,' as Engstrom claims (S. Engstrom, 'Happiness and the Highest Good in Aristotle and Kant,' in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, ed. S. P. Engstrom and J. Whiting (Cambridge: Cambridge, 1994), 128; I discuss the relationship between 'worthiness' and 'desert' in chapter 2). Nor, as Henrich seems to imply, is 'the thesis…that morality is the worthiness to be happy' an especially regular theme in Kant's 'literary remains' only (see Dieter Henrich, 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason,' in *The Unity of Reason: Essays on Kant's Philosophy*, ed. R. Velkley (London: Harvard University Press, 1994), 78). Hill's reference, in this connection, to 'remarks, sprinkled throughout [Kant's] works' (Hill, 'Wrongdoing, Desert, and Punishment', 326) comes closer to the truth, but strikes me as an understatement.

[36] Alexander Gottlieb Baumgarten, *Initia Philosophiae Practicae Primae* (Halle: 1760). Baumgarten's text is included at *KGS* 19: 7-91.

[37] *R* 6611 19: 110 (424). Kant says substantially the same, much later, in *KpV* 5: 130 (108-9).

[38] *Stammbuchblätter* (1772-7) 12: 416. The entry is dated 19 October, 1772. E. T. Langer (1743-1820) was a friend of Lessing and Goethe and librarian to Charles William Ferdinand, Duke of Brunswick.

answers the primary question of ethics, 'What ought I to do?' by asserting that 'you' ought to '[d]o that through which you will become worthy to be happy.'[39] About a decade later (ca. 1792-4), in a text that rehearses the 'moral catechism' of his later *Metaphysics of Morals* (1797), Kant describes worthiness to be happy as the 'first of all [the human being's] wishes.'[40] In the *Religion* (1793) he argues that, given 'our nature…as a substance endowed with reason and freedom,' not happiness, but only worthiness to be happy can be 'the object of our maxims unconditionally.'[41] In a related vein, Kant's *Vorarbeiten* (preparatory notes) for the *Metaphysic of Morals* refer to worthiness to be happy as the human being's 'ultimate purpose [*Endzweck*].'[42] The remarkable longevity of the notion in Kant's thinking is evident, finally, in the somewhat obscure interpolation of the phrase, '[t]o be worthy or unworthy of happiness,' into a draft passage of the *Opus Postumum* (1804).[43]

From its earliest appearance until its last, Kant's use of this idiom correlates an agent's 'morality,' on the one hand, with her 'worthiness to be happy,' on the other. 'Worthiness to be happy' serves repeatedly as a kind of gloss on 'morality.' It serves, that is, as a kind of regular elaboration of what Kant means by the latter term—at least some of the time. Sometimes Kant seems to identify or equate worthiness to be happy with morality, using 'worthiness to be happy' as an apparent synonym for expressions deploying not only 'morality' ('*Sittlichkeit*' and cognate terminology), but closely allied terms and locutions such as 'virtue' (*Tugend*),[44] 'right conduct,'[45] 'moral perfection,'[46] 'the agreement of all our maxims with the moral law,'[47] 'inner goodness,'[48] the 'perfection of freedom,'[49] and so forth.[50] Sometimes Kant makes

---

He occupied his position at the Herzog August Bibliothek in Wolfenbüttel (other famous staff included Leibniz and J. Burckhardt) from 1781, when he took over from Lessing, until 1820.

[39] *KrV* A808-9/B836-7.

[40] *R* 7315 19: 312.

[41] *Rel* 6: 46 n. (91 n.).

[42] *Vorarbeiten-MdS* 23: 402-3.

[43] *OP* 22: 129 (208).

[44] See, for example, *KrV* A315/B372; *KpV* 5: 110 (92); *ÜdG* 8: 283 (285); *R* 7196 19: 270 (461); *LEM*$_2$ 29: 599-600 (227).

[45] See, for example, *Rel* 6: 161 (182-3).

[46] *Vorlesungen-Religionslehre* 28: 1100 (429).

[47] *Rel* 6: 46 (42).

[48] *R* 6890 19: 195 (446).

[49] *R* 7197 19: 270 (461).

[50] Again, *R* 6611 19: 110 (424), a note to from the mid-late 1760s, provides a good early example. Other relatively early instances (ca. mid-late 70s) include *R* 5100 18: 87 (212) and *R* 6892 19: 195 (447). Published work and lecture notes from Kant's critical period offer a number of cognate instanc-

definitive claims about worthiness to be happy that are indistinguishable from the kinds of things that he says about morality (virtue, etc.) in other contexts. For example, Kant takes the command to do what is morally required—and so, more generally, to be moral—to be equivalent to the command to be (or to become) worthy to be happy (or to omit becoming unworthy to be happy). In the second *Critique*, for example, Kant asserts that 'morals is the doctrine…of how we are to become worthy of happiness.'[51] This echoes a key moment in the Canon section of the first *Critique* when he contrasts the pragmatic 'rule of prudence,' which '*advises* us what to do if we want to partake of happiness' with the moral law, which he defines as the law that 'commands how we should behave in order even to become worthy of happiness [*um nur der Glückseligkeit würdig zu werden*].'[52] For Kant, the association between the notions of morality and worthiness to be happy is so intimate that, mere pages later, he is able to frame his definitive reply to the fundamental question of ethics, 'What must I do?' without deploying 'moral,' or 'morality,' or any cognate terms at all: '[d]o that through which you will become worthy to be happy,' he writes.[53] And such glosses are no early-critical-period anomaly. In his much later *Metaphysics of Morals* (1797), Kant says that an agent's 'worthiness to be happy is *identical* [*ein und dasselbe*] with his observance of duty.'[54] Thus, too, Kant does not hesitate to identify the prospect of being worthy to be happy as a properly moral incentive. Indeed, as he says in the first *Critique*, the moral law is 'that [law] which is such that it *has no other motive* than the *worthiness to be happy*.'[55]

Repeatedly, then, Kant slides—and invites his readers to slide—from 'morality' and 'moral' (and these expressions' various cognates) to 'worthiness to be happy'

---

es. See, for example, *KrV* A810/B838; *KpV* 5: 130 (108-9); *KdU* 5: 450 (315); *ÜdG* 8: 278, 281 (281, 283); *MdS* 6: 481-2 (224-5); *Anthro* 7: 326 (231); *LEC* 27: 247-8, 358 (44, 136); LEM_2 29: 623-4 (242); *DWL* 24: 751 (485); *LEV* 27: 664-5, 717, 726 (399, 440, 447); LMK_2 28: 767 (407); *OP* 21: 195. See also *R* 7211 19: 286 (472); *R* 6856 19: 181 (441); *R* 6858 19: 181 (441); *R* 7202 19: 279 (467). And see Guyer, *Kant on Freedom, Law, and Happiness*, 97, 100, 117.

[51] *KpV* 5: 130 (108).

[52] *KrV* A806/B834 (translation modified).

[53] A808-9/B836-7. Cf. Smith, 'Worthiness to Be Happy': 173.

[54] *MdS* 6: 482 (225) (my emphasis). See also *LEV* 27: 718 (441); LML_1 28: 288, 301 (96, 106); *LEC* 27: 249 (46); *LEV* 27: 717 (440); *R* 4461 17: 560 (137); *R* 5100 18: 87 (212); *R* 5477 18: 193 (416); *R* 6910 19: 203 (449); *R* 6133 18: 465 (338-9); *R* 7315 19: 312.

[55] *KrV* A806/B834. For a much earlier statement of this position, likely from the mid-late 1760s, see *R* 6628 19: 117 (427). There is much evidence to suggest that Kant continues to take this approach deep into his the critical period. See, for example, *KrV* A806/B834; *Gr* 4: 450 (55); *LMM* 29: 774 (131); LEM_2 29: 623-4 (242); *R* 6133 18: 465 (338-9); *Vorarbeiten-MdS* 23: 404. Cf., however, *KpV* 5: 159 (131).

and 'worthy to be happy.' He does not draw attention to this move; he simply executes it. And the naturalness, the lack of friction, as it were, suggests that he has always already made this move, somewhere in the background of his thinking, well in advance of the contexts in which we happen to witness it.

This 'slide' from 'morality' to 'worthiness to be happy' may be characterized as a generally implicit, always frictionless '*id est*' (or '*das Heißt*') in Kant's thinking—not merely a gloss on 'morality,' as I will be characterizing it throughout this thesis, but a *habitual* one. For example, in a relatively late (1793) instance that echoes the thesis assigned to Langer nearly a quarter of a century earlier, Kant writes:

> By our *nature* as beings dependent upon circumstances of sensibility, we crave happiness first and unconditionally. Yet by this same nature of ours…as beings endowed with reason and freedom, this happiness is far from being first, nor indeed is it unconditionally an object of our maxims; rather this first object is worthiness to be happy, *i.e.*, the agreement of all our maxims with the moral law.[56]

Here, the explicitly rendered '*id est*' (actually, '*das Heißt*') captures the essence of every instance in which Kant uses the idiom. It is the unmarked, habitual channel along which his thinking slips from 'morality' to 'worthiness to be happy' and back again.[57] Surely 'worthiness to be happy' is not simply *equivalent* to 'the agreement of all our maxims with the moral law.' And yet, at first glance, this seems to be what Kant is saying. He does not claim, as we might expect him to, that the primary 'object of our maxims' is 'the agreement of all our maxims with the moral law' (although he would of course affirm this). He does not simply and straightforwardly claim, either, that given that we are 'beings endowed with reason and freedom,' our most fundamental aim is simply 'to be moral,' or 'to pursue morality,' or 'to undertake moral actions,' or 'to live a moral life' (although he would affirm all of this as well). Nor does Kant treat the expression, 'the agreement of all our maxims with the moral law' (*die Übereinstimmung aller unserer Maximen mit dem moralischen Gesetze*) as a gloss on something like 'our morality,' or 'the absolute moral worth of our actions,' or 'the absolute, unconditional goodness of the subjective ground of our

[56] *Rel* 6: 46 n. (41-2 n.) (Kant's emphasis modified).

[57] For the most part, this '*id est*,' or '*das Heißt*' (also '*das ist*'), is implicit (but see, for example, Kant's 'd[as] i[st]' at *ÜdM* 8: 257 n. [26 n.] and *R* 6610 19: 107 [422-3]). Where it is lacking it easily interpolated—as, for example, when Kant uses parenthetical constructions like 'virtue (as worthiness to be happy)' (*KpV* 5: 110 [92]). Moreover, the gloss is commutative. In the instance cited here, the explicit '*das Heißt*' works in the opposite direction: Kant glosses 'worthiness to be happy' as 'morality,' under the description, 'the agreement of all our maxims with the moral law.'

good deeds' (the *Groundwork*'s 'good will,'[58] say). In short, Kant does not say that as beings endowed with reason and freedom the first object of our maxims is *morality*—and then go on to gloss *that*, by way of his '*das Heißt'*, as 'the agreement of all our maxims with the moral law.' He does not do this, even though, at the time in question, he does of course hold this view. And he does not do this, even though a locution like 'the agreement of all our maxims with the moral law' seems more obviously and directly to explicate and clarify what Kant means by 'morality' than it seems to explicate and clarify what he means by 'worthiness to be happy.'

Instead, Kant uses this occasion, as he does every instance in which he connects morality in this way with worthiness to be happy, to assert, in a remarkably offhand manner, that happiness is a state of the human being's empirical affairs that is related in some sense to 'the agreement of all [her] maxims with the moral law' (etc.), that is, to her morality. His use of the idiom says nothing explicit *about* this relationship (although, as I will show in subsequent chapters, taken together with other considerations, it points us obscurely in a certain direction). It merely affirms that Kant takes there to be one. The immediate directness of the gloss—its frictionless '*id est*'—gives the impression of something rather unremarkable. It is as though Kant were merely glossing 'morality' as 'goodness in all circumstances,' or some such construction. His offhand mode of expression, here, not to mention the fact that he never thematizes and justifies his use of the concept, gives the impression that his repeated interpolations of the expression 'worthiness to be happy' into his discussions of morality, do not interrupt, or import extraneous material into, his discussions of the latter. But these interpolations *do* advert to something extraneous here. Worthiness to be happy is not simply equivalent to morality.

No matter how vast and deep, a catalogue of the numerous instances in which Kant uses this idiom shows little more than that a particularly stable mode of expression appears with some frequency across the various stages of Kant's thinking about morality. Of course, my claim is rather more robust than this. I claim that the stability of the idiom corresponds to the stability of an *idée fixe* that plays a consistent role throughout these stages. And I suggest that the frequency of the idiom's deployment

[58] *Gr* 4: 393 (7).

is a sign of the importance of Kant's commitment to the theoretical and practical terrain of which the idiom is a kind of extremely concise map.

In the next section, I trace the outlines of that map. I defer the task of filling in its details until chapter 2.

'Worthiness to be happy': a three-place relation

As we saw above, Kant sometimes writes as though he takes worthiness to be happy to be equivalent or identical to morality. He can hardly mean this in a straightforward sense, however. On its face, 'moral' (like 'virtuous' and 'good') is a one-place predicate.[59] To qualify an agent, her will, or her deeds as morally good is not to specify, by way of the same epithet, her (or her will's, or her deeds') relation to anything extraneous to her, *qua* moral subject, or extraneous to her will, or to her deeds and their immediate consequences.[60]

Instead, the predicate 'worthy to be happy' refers to a three-place relation in which (1) a finite, rational will (i.e., a practical-rational human subject) is related to (2) the realization of (a maximally integrated complement of) her empirical, non-moral ends (her happiness), by way of (3) a mediating basis or condition for the latter. For Kant, the sole objectively and universally necessary 'worthiness-basis'[61] is an agent's morality, which is here, ultimately, the moral goodness of her will or, equivalently, the goodness of her will's radically basic disposition vis-à-vis the moral law.

Note that, short of my explication of the 'worthiness-basis' (morality) which Kant takes to condition happiness, none of (1), (2), or (3) makes explicit reference to either 'worthiness,' or 'morality.' All that we have, in accordance with this schema,

---

[59] Other one-place predicates include adjectives such as 'blue,' 'ragged,' and 'omnivorous,' while 'taller than,' 'larger than,' and 'the father of' are two-place predicates. *N*-place predicates, where $n \geq 2$, give expression to relations. Some such relations hold between triples of *relata*: thus Calgary *is between* Vancouver *and* Toronto, my mother *travels to* France *by* train, the sunset *deserves* admiration *in view of* its beauty, the murderer *deserves* death *in view of* his crime, and Jones *is worthy of* happiness *in view of* his morality. For a different view see Hill, 'Is a Good Will Overrated?', 50 and cf. G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903); G. E. Moore, *Ethics* (Oxford: Oxford University Press, 1912).

[60] This, I take it, is one implication of Kant's claim that a good will 'is good in itself' and that its '[u]sefulness or fruitfulness can neither add anything to [its] worth nor take anything away from it' (*Gr* 4: 394 [8]). See also Engstrom, 'Happiness and the Highest Good', 112.

[61] This is deliberately reminiscent of the language of desert and, in particular, the notion of a 'desert-basis.' For a basic discussion see Louis P. Pojman and Owen McLeod, eds., *What Do We Deserve?: A Reader on Justice and Desert* (Oxford: Oxford University Press, 1999), 61-2.

is the formal idea of a relationship between something ('happiness') and one of it conditions (whatever this may be): (1) enjoys (2) on condition that (3) holds. Much turns, however, upon how (3) is taken to condition (2).

'Morality,' then, is not simply synonymous with the expression, 'worthiness to be happy.' Rather, 'worthiness to be happy' is a three-place predicate, or three-place relation, in which morality appears as one element, one of three *relata*, along with a subject (the agent) and a second predicate ('happiness'). The expression's main effect is not to identify or equate morality with worthiness to be happy (as it sometimes seems to do[62]), nor to advert to any of the particular ways in which Kant understands 'morality.' Rather, its main effect is to represent the moral agent's morality (*whatever* Kant takes this to be), on the one hand, and her (mainly prospective) happiness, on the other, as distinct states of her affairs that stand in a particular relationship to one another. Abstractly put, on Kant's representation of matters, moral agents stand in a 'worthy-of' relation to their (mainly prospective) happiness.[63] In the abstract merely, too, this is a particular kind of 'conditioned-by' relation: one in which one thing (an agent's morality) relates to another (her happiness) as one of its (in some sense) necessary conditions. Kant's glossing of 'morality' as 'worthiness to be happy' is his favoured way of adverting to this conditioned-by relation. Whenever he deploys the idiom he identifies morality as the 'worthiness-basis' *sine qua non* for happiness. This, I take it, is the force of what sometimes appears syntactically to be a straightforward identification.

The idiom's typical form, however, tends to draw Kant's readers' attention away from the fact that the gloss *adds something* to 'morality.' At times Kant's usage seems to imply—supposing it goes too far to say that morality just *is* in some sense worthiness to be happy—that the referent of 'worthiness to be happy' is nevertheless somehow internal to morality, or an immediately obvious corollary of it. But this impression is never justified.

---

[62] See, for example, Guyer, *Kant on Freedom, Law, and Happiness*, 119.

[63] Of course there are, for Kant, any number of things, apart from happiness, to which agents qualified, too, as moral, can stand in a worthy-of relation. It is not as though 'worthiness' has to be 'worthiness to be happy' merely. Agents (or their actions) can be worthy of worship, praise, emulation, and such things as 'approval' (*R* 6858 19: 181 [441]),'esteem' (*LEC* 27: 357 [135]), 'liking' or 'love' (*LEC* 27: 420-1 [183]), 'friendship' (*LEC* 27: 429 [189]), or 'humanity' (*LEV* 27: 664-5 [399]).

‘Worthiness to be happy’: the *relata*

In this section I briefly discuss Kant’s notions of morality and happiness. First, I discuss morality. Then I discuss happiness. In each case, I restrict my discussion to a minimal, core characterization of these matters, which leaves the two elements disentangled. (I discuss ways in which they might be entangled in chapter 2 in order to show that Kant’s notion of worthiness to be happy does not refer to morality or happiness under any description that entangles them.) I then argue that the sense of Kant’s gloss is stable throughout his career and that its sense is independent of any of the particular answers that he gives to the primary, forward-looking question of ethics, ‘What ought I to do?’

Morality

In the present context, all that we really need to know about morality is that the term ‘morality’ (and cognate expressions) refers to the way that an agent has to be in order to be worthy to be happy. As we shall see in the next chapter, according to the representation inherent in Kant’s gloss, *nothing* connects happiness and morality directly—certainly nothing about *morality*—and only an extraneous ‘forging’ connects immorality (assuredly) to unhappiness. That said, the only thing that we really need to show about Kantian morality, in connection with the present topic (*viz*., Kant’s gloss’s representation of the *relationship* between happiness and morality), is that, in some basic sense, Kant always already regards it as a property that can be ascribed to agents and not merely, say, to actions.

Beyond this, we can concede that Kant’s account of morality develops and changes in various ways in the course of his career and that the extent to which ‘morality’ and ‘absolute goodness,’ along with ‘immorality’ and ‘radical evil,’ become radically affixed to the agent’s absolutely discrete, totally consolidated, autocratic, and also self-legislating will, is a product of Kant’s thinking, not its explicit point of departure. In advance of his critical period—but of course not in advance of the instauration of his habit of glossing ‘morality’ as ‘worthiness to be happy’ which predates the latter by a good margin—Kant’s understanding of morality resembles, in key respects, the thinking of some of his British predecessors and is also marked by the influence by Rousseau.

In the earlier type of view morality is just cooperation amongst human beings, as such, the coordination and hierarchization of their ends, community.[64] The moral law is an empirical discovery and pertains to the rational, but still earthly pursuit of natural ends: as Ward puts it, 'the richest possible development of one's powers, consonant with the greatest possible social harmony.'[65] It is an 'anthroponomy' that organizes structures of empirical psychology and anthropology. This is not yet autonomy, but rather 'the autocracy of freedom,' here identified as the '*principium* of morals,' and conceived in terms of the agent's self-governance 'with regard to…the epigenesis of happiness,'[66] but still in accordance with heteronomous principles.

Of course, Kant's pre-critical notion of morality, which is grounded in nothing more radical than contestable observations about the typical features of the human being, is later superseded. What had been pre-critically conceived of as 'morality' turns out to be a qualified mode of prudence, still heteronomy, a source of merely hypothetical imperatives. What had appeared to be laws turn out to be mere rules, 'universally' applicable to the anthropological 'universe' only, to the changing conditions of 'humanity' alone.

As Kant asks, 'by what right could we bring into unlimited respect, as a universal precept for every rational nature, what is perhaps valid only under the contingent conditions of humanity?'[67] Kant's critical position puts him in conflict, not merely with the earlier, largely British tradition of thinking for which morality has constitutive, if carefully qualified, reference to the emotions and within which freedom is understood with reference to the specificity of human nature (whatever that turns out to be).[68] It also puts him at odds with Rousseau, an important influence just in advance of the critical period,[69] and contemporaries such as Schiller and Herder.[70]

---

[64] See, for example, *R* 7199 19: 272 (462-3). See also *R* 7052 19: 235, 272-3 (458, 462-3); *R* 7200 19: 274 (463). See also D. O'Connor, 'Kant's Conception of Happiness,' *The Journal of Value Inquiry* 16, no. 3 (1982): 193.

[65] Keith Ward, *The Development of Kant's View of Ethics* (Oxford: Blackwell, 1972), 85.

[66] *R* 6867 19: 186 (444).

[67] *Gr* 4: 408 (20-21). In other words, on Kant's ultimate view, morality cannot pertain to what holds true of mankind *alone*. See Henry E. Allison, *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge: Cambridge University Press, 1996), 117; Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960), 82; Tom Sorell, 'Kant's Good Will and Our Good Nature: Second Thoughts About Henson and Herman,' *Kant-Studien* 78, no. 1 (1987): 95-6; Allen Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), 69-70, 76.

[68] See *Gr* 4: 406ff. For a compact overview that situates Kant in relation to various key predecessors both in Britain and on the Continent see David Fate Norton and Manfred Kuehn, 'The Foundations of

None of these influences gets as far as the radically agent-centred approach to which Kant ultimately attains.[71] Kant comes to that point rather later (in his *Religion*, ultimately, of course). And yet, I suggest, this focus is already present in the thinking that is expressed by his gloss. From start to finish, whatever else 'morality' signifies, to be moral is to be worthy to be happy. This entails, at least, that Kant regards morality as a property, not merely of particular actions, motives, maxims, and so forth, but of agents. The main development in this regard is a deeper penetration of the predicate, 'moral,' to the very core of the agent and the absolute consolidation of the latter's identity.

This focusing in of 'morality,' however, is not always obvious. The unevenness and differentiations that characterize Kant's thinking as it develops entail that his ultimate views on morality can be explicated in a number of ways. To be moral is to undertake deeds that give expression to maxims that have a particular form (i.e., that assert the unconditional practical necessity and universality of what they prescribe). To be moral is to be motivated in a particular way, to be possessed of an unqualifiedly 'good will,' or to enjoy a particular mode of integrity in the radical source of one's practical life as a whole, in one's character (also regarded as the good character of one's good will).

In general, Kant predicates 'morality' under the rubric of 'moral worth,' and predicates it of an apparent variety of subjects. He predicates it of particular actions, of the choices that these express, of the maxims that they instantiate and so, in a sense, of 'types' or classes of actions,[72] of the will from which these flow, of the

---

Morality,' in *The Cambridge History of Eighteenth-Century Philosophy*, ed. K. Haakonssen (Cambridge: Cambridge University Press, 2005). Walker is also helpful (Ralph C. S. Walker, 'Achtung in the *Grundlegung*,' in *Grundlegung Zur Metaphysik Der Sitten: Ein Kooperativer Kommentar*, ed. Ottfried Höffe (Frankfurt am Main: Vittorio Klostermann, 1989), 108). Cf. H. H. Schroeder, 'Some Common Misinterpretations of the Kantian Ethics,' *The Philosophical Review* 49, no. (1940): 427.

[69] See, for example, *Bem* 20: 45, 56, 120-1. Velkley offers helpful commentary in Richard L. Velkley, *Freedom and the End of Reason: On the Moral Foundation of Kant's Critical Philosophy* (Chicago: University of Chicago Press, 1989), 62-9.

[70] With respect to the former break see J. A. Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue,' *Canadian Journal of Philosophy* 27, no. 4 (1997): 529; with respect to the latter, Wood, *Kant's Ethical Thought*, 229.

[71] See C. M. Korsgaard, 'Morality as Freedom,' in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 186 n. 21.

[72] See H. B. Acton, *Kant's Moral Philosophy*, New Studies in Ethics (London: Macmillan, 1970), 12; H. J. Curzer, 'From Duty, Moral Worth, Good Will,' *Dialogue* 36, no. 2 (1997): 307; Allen Wood, *Kant's Moral Religion* (Ithaca: Cornell University Press, 1970), 45.

character of that will, and of the manner in which the latter is motivated (i.e., of the class of incentives to which the agent's effective, actually enticing incentives belong). Of course, none of these things is altogether discrete: Kant's understanding of each is bound up in his understanding of the others. The various concepts in play here are interdependent.[73] They form a complex whose various elements correspond to the variety of ways in which the subject of 'morality' or 'moral worth' is specified.

This variety is a matter of purpose and focus. Kant illustrates this interdependence when, in the second *Critique*, he writes that '[w]hat is essential to any *moral worth of actions* is *that the moral law determine the will immediately*.'[74] Later, treating the will's deep, underlying character as a kind of fundamental maxim to act always and only in accordance with, or in contravention of, the moral law, he identifies this '*maxim* [as the ultimate factor] by the goodness of which all *the moral worth of the person* must be assessed.'[75] Similarly, Kant describes a jaded philanthropist who, though 'no longer incited to [beneficence] by any inclination…nevertheless tears himself out of this deadly insensibility and does the [required] action without any inclination, simply from duty.' And, Kant goes on to say, it is 'then [that] *the action first has its genuine moral worth*' and, at the same time, that '[i]t is just then that *the worth of character* comes out, which is moral and incomparably the highest, namely that he is beneficent not from inclination but from duty.'[76]

Again, however, to the extent that 'worthiness to be happy' is a predicate of the agent, and to the extent that morality is thereby construed as a condition that bears on that same agent's happiness, Kant's thinking tends towards an account of 'morality' that focuses, ultimately, on the agent: 'the one that did it.' Actions or maxims are not worthy to be happy, after all; agents are. From this agent-centred point of view, morality and immorality are properties that subsume the overall organization (nature,

[73] Allison, for example, sets forth a closely connected constellation of key concepts, writing that Kant's 'view is that we can be said to have a good will just in case we act from duty alone or, equivalently, just in case our actions possess moral worth' (Henry E. Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 108). See also the less succinct Paul Benson, 'Moral Worth,' *Philosophical Studies* 51, no. 3 (1987): 375-6, 379-80.
[74] *KpV* 5: 71 (62) (Kant's emphasis modified).
[75] *Rel* 6: 30 (78) (my emphasis).
[76] *Gr* 4: 398 (12) (my emphasis). See also *R* 6133 18: 465 (338-9).

character, disposition) of the moral or immoral agent's will, that is, her very self[77] regarded as a consolidated, identical 'thing' from which all of her particular choices and actions flow. The overall character of her choices and actions (simply good, or simply evil[78]) is a function of this deeper fact about her. An agent whose will, character, or motives are 'moral' is an agent whose actions have 'moral worth.' But her actions have 'moral worth' precisely because they are the actions of an agent whose will, character, or motives are good ones.[79]

Even as I come down, firmly, in favour of this ultimately agent-centred reading of Kant, I do not want to trivialize the interpretive challenges that come up on a regular basis, in connection with Kant's discussion of 'moral worth.'[80] I do not want to simply elide the subtle ways in which this concept is entangled with a whole host of other moral qualities: properties such as 'praiseworthiness,'[81] for example, or moral 'rightness,'[82] 'merit,'[83] 'value,'[84] or 'virtue.'[85] Nor do I want to simply overlook the

[77] See Robert Paul Wolff, *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysic of Morals*, Harper Torchbooks (New York: Harper & Row, 1973), 59.
[78] See *Rel* 6: 22-5 (71-3). See also Curzer, 'From Duty': 307.
[79] In general, however, it is unusual for Kant or his commentators to predicate 'moral worth,' specifically, of *agents*. See, however, Guyer, *Kant on Freedom, Law, and Happiness*, 298, 329 and cf. Judith Baker, 'Do One's Motives Have to Be Pure?,' in *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, ed. Richard E. Grandy and Richard Warner (Oxford: Oxford University Press, 1986), 459-60, 463, 465; S. Sverdlik, 'Kant, Nonaccidentalness and the Availability of Moral Worth,' *Journal of Ethics* 5, no. 4 (2001): 296.
[80] As various scholars have observed, Kant's thinking about 'moral worth' is sometimes as difficult to decipher, as it is important for understanding his moral philosophy as a whole. See, for example, Curzer, 'From Duty': 287; H. Jensen, 'Kant on Overdetermination, Indirect Duties, and Moral Worth,' in *Proceedings of the Sixth International Kant Congress*, ed. G. Funke and T. M. Seebohm (Washington, D.C.: University Press of America, 1989), 161; Sverdlik, 'Kant, Nonaccidentalness and the Availability of Moral Worth': 294.
[81] See Norman O. Dahl, 'Obligation and Moral Worth: Reflections on Prichard and Kant,' *Philosophical Studies* 50, no. (1986): 391; Richard G. Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action,' *Philosophical Review* 88, no. 1 (1979): 42; W. E. Schaller, 'Kant on Virtue and Moral Worth,' *Southern Journal of Philosophy* 25, no. 4 (1987): 559.
[82] See Curzer, 'From Duty': 291; Dahl, 'Obligation and Moral Worth': 369; P. Laska, 'Kant on Moral Worth: A Reply to Murphy,' *Kant-Studien* 59, no. (1968); J. G. Murphy, 'Kant's Concept of a Right Action,' *Monist* 51, no. 4 (1967); Mark Timmons, 'Motive and Rightness in Kant's Ethical System,' in *Kant's Metaphysics of Morals*, ed. Mark Timmons (Oxford: Oxford University Press, 2003).
[83] See especially Robert N. Johnson, 'Kant's Conception of Merit: 'Metaphysics of Morals' and Evaluating Actions,' *Pacific Philosophical Quarterly* 77, no. 4 (1996): 311-12) and cf. *Gr* 4: 424 (33).
[84] See Wood, *Kant's Ethical Thought*, 27. Cf. Schroeder, 'Some Common Misinterpretations of the Kantian Ethics': 429-430.
[85] Kant's thinking about the relationship between *virtue*, on the one hand, and such things as moral goodness, moral character, and moral worth, on the other, is a particularly vast topic in its own right. See especially *MdS* 6: 383, 392, 405, 477 (148, 155, 164, 221); *Vorlesungen-Religionslehre* 28: 1075 (409). For a typical formulation of the relationship see Susan Neiman, *The Unity of Reason* (Oxford: Oxford University Press, 1994), 133. Further helpful discussion may be found in Lara Denis, 'Kant's Conception of Virtue,' in *The Cambridge Companion to Kant and Modern Philosophy*, ed. Paul

ambiguity that arises from the apparent, but inconsistent, interchangeability, of some of the main terminology that Kant deploys in his discussions of morality.[86]

These difficulties acknowledged, my main point is this: Kant's inquiries in this area are affected by the fact that, antecedently to posing any of the questions to which his thinking about 'moral worth' offers answers, he takes the human being's morality to be the 'worthiness-basis' for her happiness. Thus, to the extent that he associates morality with worthiness to be happy, his focus falls upon the agent herself, first, and then entails an account of the morality of her maxims, choices, and actions that is relative to this focus. But this is so *only* to the extent that he makes this association; the latter does not determine the whole of his thinking about morality. The great variety of ways in which his commentators are able to characterize Kant's thinking about 'moral worth' attests to the independence of his thinking about *that* from the thinking that is conditioned by his antecedently fixed take on the relationship between morality and happiness. This variety is a function of the fact that, as soon as the association of morality with worthiness to be happy is bracketed out (as it almost always is by Kant's readers), Kant's thinking—and his commentators'—evinces a considerable degree of latitude. His antecedent commitment to a particular conception of the relationship between morality and happiness, rather than any other factor, necessitates an agent-centre approach. Such an approach is possible and sensible and present in Kant, independently of this commitment, but it is not demanded or called for, otherwise, with the same urgency.

Independently of this demand, Kant's understanding of moral worth can be construed in a manner that swings all the way to the other extreme. In other words, it is certainly possible to argue that, for Kant, morality *qua* 'moral worth' is a property of actions alone, *rather than* a property of an agent's character, or will, or any other fea-

---

Guyer (Cambridge: Cambridge University Press, 2006), 513-14; Engstrom, 'Happiness and the Highest Good', 105; Stephen Engstrom, 'The Inner Freedom of Virtue,' in *Kant's Metaphysics of Morals*, ed. Mark Timmons (Oxford: Oxford University Press, 2003), 292-3; R. Z. Friedman, 'Virtue and Happiness: Kant and Three Critics,' *Canadian Journal of Philosophy* 11, no. 1 (1981): 95-6; Robert B. Louden, 'Kant's Virtue Ethics,' *Philosophy* 61, no. 238 (1986): 478; Murphy, 'Kant's Concept of a Right Action': 594-5.

[86] Schroeder, for instance, lists 'virtue,' 'good will,' 'morality,' 'moral disposition,' 'regard for duty,' and 'respect for the law' (Schroeder, 'Some Common Misinterpretations of the Kantian Ethics': 435).

ture proper to her.[87] In the main, however, Kant's commentators offer a more nuanced reading, seeing 'moral worth' as a predicate whose immediate subject is this or that particular action, to be sure, but which also refers, more basically, to maxims,[88] or motives,[89] or to the will[90] or to the will's character.[91]

The dominant line of interpretation converges on the view that, for Kant, moral worth is a property that actions have in virtue of the way that they are motivated.[92] An action has moral worth if it both accords with what duty requires (if it is legal or

[87] See especially Barbara Herman, 'On the Value of Acting from the Motive of Duty,' *Philosophical Review* 90, no. 3 (1981): 371. See also Marcia Baron, *Kantian Ethics Almost without Apology* (Ithaca: Cornell University Press, 1995), chapter 4.

[88] See, for example, Michalson, *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*, 34. Cf. Denis, 'Kant's Conception of Virtue', 514; Sorell, 'Kant's Good Will and Our Good Nature: Second Thoughts About Henson and Herman': 87.

[89] See, for example, Hill, 'Punishment, Conscience, and Moral Worth'.

[90] The *Groundwork*'s famous opening assertion concerning the unique, absolute goodness of a good will (*Gr* 4: 393 [7]) suggests that the fundamental object of moral evaluation is the will. On this view, if actions are good then this is because, in some sense, they display the fundamental goodness of the will from which they arise (see Curzer, 'From Duty': 288, 307; Dahl, 'Obligation and Moral Worth': 379; Engstrom, 'Happiness and the Highest Good', 111; Nelson Potter, 'Kant and the Moral Worth of Actions,' *Southern Journal of Philosophy* 34, no. 2 (1996): 227; T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (London: Belknap Press, 2008), 102; Keith Simmons, 'Kant on Moral Worth,' *History of Philosophy Quarterly* 6, no. 1 (1989): 87, 93-4). See also Henrich's careful qualification of this view in Dieter Henrich, 'Ethics of Autonomy,' in *The Unity of Reason*, ed. R. Velkley (London: Harvard University Press, 1994), 95. See also Hill, 'Is a Good Will Overrated?', 44; Christine M. Korsgaard, 'Aristotle and Kant on the Source of Value,' *Ethics* 96, no. 3 (1986); Sonia Sikka, 'On the Value of Happiness: Herder Contra Kant,' *Canadian Journal of Philosophy* 37, no. 4 (2007): 530-1; Wood, *Kant's Ethical Thought*, 22.

[91] Character is a predicate, first, of reason (in the first *Critique*) regarded as a cause, then later (in the *Religion*) of the will, or rather of 'the power of choice' (*Willkür*). If, as Ameriks suggests, Kant regards 'unconditional goodness [as] a matter of goodness in all contexts' then this idea is best served by equating 'having a good will…to having a good character' (Karl Ameriks, 'Kant on the Good Will,' in *Grundlegung Zur Metaphysik Der Sitten: Ein Kooperativer Kommentar*, ed. Ottfried Höffe (Frankfurt am Main: Vittorio Klostermann, 1989), 54). See also Allison, *Theory of Freedom*, 136; Anne Margaret Baxley, 'The Practical Significance of Taste in Kant's Critique of Judgment: Love of Natural Beauty as a Mark of Moral Character,' *Journal of Aesthetics and Art Criticism* 63, no. 1 (2005): 35; Curzer, 'From Duty': 307; Dahl, 'Obligation and Moral Worth': 379; Engstrom, 'Happiness and the Highest Good', 111; Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action': 40-2, 52; Herman, 'On the Value of Acting from the Motive of Duty': 362; Hill, 'Is a Good Will Overrated?', 39, 52; Jensen, 'Kant on Overdetermination, Indirect Duties, and Moral Worth', 162; Louden, 'Kant's Virtue Ethics': 474; Potter, 'Kant and the Moral Worth of Actions': 226, 237, 288; Simmons, 'Kant on Moral Worth': 87, 93-4; Sverdlik, 'Kant, Nonaccidentalness and the Availability of Moral Worth': 296. Cf. O'Neill's reconciliation of Kant as purveyor of 'an ethic of virtue' with Kant as an advocate of 'an ethic of rules' (Onora O'Neill, 'Kant after Virtue,' *Inquiry* 26, no. 4 (1984): 397). But contrast these views with Herman's claim that Kant's notion of moral worth has no reference whatsoever to 'the permanent structure of an agent's motives' (Herman, 'On the Value of Acting from the Motive of Duty': 371 n.).

[92] See, for instance, Dahl, 'Obligation and Moral Worth': 369; Herman, 'On the Value of Acting from the Motive of Duty': 371, 375; Jensen, 'Kant on Overdetermination, Indirect Duties, and Moral Worth', 162; Laska, 'Kant on Moral Worth: A Reply to Murphy': 374, 377; Henry Sidgwick, *Outlines of the History of Ethics for English Readers* (Boston: Beacon Press, 1960), 272. For a dissenting approach see Murphy, 'Kant's Concept of a Right Action': 577.

law-conforming in the broadest sense) and if it is actually done '*from*' duty.[93] In this connection too, however, Kant's thinking evinces a considerable variety of emphases and qualifications, which is reflected in the host of readings that his thinking about duty invites.[94] Simply put, its various instantiations aside, Kant's ultimate, more or

[93] See *KpV* 5:71 (62); *Rel* 6: 37, 47 (84, 92); *Streit* 7: 91-2 (307). See also C. D. Broad, *Five Types of Ethical Theory*, International Library of Psychology, Philosophy, and Scientific Method. (London: Routledge & K. Paul, 1930), 116; P. Cicovacki, 'The Illusory Fabric of Kant's True Morality,' *Journal of Value Inquiry* 36, no. 2-3 (2002): 387; Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action': 40-2; L. M. Hinman, 'On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty,' *Monist* 66, no. 2 (1983): 251; Rae Langton, 'Duty and Desolation,' *Philosophy* 67, no. 262 (1992): 495; Robert B. Pippin, 'Kant's Theory of Value: On Allen Wood's Kant's Ethical Thought,' *Inquiry* 43, no. 2 (2000): 239; W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), 5; Schaller, 'Kant on Virtue and Moral Worth': 559; Sidgwick, *Outlines of the History of Ethics for English Readers*, 272; Simmons, 'Kant on Moral Worth': 85-7; Sorell, 'Kant's Good Will and Our Good Nature: Second Thoughts About Henson and Herman': 89.

[94] A highly charged area of debate emerged early on and persists to this day in connection with the question whether the mere presence of an inclination to perform some action affects that action's moral worth in every instance. Not every reader finds it possible to forgo what Pippin characterizes as a 'familiar and rather cartoonish image' (Pippin, 'Kant's Theory of Value: On Allen Wood's Kant's Ethical Thought': 239) of Kant according to which he takes moral worth to be a matter, not merely of acting from duty, but of acting from duty in the *absence* of any cooperating inclination whatsoever. Paton is correct to take this sort of reading to be 'a distortion of [Kant's] view' (H. J. Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy* (Philadelphia: University of Pennsylvania Press, 1971 [1947]), 49). And yet even some of his more 'sympathetic' readers, as Herman points out, have identified this view in Kant and have found its presence frustrating (Herman, 'On the Value of Acting from the Motive of Duty': 359). See Mary J. Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten* (Oxford: Blackwell, 1963), 76; Guyer, *Kant on Freedom, Law, and Happiness*, 291, 301. On the supposed *locus classicus* for this view (in Friedrich Schiller [1759-1805]) see Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 48; but cf. Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue': 513; Bertram Kienzle, 'Macht Das Sittengesetz Unglücklich?,' in *Was Ist Und Was Sein Soll*, ed. Udo Kern (Berlin: Walter de Gruyter, 2007); S. Sedgwick, 'Hegel, Mcdowell, and Recent Defenses of Kant,' *Journal of the British Society for Phenomenology* 31, no. 3 (2000). See also G. W. F. Hegel, *Phenomenology of Spirit*, trans., A. V. Miller, New ed. (Oxford: Clarendon Press: Oxford University Press, 1979), § 605, 619 (370, 376); Arthur Schopenhauer, *On the Basis of Morality*, trans., Eric F. J. Payne (Indianapolis: Bobbs-Merill, 1965), 49 (cf. Gerard Mannion, 'Kant and the Defeat of Egoism: Schopenhauerian Concerns and Some Reappraisals and Rejoinders,' *Kant-Studien* 99, no. 2 (2008): 220); Baker, 'Do One's Motives Have to Be Pure', 458, 463-5; Curzer, 'From Duty': 288; C. D. Meyers, 'The Virtue of Cold-Heartedness,' *Philosophical Studies* 138, no. 2 (2008): 236). For a defense of Kant against these detractors see Ameriks, 'Kant on the Good Will', 50; John Atwell, *Ends and Principles in Kant's Moral Thought* (Dordrecht: Martin Nijhoff, 1986), 210ff; Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue': 529; Guyer, *Kant on Freedom, Law, and Happiness*, 301; Christine M. Korsgaard, 'The Right to Lie: Kant on Dealing with Evil,' *Philosophy & Public Affairs* 15, no. 4 (1986): 327; Langton, 'Duty and Desolation': 485; Meyers, 'The Virtue of Cold-Heartedness': 233; Andrews Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination,' in *Agency and Autonomy in Kant's Moral Theory: Selected Essays* (Oxford: Oxford University Press, 2006), 16; Schaller, 'Kant on Virtue and Moral Worth': 560, 568-70; Keith Ward, 'Kant's Teleological Ethics,' *Philosophical Quarterly* 21, no. 85 (1971); Victoria S. Wike, 'Does Kant's Ethics Require That the Moral Law Be the Sole Determining Ground of the Will,' *Journal of Value Inquiry* 27, no. 1 (1993): 85, 87). On the possible moral worth of actions that are overdetermined through the cooperation of respect for the moral law and inclination in the motivating grounds of an action see Baker, 'Do One's Motives Have to Be

less canonical, recourse to the idea of action '*from*' duty tends naturally towards a focus on the agent as well. Moral worth is a property that actions have when they are motivated in a particular way. But ultimately, an agent's being motivated to her particular actions in precisely this or that way is a function of her having a particular kind of will (a good or an evil one).

This approach tends to focus on the agent, then, to the extent that it focuses on the manner in which her will is moved and entails her having and, in as sense, *being* a particular *kind* of will. But there is no theoretical demand, here, that such a will be like this—be a will so motivated—rigorously, always and only. For Kant, some argue, whatever value duty-conforming, but non-moral actions (or agents) may have, they have this value only *accidentally*.[95] Thus duty-conforming and actually moral actions, on this account, are moral in virtue of something non-accidental: the agent's free self-alignment (on the occasion in question) with respect to the moral law[96] or her rational interest (at the time) in doing what it is actually good to do.[97] But this can be construed as an occasional matter—an occasional goodness of the agent (or her will) that her actions inherit and put on display, but a property that is no more stable or permanent than the actions themselves.

In the main then, *independently* of Kant's antecedent commitment to the idea that an agent's morality is a condition bearing (somehow) on her happiness, his attention is focused upon questions about the morality of particular actions and about their motivation (on this or that occasion). The underlying, stable structures of character and will are in view as well—even prominently so. But here, in connection with questions about what it is unconditionally good to do or to have done, the will and its

---

Pure', 458; Baxley, 'The Practical Significance of Taste': 35; Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action'; Herman, 'On the Value of Acting from the Motive of Duty'; Hinman, 'On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty': 263, 265-6; Jensen, 'Kant on Overdetermination, Indirect Duties, and Moral Worth', 170; ibid; Langton, 'Duty and Desolation': 495; James Reid, 'Morality and Sensibility in Kant: Toward a Theory of Virtue,' *Kantian Review* 8, no. (2004): 100; Philip Stratton-Lake, *Kant, Duty and Moral Worth* (New York: Routledge, 2000), 6, 11-12; Wike, 'Does Kant's Ethics Require That the Moral Law Be the Sole Determining Ground of the Will': 91. For recent, useful overviews of Kant's anthropological views on the problem of motivation (the role of pure reason *versus* the given, nature) see Patrick R. Frierson, *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge: Cambridge University Press, 2003); Brian Jacobs and Patrick Kain, eds., *Essays on Kant's Anthropology* (Cambridge: Cambridge University Press, 2003).

[95] See Stephen Darwall, 'Norm and Normativity,' in *The Cambridge History of Eighteenth-Century Philosophy*, ed. K. Haakonssen (Cambridge: Cambridge University Press, 2005), 1020.

[96] Benson, 'Moral Worth': 376.

[97] Herman, 'On the Value of Acting from the Motive of Duty': 366.

character are not the absolutely urgent focus that they get to be in Kant's ultimate thinking (in the *Religion* and its critical period foreshadowings) about the possibility of imputing the agent's most radical '*Gesinnung*' to her and his worries about the consequences that would accrue if this were somehow disallowed. Again, this lack of urgency is possible because there really is a current in Kant's thinking about morality whose dynamism is conditioned only by the primary question of ethics, 'What ought I to do?' and therefore runs its course *independently* of the thinking that is expressed by his habit of glossing 'morality' as 'worthiness to be happy.'

In other words, Kant's specific account of what counts as a will's or an agent's 'morality,' or of what constitutes the 'moral worth' of her motives, maxims, or actions, develops independently of, but also sometimes coincides with, his account of the conditions that have to be met if this agent or will is to be identified, wholly and simply, not only as the unique source of particular deeds, but as their cause, so to speak, according to their kind (*qua* moral or immoral)—that is, such that their imputation subsumes and ascribes to the agent herself the very character (of her will) relative to which her deeds are deemed good or bad ones in the first place. *This* is the outcome that is so urgent, for Kant, given his understanding of the relationship between morality and happiness, to the extent that this understanding finds expression in his habit of glossing 'morality' as 'worthiness to be happy.'

From both the point of view of the primary question of ethics, then, and from the point of view of Kant's secondary, but no less urgent, worries concerning the relationship between morality and happiness (his thinking about that relationship's nature and his worries about its *actualization*), the agent, or rather the will, which is 'in a sense…the very person himself,'[98] stands at the nearer end of an evaluative continuum that begins with what is most apparent, but least decisive—particular actions—and proceeds through maxims and motives to what is most radical, decisive, but least accessible: namely, the agent herself or, again, her will. As Henrich says, for Kant 'our esteem concerns only the will which is directed to the "good," regardless of whether it actualizes the good, or whether stronger forces thwart its intentions.'[99] There is a sense in which actions are important, then, but not fundamental for

[98] Wolff, *Autonomy of Reason*, 59.
[99] Henrich, 'Ethics of Autonomy', 95. See *Gr* 4: 394 (8).

Kant.[100] An agent is not worthy to be happy because she does good deeds. She is worthy to be happy because she has a good will, because her will has a good character, because her will is so configured that when she does her duty she is motivated to do so by the consideration that her duty binds her absolutely, because she frames and acts on maxims that legislate universally for all rational beings. Her good deeds show, or *would* show if only we could establish that they really were good (we cannot of course), that *she*—she herself—is worthy to be happy.

Happiness

What does Kant take moral agents to be worthy *of*? Well, happiness—of course.[101] But matters are not at all straightforward in this respect. Kant's notion of worthiness to be happy entails a view of happiness that, as I will show in chapter 2, does not fit perfectly with any of the theories of happiness that can be drawn out of his work.[102] For now, I will outline just enough of Kant's thinking about happiness to delineate a kind of core conception of the 'thing' of which he takes moral agents to be worthy (a core whose aspects it shares with Kant's other main conceptions of happiness). In chapter 2, I will further specify this core so as to distinguish between happiness regarded as a state of affairs of which moral agents are worthy and immoral ones unworthy, and happiness regarded (also sometimes by Kant) as something that is inconsistent with the sense of his habitual gloss.

For now, then, I distinguish between five core features of Kant's understanding of happiness. It will soon be evident that some of these elements of his thinking are

[100] See Taylor, *Sources of the Self: The Making of the Modern Identity*, 121.
[101] Or cognates: see, for example, *R* 6856 19: 181 (441); *R* 7202 19: 279 (467); *R* 6979 19: 219 (454); *LEC* 27: 373 (147).
[102] Some of his readers judge Kant's thinking about happiness imprecise, uneven, and confused. See, for example, Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*, 78, 177; H. J. Paton, 'Kant's Idea of the Good,' *Proceedings of the Aristotelian Society* 45, no. (1944-5): ix; Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 85, 105. Cf. Victoria S. Wike, 'Kant on Happiness,' *Philosophy Research Archives* 13, no. (1988): 79. Others offer a more favourable assessment, but observe that Kant entertains several distinct concepts of happiness and that his terminology poses interpretive challenges. For the claim that Kant's thinking about happiness is coherent and consistent, at least on balance, see O'Connor, 'Kant's Conception of Happiness'; Victoria S. Wike, *Kant on Happiness in Ethics*, Suny Series in Ethical Theory (Albany: State University of New York Press, 1994), xiv. With respect to conceptual and terminological diversity see Thomas E. Hill, Jr., 'Happiness and Human Flourishing in Kant's Ethics,' *Social Philosophy & Policy* 16, no. 1 (1999): 146; A. Hills, 'Kant on Happiness and Reason,' *History of Philosophy Quarterly* 23, no. 3 (2006): 245; Wike, *Kant on Happiness in Ethics*, 1.

in tension with one another; they do not offer up an entirely coherent picture. This 'core' is not monolithic by any means.

First, for Kant, 'happiness' (or '*Glückseligkeit*') refers to a state of their affairs that human beings desire universally, inevitably, ineradicably, and, just as such, innocently.[103] Second, happiness is a state of the human being's *empirical* affairs. Kant always conceives of happiness in a way that has integral reference to the experience of having at least some of one's actual inclinations gratified.[104] It is going too far to say that happiness is 'a state that Kant always associates *exclusively* with the body,'[105] but we must also reject the claim that mere moral contentment or self-satisfaction, just as such, is ever what Kant means by '*Glückseligkeit*.'[106]

In fact, some commentators have held that, for Kant, moral contentment or moral self-satisfaction is a kind of happiness, or even that this is just what Kant takes authentic happiness consists in.[107] Admittedly, in the *Religion*, Kant does contrast 'moral happiness' and 'physical happiness'[108] and in *The Metaphysics of Morals* he juxtaposes 'pathological pleasure' and 'moral pleasure.'[109] Nevertheless, Kant says (a page earlier) that the idea of 'a certain *moral* happiness not based on empirical

[103] See, for example, *KpV* 5: 25 (23); *MdS* 6: 388, 480 (151-2, 223); *Rel* 6: 46 n., 58, 134 (91 n., 102, 162). For significantly earlier expressions of the same view see *R* 4463 17: 561 (138) (ca. 1772-3); *R* 6973 19: 217 (453) (ca. sometime in the 1770s). Commentators frequently affirm the presence of this doctrine in Kant's thinking about happiness. See, for example, Allison, *Theory of Freedom*, 151; Allison, *Idealism and Freedom*, 114; Beck, *A Commentary on Kant's Critique of Practical Reason*, 82; Guyer, *Kant on Freedom, Law, and Happiness*, 213; Hills, 'Kant on Happiness and Reason': 243-4; Wood, *Kant's Ethical Thought*, 65; ibid., 66.). Cf., however, *MdS* 6: 385 (149) and Robert N. Johnson, 'Happiness as a Natural End,' in *Kant's Metaphysics of Morals*, ed. Mark Timmons (Oxford: Oxford University Press, 2003).

[104] *Gr* 4: 418 (28); *ÜdG* 8: 283 (285); *LMD* 28: 689 (390); *R* 6892 19: 196 (447). See also Denis, 'Kant's Conception of Virtue', 523. But cf. *R* 6977 19: 218 (454).

[105] George Di Giovanni, 'Freedom and Religion in Kant and His Immediate Successors: The Vocation of Humankind, 1774-1800,' (2005): 182 (my emphasis).

[106] See Wike's distinction between 'contentment' and 'inclination' conceptions of happiness and her denial that the former is ever what Kant means by '*Glückseligkeit*' (Wike, 'Kant on Happiness'; Wike, *Kant on Happiness in Ethics*, 1, 13). The author borrows the distinction from Gary Watson, 'Kant on Happiness in the Moral Life,' *Philosophy Research Archives* 9, no. (1983). See also Paul Arthur Schilpp, *Kant's Pre-Critical Ethics*, Northwestern University Studies in the Humanities (Evanston: Northwestern University, 1938), 132.

[107] See Hegel, *Phenomenology of Spirit*, § 618 (375); M. Packer, 'The Highest Good in Kant's Psychology of Motivation,' *Idealistic Studies* 13, no. 2 (1983): 118; Watson, 'Kant on Happiness in the Moral Life': 81, 83. For further discussion see Guyer, *Kant on Freedom, Law, and Happiness*, 108, 112, 115, 124, 165; Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 57; Smith, 'Worthiness to Be Happy': 182; Wike, 'Kant on Happiness': 80; Wike, *Kant on Happiness in Ethics*, 13. Cf. *KrV* B837ff; *KpV* 5: 117-18; *R* 7202 19: 281 (469); *R* 6616 19:111; *R* 6867 19: 186; *R* 6907 19: 202-3; *R* 6883 19: 191. Cf. Donald R. Keyworth, 'Kant's Concept of Happiness in the Moral Argument,' *Personalist* 43, no. (1962).

[108] *Rel* 6: 67 (109). See also ibid. 6: 75 n. (115 n.); *R* 7260 19: 296.

[109] *MdS* 6: 378 (143); and again, later, 'moral happiness' and 'natural happiness' (*MdS* 6: 387 [151]).

causes' is 'a self-contradictory absurdity.' Rather, this is 'a state that could well be *called* happiness, a state of contentment and peace of soul in which virtue is its own reward.'[110] Or again, 'moral happiness' is 'a misuse of the word happiness' and 'already involves a contradiction.'[111] At most, Kant allows that moral self-contentment is *analogous* to happiness.[112]

It is almost universally agreed, in any case, that Kant's dominant practice is to distinguish between moral contentment or moral self-satisfaction and happiness, and not to conflate them.[113] As we shall see in chapter 2, moral self-satisfaction is sometimes, for Kant, a kind of moral-psychological condition without which happiness (though not impossible) will tend to be eroded or undermined.[114] And sometimes he takes moral contentment (or, really, the well-ordered freedom that it discloses) to be a necessary formal condition for happiness.[115]

Third, some commentators have found in Kant a seemingly hedonistic sense of 'happiness' that pertains to the pleasure of the moment, the fulfillment of this or that desire, here and now, the gratification of particular inclinations at a particular time. This perspective has an ambiguous position relative to what I am characterizing as an analysis of the common 'core' of Kant's thinking about happiness. It attaches to that core, as it were, to the extent that it holds forth the first and most obvious possible sense of happiness which, as I will argue in the next chapter, might be connected with Kant's habitual gloss (I will problematize this association and propose a more nuanced possibility in chapter 4). On this view, happiness is a matter of the occa-

---

[110] *MdS* 6: 377 (142) (my emphasis).
[111] *MdS* 6: 387 (151).
[112] See especially *Gr* 5: 117-18 (98); *KpV* 5: 117-18 (98); *Vorarbeiten-ÜdG* 23: 129. See also Wike, *Kant on Happiness in Ethics*, 23. But cf. Kant's identification of moral contentment and '*moral apathy*' (*MdS* 6: 408-9 [166-7]). Cf. Engstrom, 'Inner Freedom', 314; Langton, 'Duty and Desolation': 496-7.
[113] See O'Connor, 'Kant's Conception of Happiness': 202; Schilpp, *Kant's Pre-Critical Ethics*, 135; Sikka, 'On the Value of Happiness: Herder Contra Kant': 520; Wike, 'Kant on Happiness': 85.
[114] See Friedman, 'Virtue and Happiness: Kant and Three Critics': 101; O'Connor, 'Kant's Conception of Happiness': 202; Smith, 'Worthiness to Be Happy': 174. It is worth noting that it is not always clear in such instances whether contentment is specifically connected with morality or with some other aspect of practical rationality. See, for example, Hills, 'Kant on Happiness and Reason': 245-6; Wike, *Kant on Happiness in Ethics*, 3, 17. On the topic of 'moral happiness' and the latter's connection to virtue in Kant and Aristotle see Norbert Fischer, 'Tugend Und Gluckseligkeit: Zu Ihrem Verhaltnis Bei Aristoteles Und Kant,' *Kant-Studien* 74, no. 1 (1983). On the relationship between empirical happiness and moral self-satisfaction see especially M. Forschner, 'Moralität Und Glückseligkeit in Kants Reflexionen,' *Zeitschrift für philosophische Forschung* 42, no. 3 (1988): 364.
[115] Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 50.

sional, pleasurable 'satisfaction of desire and impulse.'[116] Indeed, the commentary demonstrates that it is possible to read Kant on happiness in this way, at least some of the time; it also demonstrates that this reading is problematic.[117] With one important exception,[118] Kant's published work offers little evidence to support the view that he ever thinks about happiness in such terms.[119] If this approach is present in Kant then—not merely in the form of the odd imprecision of expression, but as a view that he holds systematically—it is problematic and, evidently, extremely muted. Nevertheless, as I said above, for reasons that will become clear in chapter 2, this notion attaches to the core of Kant's thinking about happiness as a kind of implicit, most basic, default possibility for answering the question: Of what does Kant take moral agents to be worthy and immoral ones unworthy?

Fourth, Kant observes that individual human beings have distinct 'concepts of happiness,' that each agent's concept of happiness is distinct in some respect(s) or other from her neighbours'.[120] On this view, as Paton puts it, happiness can only be regarded as 'the good for *me* or *my* good.'[121] Here, the notion of happiness depends for its content upon the particular desires that particular agents happen to have at particular times. So regarded, happiness is a 'relative' concept,[122] entirely 'subjective'[123]

---

[116] Keyworth, 'Kant's Concept of Happiness in the Moral Argument': 28.
[117] See Beck, *A Commentary on Kant's Critique of Practical Reason*, 97. Cf. Ibid., 72. and see T. H. Irwin, 'Kant's Criticisms of Eudaemonism,' in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, ed. S. P. Engstrom and J. Whiting (Cambridge: Cambridge, 1994), 67; Paton, 'Kant's Idea of the Good': ix. See Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 86 for an important qualification of his view. Hills marks a close relationship 'between seeking pleasure and seeking the satisfaction of your desire,' but notes that 'they are not the same' (Hills, 'Kant on Happiness and Reason': 245). Cf. Engstrom, 'Happiness and the Highest Good', 127. For a sustained critique of approaches that take Kant to be a hedonist with respect to happiness see Andrews Reath, 'Hedonism, Heteronomy, and Kant's Principle of Happiness,' in *Agency and Autonomy in Kant's Moral Theory*, ed. Andrews Reath (Oxford: Oxford University Press, 2006). See also Allison, *Theory of Freedom*, 102; R. B. Pippin, 'Idealism and Agency in Kant and Hegel,' *Journal of Philosophy* 88, no. 10 (1991): 537.
[118] See *MdS* 6: 480-2 (223-5) and chapter 4 of this thesis.
[119] See, however, *R* 6876 19: 189 [445] and *R* 7200 19: 274 [464]; cf. Hill, 'Happiness and Human Flourishing in Kant's Ethics': 145. Even here, an apparently hedonistic notion of happiness is limited to happiness considered without reference to morality, which would otherwise (if present) render the pleasure of the moment happiness properly-so-called.
[120] *MdS* 6: 454 (203). See also ibid. 6: 481 (224).
[121] Paton, 'Kant's Idea of the Good': viii (my emphasis). But cf. *KpV* 5: 111 (93) and Hill, 'Happiness and Human Flourishing in Kant's Ethics': 148-9.
[122] See O'Connor, 'Kant's Conception of Happiness': 191; Paton, 'Kant's Idea of the Good': xi; Sikka, 'On the Value of Happiness: Herder Contra Kant': 519.
[123] Hill, 'Happiness and Human Flourishing in Kant's Ethics': 146; Johnson, 'Happiness as a Natural End', 319.

and 'variable.'[124] More strongly put, on this view, one subject's notion of happiness is incommensurate with another's.[125] Indeed, at times, Kant seems to put the individual agent in the position of being unable to determine her *own* concept of happiness in advance, thus rendering happiness a state of affairs that she cannot really *determinately* plan for or anticipate.[126] On this view, happiness appears to be a general concept, but is not one. The idea of 'happiness' that agents hold in common, here, is not a general concept of happiness, but rather the brute idea of the state of affairs that would consists in 'all' of each individual agents' particular inclinations being satisfied.[127]

Fifth, however, Kant's dominant tendency is to construe happiness as a kind of shared ideal, 'the genuinely *a priori* concept of a systematic whole of intra- and interpersonal happiness.'[128] On this view, happiness remains 'a mere *idea*,' impossible to think concretely, and impossible for a finite being to achieve in merely natural terms.[129] This tendency in Kant's core thinking about happiness construes happiness as an ideal state of affairs that would consist in the actualization of the 'absolute whole,'[130] or as the 'the natural end of the sum of all inclinations,'[131] a scenario in which 'everything' goes one's way,[132] a continuous, uninterrupted state of affairs in which one 'always' gets what one wants.[133] Even though they do not always take it

---

[124] Hill, 'Happiness and Human Flourishing in Kant's Ethics': 146. See also G. Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen",' *Modern Schoolman* 71, no. 4 (1994): 276.
[125] Wood, *Kant's Ethical Thought*, 231.
[126] See *Gr* 4: 418 (28). Cf. Hill, 'Happiness and Human Flourishing in Kant's Ethics': 146; Hills, 'Kant on Happiness and Reason': 250, 253, 255; Paton, 'Kant's Idea of the Good': ix.
[127] Guyer, *Kant on Freedom, Law, and Happiness*, 134.
[128] Ibid., 117. See also Allison, *Idealism and Freedom*, 114; Beck, *A Commentary on Kant's Critique of Practical Reason*, 97; Hills, 'Kant on Happiness and Reason': 254). Note, however, that Kant does not take happiness to be 'an ideal of reason but of imagination, resting merely upon empirical grounds' (*Gr* 4: 418 [29]). See also Reath, 'Hedonism, Heteronomy, and Kant's Principle of Happiness', 46.
[129] *KdU* 5: 430 (297). See also Sikka, 'On the Value of Happiness: Herder Contra Kant': 519.
[130] *Gr* 4: 418 (28).
[131] *KrV* A806/B834; *KdU* 5: 434 (301). See also Hill, 'Happiness and Human Flourishing in Kant's Ethics': 147. Cf. Hegel, *Phenomenology of Spirit*, §§ 601-2 (366).
[132] *KpV* 5: 124 (104); *MdS* 6: 480 (223).
[133] *KrV* A806/B834; *MdS* 6: 480 (223); *Gr* 4: 418 (28). Cf. Allison, *Theory of Freedom*, 124; Beck, *A Commentary on Kant's Critique of Practical Reason*, 72; Denis, 'Kant's Conception of Virtue', 523; Friedman, 'Virtue and Happiness: Kant and Three Critics': 95; Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*, 78; Hills, 'Kant on Happiness and Reason': 250, 254; Johnson, 'Happiness as a Natural End', 319; Paton, 'Kant's Idea of the Good': ix; H. J. Paton, *The Moral Law: Or Kant's Groundwork of the Metaphysic of Morals*, The Senior Series (London: Hutchinson's University Library, 1947), 29; Paton, *The Categorial*

to be the *only* concept of happiness that he deploys, many commentators emphasize Kant's notion of happiness in terms of this ideal of 'an absolute whole' and continuous 'maximum of well-being.'[134]

Of course, Kant has to qualify these references to 'all,' 'everything,' and 'always' in various ways in order to render the concept to which they belong relevant *in practice*. It would be incoherent to speak of—or to strive for—the joint realization of mutually exclusive ends, for example. Instead Kant tends to see the 'thing' desired under the concept of 'happiness' as an object whose counterfactual, ultimately deferred actualization would consist in the maximal fulfillment or realization of a maximally harmonized sum, or systematized complex, or hierarchy of one's jointly realizable empirical ends.[135] Since not all of a human being's projects are mutually achievable, she must anticipate and seek (by way of what Kant characterizes as 'skill') to realize the best overall organization of her aims and so, too, of the action that aims at their realization. Because inclinations conflict with one another, they must be organized in a hierarchical scheme and harmonized under the notion of happiness.[136] Kant points out, however, that given the great complexity and obscurity of this project, the human being can have no 'determinate concept of what he really wills here.'[137]

For Kant, then, in the main, 'happiness' is not a general term covering each of the particular ends for which inclination 'strives [*strebt*]'[138] (as though they were

---

Imperative: A Study in Kant's Moral Philosophy, 85; Wike, 'Kant on Happiness': 80; Wike, *Kant on Happiness in Ethics*, 5, 7; Wood, *Kant's Ethical Thought*, 254. Cf. Irwin, 'Kant's Criticisms of Eudaemonism', 67. Somewhat more modestly, however, Kant speaks too of a 'freedom from evils [*Übeln*] and enjoyment of ever mounting pleasures' (*Rel* 6: 67 [109]), which implies nothing so vast as the realization of all of one's desires at every moment.

[134] See Allison, *Theory of Freedom*, 124; Beck, *A Commentary on Kant's Critique of Practical Reason*, 72; Denis, 'Kant's Conception of Virtue', 523; Friedman, 'Virtue and Happiness: Kant and Three Critics': 95; Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*, 78; Hills, 'Kant on Happiness and Reason': 250, 254; Johnson, 'Happiness as a Natural End', 319; Paton, 'Kant's Idea of the Good': ix; Paton, *The Moral Law: Or Kant's Groundwork of the Metaphysic of Morals*, 29; Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 85; Wike, 'Kant on Happiness': 80; Wike, *Kant on Happiness in Ethics*, 5, 7.

[135] See *Gr* 4: 417-18 (28-9) and Guyer, *Kant on Freedom, Law, and Happiness*, 117, 339-40; Wike, 'Kant on Happiness'; Wike, *Kant on Happiness in Ethics*. Cf. Beck, *A Commentary on Kant's Critique of Practical Reason*, 97; Schilpp, *Kant's Pre-Critical Ethics*, 133; J. B. Schneewind, 'Active Powers,' in *The Cambridge History of Eighteenth-Century Philosophy*, ed. Knud Haakonssen (Cambridge: Cambridge University Press, 2005).

[136] *Rel* 6: 58 (102).

[137] *Gr* 4: 418 (28). See also *KdU* 5: 430 (297-8).

[138] *MdS* 6: 481 (223).

mere instances of a kind). Rather, 'happiness' refers to this whole—whether to the inconsistent totality of such ends, or to the harmonized hierarchy that entails complex compromises between them. In either of these perspectives, there can be no actual instances of happiness *as such*. As Kant says in his philosophy of religion lectures, '[w]e do, to be sure, have an idea of the complete entirety of well-being and of the highest contentment; but we cannot cite a case *in concreto* where this idea of happiness is *entirely* realized.'[139]

Happiness, in Kant's dominant view, would depend so profoundly upon the human agent's own skill and foresight and power and, too, upon the cooperation of both nature and other agents that it would be unattainable, as such—an aspiration, merely. Happiness is out of reach, both as 'all,' 'everything,' and 'always,' and as an internally coherent, limited, but still ideal whole. All of this suggests that happiness lies always out of the human being's reach, beyond our ability to think it and beyond our ability practically to achieve it.[140] '[W]e have no concept of the whole [of all ends],' Kant argues. It is unthinkable by us. We certainly 'cannot direct our actions according to [it].'[141] Indeed, as Wood points out, there is a sense in which, on Kant's thinking, *unhappiness* alone is assured so that the latter must be our lot, 'even when our present needs are satisfied.'[142]

Note, however, that happiness' impossibility, here, is not a matter of the human being's failure to be moral. On this account, not morality, but the cooperation of external (natural and social) factors, in combination with the right kind of skill, foresight, and power are necessary and sufficient conditions without which happiness turns out to be impossible. Given these conditions, happiness would be assured. The human being's finitude, rather than her immorality, militates here against her happiness. Thus far, given the degree to which I have isolated these core components of Kant's thinking about happiness from the rest of his theoretical concerns, it would appear that if the human being were a kind of adequately skilled, insightful, and

[139] *Vorlesungen-Religionslehre* 28: 1080 (412). See also ibid. 28: 1080 (413).
[140] See especially *KdU* 5: 430 (297-8). See also *R* 7202 19: 278-9 (467).
[141] *Vorlesungen-Religionslehre* 28: 1057-8 (394-5).
[142] Wood, *Kant's Ethical Thought*, 254. This is the sense of 'happiness' relative to which it is accurate to say that Kant is 'the first critic of the *concept* of happiness' (Henrich, 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason', 77). By making Kant too definitively a critic of the concept of happiness, however, Henrich elides the fact that Kant's thinking about happiness is far from monolithic and that Kant, from time to time, rather uncritically *does* speak of happiness as though the concept is a coherent and useful one.

powerful demigod, and if she entered into an adequately far-sighted compact with adequately cooperative, pragmatically motivated companions, and if nature happened to be adequately amenable or pliant, then she could be happy—even if, as demigods too often are, she were profoundly immoral.

But in some sense Kant takes morality to be a *necessary* condition for happiness. As I show in chapter 2, to the extent that Kant glosses 'morality' as 'worthiness to be happy,' he regards the happiness of which moral agents are worthy and immoral ones unworthy to be something to which they might possibly attain, morality aside. Specifically, however, he regards this happiness as a state of affairs that would have to be forged, then, on behalf of such agents, by God—or occluded (just in case they had attained the rank, as it were, of immoral demigods). The notion of worthiness to be happy, therefore, has a fundamentally eschatological orientation (see chapter 4). It also entails a particular specification of Kant's concept of happiness, beyond what we have been exploring here. Again, I take this up in the next chapter.

The independence of the idiom

In the foregoing, I have at times already touched upon the subject matter of this subsection. I will now consolidate and explicate the claim that, while Kant's thinking about morality (in its primary, forward-looking sense, in constant connection with the question, 'What ought I to do?') varies and develops in the course of his career, his understanding of the relationship between morality and happiness—*to the extent that this is encapsulated in his regular glossing of 'morality' as 'worthiness to be happy'*—remains constant. In other words, the expression, 'worthiness to be happy,' always has more or less the same sense.

This means that even as he is working out his 'critical' answer to the primary question of ethics, 'What ought I to do?', Kant's gloss indicates, repeatedly, that he is always already committed to a particular answer to the question: 'How are morality and happiness related?' Obliquely and obscurely, Kant answers this question in the same manner every time he glosses 'morality' as 'worthiness to be happy.' While his thinking about morality (in its primary, forward-looking sense) develops and deepens, the core of his thinking about the relationship between morality and happiness does not. The stability of this core is evident in the stability and frequent de-

ployment of the expression itself. The stubborn incorrigibility of its referent, however, is evidenced by the expression's regular interpolation into contexts where it adds nothing to the discussion at hand. Its obscurity—its virtually submarine character—is on display in the fact that these interpolations make no waves, neither in the immediate context of Kant's text, nor in the reading of most of his commentators.

Kant's assumption concerning the relationship between happiness and morality, as encapsulated in his habit of glossing 'morality' as 'worthiness to be happy,' is both deep and silent. It is logically antecedent to almost all of his thinking about both morality and happiness. There are, however, a number of exceptions. Kant's assumption affirms, first, that irrespective of any of the other uses to which he puts 'moral,' 'immoral,' 'happy,' and 'unhappy,' it is possible for immoral agents to be happy (see chapter 2). It affirms, second (as we have seen in this chapter), that irrespective of any of the other uses to which he puts 'moral' and 'immoral,' Kant is able to turn these notions around, disentangle them from their primary reference to the possible maxims, motives, incentives, and actions that would answer to the forward-looking questions, 'What ought I to do?' and 'What ought I not to do?', and assign them to antecedently established *actual* properties of active agents—of their wills and their will's characters. And it affirms, third, that immoral agents ought not to be happy (which affirmation has the major entailments that I discuss in the rest of this thesis).

The thoughts that find implicit expression in Kant's habitual gloss pertain to a *topos*, then, whose special character does not depend upon any particular outcome of the primary, forward-looking mode of moral inquiry whose fundamental question is: *What ought I to do?* By contrast, the thoughts that find expression in Kant's habitual gloss bear profoundly on the outcome of the secondary, backward-looking mode of moral inquiry whose fundamental questions are not only, 'How are happiness and morality (in its forward-looking sense) related?', but also 'Who ought actually to be unhappy?'

Indeed, this is not *quite* right. Kant's gloss does not exactly answer these questions. Rather, it declares, in the obscure, almost silent manner with which this thesis must contend, that they have already been answered, in advance.

Thus, to put a sharp point on my claim, the representation of the relationship between morality and happiness that is implicit in Kant's glossing of 'morality' as

'worthiness to be happy' has priority over all of the particular claims that Kant makes about morality and happiness in his 'critical' period. This priority has significant consequences. It is in the nature of the 'worthy-of' relation to which his gloss refers that it is possible to elucidate the latter without having to say very much at all about what 'morality' consists in. 'Morality' can be allowed to denote a variety of possible states of an agent's practical affairs (obviously, however, not just *any* old state of her practical affairs), without this variation's affecting morality's basic function—as represented by Kant's use of this idiom—in relation to happiness: that is, its role as a necessary condition for the latter. What then of this necessity, this conditioned-by relation? What exactly does Kant mean when he affirms that moral agents are *worthy* and immoral ones *unworthy* of happiness? I answer these questions in chapter 2.

'Worthiness to be happy' in the commentary

In the course of this section I will refrain, as far as possible, from referring to the various ways in which this thesis' position on Kant's use of the 'worthiness to be happy idiom' differs from the positions espoused by the handful of commentators that take the matter up for discussion. I do not want to jump too far ahead by making substantive claims about my understanding of Kant's gloss—a sense of the latter that will be best clarified by a process of accrual over the course of this work's subsequent chapters. Instead, this chapter's conclusion will offer a bare sketch of the main ways in which I take these other readers' interpretations to differ from mine. But it will be a task of the rest of this thesis to demonstrate, bit by bit, the extent to which the readings of these other commentators are mistaken, or inadequate, or beside the point.

Before beginning, I want to take a moment to deplore the *paucity* of such readings—to note this lacuna and to take issue with it. Aside from the very few exceptions that I discuss below, the commentary pays no special attention to the Kantian habit with which this thesis is concerned. In general the idiom encounters a kind of oblivious complacency. Reader's whose citations of the Kantian text include the expression, 'worthiness to be happy' (or any of its cognates), read right *through* the latter as though it were not there at all. They do not pause to interrogate it, to ask what it is doing there in the midst (as it so often is) of Kant's discussions of morality in the

latter's primary, forward-looking sense. They do not ask Kant to explain himself. They simply proceed as Kant does, sliding without noticeable friction from the language of 'morality,' 'moral worth,' 'inner worth,' 'good will,' (moral) 'character,' and 'virtue' to talk of 'worthiness to be happy.'[143] It must be (I surmise) that these readers take the gloss to effect a theoretical move so obvious and of such peripheral significance that it merits no discussion. Or perhaps they do not even notice it. Perhaps they really do simply take the expressions 'morality' ('virtue,' etc.) and 'worthiness to be happy' to be simply and unremarkably *synonymous.*

The immediacy of this inherited, habitual procedure, however, gives the impression of endorsing a theoretical move: Kant's repeated affirmation of a particular connection between happiness and morality. But this move is not obviously warranted at all. And it is incumbent on Kant's readers not to make it—unless they are prepared to make it explicitly. It is simply not obvious that the 'inner worth' of my will, in Kant's sense—the worth of my will 'as compared not with others, but with the moral law'—is simply *equivalent* to my worthiness to be happy.[144] This is what Kant says, to be sure; but what does he mean? Much turns on the nature of the 'worthy-of' relation in which Kant takes moral agents to stand to their own (perhaps mainly prospective) happiness. Perhaps nothing interesting lies that way; perhaps something very interesting does. In any case, this ought to be carefully investigated *before* Kant's move is endorsed. In general, though, it is not remarked upon at all.

---

[143] Allison, for example, simply follows the pattern set by Kant, leaving this move entirely implicit, unexamined, and unjustified. He says that moral laws do not merely tell us what we ought unconditionally to do; they 'tell us what we have to do to be worthy of happiness' (Henry E. Allison, 'The Concept of Freedom in Kant's Semi-Critical Ethics,' *Archiv Fur Geschichte Der Philosophie* 68, no. 1 (1986): 101; Allison, *Theory of Freedom*, 66). Similarly, Beck makes the 'state of being worthy of happiness' an 'a priori condition' of happiness without explicating the conceptual relationship between 'worthiness to be happy' and morality (Beck, *A Commentary on Kant's Critique of Practical Reason*, 215). Or again, simply following along after Kant, Carnois avers that '*moral laws* absolutely command the way we must act if we would make ourselves worthy of happiness' (Bernard Carnois, *The Coherence of Kant's Doctrine of Freedom* (Chicago: University of Chicago Press, 1987), 26). In a related vein, Sverdlik conflates 'morality' as the doctrine of 'what we ought to do' with 'morality' as the doctrine of 'what we deserve praise for doing' (Sverdlik, 'Kant, Nonaccidentalness and the Availability of Moral Worth': 293). See also Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action': 42; Murphy, 'Kant's Concept of a Right Action': 591. Gunkel comes close to thematizing Kant's habit, but his implicit observation that, for Kant, 'morality' is somehow 'morality *as* worthiness to be happy' is simply an occasion for his critique of the role that Kant assigns (in the Canon) to belief in God and immortality, in connection with morality (Andreas Gunkel, *Spontaneität Und Moralische Autonomie: Kants Philosophie Der Freiheit*, Berner Reihe Philosophischer Studien (Bern: Verlag Paul Haupt, 1989), 101-2).

[144] Thus Allen Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2008), 220.

As I said above, however, a small handful of commentators do encounter a kind of friction when they read through Kant's gloss and recognize that it stands in need of explanation, at least.

Dieter Henrich, to begin with, in his essay on 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason,' elucidates 'the theoretical reason for the thesis…that morality is the worthiness to be happy.' As an 'ideal,' he argues, happiness

> has its source in a longing of reason for unconditioned unity even in action. Because this unity cannot be achieved on the basis of content, reason, in order to satisfy its principle, must place happiness under a condition which does not take into consideration the accidental distribution of luck and the contradiction of desires. This condition is that happiness should be distributed only in the form of rational universality for the sake of the idea of universality. Reason places all striving for happiness under the condition that it must correspond to the form of rational order. This is the theoretical reason for the thesis that continually returns in Kant's literary remains, namely, that morality is the worthiness to be happy.[145]

I will make two main observations here and draw a conclusion from them. My first observation is that Henrich's argument depends upon two distinct notions of happiness. On the one hand, there is happiness as an 'ideal' that 'has its source in a longing of reason for unconditioned unity even in action.' On the other hand, Henrich adverts implicitly to a distinct possibility: happiness as a product of mere 'luck' and as a state of affairs that would haphazardly gratify (at least some of) the ultimately contradictory desires that human beings just happen to have.

My second observation is that Henrich refers to the relationship between happiness and 'the form of rational order' in a manner that has direct reference to happiness' 'distribution'—and that there seem to be two senses of this term in play as well. On the one hand, there is 'the accidental distribution of luck' (which is correlated with 'the contradiction of desires'). On the other hand, there is 'distribut[ion]…in the form of rational universality.' This 'form' contrasts, of course, with the 'accidentalness' and 'contradiction' of the merely empirical 'content' that contaminates reason's 'ideal' of happiness, which is antecedently grounded in its pure 'longing,' and conceived of as the upshot of 'unity even in action' from the outset.

Reason places happiness or, more precisely, 'all striving for happiness' under a condition such that, if the aim of this striving were realized (by the agent herself), then the agent's happiness just *would be* distributed along the contours of morality

[145] Henrich, 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason', 77-8.

and 'for the sake of the idea of universality.' Henrich does not use the term 'morality' here, of course, at all, but refers instead to such things as 'rational universality,' 'the idea of universality,' and 'the form of rational order.'

Thus, I conclude, Henrich defines happiness, from the outset, in such a way that the happiness of which moral agents are worthy, on his reading of Kant's expression, can only be happiness *qua* pleasing to reason. In other words, the happiness of which moral agents are worthy is, at the same time, happiness that is distributed (or rather striven for and achieved) in a manner that answer's to reason's antecedent 'longing.' Happiness in *this* sense is correlated with morality from the outset. The baser state of affairs that is 'distributed' haphazardly, by mere chance, along the discontinuous, frayed contours of contradiction-riven desire is not the state of affairs at which the unified action (or 'striving') of moral agents aims in the first place.

This suggests that Henrich's use of 'distributed,' or 'distribution' here, is not univocal. Happiness, in the sense of the 'ideal' that 'has its source in a longing of reason for unconditioned unity' is not something that even *could* be 'distributed' to immoral agents; nor could their striving count as striving for *that*. Thus, Henrich makes morality internal to the happiness of which moral agents are worthy. Reason's strategy for getting what it wants, by placing 'all striving for happiness under the condition that it *must* correspond to the form of rational order,'[146] empties Kant's identification of morality as worthiness to be happy ('the thesis…that morality *is* the worthiness to be happy') of the normative significance that it appears, at least, to have. On Henrich's reading, to say that an agent is worthy to be happy is to say that her *striving* for happiness has a form such that, if she were ever to attain what she sought, then what she had attained would satisfy her—and otherwise not. On this reading, the haphazardly distributed/achieved 'happiness' of immoral agents would be too odd an incarnation of the thing (by not being *satisfying* to them) to count as happiness at all.

While Henrich has a ready explanation for Kant's 'thesis…that morality is the worthiness to be happy,' Morris Cohen finds it puzzling—at least at first. 'Why,' he asks, 'after Kant has gone to so much labor to prove that we must do our duty for duty's sake and for no other reason, does he in the end spring the demand that virtue

[146] Ibid., 78.

be rewarded in accordance with "worthiness to be happy."'[147] Evidently, Cohen's first thought, when faced with Kant's use of the idiom, is that the latter's primary role is to represent the potential for reward that inheres in doing what is right—a possibility that he takes Kant to be holding forth as an apparent incentive (something that Kant ought not to be doing). More than this, Cohen seems to take Kant to hold that virtue in some sense '*demands*' reward.

Cohen then proceeds to the rather judicious observation that 'worthiness to be happy' is an 'unanalyzed concept [that] seems to be dragged in *ab extra* without any relevance to, or agreement with, Kant's other ethical ideas.'[148] Again, Cohen finds this puzzling. But then he declares that

> the puzzle is clarified when we take into account Kant's philosophy of law, according to which it is a moral imperative that offenses be punished and worthy labor properly rewarded. A society or universe in which this is not the case is not moral or just.[149]

Cohen's clarification of the originally puzzling, apparently '*ab extra*,' seemingly theoretically 'irrelevant' intrusion of Kant's notion of worthiness to be happy, is rather reminiscent of his original (puzzled) sense of the concept: Kant's deployment of this notion communicates his view that 'offenses [ought to] be punished and worthy labor properly rewarded,' which last claim corresponds to Cohen's earlier sense that the idiom refers to the idea that 'virtue [ought to] be rewarded.' Cohen's reference to Kant's notion of justice and to his 'philosophy of law' marks an important difference, though, between the puzzled reading and the clarified one.

On the former reading, the possibility of reward is held forth as an apparent enticement (or as a possible enticement, in any case, where no enticement should enter). On the latter reading the concept of worthiness to be happy reminds us that '*it is a moral imperative* that offenses be punished and worthy labor properly rewarded.' Evidently, Cohen takes Kant's concept, here, to express the view that morality *necessitates* happiness. He also implies that the notion of justice that is associated with the concept of worthiness to be happy has both a political and an eschatological character, applying to both 'society' and the 'universe.'

---

[147] Morris Cohen, 'A Critique of Kant's Philosophy of Law,' in *The Heritage of Kant*, ed. George Tapley Whitney and David Frederick Bowers (Princeton University Press: Princeton, 1939), 280.
[148] Ibid.
[149] Ibid., 280-1.

Allan Wood perceives the need for an explanation here as well. His approach is as follows. Kant refers to 'virtue,' or 'the moral good,' as 'the condition of our worthiness to be happy.' He does this because he sees the moral good (virtue) as both 'the unconditioned good' and 'the supreme condition of all else which is good.' In other words, for Kant, an agent's happiness (the sum of her 'natural ends') lacks 'goodness,' in the sense of 'moral validity,' if she is not moral or virtuous. It is virtue's 'role,' here, to 'provide the condition' without which our happiness is (objectively) bad, or morally invalid. And *this*, Wood concludes, is why Kant characterizes virtue as 'the condition of our worthiness to be happy.'[150] The question that Wood takes Kant to be answering goes something like this. Under what condition(s) is happiness a good for morality, or good from morality's point of view, or good from a point of view that is itself moral? Kant's answer, according to Wood, is that happiness 'is a good *for morality* only insofar as it is conditioned by [virtue].'[151] Evidently, at least on its face, Wood's reading conflates virtue's being a condition for worthiness to be happy with virtue's being a condition without which happiness is not a good *for morality*.

In spite of this ambiguity, Wood's approach captures fairly clearly the idea that there is an objective point of view—the point of view of 'morality' in some sense—from which the happiness of immoral agents may be judged to be a bad thing. And Wood opens a path, at least, to Kant's idea that, from this point of view, while they *can* be happy, immoral agents *ought not to be*. He does not make morality a condition that is simply internal to happiness in the way that Henrich does. Nor does he promote the idea (present in Cohen's reading) that, from the point of view of Kantian 'morality,' virtue positively *demands* the happiness of its bearers.

Paul Guyer offers the most sustained discussion of Kant's practice in regard to this idiom. He begins by characterizing Kant's 'frequently reiterated characterization of virtue as the worthiness to be happy' as a 'profound mystery in [his] ethics' and identifies the notion of worthiness to be happy as the main concept 'through' which Kant represents the connection between 'virtue and happiness.'[152] Guyer observes, too, that 'by the time of his published writings' Kant's 'equation of virtue with wor-

[150] Wood, *Kant's Moral Religion*, 80.
[151] Ibid., 84.
[152] Guyer, *Kant on Freedom, Law, and Happiness*, 117.

thiness to be happy…had come to seem so obvious to him as to need no real explanation.'[153]

In the course of clearing up this 'profound mystery,' Guyer makes at least the following five useful points. First, he recognizes that Kant's use of the 'worthiness to be happy' idiom expresses a distinctive point of view, one that Kant simply takes for granted. Second, Guyer sees that the idiom's sense is far from obvious and that an explanation is in order. Third, he recognizes that the idiom comes into currency well in advance of Kant's critical period. Fourth, Guyer understands that it gives expression to a single (if complex), more or less consolidated, intellectual commitment. Fifth, he recognizes that the notion of worthiness to be happy has something to do with desert.[154]

Guyer favours a reading that makes Kant's use of the idiom the end product of a relatively straightforward line of antecedent reasoning. And he sees this reasoning as focused on the conditions under which agents *deserve* to be happy. He interprets Kant's 'equation' of virtue with worthiness to be happy as the expression of a certain construal of distributive justice; a notion, here, that has reference to 'merit,' right,' and 'entitlement.' And he identifies two main ideas, prevalent across a number of *Reflexionen* from the mid-late 1770s, which appear to be the basis for Kant's 'equation of virtue with the worthiness to be happy.'[155] First, there is

> the general claim that worthiness to enjoy a good, whether natural or otherwise, presupposes that one *merits* it, or has earned the right to it *by one's own actions*, and that as the product of our only genuine *free* actions, or, in other words, as one of the two possible outcomes of our only genuine *actions* at all, virtue is the only ground for any merit at all.[156]

Guyer's interpretation assimilates 'worthiness' to 'merit.' He glosses 'merits it' as 'has earned the right to it,' specifies this 'right' as a consequence of 'one's own actions,' and characterizes the relevant actions as 'free' ones. The thrust of Guyer's interpretation to this point may be summed up in what he takes to be Kant's dictum

[153] Ibid., 119. Guyer does not balk at referring to Kant's gloss as an 'equation,' which, strictly speaking, is misleading, although this does of course reflect Kant's own way of expressing himself.
[154] Guyer goes to the heart of the matter when he asks '[w]here…the element of *desert* that it seems natural to associate with the idea of worthiness come from?' (ibid., 118).
[155] Ibid., 119. See, for example, *R* 7204 (19: 283 [470]); *R* 7211 (19: 286 [472]); *R* 6280 (18: 547 [352]); *R* 7058 (19: 237).
[156] Ibid.

here: that 'it is from an agent's *free actions* that his entitlement to happiness arises.'[157] The second idea is that

> since universal happiness or a system of happiness is not a merely natural good or a product of merely natural behavior but something that even under the best of circumstances would be produced only by virtuous action, enjoyment of one's own share of universal happiness is the *appropriate reward* for genuine merit—payment in kind, as it were.[158]

Guyer's reading (and indeed, presumably, the pre- or semi-critical thinking that he is interpreting) makes morality internal to happiness—at least, that is, to 'universal happiness.' On this reading, the happiness of which virtuous agents are worthy is not their 'own' happiness (which, presumably, *could* be attained by immoral means[159]), but their 'own share of universal happiness.' Guyer construes this as 'reward' and 'payment in kind.' Virtuous action, defined as action that aims at the happiness of all, entails 'genuine merit.' And 'virtue is worthiness to be happy *precisely because* virtue concerns the universal distribution of a good in which one is then entitled to one's own fair share.'[160] As Guyer puts it earlier, 'only insofar as our own happiness as part of this larger whole is a product of our own free will do we in any sense deserve it.'[161]

Here again, as we saw in the case of Henrich's reading of Kant's idiom, we have a notion of 'distribution' that is not really what it appears to be. The 'system of happiness' that Kant has in mind is here conceived of as 'the systematic *distribution*' of happiness,'[162] but the object of the distribution (happiness *qua* the happiness of which moral agents are worthy, or happiness in the sense of one's 'own share of universal happiness') is such that full participation in the universally binding task of making that distribution a reality, is internal to it. If virtue '*concerns*' this 'universal distribution,' it is a matter of virtuous action's consisting in a striving *for* that, in effort aimed *at* that state of affairs. Happiness 'is…the appropriate consequence of vir-

---

[157] Ibid. Cf. *R* 7058 (19:27).
[158] Ibid. (original emphasis modified).
[159] As Guyer points out later, there is a major difference between 'aiming at one's *own* happiness alone' and 'aiming at the happiness of *all* including oneself' (ibid., 344). Indeed, practical reason's highest aim (i.e., the highest good) 'includes as an *indispensible component* the systematic happiness of mankind rather than the self-centered happiness of the individual' (ibid., 9; my emphasis).
[160] Ibid., 123 (my emphasis). Guyer cites *R*'s 6989, 7049, 7197, and 7202 (19: 221, 235, 270-1, and 279 respectively).
[161] Ibid., 120. Cf. *R* 6971 19: 216-17, the note on which Guyer is commenting. See also *R* 7058 19: 237 (458); *R* 6867 19: 186 (444). See also *R* 7199 19: 272-3, a note from the early 1780s, and ibid., 103.
[162] Ibid., 100 (my emphasis).

tue *because* it would be an inevitable consequence of virtue under ideal circumstances—that is, if virtuous action necessarily had its intended effect.'[163] Thus Guyer avers that, on the pre-critical view expressed in Kant's ongoing deployment of the 'worthiness to be happy' idiom, it is *because* happiness would follow inevitably from virtue 'under ideal circumstances' that happiness is connected to virtue as its 'appropriate consequence.' And, as we saw above, he avers that it is '*precisely because* virtue concerns the universal distribution of a good in which one is…entitled to one's own fair share,' that is, just to the extent that one has also freely and fully worked for that universal distribution, that ' virtue is worthiness to be happy.' In each of these formulations, Guyer's interpretation makes morality internal to happiness. But he also represents the relationship between morality and happiness—the sense of Kant's concept of worthiness to be happy—as a variety of desert.

To sum up, Guyer takes Kant's idiomatic 'equation' of virtue with worthiness to be happy to express the idea, pre-formed just in advance of the critical period, that happiness is the 'appropriate reward' for virtue, that virtue entails 'genuine merit,' that virtuous agents have a 'right' to happiness, that happiness is 'payment in kind' and something to which they are 'entitled.' And this view tends, at the same time, to shade off into the view that happiness is an 'appropriate *consequence*,' or 'effect' of morality, to the extent that, 'under ideal circumstances' (given the presupposition that nature and its laws are the work of a wise and omnipotent author whose own 'holy' nature it is, also, to embody and promulgate the moral law), the happiness of virtuous agents would be simply 'an inevitable consequence' of their morality. But the form in which this latter view is expressed, as I have suggested, belies the fact that Guyer interprets Kant, here, in a manner that makes morality and happiness not extrinsically related (as cause and effect, or as normative condition for distribution and object thereof), but intrinsically so.[164] Guyer takes this to be the point of view that we encounter here, first, in these *Reflexionen,* in the process of its formation.

[163] Ibid., 120 (my emphasis). This same mechanism is expressed in Kant's claim, in the *Groundwork*, that 'a kingdom of ends would actually come into existence through maxims which the categorical imperative prescribes as a rule for all rational beings, *if these maxims were universally followed*' (*Gr* 4: 438 [45]). Wood is right to read this as a claim pertaining to Kant's *summum bonum* as well, to an ultimate scenario in which all of the agents participating in it would be both moral and happy (Wood, *Kant's Moral Religion*, 127).

[164] In chapter 2, I discuss a different facet of Guyer's reading of Kant's 'worthiness to be happy' idiom—one that *does* represent happiness and morality as extrinsically related.

And he understands it to be a point of view that Kant simply takes for granted henceforth.

Finally, Steven Smith reads Kant's regular glossing of 'morality' as 'worthiness to be happy' as a more or less well-defined strategy for dealing, preemptively, with a particular problem. Smith argues that

> Kant has anticipated the objection that the interests of reason [i.e., pure reason's interest in morality and embodied reason's interest in the unification, harmonization, and realization of the particular agent's empirical interests, in short, her happiness] are simply heterogeneous and irreconcilable by consistently formulating the moral law with reference to happiness; he never speaks of moral worth merely, but always of 'worthiness to be happy.'[165]

More pointedly, Smith takes this consistent practice in respect of Kant's formulations of the moral law to express the latter's view that 'moral worth is *necessarily* interpreted by moral reason as "worthiness to be happy,"'[166] that 'a rational agent approves *a priori* of a strict proportion between virtue and happiness,'[167] and that this means that 'some unexplained scheme of deserving is presupposed.'[168]

Smith observes, in short, that Kant's idiomatic use of the notion of worthiness to be happy shows that Kant takes there to be an 'obvious' and 'ideal linkage between virtue and happiness.'[169] But this means, more specifically, that '"worthiness to be happy" implies *desert*' and that this 'in turn implies some kind of contractual scheme.'[170] Again, though, Smith takes this presupposed 'scheme of deserving' to be precisely '*unexplained*.' He goes on to explain it, making explicit what he takes to be Kant's implicit strategy.

He argues that '[o]n Kant's view, virtue is deserving, even though it is not undertaken in order to earn.'[171] He argues, too, that the implied 'scheme' is one 'in which rewards can be earned.' And he observes that this then warrants the introduction of 'God,' or a being who 'reward[s] those who have made themselves worthy of happiness, i.e., [those who] have earned it.'[172] Smith observes, however, that in his *Lectures on the Philosophical Doctrine of Religion* 'Kant rejects precisely this idea' (i.e., the idea that, by their morality, human beings can put God in their debt, so that

[165] Smith, 'Worthiness to Be Happy': 177.
[166] Ibid., 169 (my emphasis).
[167] Ibid.
[168] Ibid., 184.
[169] Ibid., 173. Smith observes, correctly, that this 'linkage' is not obvious and that '[a]n explanation is required' (ibid.).
[170] Ibid., 185.
[171] Ibid., 186.
[172] Ibid., 185.

God is *bound* to reward them for it).[173] Smith resolves this apparent tension by restricting Kant's notion of reward for virtue (precisely, that is, in connection with the concept of worthiness to be happy) to a primarily political frame of reference. He defines the 'scheme of deserving' with which he began in terms of a primarily *human* community—of which God is a possible member (but such that God's membership remains an open question)—in which virtue does not after all *necessitate* happihappiness, but in which the latter is rather an expression of kindness and a correlate of praise (of the moral subject). He argues that

> [t]he judgment that God should reward a virtuous person with happiness is cognate with the desire to praise him [i.e., the virtuous person]; it is a completion of his [moral] gesture, and is morally motivated insofar as it expresses respect for and concurrence in the virtuous maxim of the act of the one praised.[174]

The 'true basis of the idea of virtue as "worthiness to be happy,"' writes Smith, given the latter concept's implications with respect to desert, of an anterior 'contract,' or 'a solicited concurrence of free will,' is that there is a kind of 'moral contract offered by the universalizing will'—a contract to reward virtue with praise and, to whatever extent possible, with happiness. This is a contract whose execution by the moral community may or may not be realized with God's assistance—just depending upon whether the latter exists or not. If God is taken to exist, however, then the sense of Kant's gloss is that '[w]hen the order of things represented by the concept of the highest good is realized, only righteous purposes will be crowned with success, and thus people will only be happy when they are moral (although not necessarily happy *because* they are moral).'[175]

Conclusion

In this chapter, I executed four main tasks. First, I established the originary presence, pervasiveness, and longevity of Kant's habit of glossing 'morality' as 'worthiness to be happy.' Next, I discussed the expression's logical structure. I characterized 'worthiness to be happy' as a particular kind of predicate, that is, as a three-place relation. Third, I distinguished the *relata* whose connection the idiom represents and argued that the variety of ways in which Kant represents 'morality,' in the latter's primary, 'forward-looking' sense are independent of the idiom's representation of the rela-

[173] Ibid.
[174] Ibid., 187.
[175] Ibid., 180.

tionship between that (morality, irrespective of what that turns out to be for Kant) and happiness. Fourth, I pointed to the paucity of commentary on Kant's use of the 'worthiness to be happy idiom,' problematized this oversight, and discussed its significance. I also acknowledged that there are, nevertheless, a handful of readers who have addressed this topic directly. I discussed their explanations of Kant's use of the concept in question. Now, by way of concluding this chapter and setting the scene for what follows, I will briefly note some of the key features that distinguish my approach from theirs.

Taken together, these commentators provide useful insight into the background, sense, and purpose of Kant's original and ongoing deployment of his concept of 'worthiness to be happy.' Their readings do not, however, penetrate deeply enough into what I take to be the bedrock of Kant's critical period thinking about freedom and moral accountability. The constitution of that substrate is more obscure than their readings suggest. The substrate itself is also more deeply set. This thesis as a whole contends with this obscurity and seeks to clarify what it conceals. For now, however, I simply offer, without embellishment, a series of counterclaims, which will be familiar from my introduction, but which I repeat here in direct response to these other commentators (or rather my reading of them).

First, the happiness of which moral agents are worthy is not distinct from the happiness of which immoral ones are unworthy. To the extent that Kant glosses 'morality' as 'worthiness to be happy,' he does not take morality to be internal to happiness. Kant's gloss shows that he presupposes that immoral agents *can* be happy (in the same sense that moral ones can) and it promulgates his strong view that they ought not to be.

Second, the gloss represents a certain order of necessitation: it reflects Kant's view that immoral agents ought *necessarily* (i.e., categorically) to be unhappy. This is the point at which Kant's notion of desert shows up in relation to the concept of worthiness to be happy. Kant's gloss does not communicate the idea that virtue ought necessarily to be rewarded. Kant's 'philosophy of law' *does* subsume the view that offenses *must* (necessarily) be punished and that labourers, for example, who have fulfilled the terms of their contracts must (necessarily) receive the remuneration *promised* them; but the relationship, in that thinking, between contracted labour and

remuneration, on the one hand, and morality and happiness, on the other, is not symmetrical.

Third, Kant's gloss expresses a concern for something beyond the mere 'moral validity' or 'moral invalidity' of the human being's happiness. The main idea that is implicit in Kant's use of the 'worthiness to be happy' idiom is that 'morality,' in Kant's *secondary*, *backward-looking* sense, demands not only that happiness be 'morally valid,' but positively demands the unhappiness of immoral agents, demands that this be actualized, and declares (prospectively at least) that this *unhappiness* is good and morally valid in turn.

Fourth, Kant's use of his 'worthiness to be happy' idiom does give expression to a single (albeit complex), more or less consolidated, intellectual commitment (I have detailed some of its main parts in the three foregoing paragraphs). But it expresses something more immediate and basic as well. It also gives expression to Kant's antecedent, fundamental commitments to the *practices* of (retributive) punishment and (retribution-sensitive) imputation.

Fifth, Kant's gloss is not a self-conscious strategy for forestalling the objection that morality and happiness constitute 'heterogeneous' and 'irreconcilable' ends. It achieves this aim, to be sure, but it does not do so as a piece of reasoning on Kant's part. If it may be regarded as a 'strategy' at all, then it is one that serves the practical commitments that Kant has *in advance* of the development of his 'critical' moral theory's even getting under way. Kant's gloss is a habit and, like all habits, it is of course also a strategy. It *shows up* in his argumentation, however, as an intractable, axiomatic body of implicit thoughts and commitments. But it is there in advance of Kant's reasoning about the relationship between morality and happiness. It is not a result of the latter.

In this chapter I have argued that Kant's gloss represents morality and happiness as two states of the human being's affairs that stand in a particular relationship to one another. In the next chapter I discuss this relationship and so explicate the *significance* of Kant's habit of glossing 'morality' as 'worthiness to be happy.'

CHAPTER TWO

'Worthiness to be Happy' and the Relationship between Morality and Happiness

> [T]he system of morality is…inseparably combined with the system of happiness.[176]

Introduction

It is generally agreed that resolution of the apparent tension between morality and happiness is a centrally important problem for moral theory.[177] In the first *Critique*, Kant refers to his basic position on this matter when he asserts that 'the system of morality is…*inseparably combined* with the system of happiness.'[178] As a whole, however, Kant's body of work offers resources for several disparate answers to the question how this 'combination' and its 'inseparability' are to be construed.

This thesis focuses on one way in which Kant answers this question: the way that is encapsulated in his habit of glossing 'morality' as 'worthiness to be happy.' This chapter, in particular, presents the rudiments of this answer by making a number of key distinctions. In the next two chapters I will fill this picture out and argue that, within the ambit of this habit—in it and under it as it were—Kant really does, most *fundamentally* (albeit not always and only), take morality and happiness to be related

[176] *KrV* A809/B837.
[177] See, for example, Friedman, 'Virtue and Happiness: Kant and Three Critics': 95.
[178] *KrV* A809/B837 (my emphasis).

in the way that the gloss suggests. The aim of this chapter is a first, preliminary step in that direction: to show that Kant's gloss expresses the presupposition that immoral agents *can* be happy (in the same sense that moral ones can) and that it promulgates his strong view that they ought not to be.

In order to demonstrate my claim that Kant's gloss represents happiness as a state of affairs such that immoral agents *can* enjoy it, but *ought not to*, this chapter executes three main tasks. First, I show that Kant's gloss represents morality and happiness as states of an agent's affairs that are *extrinsically* related. Second, I specify this claim more closely by arguing that Kant's gloss represents these states of an agent's affairs as ones that are not only extrinsically, but also *necessarily* and *normatively* related. Third, I argue that Kant's notion of worthiness to be happy points to an extrinsic, normative, necessary relation that holds, not between morality and happiness, but between *unhappiness* and *immorality*. In other words, I argue, Kant's gloss pertains to the normative necessity of a particular kind of desert—the immoral agent's desert of unhappiness. However, as I also show, worthiness to be happy is not desert of happiness. This is a key asymmetry in Kant's thinking, generally overlooked. Indeed, Kant's use of the expression, 'worthiness to be happy,' does not, strictly speaking, represent happiness and morality as elements that are *immediately* related at all. Rather, it evinces a thought about the relationship between morality and the deontic *possibility* of happiness, a possibility that, in its ultimately theological inflection, requires benevolence (in addition to the normative dictates of pure practical reason) for its realization. The deontic possibility of happiness (but only given morality) is formally equivalent, as I show, to the deontic *necessity* of unhappiness (given immorality). Thus Kant's habit of glossing 'morality' as 'worthiness to be happy' is able to represent this dreadful necessity without adverting to it explicitly.

It might be objected here that, in a very un-Kantian spirit, I am playing on the asymmetry that characterizes the 'practical-*rational*' distinction between 'deontic possibility' and 'deontic necessity' in order to assert that there is an associated asymmetry between the '*sensible*' notions of 'happiness' (the production of which, with my various qualifications, I claim is *permitted* just in case the object of that 'production' is a moral agent) and 'unhappiness' (the production of which state of

affairs I take to be *commanded*, that is, just case in the object of said 'production' is an immoral agent—given, especially, that the making- or leaving-happy of immoral agents would be *forbidden* to a particular kind of agent). Given that this objection is on the horizon, here, I must pause to address the possible misunderstanding upon which such an objection would turn.

Let me begin this *excursus* by pointing out, once again, that it is Kant's explicit contention that the law of *punishment*—where the latter is regarded as a law commanding that a particular course of action be executed (by the sovereign or her legal representative) within the *polis*—is a categorical imperative.[179] This means, in Kantian terms, that the course of action that the law of punishment prescribes actualizes the more fundamental command that a particular *end* (namely, the punishment of criminals) be adopted *in general*. Here, however, in Kant's thinking about the relationship between crime and punishment—without even referring, that is, to the relationship that I claim that Kant takes to hold between *immorality* and *unhappiness*—we are already faced with the very asymmetry that the objection (set forth above) would put into question.

To reward those who abide—or do more than is required—by the laws of the land is an *option* for the sovereign, it is permitted him or her. It is 'deontically possible' for the latter. The sovereign is *permitted* to make the happiness of his or her subjects an end and so, too, to make them actually happy—given the law-conformingness, or the supererogatory character, of their concrete participation in the life of the *polis*. But there is no constraining, necessitating 'law of reward' (although, for Kant, *rightful* 'remuneration'—for contracted effort, or effort for which payment is *promised*—is mandated *a priori*). In any case, if such a principle (of reward) *were* instituted by way of positive law, this principle would be no genuinely *categorical* (i.e., moral) law.

By contrast, if the law of punishment (in the political context of its enactment) *is* a categorical imperative, as Kant says it is, then, while the sovereign is *permitted* to take action to render particularly meritorious or even simply law-abiding citizens happy (so that such action and the having of the end at which such action aims is *deontically possible*), the same subject (the sovereign) is *commanded* to take action to

[179] *MdS* 6: 331 (105).

insure a *forfeiture* of happiness (to the degree required by the *ius talionis*) on the part of law-breakers.[180] And this forfeiture, which is instantiated as an empirical state of the punished subject's practical affairs (addressed, that is, to the will that wrongfully wills *both* crime and happiness, in the form of a coercive limitation of such a will's capacity for the realization of its empirically given desires), effects an end, the having and pursuing of which (i.e., just in case the law of punishment really is a categorical imperative, as Kant claims) is *deontically necessary*. This already involves a kind of 'transfer,' then, of the asymmetry that I identified above from the 'practical-*rational*' to the 'sensible.' In other words, when it comes to the sovereign's aim of actualizing, or of making *concrete*, the *a priori* relationship that Kant takes to hold between the notions of 'transgression' (here, for the moment, 'political' crime) and 'punishment,'[181] we are already presented (by Kant himself) with an intimate association, on the one hand, between 'deontic possibility' and 'reward,' which (regarded under the rubric of the action that offers the latter) effects, funds, enables happiness, and, on the other hand, between 'deontic necessity' and 'punishment,' which (regarded under the rubric of the action that metes out punishment-for-transgression) effects the punished agent's unhappiness.

The 'transference' that allows me to say that, by Kant's use of the 'worthiness to be happy' idiom, he signals his commitment to the idea of a 'law of unhappiness,' is grounded in the following two considerations. I set these forth here, now, in the most rudimentary fashion and in service of my attempt to forestall the objection that my argument surreptitiously conflates the 'rational' and the 'sensible.' The first of these considerations, then, is that for Kant the concept of death-for-murder has a special status; that is, that unique among politically situated punishments, death-for-murder punishes the 'inner wickedness' of the murderer and, as such, is aimed not at her punishable *deed* only, but at *her*, at her *will*, given the latter's radically immoral character. The second consideration is that, in Kant's view, to suffer punishment is, in general, to be rendered unhappy.

---

[180] I discuss Kant's definition of 'punishment' in terms of this 'forfeiture' both in this thesis' introduction and then, at greater length, in the first section of chapter 3, below.

[181] See *KpV* 5: 37 (34), where Kant simply asserts that 'there is *in the idea of our practical reason* something further that accompanies the transgression of a moral law, namely its deserving punishment' (emphasis modified). I explore the significance of this claim at greater length in chapter 3.

The move by which I shift from talking about an explicitly Kantian 'law of punishment' to talking about an equally Kantian, but mostly implicit 'law of unhappiness' turns on the definitional proximity—indeed the identity, I suggest, when properly qualified in terms of the sovereign or divine action that aims (or would aim) at their actualization—of the notions of 'being-rendered-unhappy' and 'being-punished'; and on the extremely close definitional proximity of the notions of 'immorality' and 'crime,' specifically in Kant's thinking about the crime of murder. The move by which I make this shift also turns on an analogy between the *notional* relationship (which Kant clearly construes as one that is cognized *a priori*, that belongs to pure practical reason itself) between politically-contextualized 'transgression' and 'punishment,' on the one hand, and between eschatologically- or ethically-contextualized 'transgression' and 'punishment,' on the other. I suggest that, on the Kantian scaffold, we are confronted, however, with a relationship that exceeds the merely analogical, since the practice that actually enacts the 'thought' that crime and punishment, immorality and unhappiness, are normatively related *a priori* 'literally' *identifies* the respective analogues (the pairs 'crime-punishment' and 'immorality-unhappiness') with one another.

The main sticking point, here—the point upon which the objection would largely turn—pertains largely to the manner in which I am associating the notions of *punishment* and *unhappiness* (the claim that Kant identifies criminality and immorality in the limiting case of murder—but not elsewhere—is likely to be more readily accepted). Once it is recognized, however, that an 'asymmetry,' which falls between 'sensible' states of affairs whose willing and production are deontically possible and deontically necessary respectively, is already implicit in Kant's committed thinking about the relationship between political crime and punishment (and between law-abidingness and reward), it still remains to show that the 'thing' that is deontically necessitated, given the ostensibly *a priori* notional connection of crime and punishment—namely the execution of (proportionate) *punishment* on criminals in a political context—is equivalent to the 'thing' that is deontically necessitated by the *a priori* notional connection that I claim holds, for Kant, between immorality and unhappiness. In other words, it remains to be argued that a criminal's being-punished is

in some important sense *equivalent* to her being-rendered-unhappy (and so, too, that the having of these states of affairs as ends is also equivalent in the relevant sense).

Now, I claim that Kant's deeply rooted habit of glossing 'morality' as 'worthiness to be happy' shows in a somewhat oblique manner that he takes there to be an *a priori* notional relationship between immorality and unhappiness. And I claim that his mostly implicit thinking about this relationship endorses and, eschatologically speaking, extends the 'reach' of his *explicit* thinking about the relationship between politically-situated 'transgression' and politically-contextualized 'punishment.' I also assert that Kant appears to simply take the *aprioricity* of this thinking about the relationship between immorality and unhappiness for granted. He asserts, but does not demonstrate, in other words, that reason, rather than nature, or habits of judging socially acquired, is the source of this thinking.

Note, however, that these major claims (core elements of the subject matter of this thesis) are independent of the one that I am advancing now. For the moment, in anticipation, again, of the possible objection to which I adverted above, I am proposing only that a certain asymmetry divides what the political sovereign is (morally) permitted to do in behalf of law-abiding agents from what this same subject is (morally) *constrained* to effect in the experience of law-transgressing ones. This is what it means to claim that Kant takes 'the law of punishment' to be a categorical imperative (although he never *shows* that it is), while he does *not* take there to be any such thing as a 'law of reward' (with the *caveat* regarding 'rightful remuneration' cited above). And I am proposing that the 'thing' that is (morally) practically necessary under the rubric of 'punishment,' takes shape in a 'sensible,' spatio-temporal state of affairs.

In order to show that, given the overlap between the political and the ethical that belongs to his conception of what takes place in cases of capital punishment for murder, Kant also holds that the principle that demands the 'making-unhappy' of immoral agents (by 'God,' finally) may be regarded as a categorical imperative (*just in case*, Kant's 'law of punishment' can be—which perhaps it cannot), I do not have to justify the claim that 'the law of unhappiness' really *is* a categorical imperative. I only have to show that there is an equivalent asymmetry between what is thought 'in' Kant's *explicit* glossing of 'morality' as 'worthiness to be happy' and in his *implicit*

association of 'immorality' with 'unworthiness to be happy.' The fact that 'happiness' and 'unhappiness' refer to states of agents' 'sensible' affairs is beside the point—since this is also true for Kant of 'punishment' and since the latter refers to a particular mode of 'being-rendered-unhappy.'

Again, then, punishment, is an *empirical* and *sensible* matter—precisely because, from the point of view of the agent that enacts it, it unilaterally coerces its empirical object (the political subject) in such a way that she forfeits happiness; punishment renders its object unhappy, that is, by occluding access to happiness (or to the means to happiness). For certain purposes, we may prefer to regard unhappiness as distinct from punishment; in particular, we may prefer to regard unhappiness as punishment's *effect*. But this conceptual distinction remains ambiguous when actual instances of punishment come into view. From the perspective of the *patient*, to suffer punishment is simply to be rendered unhappy; to punish (in a political context) is to render-unhappy. The asymmetry between what the law *permits* and what the law *commands*, here—namely (and respectively), the sovereign's making it her end that some law-abiding agent or other be happy and that some criminal or other be unhappy—*does* involve a kind of 'transfer' from the 'practical-*rational*,' which is the supposedly *a priori*, notional connection between the concepts of transgression and punishment (both of which are actualized empirically), to the 'sensible,' which is the *actual* punishment of *particular* criminals. But this move is present, already, in Kant. I am not forcing it onto his thinking. To the extent that there is something 'un-Kantian' about my way of proceeding, here, this is simply another instance of the tension that characterizes Kant's own rather varied and uneven thinking about the relationship between crime and punishment—and between immorality and unhappiness.

I will now proceed with my analysis of Kant's thinking about the latter, but must do so indirectly, by way of an exploration, rather, of his thinking about the relationship between *morality* and *happiness*.

Morality and happiness as extrinsically related

When it comes to answering the question how morality and happiness are related, we are presented with at least the following four options, each of which implies distinct

conceptualizations of its *relata*. Either (1) morality is in some sense identical to happiness; (2) morality (or moral deeds) is (are) a (physical) cause of happiness; (3) morality (along with some other necessary, but material conditions) is a formal/transcendental condition of the possibility of happiness; or (4) morality is a deontic-normative condition for the permissibility (the deontic possibility) of action that aims at happiness (or its distribution).

Now, it is generally agreed that Kant's thinking about the relationship between morality and happiness differs markedly from the views of his predecessors, both ancient and early modern.[182] There are two clear ways in which it does so. First, it is a central and distinctive feature of Kant's moral theory, at least in its ultimate form, that he denies happiness any constitutive role (as incentive, motive, interest, or whatever) that would make the latter a condition of the very possibility morality (or moral deeds) in the first place.[183] Second, it is clear that, unlike the Stoics and Epicureans who, each in their own way, took happiness and morality to be identical, Kant never takes the relationship between morality and happiness to be an analytic one (i.e., as in [1], above).[184] Apart from these two core features of his moral theory, however, Kant's thinking about morality and happiness invites a significant variety of readings.

Kant is sometimes taken to hold that morality and happiness are entirely heterogeneous and irreducibly alien to one another, not only conceptually distinct, but such that they share no deeper affinity of any kind at all (in contrast, say, to triangularity and trilaterality).[185] Some view Kant's treatment of happiness in relation to morality

---

[182] Scholarly consensus holds that Kant's ultimate views about the relationship between morality and happiness are a repudiation or, at least, a radical rereading, of the main elements of prior accounts. See *KpV* 5: 111-12 (93-4); *LEC* 27: 247-50 (44-6); and *LEM*$_2$ 29: 602-4. See also Denis, 'Kant's Conception of Virtue', 506-7; Hill, 'Happiness and Human Flourishing in Kant's Ethics': 149-51; Irwin, 'Kant's Criticisms of Eudaemonism'; Allen Wood, 'The Supreme Principle of Morality,' in *The Cambridge Companion to Kant and Modern Philosophy*, ed. Paul Guyer (Cambridge: Cambridge University Press, 2006). But cf. Schopenhauer, *Basis of Morality*, § 3 (49).

[183] To the contrary, as Kant puts it in one particularly compelling turn of phrase, when it comes to morality the 'sirens-song of happiness [*Sirenenstimme der Glückseligkeit*]' (*R* 7315 19: 312) poses an ever-present *risk* to morality.

[184] See, for example, *KpV* 5: 111-12 (93-4); also Guyer, *Kant on Freedom, Law, and Happiness*, 348; Wood, *Kant's Moral Religion*, 85.

[185] See, for example, Engstrom, 'Happiness and the Highest Good', 122; Gene Fendt, *For What May I Hope?: Thinking with Kant and Kierkegaard*, American University Studies (New York: P. Lang, 1990), 83; Keyworth, 'Kant's Concept of Happiness in the Moral Argument': 23; Mannion, 'Kant and the Defeat of Egoism: Schopenhauerian Concerns and Some Reappraisals and Rejoinders': 224; N. Ranasinghe, 'Ethics for the Little Man: Kant, Eichmann, and the Banality of Evil,' *Journal of Value*

as distorted and inhumane.[186] Others allow that morality might well entail unhappiness, but that it need not do so.[187] Others argue that, while Kant restricts the role of happiness in moral matters, he does not disregard it altogether.[188] Yet other readers understand Kant in a manner that implies, at least, that he takes morality to be 'internal' to happiness. As Guyer puts it, 'virtue and happiness are not separated but at least ideally joined at the hip';[189] conceptually, at least, there is an intimate relationship between them.[190] In 1944 H. J. Paton found it necessary to assert that '[h]appiness…play[s] a much greater part in Kant's moral philosophy than is commonly recognized.'[191] The first part of Paton's claim (concerning the significant role of happiness in Kant's moral theory) remains apposite today, of course, but the problem that he identifies is no longer so pressing.

In this section, I argue that Kant's gloss represents morality and happiness as distinct states of an agent's affairs that are *extrinsically* related.[192] An important implication of this claim is that Kant's regular use of the 'worthiness to be happy idiom' shows that he takes the notion of an immoral, thus unworthy, but nevertheless happy agent to be perfectly coherent. An immoral agent's happiness is *deontically* impossible, but not impossible in any other (inherent) sense. It is impermissible (in a sense to be explained below), or 'impossible' in the sense that my driving an eight-

---

Inquiry 36, no. 2-3 (2002): 311; Wood, *Kant's Ethical Thought*, 311-12. Hegel appears to take Kant along these lines (Hegel, *Phenomenology of Spirit*, § 599 (365)); cf. Terry Pinkard, *Hegel's Phenomenology: The Sociality of Reason* (Cambridge: Cambridge University Press, 1996), 288-94; Charles Taylor, *Hegel* (Cambridge: Cambridge University Press, 1975), 376-8, 385-8; and Schopenhauer writes (in what he takes to be a faithfully Kantian spirit) that 'virtue is obviously quite foreign to happiness' (Schopenhauer, *Basis of Morality*, § 3 (49)). See *Gr* 4: 393, 400-401 (7, 13-14); *ÜdG* 8: 283 n., *KdU* 5: 431 (298-9), *MdS* 6: 387-8 (153-4).

[186] See Richard Taylor, *Good and Evil* (New York: Macmillan, 1970), chapter 8.

[187] See, for example, Kienzle, 'Macht Das Sittengesetz Unglücklich?'.

[188] See Beck, *A Commentary on Kant's Critique of Practical Reason*, 10; Cicovacki, 'Illusory Fabric': 395; Hill, 'Happiness and Human Flourishing in Kant's Ethics': 148, 150-2, 155; Hinman, 'On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty': 252; Mannion, 'Kant and the Defeat of Egoism: Schopenhauerian Concerns and Some Reappraisals and Rejoinders'; Schroeder, 'Some Common Misinterpretations of the Kantian Ethics': 436; Wike, *Kant on Happiness in Ethics*, xix.

[189] Guyer, *Kant on Freedom, Law, and Happiness*, 379.

[190] Ibid., 117, 378.

[191] Paton, 'Kant's Idea of the Good': xix. By way of contrast, cf. A. Kelley, 'Kant on Freedom, Happiness, and Peace,' in *Spiritual and Political Dimensions of Nonviolence and Peace*, ed. D. Boersema and K. G. Brown (Amsterdam: Editions Rodopi B. V., 2006), 175.

[192] With respect to this basic claim see Engstrom, 'Happiness and the Highest Good', 106; Sikka, 'On the Value of Happiness: Herder Contra Kant': 539 and *KpV* 5: 110-11 (92-3); cf. *KdU* 5: 208-9 and *KrV* A813/B841.

een-wheeler is impossible, given that I am not licensed by my government to operate a vehicle with air brakes.

Now, someone might object that I am sliding in an unjustifiable, or at least suspect, manner from the notion of 'impermissibility' to that of 'impossibility'—and vice versa. Note, however, that this is a matter of definition. The 'impossibility' in question is deontic, not material, and is directly connected with the *normativity*, for Kant, of the representation of the relationship between morality and happiness that is expressed by his habit of glossing 'morality' as 'worthiness to be happy.' In the pre-critical period, as I will show below, Kant sometimes represents the relationship between happiness and morality as intrinsic in the sense that he sees morality as a condition of the possibility of happiness (which entails a particular account of happiness; an account of happiness, in other words, that regards morality as *internal* to it). In the critical period, however—particularly in the second *Critique*—he tends to see the relationship between morality and happiness as an *extrinsic* one. He thinks that we are permitted, at least, to *regard* moral action as action that would lead, on the whole, to happiness, under natural law—so that moral deeds, which are empirical events (but for Kant, of course, not only that), are regarded as the *cause* of happiness (which entails a teleological account of the laws of nature that sees the latter as the legislation of a divine author).

On neither of these accounts is the relationship between morality and happiness represented in a manner whose explication calls for the notions of 'deontic possibility,' 'deontic impossibility,' and 'deontic necessity' that I am deploying now. The 'extrinsicness' or 'intrinsicness' of the relationship between morality and happiness is not the decisive factor, here, just as such. Rather, again, the decisive factor is the *normativity* of the representation of the ostensibly *a priori* notional relationship between (im)morality and (un)happiness that is expressed by Kant's gloss—a relationship whose empirical instantiation is conceived of in terms that are neither analytic, nor natural-causal. By contrast with these other approaches, Kant's habit of glossing 'morality' as 'worthiness to be happy,' which both predates and persists throughout his critical period, gives expression to a conception of the relationship between morality and happiness in whose elucidation the notion of 'deontic impossibility,' for example, regarded as a kind of 'impermissibility,' does have a role to play.

This is the tendency in his thinking to which Kant's gloss gives expression. But again, Kant's gloss gives expression to just one of several such tendencies. Of course I claim that Kant's gloss represents the relationship between morality and happiness in a distinctive (i.e., extrinsic *and* normative) manner—but I do not claim that this is the *only* way in which he represents the relationship between them. Again, in some unusual instances, for example, Kant's thinking about happiness does entail that happiness is constituted in such a way that morality is internal to it (as we have seen, Guyer and others put this approach in the foreground, but mistakenly connect it with the 'worthiness to be happy' idiom). This has at least one important implication. If morality is internal to happiness, then of course the idea of an immoral, but happy agent is simply incoherent.

But some of the compelling force of Kant's repeated gloss lies in the fact that it evinces a strong (mainly prospective) disapprobation of the happiness of immoral agents. This would make no sense if Kant held, in general, that immoral agents were simply incapable of being happy. Their unhappiness would be assured, in that case, from the outset. Nevertheless, Kant does sometimes take such an approach. Before taking up the primary claim of this section, then, I will acknowledge and assess the significance of this other tendency in Kant's thinking.

Morality and happiness: intrinsically related?

Some of Kant's rather varied thinking about happiness proceeds independently of the thought expressed by his gloss: that somewhere, somehow, it might turn out to be the case that immoral, unworthy agents were numbered among the happy. Indeed, Kant sometimes seems to regard happiness in a way that would block immoral agents' access to it, not by their happiness' occlusion by some extraneous force, say, but just in the nature of the phenomenon. Here, emphasis falls heavily upon the issue of what an agent has to be like even to be *capable* of happiness.

Morality as a transcendental condition of the possibility of happiness

The main source for this earlier approach is an important fragment (*R* 7202) from the so-called *Duisburg Nachlass*.[193] In it, while also insisting that happiness 'consists in

[193] This is the main showpiece in a whole body of notes, originating in the 1770s, in which, according to Guyer, 'Kant hardly separates morality and happiness' (Guyer, *Kant on Freedom, Law, and*

[empirical] well-being,' Kant identifies 'the *a priori* condition under which alone one can be *capable* of happiness,'[194] 'its [i.e., happiness'] possibility and its idea,'[195] or 'the necessary condition of its possibility and its essence.'[196]

Obviously, this is rather reminiscent of Kant's critical project of uncovering the *a priori* conditions of the possibility of experience more generally. While happiness is an *empirical* state of the agent's affairs, Kant takes it to consist, from the outset, in an arrangement that 'is not externally contingent, also not empirically dependent, but [that] rests on our own choice.'[197] Happiness is empirical well-being that satisfies. But nothing is more satisfying to the human being than the exercise of her freedom, or the '[c]onsciousness of one's own power.' Indeed, this is both an intellectual 'pleasure' and an 'essential formal condition of happiness.'[198] Elsewhere, Kant avers, too, that pleasures (comforts, gratifications, etc.) that one achieves freely, for oneself, on one's own, are satisfying; those that fall to one by mere chance far less so—indeed they may even be shameful.[199] On the present account, however, gratifications that fall to one's lot by luck do not count as instances of what Kant means by happiness (here) at all.[200]

On the account that we are now rehearsing, the mere fact that one achieves what one desires, freely, is not sufficient either, just as such, for happiness. Immoral deeds

---

Happiness, 100). Schilpp calls *R* 7202 'a most remarkable document' (Schilpp, *Kant's Pre-Critical Ethics*, 138) and, following Menzer, dates the fragment to around 1775 (ibid., 127). He characterizes its content as 'semi-critical,' at least (ibid.), and rues the fact that the fragment is not 'better known,' confident that if it had been 'taken seriously' early on, then a great deal of 'nonsense' would have been avoided (ibid., 138). Beck, on the other hand, thinks it originates in a later period, just after the publication of the first *Critique* (Beck, *A Commentary on Kant's Critique of Practical Reason*, 11, 215), but downplays its significance (ibid., 215-16). See also Forschner, 'Moralität Und Glückseligkeit in Kants Reflexionen'; Guyer, *Kant on Freedom, Law, and Happiness*, 106, 115; O'Connor, 'Kant's Conception of Happiness': 202-3; Wike, 'Kant on Happiness': 81; Wike, *Kant on Happiness in Ethics*, 15-16. *R 7202* belies Watson's claim that 'there is no direct evidence' for his view that, for Kant, happiness is equivalent to moral contentment (Watson, 'Kant on Happiness in the Moral Life': 83).

[194] *R* 7202 19: 278-9 (467) (my emphasis). Cf. Beck, *A Commentary on Kant's Critique of Practical Reason*, 215.

[195] *R* 7202 19: 276 (465).

[196] *R* 7202 19: 277 (465).

[197] *R* 7202 19: 277 (466). See also *R* 6910 19: 203 (449); *R* 6849 19: 178 (439); *R* 7149 19: 258.

[198] *R* 7202 19: 276-7 (465). See Schilpp, *Kant's Pre-Critical Ethics*, 131. For more on why freedom—or its unifying capacity—satisfies see Guyer, *Kant on Freedom, Law, and Happiness*, 98, 117; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 278, 283.

[199] See, for example, Kant's discussion of games of chance at *Anthro* 7: 238 (134). See also the translator's note to *KdU*, § 87 (395 n. 33); *Idee* 8: 20; *Anthro-M* 25: 1334-5; and *Menschenkunde* 25: 1142-3. Cf. Howard Caygill, *A Kant Dictionary*, The Blackwell Philosopher Dictionaries (Oxford: Blackwell Reference, 1995), 222.

[200] Cf., however, ibid; O'Connor, 'Kant's Conception of Happiness': 189-90.

are as free as moral ones. Thus the *manner* in which freedom operates in the course of an agent's striving for happiness is of the essence as well. The integrity with which one proceeds extends further too, however, than the merely technical integrity of skill.[201] Freedom must stand under '*a priori* laws of its consensus with itself.'[202] As Kant puts it in his philosophy of religion lectures, 'self-contentment is a pleasure in one's own freedom,' but he glosses this as 'the quality of one's will' and 'the consoling consciousness of rectitude.'[203] In short, the exercise of freedom in the attainment of one's desires is only satisfying when its operations are properly ordered, if its exercise in the pursuit of happiness is *moral*.[204] At the relevant stage in Kant's thinking, this means, for example, that this exercise aims at 'the systematic and therefore maximal happiness of all,'[205] or at a 'maximally consistent system of purposes.'[206] The *a priori* unity that Kant identifies as a condition of the possibility of happiness turns out to be *morality*, then, which he glosses as 'freedom under universal laws of the power of choice.'[207]

The form and matter of happiness

As I have already intimated, Kant's approach, here, turns on a distinction between the 'form' and the 'matter' of happiness—the latter 'sensible,' the former, 'intellectual.'[208] The mere gratification of desires, the immediate feeling of enjoyment, no matter how superlative, does not count as happiness.[209] Just as on Kant's 'critical' view intuition cannot be regarded as a moment *within* (discursive) cognition, but is nevertheless an 'element' of it, so mere pleasure or sensual enjoyment is an element of happiness, but not happiness as such.[210] Inclinations, as Kant says in the second

---

[201] See *Gr* 4: 414-16 (25-27). See also Beck, *A Commentary on Kant's Critique of Practical Reason*, 97; Wood, *Kant's Ethical Thought*, 54, 65.
[202] *R* 7202 19: 276 (465). See also *Streit* 7: 87 n. (303 n.).
[203] *Vorlesungen-Religionslehre* 28: 1090 (420).
[204] In this regard see, for example, *R* 6805 19: 167; *R* 7049 19: 235 (457); *Vorlesungen-Religionslehre* 28: 1057 [394]).
[205] Guyer, *Kant on Freedom, Law, and Happiness*, 100-1.
[206] Ibid., 94.
[207] *R* 7202 19: 277 (465). See O'Connor, 'Kant's Conception of Happiness': 201; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 279-80.
[208] *R* 7202 19: 276 (465). See also *R* 6820 19: 172 (437-8).
[209] *R* 7202 19: 276 (465). See also *R* 6910 19: 203 (449); *R* 7200 19: 274 (463); *R* 6820 19: 172; and Guyer, *Kant on Freedom, Law, and Happiness*, 104; Ward, *The Development of Kant's View of Ethics*, 56.
[210] Römpp notes the same parallel (Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 276-7). See also Schilpp, *Kant's Pre-Critical Ethics*, 131, 133.

Critique, are 'blind and servile,'[211] their immediate gratification formless. Again, just as intuition—a mode of representation that, while being representation indeed, is nevertheless antecedent to thought—is only constitutive for cognition under the understanding's guidance; so, too, inclination requires guidance if action that aims at the fulfillment of one's desires is to count as action aiming at happiness, rather than at the chaotic pursuit of momentary pleasures.[212]

On this account, morality gives the form of happiness to one's empirical satisfactions and, too, the form of the pursuit of happiness to one's particular, desire-targeted actions. In one sense, Kant could not be clearer: morality 'makes happiness as such possible'; morality is 'the original form of happiness'; and happiness 'originate[s] in an *a priori* ground of which reason approves.'[213] But how does morality play this role? Kant's answer is, first, that 'the principle of self-satisfaction *a priori* [is] the formal condition of all happiness' and, second, that self-satisfaction runs in 'parallel with apperception.'[214]

'I am moral'

These references to 'self-satisfaction *a priori*' and 'apperception' are fairly unhelpful at first glance. But Kant's meaning can be clarified with help from other texts. Notes from Kant's anthropology lectures of 1772-3 assert that 'a creature which cannot say "I,"' even if it 'can suffer much pain,' cannot be unhappy. And '[o]nly through [the] "I" are we capable of happiness and unhappiness.'[215] Kant says, too, that 'pure reason is universally necessary for happiness' and that the latter is 'not something felt but thought.'[216] This emphasis on the necessity of the 'I' does not get us as far as an answer to the question how morality 'makes happiness as such possible,' of course. But the ability to 'say "I"' does have direct reference to the human being's capacity

[211] *KpV* 5: 118 (99).
[212] Cf. Langton, 'Duty and Desolation': 497-8.
[213] *R* 7202 19: 277 (465).
[214] *R* 7202 19: 280 (467) (my emphasis).
[215] *Anthro-C* 25: 11-12. Cf. O'Connor, 'Kant's Conception of Happiness': 192. See also *Anthro-P* 25: 422 and Wood, *Kant's Ethical Thought*, 54.
[216] *R* 7202 19: 279 (467). In a sense, this goes too far. I take it, however, that Kant is simply putting as sharp a point as possible on a key distinction. See also *R* 6973 19: 217 (453-4); *R* 6910 19: 203 (449); *LEV* 27: 647-8 (385).

for 'prudential reasoning,'[217] which gets us part way to morality, as it were, and into close proximity to the topic of happiness.[218]

In 'parallel with apperception'

As I mentioned above, the second aspect of morality's role as 'the original form of happiness' is that moral self-satisfaction (the affective correlate of the reflexive judgment 'I am moral') runs in 'parallel with apperception.'[219] Moral self-satisfaction is 'a spontaneity of well-being.'[220] In terms of Kant's first *Critique*, transcendental apperception is the absolutely spontaneous 'act' that unifies consciousness—that founds the subject and grounds the objectivity of objects. Analogously, moral self-satisfaction is not happiness, but is (or discloses) the activity of freedom 'before' and 'in' happiness that constitutes happiness' unity, the conceptual unity of its diverse particular instances (e.g., the enjoyment of *this* ice-cream cone, of *that* sunset, of the company of *this* friend, etc.).[221] In *R* 7202 Kant designates this satisfaction, which is a feeling, but which is also grounded in morality's freedom, 'happiness *a priori*.'[222] But he is quick to say that 'there is nothing real' in it. It is only 'the formal condition of unity, which is essential to it.'[223]

In this way, Kant implies that if all of the particular expressions of one's free agency have the right form (if they are moral), then there is a sense in which one already knows what happiness is *like*, even without reference to the 'matter' of happiness. Just as the pure concepts of the understanding 'extend further than sensible intuition, since they think objects in general,'[224] the 'original form of happiness,' morality—in the affective guise of moral self-satisfaction—extends further than all of

---

[217] Sikka, 'On the Value of Happiness: Herder Contra Kant': 518-19.
[218] Prudence is already a source of action that is 'highly intelligible' (Karl Ameriks, *Kant and the Historical Turn: Philosophy as Critical Interpretation* (Oxford: Clarendon Press, 2006), 100). But cf. Roger J. Sullivan, 'The Categorical Imperative and the Natural Law,' in *Proceedings of the Sixth International Kant Congress*, ed. G. Funke and T. M. Seebohm (Washington, D.C.: University Press of America, 1989), 228.
[219] *R* 7202 19: 280 (467). See also *R* 6861 19: 183 and Guyer, *Kant on Freedom, Law, and Happiness*, 115.
[220] *R* 7202 19: 278 (466). Cf. Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 277.
[221] By contrast the 'material' of happiness owes everything to nature and other agents, empirical conditions that 'create differences' rather than unity (*R* 7204 19: 284 [470]).
[222] *R* 7202 19: 279 (467). See also *R* 6911 19: 203-4 (449); *R* 7029 19: 230-1; *R* 7204 19: 284 (470). Cf. *Gr* 5: 117-18 (98).
[223] *R* 7202 19: 278 (466).
[224] *KrV* A253-4/B309.

the particular instances in which one's inclinations are gratified. Moral self-satisfaction allows us to represent the satisfactions of a rational, but also inclined, being in general, which representation is the concept of happiness, in general. In other words, on this account (still pursuing my 'critical' analogy), just as the paradigm for *any* object of experience lies, already, in the categories through which, even without the 'matter' of intuition, one is able to think an object 'in general,' the paradigm or reference point for happiness is moral self-satisfaction. The latter, rather than individual instances of mere pleasure or gratification tell us what counts as happiness. The moral agent is, in a sense, virtually happy: she 'contains happiness' in herself, as Kant says elsewhere.[225]

Unity

Again, in the feeling of moral self-satisfaction we have a paradigm for the satisfactions of inclined beings (such as we are) in general. But what is it about morality that shines through here, as it were, and gives this unity to the concept?

On the present account, 'transcendental unity in the use of freedom'[226] is the condition through which the empirical well-being that I enjoy is not merely the *experience*, but the *happiness*, of someone in particular: it is mine, it belongs to the same 'one' whose disparate inclinations, where satisfied, provide the matter through which happiness is given. It is only through the unity of the one 'desirer' (the 'I' in 'I am moral'), that each and every one of these instances (e.g., again, enjoying *this* ice-cream cone, admiring *that* sunset, enjoying the company of *this* friend, etc.) counts as an instance of one *kind* of thing, that is, as something that *I* have desired and foreseen in a consistent, integrated manner, now realized under the aspect of diverse particulars. To echo a well-know thesis of the first *Critique*, an 'I am moral' must be able to accompany each of an agent's particular empirical satisfactions if these are to count as instances of the one kind of thing, ('my') happiness. Any satisfaction that does not evince this unity with all other possible satisfactions is just not an instance of happiness at all.

[225] *R* 6867 19: 186 (444). See also *R* 7202 19: 277 (465); *R* 7204 19: 283 (470).
[226] *R* 7204 19: 284 (470). See also *R* 7204 19: 283 (470).

This unity is an achievement of the subject—again, 'in parallel with apperception.' It is an integrity that the subject herself must forge. But what is required here is not merely the integrity of the subject of experience, which is nothing specifically moral, but the 'integrity of willing'[227] and the connection of this integrity to happiness.[228] This involves a single, but complex task. The human agent must organize and standardize her own free activity;[229] forego mere habit and instinct; reflect on, harmonize, and prioritize her inclinations; rationally order the efforts that she makes for their realization;[230] take care to avoid the acquisition of new (inevitably diverse and mutually conflicting) desires; harmonize her practical activity with the practical activity of the other members of the human community; and (to extend matters as far as Kant's ultimate point of view) see to it (at all times) that her actions conform to the norms that would be constitutive for *any possible* community of rational subjects (i.e., she must always do her duty from an absolutely objective motive).[231] The form of happiness, then, is also the form of a particular gathering-together of the subject, a consolidation of ends and an integration of action that already resembles morality at various points in its articulation—and might be mistaken for it (in advance) to the extent that it entails the operation of reason from the outset and tends towards greater and greater degrees of rationality.[232]

---

[227] See O'Connor, 'Kant's Conception of Happiness': 201.
[228] Kant brings these themes together quite clearly in *R* 7204 19: 284 (470).
[229] This is a minimal requirement for all rational action: 'The faculty of order, *ratio*,' as Henrich says, 'possesses a *horror vacui*, a fear of all that is without a rule' (Henrich, 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason', 75). See also Beck, *A Commentary on Kant's Critique of Practical Reason*, 245; Smith, 'Worthiness to Be Happy': 184.
[230] Kant's definition of 'the doctrine of prudence' at *KrV* A800/B828 is particular clear and apposite.
[231] *R* 7197 19: 270 (461-2) marks a particularly clear division of the task along these lines. See also *R* 7202 19: 279-80 (467); *R* 7199 19: 272 (463); *LEV* 27: 648 (386).
[232] Hills, for example, observes that in its regulative role the concept of happiness already resembles the idea of duty (Hills, 'Kant on Happiness and Reason': 251). See also Reath, 'Hedonism, Heteronomy, and Kant's Principle of Happiness', 49; Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', 18. Various other commentators make closely allied observations. See, for example, Beck, *A Commentary on Kant's Critique of Practical Reason*, 98; Engstrom, 'Happiness and the Highest Good', 106, 127; Hill, 'Is a Good Will Overrated?', 47; Hills, 'Kant on Happiness and Reason': 252; O'Connor, 'Kant's Conception of Happiness': 203; Viggo Rossvaer, *Kant's Moral Philosophy* (Oslo: Universitetsforlaget, 1979), 168; Smith, 'Worthiness to Be Happy': 172; Victoria S. Wike, 'The Role of Happiness in Kant's *Groundwork*,' *Journal of Value Inquiry* 21, no. 1 (1987): 76; Wike, 'Kant on Happiness': 87; Wood, *Kant's Ethical Thought*, 67.

Critical period instances of the transcendental construal of the conditioned-by relation

Apart from *R* 7202 (and related notes), one finds hints of a similar conception of the relationship between morality and happiness later, too, well within the limits of Kant's critical period. In the second *Critique*, for example, Kant argues that 'virtue' is 'the *supreme condition* of whatever can even seem to us desirable and hence of all our pursuit of happiness.'[233] Here we are presented with the idea that our desires appear to us, upon reflection, in direct relation to the primary question of ethics, 'What ought I to do.' As Römpp puts it, '[t]he viewpoint of happiness projects a horizon from which the particular satisfactions of needs and desires win a new significance and are rated by a new measure.'[234] Our satisfactions do not satisfy us, just *as such* (as merely empirical gratifications). They (or our enjoyment of them) can be called into question. In the *Metaphysics of Morals*, too, Kant asserts that happiness is '*satisfaction* with what nature bestows, and so with what one enjoys as a gift from without.'[235] To be sure, then, there is, on the one hand, 'what nature bestows,' the external gift of the moment, which involves a certain possibility of pleasure or enjoyment. On the other hand, however, there is '*satisfaction*' with this 'gift,' which is *not* something that 'nature bestows.' Happiness has its material and formal elements here too then: the 'gift from without,' on the one hand, and this satisfaction 'with' it, on the other.

If the prospect of empirical contentment cannot even '*seem* desirable' when pursued independently of moral considerations, then this is because it does not constitute something that a rational being (of any kind) could pursue in a systematic way. The only thing that has the right kind of unity and integrity is the pursuit of practical-systematic unity as such, or morality, which subsumes happiness. Without morality in *this* sense—having no reference to the peculiarly human, let alone to that which is merely 'mine' *qua* individual animal—happiness is no 'intelligible' system. Instead it is a mere concatenation of pathologically contingent instances of idiosyncratic satisfaction, or (at best) socially, and so for Kant only contingently, sanctioned ones.

---

[233] *KpV* 5: 110 (92).
[234] Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 281.
[235] *MdS* 6: 387 (151) (my emphasis).

The ultimately satisfying 'thing' (unity, organization) subsumes happiness as the end that consists in the harmonized totality of mutually realizable empirical ends.

On the foregoing account (or, to speak more loosely, account-type) the human being's happiness is impossible without the antecedent satisfaction of reason. This is a mode of satisfaction that has no reference to the human being's empirical state, but which is nevertheless felt in it (as, too, the feeling of respect for the moral law): self-satisfaction as consciousness of one's morality. Here the incapacity of the immoral agent for happiness is implicit in the axiomatic elements of the transcendental logic of 'happy experience,' as such. On this view, if an immoral agent's happiness is impossible, this is not because the latter would have to be constituted all at once as an 'absolute whole' and realized *per impossibile* at some particular time and place (a possibility—or impossibility—that we encountered in chapter 1), but because the agent that seeks the gratification of her inclinations is not moral.

On this account, too, the claim that moral agents were worthy to be happy would have reference, at most, to the material conditions that would have to be added to their morality and the latter's immediate affective *sequelae* (i.e., added to their antecedently established moral self-satisfaction) in order for the happiness of which they were formally capable (given that they were moral) to be actualized. In other words, the idiom would have reference to those 'other material conditions'[236] over which, in contrast to her morality, the moral agent would remain powerless.

But the notion of 'unworthiness to be happy' would be simply superfluous. The claim that immoral agents were unworthy to be happy would be oddly out of place. Given their immorality, such agents would lack any capacity for happiness in the first place. This would be so even given the realization of the relevant material conditions (again, the satisfaction of an agent's particular desires). Here, to declare an agent immoral would be to declare her unhappy and to declare that she could not help but *remain* unhappy as long as she was immoral—irrespective of any merely empirical satisfactions that she happened to enjoy. The idea of a happy, but immoral, agent would be simply incoherent. On this view, not only the agent herself, but God as well, would be powerless to bring it about that an immoral agent is happy.

[236] See *R* 7202 19: 276-7 (465). See also Schilpp, *Kant's Pre-Critical Ethics*, 131.

If this were the end of the story, then Kant's habit of glossing 'morality' as 'worthiness to be happy,' would be rather excessive, or (barely better) an empty rhetorical habit. If the happiness of which immoral agents are 'unworthy' were, in the end, something of which they are constitutively *incapable*, then the notion of worthiness to be happy would be at best a kind of inept juridical metaphor that expresses an ultimately illusory feeling, based in a mistake in thinking about the nature of happiness.

To the contrary, however, to the extent that Kant's habit of glossing 'morality' as 'worthiness to be happy' expresses his thinking about morality and happiness, he takes it that human beings in general, and immoral ones in particular, can be happy. He takes it that the aspiration to happiness is realizable, in principle, given the right mix of prudent effort, natural circumstances, and assistance from other agents. As I show below, only a *normative* conception of the relationship between morality and happiness makes any sense of Kant's use of this idiom. The transcendental conception of this relationship, in any case, is ill served by it.

Morality and happiness as extrinsically related states of an agent's affairs

The 'extrinsicness' thesis as the claim that an immoral, but happy agent is possible

To the extent that he deploys it, Kant mediates his references to morality (or 'virtue,' or 'goodness'), the '*conditionem sine qua non*' for happiness,[237] by way of his immediate use of the 'worthiness to be happy' idiom. The human being is worthy to be happy only *on condition* that he is moral—in possession of a 'good will,'[238] a 'good man,'[239] an agent who enjoys a 'pure disposition of…heart.'[240] But the mere assertion that *something*—'morality' (however construed)—is a condition of the possibility of happiness does not settle the question whether this relationship, even qualified

[237] See *Anthro* 7: 326 (231); *R* 5477 18: 194 (416); *R* 6133 18: 465 (338-9); *R* 6317a 18: 632 (373); *LEV* 27: 717 (440).
[238] See *Gr* 4: 393 (7); *MdS* 6: 482 (225); *R* 6890 19: 194 (446); *R* 7217 19: 288 (472); *R* 7315 19: 313.
[239] *R* 7315 19: 313.
[240] *R* 6858 19: 181 (441).

as one that is subject to a 'fundamental principle of reason,'[241] is a transcendental one, a causal one, or a normative one.

Taken at face value, the assertion that *worthiness to be happy* is the condition in question settles this matter outright: it points to a normative conception of morality's relation to happiness. The nature of happiness is also settled—that is, specifically as far as its being a possibility for immoral agents goes. In the idea of a unity of ends, for example, secured through far-reaching prudence, prudence that considers what is true of human beings in general, but without giving special priority to their capacity for pure practical rationality, we have already encountered the idea of happiness as *both* a unified whole of empirical gratifications *and* a state of the human agent's empirical affairs that is compatible with her immorality in Kant's ultimate sense. Here, moral self-satisfaction, or moral contentment as the consciousness of one's deeds' ascribability to oneself (their 'negative' freedom) and of their conformity to the moral law (in the manner of their being motivated, etc.) and, too, of this conformity's ascribability as well, is not integral for happiness, as such.

Disapproval of the happiness of immoral agents is disapproval of something for which the ground of this same disapproval (i.e., the immorality that gives rise to self-contempt) is not, of itself, already a sufficient negating condition. It is disapproval of happiness where the latter is regarded as something that ought, perhaps, to be occluded (where morality is lacking), but that would have to be blocked extraneously.

Thus the claim that morality and happiness are extrinsically related, on the one hand, and the claim that immoral agents are capable of happiness, on the other, are reciprocally related. But further qualifications must be put into play. How should we characterize the happiness of which both moral and immoral agents are capable, but not equally worthy?

Does Kant's gloss imply a particular theory of happiness?

We saw in chapter 1 that Kant sometimes represents happiness as an ideal and hence perpetually deferred state of affairs. If happiness were something to which *no one* could attain, then it would follow *a fortiori* that it was something of which immoral agents were incapable. We can reject this representation of happiness out of hand,

[241] See *R* 5477 18: 194 (416).

then, to the extent that we allow that Kant's gloss expresses his view that immoral agent's are capable of happiness. We need not *quite* agree, however, that 'for Kant happiness is a subjective state of satisfaction *rather than* the achievement of some objective ideal.'[242] Kant evinces both views, even if he emphasises the activity of *thinking* particular instances of empirical gratification under this 'objective ideal,' over the merely (practically) regulative notion of the ideal's (at best perpetually deferred) 'achievement.'

In fact, Kant does not explicitly articulate a theory of happiness that is consistent with his use of the 'worthiness to be happy' idiom. Rather, his habit *implies* three main possibilities. The happiness of which immoral agents are both unworthy and capable is *either*: (1) the pleasure of the moment; (2) a mode of empirical well-being constituted in such a way that *skill* and *prudence* and *freedom* are internal to it; i.e., the rational organization of the agent's empirical desires and the organized realization of some mutually compatible subset of them that *falls short*, by definition, of the integrity that is involved in Kantian morality (whatever this integrity may entail at any stage in his forward-looking thinking about morality); or (3) a mode of empirical well-being that resembles (2) in terms of its empirical content, but which is forged by a third party acting in the agent's behalf, and offered her as a gift. Both (2) and (3) entail, further, that happiness is a mode of satisfaction that either (α) subsists *in spite of* the immoral agent's (lively) bad conscience; or (β) subsists *only thanks to* the deadness of her bad conscience; or (γ) subsists *only thanks to* an antecedent healing of her will and bad conscience, together, that leaves the first good and the second unburdened.

In spite of at least one prominent counter-example,[243] it seems unlikely that the happiness of which Kant takes immoral, unworthy agents to be capable is simply the pleasure of the moment. No human being, even the least disposed to morality, can really live for the moment; she cannot inhabit her gratifications immediately and suppress her interest in questions about whether she will ever be satisfied again, whether the present satisfaction will last, whether it might be improved upon next

[242] Johnson, 'Happiness as a Natural End', 319 (my emphasis).
[243] Again, *MdS* 6: 480-2 (223-5) offers the clearest example of this approach.

time, whether other satisfactions might be joined to it, or whether other satisfactions might, to the contrary, occlude (or be occluded by) the realization of this one.

What Wood calls 'the pleasure of the moment' contrasts with 'happiness on the whole'[244] and, while human beings often sacrifice the latter in order to obtain the former, this does not meant that the pleasure of the moment counts as happiness. It is hardly likely that these states of affairs are equivalent for Kant. It is more plausible to claim that, for Kant, happiness always has reference to reason's organizing capacity. As Adorno puts it in his lectures on Kant's moral theory, 'the individual [who] renounce[s] momentarily a certain amount of happiness or pleasure…gets it back with interest in terms of the rational organization of his life.'[245] Adorno's apparent conflation of 'happiness' and 'pleasure' aside, Kant makes some degree of 'rational organization' central to his thinking about happiness (again, 'on the whole'). Here as elsewhere, however, 'reason requires some sensuous material to carry out its regulative function.'[246] As the mere concept of the empirical object of human striving *in general*, happiness remains empty, a kind of blank.[247] The requisite content is 'supplied by the individual's inclinations.'[248] And where this content is supplied, and to the extent that an identical '*I* have desired this' (a kind of intellectual sigh of pleasure) is able to accompany each particular instance of empirical satisfaction that she enjoys, *there is happiness*. Thus it would seem that (2), above, is a more probable candidate than (1) for the happiness of which both moral and immoral, worthy and unworthy, agents are capable.

But it seems improbable, at best, that Kant would allow happiness in the sense of (2), above, to count, without qualification, as the apogee of empirical well-being for the human agent. It is possible for happiness in that sense to subsist with the presence of a guilty conscience. But bear in mind that the *experience* of the bad conscience is a matter of affect—of *painful* affect. It is hard to see how an agent's empirical well-being, no matter how deeply grounded in 'the rational organization of his life,' and

[244] See Wood, *Kant's Ethical Thought*, 66. See also Acton, *Kant's Moral Philosophy*, 18; O'Connor, 'Kant's Conception of Happiness': 193.
[245] Theodor W. Adorno, *Problems of Moral Philosophy* (Stanford, CA: Stanford University Press, 2001), 138.
[246] Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue': 524.
[247] Cf. Johnson, 'Happiness as a Natural End', 329
[248] Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*, 78.

no matter how much an expression of that organization, could be ultimately satisfying to him, if he had a bad conscience and suffered from it.

The problem of moral self-satisfaction

In the second Critique, Kant makes a pair of observations about the moral psychology of two kinds of human agent: a cheat, on the one hand, and an 'upright man,' on the other. On the one hand, Kant describes an agent who enriches himself through dishonesty and then avers 'to himself' both that 'I am a prudent man, for I have enriched my cash box,' and that 'I am a worthless man although I have filled my purse.'[249] On the other hand, the 'upright man' finds himself 'in the greatest distress' and is aware that 'he could have avoided [this suffering] if he could only have disregarded duty.' But, Kant argues, he is 'sustained [in his superlative state of distress] by the consciousness that he has maintained humanity in its proper dignity in his own person and honored it' and 'that he has no cause to shame himself in his own eyes and to dread the inward view of self-examination.' His consciousness that he is not 'unworthy of life in his own eyes' is 'consolation,' Kant says, but it 'is not happiness, not even the smallest part of it.'[250]

Conversely, 'the greatest distress,' 'shame,' and 'dread' that are associated with an agent's consciousness that she is 'unworthy of life,' are not simply equivalent to unhappiness. To have an empty 'purse' or an impoverished 'cash box' is of a different order from the experience of having to concede (even if only secretly) that one is 'a worthless man.' Two 'different criteri[a] of judgment' are in play.[251]

I take it that Kant is simply describing some well-attested facts about ordinary human psychology. The human being wishes to avoid 'distress.' She wants to have a full purse. But she also cares about her own 'humanity in its proper dignity'; she wants to avoid 'shame'; she 'dread[s] the inward view of self-examination' when she knows that she has acted immorally. She does not want to have to acknowledge that she is morally 'worthless.' For Kant, the relationship between 'recognition that one has done something morally wrong,' as Hill puts it, and 'painful self-reproach and

[249] *KpV* 5: 37 (34). For a compelling, earlier representation of such knavery and its relation to (or alienation from) happiness cf. David Hume, *An Equiry Concerning the Principles of Morals*, Oxford Philosophical Texts (Oxford: Oxford University Press, 1998 [1751]), 9.23-5 (155-6).
[250] *KpV* 5: 88 (74-5). The *Vorlesungen-Religionslehre* offers an interesting, cognate picture of this 'consolation and comfort' (28: 1011 [356]).
[251] *KpV* 5: 37 (34).

alienation from others' is a necessary one.[252] In other words, '[t]he tendency to suffer, though perhaps blocked in some cases, is inevitable'[253] and is an 'inherent liability'[254] for immoral agents. But this is not simply equivalent to unhappiness.

In this respect, the account of their relationship that makes morality a *transcendental* condition of the possibility of happiness, which was heavily dependent upon *R* 7202, included a decidedly ambiguous moment.[255] One of that document's main insights is that the human being's conscience and its state cannot be ignored when it comes to happiness. But it is not clear that the role that Kant gives to self-satisfaction with respect to happiness necessarily *entails* that morality is internal to happiness. On my reading, it does not. The pangs of conscience are a source of unhappiness, to be sure, but this does not mean that immorality makes happiness simply impossible. Instead, by way of the pangs of a bad conscience, immorality is a possible source of inner suffering, which tends to undermine and erode any happiness that the agent has secured. (Here, again, happiness might be regarded as either the pleasure of the moment or, more realistically in reference to Kant, as the outcome of natural factors, clever foresight, prudent action, and contributions from other agents.)

On this account, how an agent lives, in general, and how she comes by her empirical gratifications, in particular, affects whether or not she can take satisfaction in the outcomes that satisfy her inclinations. But this is so in a distinctly non-transcendental sense. The gratification of her desires as the outcome of cleverness or prudence is not in question *qua* happiness. The problem, here, is that this happiness would tend to be undermined by the equally empirical fact that her conscience torments her. Emphasis falls on this affect, not on its source. Unhappiness ensues, if her conscience is a sensitive one, as the result of a countervailing, heartfelt suffering, irrespective of the latter's ground. In short, the *absence* of a wounded conscience (either because the latter is insensate, or because it is 'healed') is internal to happiness—but this leaves the relationship between morality and happiness an extrinsic one.

---

[252] Hill, 'Wrongdoing, Desert, and Punishment', 316.
[253] Ibid., 321.
[254] Ibid., 322.
[255] See Wike's advice in regard to making too much of this note (Wike, *Kant on Happiness in Ethics*, 16-17). See also Josef Schmucker, *Die Ursprünge Der Ethik Kants in Seinen Vorkritischen Schriften Und Reflektionen* (Meisenheim am Glan: Anton Hain, 1961), 314-15. However, cf. Guyer, *Kant on Freedom, Law, and Happiness*, 108.

This view makes sense even if happiness is regarded as the mere pleasure of the moment. In general, however, this is not how Kant proceeds. Instead, as Kant puts it in his philosophy of religion lectures, happiness is an overall '[w]ell-pleasedness with one's own existence.'[256] Even regarded as pleasure, this well-pleasedness 'applies to the entirety of our existence' and not to our state at some time or other. It 'is consequently *pleasure in our state as a whole*.'[257] Thus 'consciousness of one's own dignity, or self contentment, belongs to *perfect* happiness.' But at the same time there is a distinction between 'self-contentment,' which 'arises from morality,' and happiness, which 'depends on *physical* conditions.'[258]

In *R* 7202, Kant asserts that '[a] certain basis (capital, property) of satisfaction is necessary…without which no happiness is possible.'[259] He construes self-satisfaction as 'pure happiness,' but also refers to 'the principle of self-satisfaction' as 'the condition of all happiness.'[260] In the *Groundwork*, too, Kant says that happiness is 'complete well-being,' adding that this entails 'satisfaction with one's condition.'[261] This 'satisfaction' has a special status. Like 'well-being' (*Wohlbefinden*), it pertains to the agent's affective state, but it must be carefully distinguished from the latter. In the *Religion*, Kant expresses this distinction in the strongest terms by contrasting merely empirical satisfaction with an '*unconditional good pleasure* in oneself' that is altogether 'independent of gain or loss resulting from action,' that is, 'a contentment only possible for us on condition that our maxims are subordinated to the moral law.'[262] In a draft version of *The Metaphysics of Morals*' 'moral catechism,' Kant gives the distinction a slightly different emphasis, writing that it is not enough to be 'satisfied [*zufrieden*]' with one's 'state,' but that one needs to be satisfied with *oneself* as well.[263]

---

[256] *Vorlesungen-Religionslehre* 28: 1060 (396-7).
[257] *Vorlesungen-Religionslehre* 28: 1089 (420).
[258] *Vorlesungen-Religionslehre* 28: 1089 (420).
[259] *R* 7202 19: 278 (466). O'Connor offers an insightful reading in O'Connor, 'Kant's Conception of Happiness'.
[260] *R* 7202 19: 281 (469). See Guyer's elaboration of this passage (Guyer, *Kant on Freedom, Law, and Happiness*, 115). This position shows up repeatedly, albeit sometimes obliquely, in Kant's thinking. But cf. Wike, *Kant on Happiness in Ethics*, 17.
[261] *Gr* 4: 393 (7). Cf. *R* 6892 19: 195 (447).
[262] *Rel* 6: 46 n. (90 n.).
[263] *R* 7315 19: 313. For related instances from various periods, see *R* 6760 19: 151 (433); *R* 6892 19: 195-6 (447); LEV 27: 647-8 (385); *Vorarbeiten-ÜdG* 23: 129; *LMD* 28: 689 (390). For an eschatological construal of this contentment see LMK_2 28: 770 (409).

R 7202 gives rather unsystematic expression, then, to two closely related, but distinct ideas. The first is that happiness is impossible without morality for the simple reason that the concept of happiness is subsumed under the concept of morality. The agent cannot think of herself as happy if she thinks of herself as immoral. In other words, here, the unity and integrity of practical agency that find their ultimate embodiment in Kant's critical notion of morality mark a kind of paradigm for thinking the unity of what it is that satisfies *me*, in particular (all of the diverse instances of empirical satisfaction that I happen to enjoy and that are a function of what I actually happen to desire), with what it is that satisfies agents *like* me, in general.

The second view is that empirical satisfaction is not adequate for happiness to the extent that happiness also entails 'satisfaction with one's existence.' Equivalently, Kant says the same of moral contentment, which is what allows him to say that happiness and moral contentment are analogous.[264] The difference between this perspective and the transcendental one is that happiness, which does not internalize satisfaction with oneself *qua* moral agent, is undermined and eroded by the feeling of moral self-contempt, which is a kind of painful feeling of shame.[265]

Kant's description of the cheat and the upright man brings us close to saying that morality is a condition without which immoral agents could not be happy in the first place. Of course, to the extent that we are trying to hone in on the sense of Kant's gloss, we do not want to say this. We want to say, rather, that it really is *happiness* that is eroded by a cheat's bad conscience and that such a person's happiness is not simply impossible, from the outset. But we also want to reject the idea that the happiness of which moral and immoral agents are equally capable goes on being *that* in the ongoing, sustained presence of the affective suffering that comes from having a guilty conscience. This way of relating morality and happiness is an extrinsic one to the extent that it leaves open the possibility of an immoral agent's having an insensate conscience (take, for example, the immediacy of Kierkegaard's 'aesthetic' subjectivity) and being happy in spite of her own wickedness.[266] But it also leaves open a more (for Kant) concerning possibility: the idea of an agent whose conscience is

[264] *KpV* 5: 117-18 (98). See also *Vorarbeiten-ÜdG* 23: 129.
[265] Cf. Kant's remarks concerning 'the possibility of feeling one's health as something agreeable' (*MdS* 6: 485 [227]).
[266] See, however, *MdS* 6: 438 (18).

whole, whose will has been healed and made good (in spite of, not thanks to, her discrete freedom, which was enslaved), and whose conscience has been unburdened by *forgiveness*.

This means that (α), above, is out. Thus, on Kant's gloss, the happiness of which we are *all* capable is either a mode of well-being such that any immoral agent that enjoys it has an insensate conscience (*viz*., [β] above), or such that any immoral agent that enjoys it has been 'healed' in conscience and in will, in such a way that she neither feels, nor takes herself to have *cause* for feeling, the pangs of conscience any more than a moral agent would (*viz.*, [γ] above).

Of course these possibilities integrate, in turn, with (2) and (3), above—the ideas of happiness as a mode of empirical well-being constituted in such a way that *skill* and *prudence* and *freedom* are internal to it or as an endowment that resembles the latter in its upshot, but is rather a gift than a product of the agent's freedom. If Kant's gloss represents the relationship between morality and happiness in terms of a normative connection between them, then—given either an insensate conscience or a 'healed' one—happiness as *either* a product of the agent's own optimally, rationally organized (but ultimately immoral) efforts, *or* a product (in whole or in part) of the efforts of another (hence, in part or in whole, a gift), would be happiness of which the immoral agent was *capable* and *unworthy*. In either case, we would be faced with a happiness that, on the view embodied in Kant's gloss, she *ought not* to enjoy.[267]

Morality and happiness as necessarily and normatively related

Immoral agents ought not to be happy. In some cases, the problem of an immoral agent's happiness is resolved when her guilty conscience robs her of the satisfaction that she would otherwise take in her empirical well-being. Such an agent ought not to be happy—and is in any case not happy (or not for long, or not very). The idea that immoral agents ought not to be happy has its real traction, then, in regard to the other two scenarios that I described above. I will address Kant's way of addressing these in chapters 3 and 4. First, however, it remains to be shown that Kant's gloss really does

[267] The significance of the third scenario (happiness as gift) and its supplement (happiness *along with* healing of will and conscience as gift) in respect to Kant's concept of 'worthiness to be happy' will become clearer in chapter 4.

represent morality and happiness as states of an agent's affairs that are not only extrinsically, but also necessarily and normatively related. The argument of this section will bring us to the threshold of the next one, with its claim that Kant's concept of worthiness to be happy corresponds to a particular understanding of desert.

As I did with respect to my claim that Kant's gloss shows that he takes morality and happiness to be extrinsically related, I must grant first, once again, that Kant's thinking exhibits more than one tendency when it comes to the *necessity* that characterizes the extrinsic connection that he takes to hold here. Before taking up the primary claim of this section, then, I will once again contend with and assess the significance of a second tendency in Kant's thinking.

Morality regarded as a (physical) cause of happiness

In the second *Critique*, Kant expresses the view that, given certain of our aims as practical-rational beings, the affirmation that morality is connected, immediately, with happiness may be regarded as a particular kind of synthetic judgment whose validity is grounded *a priori*. This affirmation attests, more precisely, to the representation of morality (and so, too, of freedom) as a *physical cause*, in the empirical medium of moral deeds, of happiness and so, of course, as a state of the agent's affairs that is not internal to morality (to the extent, that is, that effects are immediately connected with their causes, for Kant, but are not 'internal' to them—else true claims about causality would be analytic, not synthetic). Before coming to Kant's argument for this position, however, it will be necessary to clarify a number of distinctions without which it is possible that my reading of Kant will be misunderstood.

As we shall see, in the second *Critique* Kant expresses the view that we are permitted to infer from the statement that such-and-such an agent is a moral that this same agent will end up happy. We are permitted to *regard* the one claim as entailing the other. But if it does so (as we are permitted to think), it does not do so without the support of an antecedently granted covering law, a conditional that states that *if* an agent is moral, *then* she will be happy (my expression, '*will* be happy,' indicates that we are faced, in this context, in contrast to Kant's deployments of the 'worthiness to be happy idiom,' with a relationship between antecedent and consequent that is 'natural' and not 'deontic'). That the statement that an agent is moral entails the

statement that she *will be* happy (given that we are permitted to postulate the covering law that opens this inference) is precisely what I mean when I claim that, in the second *Critique*, Kant regards the relationship between morality and happiness as an 'extrinsic,' but not a *normative* one.

Now, in the implicit thinking that finds expression in Kant's habit of glossing 'morality' as 'worthiness to be happy' he appears to hold that the *notions* of unhappiness and immorality are bound to one another intrinsically. The respective states of an agent's affairs that would instantiate these notions (her actual immorality, as embodied in actually immoral deeds, and her actual unhappiness), however, are related *extrinsically*. In other words, here, although she *ought* unconditionally to be unhappy (eschatologically speaking) an immoral agent's happiness remains a possibility for her; nothing about her being immoral assures that she really *will* be unhappy; her unhappiness can only be unfailingly secured by being 'forged' (my term) by a kind of 'extra step' (as Guyer says of the actualization of the notional relationship that he thinks holds, for Kant, between happiness and morality[268]). The relationship between these elements—as encapsulated again strictly, however, in Kant's use of his 'worthiness to be happy' idiom—is neither a causal one, nor an analytic one. I call it 'normative' to signal the idea that, for Kant, its being actualized is *valued* and *called for* (indeed without conditions), but is never regarded an outcome whose eventual actualization we are permitted to simply take for granted (as we are, on Kant's view, where moral deeds are regarded as *causally* related to happiness—the view that I discuss below).

The only view on which the inexorable connection between immorality and unhappiness is an *analytic* one is the view according to which an immoral agent is, by definition, *incapable* of being happy—no matter what the empirical conditions are vis-à-vis the realization of the desires that she happens to have. In that case, the relationship between immorality and unhappiness, at least, would be an *intrinsic* one. The statement that an agent is immoral would entail the claim that she is unhappy. But there is a *normative* or *deontic* sense of entailment too. With respect to immorality and unhappiness, this would be expressed by saying that the statement that an agent is immoral entails the claim that she *ought to be* unhappy. Again, however,

[268] Guyer, *Kant on Freedom, Law, and Happiness*, 118.

this entailment depends upon the antecedent establishment, or granting, or presupposition (as the case may be), of a covering law, a conditional, stating that *if* an agent is immoral, *then* she ought to be unhappy.

In a moment I will turn to Kant's second *Critique* argument for the claim that we are permitted, as I said above, in view of our combined interest in morality, on the one hand, and happiness, on the other, to regard morality (and so, too, freedom) as a *physical cause*, in the empirical medium of moral deeds, of happiness. First, however, there is one other matter that requires clarification. I want to be clear, in particular, about how I understand the bearing of Kant's notion of the 'synthetic *a priori*' in this context.

As we have seen, some of Kant's thinking about happiness entails that the latter is intrinsically related to morality, or that happiness is constituted in such a way that morality is internal to it, such that the happiness of immoral agents is precluded from the outset. I am also conceding, now, that—particularly in the second *Critique*—Kant sometimes expresses the view that, with a view to certain of our fundamental aims as practical-rational beings, morality (or moral deeds) may be regarded as a *cause* of happiness—as able, at least under some ideal set of circumstances, to bring it about that moral agents are happy ones. As I pointed out in my introduction, given Kant's understanding of causality, this way of relating morality to happiness entails that the connection between them is a necessary one; but it also entails that the connection is an extrinsic one. In other words, the assertion that this connection holds is a claim to synthetic knowledge whose objective validity is grounded *a priori*.

That bolts of lightning, for example, cause peals of thunder, is not a synthetic *a priori* judgment. Nevertheless, for Kant, it is a condition of the 'objective validity' of (fallible) particular generalizations like this one that we be warranted *a priori* in holding that, *whatever* its cause may actually turn out to be, every event has one. And, as Kant argues in the first *Critique*, we are warranted *a priori* in making the general synthetic assertion that *every event has a cause*, which is to say that every event has temporally antecedent necessary and sufficient conditions—and that the relationship between any event and its conditions may be described with reference to universally binding laws (of nature).

The synthetic claim that morality causes happiness is not warranted *a priori* in the manner that the claim that every event has a cause is. But it is not like particular claims about empirical matters of fact, either: the claim that morality (or moral deeds) is (or are) the cause of (ultimate) happiness is not like the claim that bolts of lightning cause peals of thunder. With respect to the latter, it is Kant's view merely that, *if* the proposed (or supposedly observed) connection holds, *then* it holds necessarily. This is what it means to say that it is an instance of *causality*, rather than of merely happenstantial conjunction. But that there really *is* a causal connection between lightning and thunder is not something that can be ascertained *a priori* (in contrast to the *general* claim about events having causes). The view, which we will explore now, that there really is a causal connection between moral deeds and the happiness of their agents is not like this. Kant does not claim to be able to show *that* this is so, but only that we are permitted to think it is. In other words, he takes us to be permitted to *regard* the claim that moral deeds cause happiness as a claim to synthetic *a priori* knowledge—and so, to this extent, like the general assertion that every event has a cause. We are not permitted to regard specific empirical claims about causation in this way: they remain claims to synthetic *a posteriori* knowledge, even though their 'objective validity' is grounded in the synthetic *a priori* claim that—our fallibility in specific matters aside—every event does have a cause.

Now, after defining virtue and happiness as a pair of 'determinations *necessarily* combined in one concept' (i.e., in the concept of the highest good), Kant goes on to make an argument concerning such necessary combinations in general. Their elements, he argues

> must be connected as ground and consequent, and so connected that this *unity* is considered either as *analytic* (logical connection) or as synthetic (real *connection*), the former in accordance with the law of identity, the latter in accordance with the law of causality. The connection of virtue with happiness can therefore be understood in one of two ways: either the endeavour to be virtuous and the rational pursuit of happiness are not two different actions but quite identical…or else that connection is found in virtue's producing happiness as something different from the consciousness of virtue, as a cause produces an effect.[269]

Kant insists that, conceptually speaking, happiness and virtue are 'extremely heterogeneous'[270] (hence the futility of searching out an analytic connection between them)

[269] *KpV* 5: 110-11 (92-3).
[270] *KpV* 5: 111 (93). But this cannot be taken to mean that happiness and virtue are in *no way* related (see Engstrom, 'Happiness and the Highest Good', 107; O'Connor, 'Kant's Conception of Happiness':

and that if happiness is to be regarded as a *consequence* of morality (and the latter, then, as *ground*) then this will be a matter of a synthetic *a priori* relation. Obviously, it is possible to affirm that Kant takes the notion of the highest good to be a representation of the co-inhering of morality and happiness and still take it that 'the specific relation [that] they bear to one another in the highest good'[271] is a merely *extrinsic* one,[272] or to regard it as perpetually deferred to another 'world,' or to a future state of this one,[273] where it is realized as an expression of a material connection that he does not take to be discernable in advance.[274] Here, Kant's 'critical' aim is merely to block anyone's pretension to know that a causal relationship between moral actions and happiness is *impossible*.[275]

There are two immediate problems with Kant's argument, however. First, Kant does not show that morality's and happiness' conceptual connection (their being 'two determinations necessarily combined in one concept') is a necessary one. He appears to simply assert it by way of his two-part claim that, *first*, in addition to their virtue, 'happiness is also *required'* for 'the whole and complete good as the object of the faculty of desire of *rational finite beings*,' and, *second*, that this is so 'not merely' in the eyes of such beings alone, 'but even in the judgment of an impartial reason.'[276] As I will show below—with reference to another of Kant's uses of the notion of the 'impartial spectator' and in connection with this spectator's further appearance in the commentary—it is easy to misrepresent such a spectator's concurrence or acquiescence in what is deemed necessary from the point of view of the human being (the 'rational finite being') as this spectator's acquiescence in a demand of pure reason (hence necessary—as Kant indicates). The ease with which this misunderstanding arises can be attributed to a certain ambiguity in Kant's description in this context.

189-91; cf. *Gr* 4: 393 [7]; *KpV* 5: 25 [23]; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 277, 281).

[271] Engstrom, 'Happiness and the Highest Good', 122.

[272] Sikka, 'On the Value of Happiness: Herder Contra Kant': 539.

[273] See, for example, Engstrom, 'Happiness and the Highest Good', 133.

[274] Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 272.

[275] Of course, to all appearances, nature is not on our side in this respect (see *KpV* 5: 111 [93]; *KdU* § 87; Hills, 'Kant on Happiness and Reason': 254; cf. Wike, *Kant on Happiness in Ethics*, 25). But see *KpV* 5: 116 (96); *R* 6111 18: 458 (337); *LEC* 27: 304 (94). Cf. *R* 7211 19: 286 (472); *LEV* 27: 717 (440). See also Engstrom, 'Happiness and the Highest Good', 128; Fendt, *For What May I Hope?*, 70, 83; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 274-5. Cf. Friedman, 'Virtue and Happiness: Kant and Three Critics': 110; Hegel, *Phenomenology of Spirit*, §§ 620-22 (376-8); but with respect to the latter see *KpV* 5: 117-18 (98).

[276] *KpV* 5: 510 (92).

The second problem is that Kant appears to elide a *third* possibility here—a second kind of synthetic 'real connection'—and the one which, as I claim, is anticipated by Kant's habitual gloss. I say that he 'appears' to elide it since he does advert to it—in spite of himself—twice within the same text: firstly, on the two occasions on which he glosses 'morality' and 'virtue' as 'worthiness to be happy' and, secondly, when he defines 'the highest good of a possible world' (which is related to, but distinct from, 'the highest good in a person') as 'happiness *distributed* in exact proportion to morality.'[277] This third possibility is the connection or conceptual 'combination' that would be actualized by being forged by a third party (here indeterminate) acting in accordance with a categorical imperative commanding that this *practically* (hence morally), but not *physically*, necessary 'distribution' take place.

Apparently, Kant's near elision of this possibility has the effect, at times, of a complete obfuscation. When Guyer focuses on this same text, his explanation of Kant's use of the concept of worthiness to be happy slips comfortably from the *normative* focus of his reading of Kant's unpublished notes from the 1770s (where the emphasis is on questions of the fairness of happiness' distribution, or of the moral agent's right and entitlement to it) to Kant's 'critical,' *causal* construal of the synthetic *a priori* ('real') ground-consequent connection that Kant claims here for morality and happiness.[278] He simply follows where Kant apparently leads.

Guyer observes that Kant's concept of the 'highest good' has sometimes been misinterpreted as a 'composite' ideal consisting in 'two independent aims or ends,' virtue and happiness, whose independence and distinctness entails that they can only be 'reconciled' to the extent that we 'pursu[e] our natural end of happiness within the limits set by virtue.' This misunderstanding, he argues, is grounded in the assumption that virtue 'intrinsically has nothing to do with happiness.'[279] Earlier in his argument, however, Guyer points out that 'if happiness in that sense [i.e., in the

[277] *KpV* 5: 510-11 (92-3).
[278] As we saw in chapter 1, Guyer's normative reading of Kant's concept of worthiness to be happy is one whose key concepts are 'merit,' 'right,' 'entitlement,' and 'desert' (of happiness). The key to each of these (interconnected) properties is the freedom of the deserving agent. Freedom continues to play its integral role here as well. I describe this as a 'shading off' of one approach into the other in order to indicate their indistinctness from one another. This reflects the Kantian text as well: in the second *Critique* and elsewhere Kant characterizes the connection between morality and happiness in terms of this hoped-for causality and, at the same time, continues to gloss 'morality' as 'worthiness to be happy.' See *KpV* 5: 510-11 (92-3). See also *R* 7200 19: 274 (463-4).
[279] Guyer, *Kant on Freedom, Law, and Happiness*, 339. See also ibid., n. 3.

ordinary, natural sense that distinguishes the latter from moral contentment] simply has no connection with virtue, then it is hard to see why the two should be connected in the sentiments of any rational observer at all.'[280]

Of course, Kant *does* take happiness and virtue to be 'connected in the sentiments' of such an observer.[281] By turning in this direction, however, we—along with Guyer and, ultimately, Kant—are faced, again, with a notion of morality's and happiness' conceptual 'combination' that is distinct from the notion of physical causality whose relevance for answering the question (how morality and happiness are related) both Kant (and Guyer) are depending upon. Again, however, the notion of (well-ordered, morally oriented) freedom's causality with respect to happiness (i.e., under a widened *corpus* of natural laws) does not fit well with the 'sentiments' of the observer that Kant has in mind. He veers in this other direction, however, towards the 'is' and 'will be' of an inscrutable natural order (framed by his theological-practical 'postulate' of God's authorship and oversight of nature) and *away* from the normativity, the 'ought' to which those impartial sentiments have reference.

As Guyer puts it, Kant 'claims [that] the connection must be synthetic, which he takes to mean that it must be causal. But here an antinomy arises because both of the obvious candidates for a causal relation between virtue and happiness seem impossible.'[282] Kant argues, however, that it is only apparently false 'that virtue can be the cause of happiness' and that this appearance of falsehood 'aris[es] from an empirical restriction of our conception of causality to our own causality in the sensible world of appearance.'[283] In the section of the second *Critique* entitled 'The Existence of God as a Postulate of Pure Practical Reason,' Kant argues that

> there is not the least ground in the moral law for a necessary connection between the morality and the proportionate happiness of a being belonging to the world as part of it and hence dependent upon it, who for that reason cannot by his will be a cause of this nature and, as far as his happiness is concerned, cannot by his own powers make it harmonize thoroughly with his practical principles.[284]

However, Kant continues, since 'we *ought* to strive to promote the highest good (which must therefore be possible),' 'the existence of a cause of all nature, distinct from nature, which contains the ground of this connection, namely of the exact cor-

[280] Ibid., 118-19.
[281] See *KpV* 5: 110 (92); *Gr* 4: 393 (7).
[282] Guyer, *Kant on Freedom, Law, and Happiness*, 348. See *KpV* 5: 111-12 (93-4).
[283] Ibid., 349.
[284] *KpV* 5: 124 (104).

respondence of happiness with morality, is also *postulated*.'[285] Guyer follows Kant in this direction, and associates Kant's deployment of the worthiness to be happy idiom with the latter's argument for this postulate. In terms of the text, of course, Kant does associate the various elements of his causal approach (include this theological postulate) with his concept of worthiness to be happy. The latter shows up in the midst of Kant's account here, after all. Theoretically, however, the two frames of reference diverge. The thinking that Kant's gloss expresses does not drift in the direction of the synthetic *aprioricity* of *causality* (which would have to subsume even 'a causality,' embodied in 'a supreme cause of nature,' that is 'in keeping with the moral disposition,'[286] i.e., that physically 'permits' happiness only to those agents with such a disposition), but remains with the synthetic *aprioricity* of the *normative* judgment that those 'sentiments' signify.

Guyer reconciles Kant's causal approach with his notion of worthiness to be happy by naturalizing the latter. That agent *deserves* to be happy who participates fully (all the way down into the heart of her motives) in the commission of deeds which, taken in tandem with the deeds of all other such participants, would *physically cause* the happiness of each and all in an ultimate scenario where the laws of nature turned out to have been conducive to this all along (i.e., to the extent, precisely, that these laws turn out to have had the author of the moral law as their author as well). As Guyer observes, 'Kant generally conceives of the highest good as a condition to be realized *in nature*.'[287] But the difference between the 'is' of the causal conception of this 'realization' (precisely 'in nature') and the 'ought' of the normative emphasis of Kant's gloss must be smoothed over. Kant's unpublished notes offer abundant resources for achieving this.

There, for example, Kant argues that practical reason regards happiness as a state of the human agent's empirical affairs which, objectively speaking, 'is only possible through the consensus of the whole with [the individual agent's] *natural universal* [and not merely her particular, self-serving] will.'[288] Kant sees happiness as a product of the agent's will, so construed; happiness is tied *causally* to the pro-

[285] *KpV* 5: 125 (104).
[286] *KpV* 5: 125 (104).
[287] Guyer, *Kant on Freedom, Law, and Happiness*, 379.
[288] *R* 6969 19: 216 (453) (my emphasis).

gressive activity *within nature* of practically-rational subjects, but it is also imbued with a certain dimension of justice or fairness. Again, this means seeing happiness, ambiguously, as the upshot of both remunerative/distributive justice and physical causality. As Kant argues in a note to *IPP* from the latter part of the 1770s, the virtuous agent is equipped with principles in accordance with which she and any other rational being could come to consensus concerning what she ought to do in the future (i.e., moral principles in the primary, forward-looking sense). But he adds that these agents are equipped, at the same time and for this very reason, with 'the *principium* of the *epigenesis* of happiness.'[289] Virtue is connected with happiness to the extent that the universal happiness of moral agents is a *direct consequence* of what the moral law actually enjoins. But there is a sense, then, in which this happiness is enjoined as well.[290]

It is important to recognize that, whether in these *Reflexionen* or in his published work, when it comes to this causal construal of the connection between morality and happiness, Kant does not make *God* the cause of moral agents' happiness. Rather Kant approach is to represent the connection between morality and happiness as an ultimate outcome that would be brought about by moral agents themselves, that is, by actions (of theirs) that put into empirical practice their antecedent, irrevocable, free adherence to the moral law. This felicitous outcome is entailed, indeed *necessitated*, by the laws of nature (authored by God), but *caused* by these agents—or it *would* be, ultimately, on the assumption, again, that these laws include some (unknowable by us) in conformity with which such agents' free and moral deeds would, in some ultimate scenario, turn out to have been necessitating their happiness all along.

Kant, in any case, does not outright cheat and naturalize a necessary, *a priori* 'combination' that is better regarded as normative. These two tendencies, one of which antedates the other, simply coexist in his thinking. Guyer, as we saw in chapter 1, is struck by the 'profound mystery'[291] of Kant's constant articulation of his worthiness to be happy idiom and tries to harmonize it with Kant's dominant critical period thinking. This is an unhelpful strategy, however. Kant's gloss represents a dis-

[289] *R* 6867 19: 186 (444).
[290] Guyer, *Kant on Freedom, Law, and Happiness*, 340, 345.
[291] Ibid., 117.

tinct point of view on the connection between happiness and morality, one of several tendencies in his thinking and the most subterranean. In this case, we are faced with a tendency that constantly (as it were, parenthetically) transects arguments in which it plays no substantive role. The really mysterious thing is that Kant never separates it out and thematizes it as such—in spite of its remarkable pervasiveness.

The second *Critique* account of the relationship between morality and happiness that I have discussed above is obviously a view that Kant holds. But it does not really express the thinking to which his gloss refers—even though the gloss appears in it. Again, Kant's work is pierced repeatedly by uses of this idiom that do not always connect directly—or not well—with their immediate context. There are degrees of fit here, however. The causal account relates the two elements, happiness and morality, extrinsically and necessarily, which brings this account into the conceptual vicinity of Kant's gloss. The sticking point, however, is that the causal account does not relate happiness and morality in a truly *normative* manner. Indeed, because it turns on a particular notion of natural causality, it precludes this normative dimension altogether. And this is of the essence. In the next sub-section I explain why.

Morality as a condition without which happiness is deontically impossible

Kant's second *Critique* account of morality's and happiness' relationship entails that their relationship is a peculiar kind of physical fact (peculiar because inflected by Kant's theological 'postulate' and hence outside the purview of anything that might be brought within the ambit of empirical enquiry). It is a matter of the 'synthetic' combination, in a 'real connection,' of elements that are conceptually combined *a priori* (in the idea of the *summum bonum*). But its upshot is natural and factual. The notion of necessity that is at work in Kant's causal account of the relationship between morality and happiness makes the ultimately perfect disposition of these elements *physically inexorable*.[292] If it were really so that the moral deeds of rigorously moral agents could bring it about, within the ambit of nature, that *all were happy*

[292] See note 30, above.

then not only their happiness, but even (if we take 'all' in as broad a sense as possible) the happiness of immoral agents would be an inevitable fact.[293]

The normativity with which Kant imbues this outcome—and which Guyer traces and explicates in remunerative/distributive terms—is not the normativity that is expressed in Kant's habit of glossing 'morality' as 'worthiness to be happy'. Guyer sees that Kant's use of this idiom antedates the critical period, but then tries to make it fit more comfortably within the latter framework than is either necessary or possible. Instead, Kant's gloss directs us to a way of conceiving of the relationship between morality and happiness where the relevant necessity is not physical (not even in the peculiar sense detailed above), but rather *normative*; and where the ultimately perfect disposition of these elements would have to be *forged*, by being put into practice as it were, by a third party for whom this task was merely *practically* necessary.

However, as I will show below, the connection whose 'forging' Kant's gloss represents as necessary holds (or ought to hold), not between morality and happiness, but between immorality and unhappiness. In short, Kant's gloss expresses the idea that immoral agents are perfectly capable of happiness (my 'extrinsicness' claim), but that they *ought to be* unhappy (my normativity claim) and, indeed, that they unconditionally *must* be unhappy (my necessity claim).

This abrupt transition from talk about morality and happiness to talk about *unhappiness* and *immorality* is no accident. It is an inevitable requirement for properly understanding the expressive force of Kant's gloss, as I will show. The normativity that Kant's gloss signifies pertains to morality and happiness, to be sure, but it pertains in a different sense and with even greater *urgency* to unhappiness and immorality.

Guyer associates the concept of worthiness to be happy with a physically necessitated state of distributive or remunerative justice in which all deserving agents are assured of their happiness to the extent that they have aimed at the happiness of all. On this view, there is no reason that the happiness that morality (causally) entails

[293] This suggestion is not at all integral to my argument, but mentioned, really, in passing as a point of interest. I am not claiming that Kant actually *says* this, of course, but rather, following Guyer's reading of Kant's pre-critical notes, that he implies as much (i.e., in those notes). I am simply noting a consequence of Kant's claim (in this context) that action is moral when it aims at 'the happiness of all,' taken together with his account of what would have to be the case in terms of the fit between moral action and laws of nature if this universal happiness were to be actually realizable.

must exclude immoral agents.[294] Here, in fact, the (here physical) possibility of an immoral, but happy, agent is not ruled out. Guyer aligns his reading of Kant's *Reflexionen* with this causal account and comes up with the idea that, again, to be worthy to be happy just means that an agent has contributed to this ultimate happiness. But this is the happiness, precisely, '*of all*.' Moral agents are worthy of it and immoral agents—also caught up in it[295]—unworthy.

This comes closer to what I take Kant's gloss to represent. The happiness that morality causes, by being the happiness of all, subsumes the happiness of moral and immoral agents alike. There is just the one happiness—and immoral agents are susceptible to it. Again, the (physical) possibility of an immoral, but happy, agent is not ruled out. If morality—or maximally (hence morally) ordered freedom—were to turn out to be the cause of the happiness of all within some order of things overseen by God (i.e., if *this*, the world that we know, that is, nature, were to turn out to be the order in question), it might be possible for immoral agents to get caught in this happiness' net. Then, they would not deserve it—in the sense that they had not originally participated in its actualization. But this would not entail that their happiness ought to be taken away. This is a significant oversight in Guyer's explanation of Kant's concept of worthiness to be happy. Kant's habitual gloss shows that he takes the possibility of immoral, hence unworthy, and yet happy agents to be a worrying, morally problematic one.

Without intending to, Vleeschauwer offers us a way back to the normativity of Kant's gloss and the worried 'sentiments' of Kant's impartial spectator. This entails reading this commentator against the grain, however. He writes,

> The autonomy of the will forbids us to make an immediate synthetic connection between the two constituents of the practical nature of man [the two ends, the moral good and happiness, the objects of pure practical reason and inclination respectively]. It does allow nevertheless *a mediate synthetic connection*, *an external connection* between them, that is, a connection

[294] Note that I am not using 'entails' here in a technical, logical sense. I mean it in the spirit of the claim that 'smoke entails a fire.' But then again, given that Kant takes the laws of nature to be laws that are legislated for the nature and relations of the objects of our experience by us, that is, by the understanding through which we 'think' the basic nature and relations of those objects, then it might be permissible to say that physical causes *do* 'entail' their effects, that is, 'logically,' in accordance with what Kant takes to be the 'laws of thought.'

[295] This, I take it, is entailed by Guyer's argument in Guyer, *Kant on Freedom, Law, and Happiness*, 392.

> which does not arise from the very nature of man but from *the intervention of another being* who would have nature within his power.[296]

Of course, this is Vleeschauwer's gloss on Kant's causal resolution of the problem of morality's and happiness' relationship, the account of the second *Critique*. However, he construes God as the cause of the happiness of moral agents—rather than as the author of the laws in accordance with which *they*, by means of their moral deeds, bring their happiness into being. Here, as elsewhere, Kant's gloss creates confusion. It is possible, though, to read Vleeschauwer's interpretation as a declaration concerning what the gloss itself signifies: the notion of '*a mediate synthetic connection*, *an external connection between [morality and happiness]*…a connection which…[could only arise by way of] *the intervention of another being* who would have nature within his power.' There is one other thing that would have to be added here, however: this being would have to have a *reason* to forge this connection. Kant does not take *pure* reason to offer one. Kindness, however, does. By contrast, the forging of an 'external connection' between *immorality* and *unhappiness*, the forging of 'a mediate synthetic connection' between *these* elements is another matter. For Kant, I claim, pure practical reason does not demand that moral agents be happy. It does, however, demand that immoral ones be unhappy. Again, this is the necessity to which the gloss adverts.

As we saw above, Kant's second *Critique* discussion of virtue *qua* (physical) cause of happiness makes reference to what holds concerning the happiness of moral agents 'in the judgment of an impartial reason' and with respect to 'the perfect volition of a rational being that would at the same time have all power.'[297] I claimed that Kant's ambiguous description seems to represent his impartial spectator's concurrence or acquiescence in the happiness of the moral agent as something that is deemed *necessary*, not merely from the point of view of the 'rational finite being,' but from the *impartial* point of view of this spectator as such. I claimed, in other words, that Kant represents this concurrence as acquiescence in a demand of pure reason.

The spectator that Kant has in mind, however, is—to be precise—*impartial*. If this observer is also a kind one, then nevertheless, to the extent that her judgments

[296] J. H. Vleeschauwer, *The Development of Kantian Thought: The History of a Doctrine*, trans., A. R. C. Duncan (London: T. Nelson, 1962), 122 (my emphases).
[297] *KpV* 5: 110 (92-3).

are 'the judgment[s] of impartial reason'[298] then even her kindness, which is a particular kind of inclination, is set to one side. Kant's *Groundwork*'s formulation of this spectator's fundamental interest is far clearer than what we encounter in the second *Critique*. He writes:

> Power, riches, honor, even health and the complete well-being and satisfaction with one's condition called *happiness*, produce boldness and thereby often arrogance as well unless a good will is present which corrects the influence of these on the mind and, in so doing, also corrects the whole principle of an action and brings it into conformity with universal ends—not to mention that an impartial rational spectator can take no delight in seeing the uninterrupted prosperity of a being graced with no feature of a pure and good will, so that a good will seems to constitute the indispensable condition even of worthiness to be happy.[299]

At least a couple of prominent commentators make the mistake of thinking, not only on the basis of Kant's comments in the second *Critique*, but on the basis of the foregoing as well, that, in Kant's view, a rational, impartial spectator would be 'pained' or 'dismayed' at the sight (or mere prospect) of a *moral*, but *unhappy* agent. But this is not what Kant thinks.[300]

Guyer, for example, asks 'if freedom or autonomy has nothing to do with happiness at all, why should virtue be equated with the worthiness to be happy, and *why should it seem unacceptable to a rational being to see virtue unaccompanied with happiness*.'[301] More pointedly, he observes a bit later that, in the second *Critique*, Kant

> first proceeds as if virtue and happiness are two entirely separate goods connected only by what has recently come to be called a "reactive attitude": his idea seems to be that although there is no intrinsic connection between the moral good of virtue and the natural good of happiness, *it would nevertheless pain an impartial rational observer to see someone who is successful in striving after the moral good of virtue nevertheless be frustrated in her independent but acceptable natural desire for happiness*.[302]

Korsgaard, referring to both the second *Critique*'s and the *Groundwork*'s representations of this purely objective point of view, observes that while the 'impartial observer' does of course disapprove of the idea of an immoral, hence unworthy, but

---

[298] *KpV* 5: 110 (92).
[299] *Gr* 4: 393 (7).
[300] In addition to the foregoing, see also *R* 6280 18: 547-8 (352).
[301] Guyer, *Kant on Freedom, Law, and Happiness*, 97 (my emphasis).
[302] Ibid., 118 (my emphasis). See *KpV* 5: 110 (92-3). For the original appearance and use of the term, 'reactive attitude,' see P. F. Strawson, 'Freedom and Resentment,' in *Freedom and Resentment and Other Essays* (London: Methuen, 1974). For further clarification and development of the notion see R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press: 1994).

nevertheless happy agent, 'the impartial observer is *equally dismayed* by the idea that the virtuous person be without happiness.'[303]

Either these commentators err—or they need to be much more precise. Kant's point in the second *Critique* is not that 'an impartial rational observer,' just as such, would experience a kind of 'pain,' as Guyer puts it, at the sight of a virtuous agent who was 'frustrated in her independent but acceptable natural desire for happiness.' Kant's point—not clearly made—is that such an observer would acquiesce in the goodness of this desire, given the same observer's antecedent acquiescence in that desire. In the second *Critique* 'the perfect volition of a rational being that would at the same time have power' calls out for happiness, but this same being (as described here only) is also apparently one that 'need(s) happiness' and is 'also worthy of it.'[304] This seems to be what Kant is saying, although his sense is ambiguous. Another less specific possibility is that this 'rational being' simply has an *interest* in happiness (its own or others'). The ambiguity, here, inheres in Kant's reference to 'the judgment of an impartial reason,' on the one hand, and his description of an all-powerful 'rational being' that has an interest in happiness (in short, a self-loving, or a *benevolent* one), on the other. The main point, however, is that *impartial* reason, just as such, has no interest of *its own* in happiness, but will *assent* to the happiness of moral agents, at least, and, only where supplemented by an interested benevolence, will *will* it.

Korsgaard's evident misreading is particularly problematic in that she refers to both the ambiguous second *Critique* text and the rather clearer *Groundwork* one. She, in particular, resolves an obvious asymmetry in Kant's thinking, favouring an impartial spectator that is *just as* 'dismayed' by the unhappiness of moral agents as by the happiness of immoral ones. Korsgaard's misreading of Kant's position is typical of the style of her overall project (where the latter pertains to Kant), which articulates and defends a 'Kantianism' that has rather cleaner edges and neater lines than one finds in Kant himself. Guyer's misreading, I suggest, is a function of the fact that he has already interpreted worthiness to be happy as *desert* of happiness and thus as

[303] Korsgaard, 'Kant's Formula of Humanity' (my emphasis). For cognate instances of this reading see Beck, *A Commentary on Kant's Critique of Practical Reason*, 245; Broad, *Fives Types*, 134; A. C. Ewing, 'Paradoxes of Kant's Ethics,' *Philosophy* 13, no. 49 (1938): 41; Smith, 'Worthiness to Be Happy': 177; J. E. Walter, 'Kant's Moral Theology,' *Harvard Theological Review* 10, no. 3 (1910): 29.
[304] *KpV* 5: 110 (92-3). See also *LEC* 27: 250-1 (47).

an outcome that is practically necessary, as a matter of fair distribution, entitlement, and right.

Pace Guyer, something very much *like* a reactive attitude is at work in the heart of Kant's 'impartial rational spectator.' In a sense, Kant offers no complaint in the face of such a reading. The 'judgment of an impartial reason'[305] *does* find expression in an 'attitude,' indeed in a feeling: a feeling of disapprobation. Kant will simply claim (see chapter 4) that this feeling, like the feelings of respect for the moral law, the humiliation of 'self-conceit' before it, moral self-satisfaction through conformity to it, and moral self-contempt through disobedience of it, is a feeling whose ground lies in pure practical reason itself. For Kant, these attitudes are *active*—a sign of the pure activity of judgment—not 'reactive,' but they are still affective in their expression. I address this topic more fully in chapter 4. For now, it will suffice to say something about the *judgment* that the impartial rational spectator's attitude of disapprobation expresses.

In essence, the judgment of impartial reason is that the happiness of immoral agents is absolutely, objectively *bad*. It is the judgment that immoral agents *ought not to be* happy. And, consequently, it is the judgment that such agents' unhappiness is a *good* that ought to be realized.

Of course, this does not follow *immediately* from Kant's claim that 'an impartial rational spectator *can take no delight* in seeing the uninterrupted prosperity of a being graced with no feature of a pure and good will.' All that Kant explicitly claims here is that, whatever else such a being might judge concerning it, or feel about it, the 'uninterrupted prosperity of a being graced with no feature of a pure and good will' can provoke no feeling of 'delight' in such a being's heart (supposing that she is endowed with a faculty for such feeling). It does not follow from this, just as such, that some *other* response (disapprobation, repugnance) is provoked instead. It is compatible with such a being's inability to delight in the happiness of an immoral agent, that the happiness of such agents is simply a matter of *indifference* to her. Guyer and Korsgaard, however, clearly take themselves to be warranted in asserting that Kant's 'impartial rational spectator' is *not* indifferent to the happiness of immoral agents. They overlook the asymmetry that holds between this spectator's value

[305] *KpV* 5: 110 (92).

judgment (and the 'pure,' moral 'feeling' that is correlated with it) concerning the unhappiness of moral agents and the happiness of immoral ones, but they are right, nevertheless, to affirm that Kant's spectator is 'pained,' at least by the latter prospect.

Of course, this 'delight' and 'pain' are *affect*—the 'spectator' that experiences these feelings is passively affected and not simply engaged in the activity of rendering moral judgments. Nevertheless, for Kant, 'moral feelings' do correspond directly to a certain class of judgment—moral ones. The impartial rational spectator's 'pain' or 'dismay,' here, is of a piece with her 'respect' for the moral law—affect, again, but affect that arises from a certain stimulation of her general receptivity to being-affected, by considerations, in this case, of pure practical reason. The 'impartial rational spectator's' 'delight' (or otherwise) is not equivalent to the judgment that the thing delighted in is good—but it nevertheless signifies this judgment.

Taken by itself, this passage from the opening of the *Groundwork* does not allow us to say definitively that, for Kant, the rational impartial spectator finds the notion of an immoral, but happy agent repugnant and that this affect arises from and gives empirical expression to a 'pure' judgment of practical reason. But, taken together with the rest of his thinking about this matter, it is clear that Kant does not mean to leave open the possibility that this spectator is simply apathetic when it comes to the prospect of an immoral agent ending up a happy one (delighting in the happiness, and suffering from the unhappiness, of moral ones, but indifferent to the happiness or unhappiness of immoral ones). Rather, along with respect for the moral law, the 'humiliating' abasement of self-conceit before it, and moral self-contempt in the wake of disobedience of it, Kant takes the disapprobation of the happiness of immoral agents and the demand that they be unhappy to be a feeling and a demand that are grounded in pure practical reason

It is in tandem with other key elements of Kant's thinking and writing, then—his assertion, above all, that the law of punishment is a categorical imperative, his identification of punishment with the 'forfeiture' of happiness, his affirmation of that 'necessity' of punishment that warrants an 'inference to a future life'—that this opening passage of the *Groundwork* supports my claim that it is *disapprobation* and not *apathy* that the impartial rational agent, also receptive to the modification of affect, feels at the very prospect of an immoral, but happy, agent. She *delights* in the happiness of

the moral agent and this delight is both grounded in kindness and a sign that she judges this happiness to be good (she does not judge it to be *absolutely* good—if she did, then kindness would not be required to motivate her to will it). On the other hand, her disapprobation of the happiness of immoral agents is not, for Kant, grounded in the negative of kindness, which would be malice, or a desire for revenge, or some such thing. If it were, then it would not be possible to regard the ethical-eschatological analogue of the 'law of punishment,' Kant's 'law of unhappiness' (the law, as I put it earlier of the 'making-unhappy' of immoral agents) as a categorical imperative. But it is. The happiness of immoral agents is not merely bad from a parochial point of view; it is absolutely bad. Eschatologically speaking, the punishment of an agent whose will is qualified by 'inner wickedness' is *necessary*.

Now, it is common to characterize morality, virtue, possession of a good will as the source or condition of happiness' goodness or value. Korsgaard, for example, argues that '[a] conditionally good thing, like happiness, is objectively good when its condition is met in the sense that it is fully justified and the reasons for it are sufficient. Every rational being has a reason to bring it about.'[306] Korsgaard thinks that this applies to both the happiness of our fellows and the highest good. On this reading, the primary, forward-looking moral law is regarded as a law that tells us both what to do, and prescribes that our doing *that* be the condition *sine qua non* bearing on the possibility of happiness being judged good, or of its being deemed valuable at all from the point of view of the kind of being for whom the moral law is actually binding.[307] The unconditional goodness of morality, the good will or its goodness, is 'the condition of the value of other good things' and 'the source of value.' Happiness is not 'fully justified,' then, unless it 'stand[s] in the right relation to' that which is 'unconditionally good.'[308] If we *do* meet this condition, however, as Wood puts it, 'it is objectively *good* that we be happy.'[309]

But the upshot of my argument, above, is that happiness is never absolutely, objectively good, for Kant. Happiness can never be 'fully justified,' as Korsgaard puts it, in a sense that would entail that its production was practically *necessitated*. Grant-

[306] Korsgaard, 'Kant's Formula of Humanity'.
[307] See, for example, Allison, *Idealism and Freedom*, 114; Wood, *Kant's Ethical Thought*, 42.
[308] Korsgaard, 'Aristotle and Kant': 488. See also Korsgaard, 'Kant's Formula of Humanity', 118; Sikka, 'On the Value of Happiness: Herder Contra Kant'.
[309] Wood, *Kant's Moral Religion*, 167.

ed: happiness is not unconditionally good. The impartial spectator makes a judgment concerning the value of happiness. But if the judgment that something is good gives the judging agent a reason for bringing it into existence, then the impartial spectator does *not* judge that the moral agent's happiness is good in a straightforward sense. The impartial rational spectator, just as such, has to be endowed with *benevolence* as well as reason, in order to have a reason for bringing the moral agent's happiness into existence. Rather, the impartial rational spectator's judgment is a negative one: it would *not* be an objectively bad thing for the moral agent to end up happy, *given* that someone (she herself, say, or God) had a reason to make her happy. But pure reason does not have any interest here; only 'impure' reason, reason inflected by, and inclined to, kindness does.

I will argue all of this more closely in the coming chapters, in discussing the law that commands that immoral agents be unhappy (chapter 3) and the practices and agents that this law prescribes and binds (chapter 4). For now, I am simply setting these matters out for initial review. Here is what Kant's impartial reason judges absolutely, objectively good: the unhappiness of the immoral agent; objectively bad, his happiness. Here is what serves the interests of pure reason, on Kant's view: the unhappiness of immoral agents. Happiness is only an objectively good thing from the point of view of pure practical reason to the extent that the latter is inflected by benevolence. Pure practical reason judges the unhappiness of immoral agents a good *without any analogous inflection at all*.

But of course the unhappiness of immoral agents is a good, too, from the subjective, parochial points of view of malice, and pathological self-contempt, and vengeful wrath as well. Kant's habit of glossing 'morality' as 'worthiness to be happy' is in play long before he gets to the point of distancing this habit, theoretically, from such attitudes; long before he gets to the point of theoretical efforts that serve the gloss by disposing (but not programmatically, or well) of the possibility that something 'impure' lies in back of it from the outset. In a sense, in its originary function, Kant's gloss is the incipient form of a kind of ultimately explicit theoretical patrimony: that the judgment that immoral agents deserve to be unhappy is warranted *a priori* and that the feeling of disapprobation with which we greet the prospective happiness of an immoral, hence unworthy, but happy agent (oneself included), is a

feeling that it is incumbent on us to feel (just as, finally, Kant will imply, at least, that it is incumbent on us to respect the moral law—in addition to acting in accordance with it).

On Kant's account, the judgment that something is good gives the judging agent a reason to actualize it (see chapter 4). The judgment that something is objectively bad gives the judging agent a reason to omit bringing it to pass and, perhaps, to occlude its arrival on the scene. The judgment that the moral agent's happiness is good is really the judgment that nothing militates against it. The judgment that an immoral agent's happiness is objectively bad means that something ought to be done about it—*in a special sense* (see chapter 4). For now, suffice it to say that morality is practical reason's *a priori* condition for approving of a hypothetical distribution (grant or apportionment) of happiness, for assenting to the state itself (were it realized), or for approving of the prospective wish for happiness, or of effort or work aimed at it.[310] The realization or distribution of what is wished and worked for, here, is regarded as a real possibility—even without consideration for morality. The notion that human effort (integrally supplemented, perhaps, by cooperative divine intervention) can bring happiness into existence—while disregarding the deeply 'authentic' morality of disposition—is entirely coherent. But, on Kant's critical account, no rational and impartial being (including the agent herself, to the extent that she judges in accordance with pure reason as such) would ever *approve* of this action (or its end). In short, the objectivity of the condition is strictly *normative*, not *causal*.

Worthiness to be happy, unworthiness to be happy, and desert

In this final section I further explicate my claim that Kant's habit of glossing 'morality' as 'worthiness to be happy' does *not* express the view that moral agents ought simply to be happy. More specifically, I argue that his gloss does *not* express the view that moral agents deserve to be happy in any sense of 'deserves' that subsumes the notion of practical necessity. By contrast, I show that his gloss signifies the view that immoral agents ought and deserve to be unhappy.

[310] See *Rel* 6: 46 n. (91 n.); *KpV* 5: 130 (108-9); *MdS* 6: 481 (223); *R* 4463 17: 561 (138); *R* 6989 19:221 (455); *LEC* 27: 418 (180-1); *R* 7093 19: 247 (460). See also Dews, *Idea of Evil*, 21.

Kant's concept of worthiness to be happy represents happiness and morality as elements that stand in a particular kind of extrinsic relationship to one another. The connection between them *looks*, at least, like a synthetic and necessary connection *a priori*. Kant tries to make sense of this appearance by developing an account with the notion of causality at its centre. But this is not the connection that his gloss signifies. With respect to some of its instances it would be egregiously anachronistic to say that the synthetic *a priori* connection that the gloss signifies is a normative one. Here, rather, the theoretical claim that something ought necessarily to be the case gives expression to an antecedent sensitivity to the inexorability of something (a practice) that cannot be given up, of a complete failure of imagination with respect to alternative courses of action. Ultimately, this coalesces in the appearance of what Kant—in his thinking both about causality and about the primary, forward-looking moral law—will come to identify as the synthetic *a priori*.

In the course of this section I specify the normativity that marks the nearer edge or surface of this inexorable 'something' (i.e., the practices of punishment, that is, of retribution, and of 'retribution-sensitive' imputation). My point of departure, here, is my claim that the rational, impartial spectator's displeasure at the prospect of immoral, unworthy, but happy agents is not balanced, on the other side, by similar consternation (Guyer's 'pain,' Korsgaard's 'dismay') at the prospect of moral and yet unhappy ones. I argue, that there is a corresponding asymmetry in the case of desert. More than this, I show that there is a sense in which reason *rejoices* at the unhappiness of immoral agents in a way that it cannot—on the view that is encapsulated in Kant's gloss—in the case of the happiness of moral ones.

Desert

Obviously, there are many uses of 'desert,' none without its associated controversies.[311] My discussion treats desert as a state of an agent's affairs that, practically

[311] The topic of desert is an enormous one. For a good overview of recent thinking in this area see Serena Olsaretti, ed. *Desert and Justice* (Oxford: Clarendon Press, 2007). Further useful reading may be found in Richard Burgh, 'Do the Guilty Deserve Punishment?,' *The Journal of Philosophy* 79, no. (1982); Geoffrey Cupitt, 'Desert and Responsibility,' *Canadian Journal of Philosophy* 26, no. (1996); J. L. A. Garcia, 'Two Concepts of Desert,' *Law and Philosophy* 5, no. 2 (1986); Hyman Gross, *A Theory of Criminal Justice* (Oxford: Oxford University Press, 1979); A. T. Nugen, 'Just Desert,' *Journal of Value Inquiry* 31, no. (1997); James Rachels, 'What People Deserve,' in *Justice and Economic Distribution*, ed. J. Arthur and W. H. Shaw (Englewood Cliffs, NJ: Prentice-Hall, 1978);

speaking, *necessitates* a further course of action by a third party. The basic idea may be laid out schematically in a manner that resembles my definition of worthiness to be happy in chapter 1. Desert is a three-place relation between a subject (*A*), a thing or state of affairs that *A* deserves (*x*), and a 'desert-basis' (*y*), or the property of the agent in virtue of which she is deemed to deserve *x*.[312] One says, then, that '*A* deserves *x* in virtue of *y*.'[313] It is key that *x* and *y* be related in an appropriate way (which point, too, will be a matter of debate). In the case of worthiness to be happy, the property that constitutes *y* is a fact about the deserving subject *qua* free agent, her morality. The same applies here. The claim that '*A* deserves *x*' must have reference to a fact about *A* and, more specifically, must be such that *A* is somehow the source or author of the desert-basis (*y*).[314]

It is a basic assumption of the theory of desert that desert claims translate into 'ought' ones.[315] I deploy this assumption here. If *A* deserves *x* in virtue of *y*, then, given *y*, A ought to have (or perhaps be) *x*. If immoral agents deserve to be unhappy, then immoral agents ought to be unhappy. To assert that something ought to be the case, moreover, is to regard this state of affairs (that ought to be the case) as the possible effect of the possible action of some agent—and it is to regard that action as a practically necessary one.[316]

On this score, in relation to happiness and unhappiness, Kant's thinking evinces a stark asymmetry between 'positive desert' (desert of 'something desirable,' or 'nice treatment'), on the one hand, and 'negative desert' (desert of 'something undesira-

---

George Sher, *Desert* (Princeton: Princeton University Press, 1987); James Sterba, 'Justice and the Concept of Desert,' *The Personalist* 57, no. (1976).

[312] Another possibility is 'desert-ground' Joel Feinberg, 'Justice and Personal Desert,' in *Doing and Deserving* (Princeton: Princeton University Press, 1970).

[313] See Pojman and McLeod, eds., 61-2.

[314] Of course there are other notions of desert. A desert-basis need not be tied to the deserving subject's morality at all. Moral desert is not the only kind of desert that there is. We speak, for example, of sunsets that deserve to be admired. And we also think that victims of various kinds deserve compensation—not on the basis of what they have done, but on the basis of what they have *suffered* (ibid., 63).

[315] See ibid., 62.

[316] Guyer clearly sees that as soon as the notion of an extraneous 'forging' comes into view, then we have to conceded that Kant's equation of 'virtue' with 'worthiness to be happy' references 'the ordinary, natural sense of happiness, something that is not an immediate consequence of virtue but needs to be connected to it by an extra step' (Guyer, *Kant on Freedom, Law, and Happiness*, 118). Note, however, that I am eliding the fraught business of distinguishing between the notions of 'what agents ought to do' and 'what ought to be the case,' along with the further task of showing how they—and talk about them—are related. On this score see John Francis Horty, *Agency and Deontic Logic* (Oxford: Oxford University Press, 2001), 1ff..

ble,' or 'nasty treatment'), on the other.[317] Conceived in terms of an obligation to act (i.e., an obligation to bring it about that an agent is happy or unhappy), worthiness to be happy is not desert of happiness. Rather, it is the negation of desert of unhappiness.[318]

Worthiness to be happy is not desert of happiness

Moral merit and morally necessitated reward

For Kant, there is no such thing as moral merit and hence no moral analogue of 'rightful' reward (i.e., remuneration). To fulfill the demands of the moral law is not even praiseworthy. Virtue, no matter how great, is never 'deserving of wonder,' Kant says; it is neither 'extraordinary' nor 'meritorious.'[319] In matters of morality (as opposed to mere legality) the human agent can produce no 'surplus over and above what he is under obligation to perform.'[320] Indeed, as Kant puts it in a note from the early 1780s, when it comes to morality (in its wide sense, as virtue), '*everything is indebtedness*.'[321]

Some of Kant's clearest expressions of this position are couched in theological language. Others are expressed in terms of the demands and judgments of conscience. These are major topics that I will take up more fully in chapter 4. A couple of examples will suffice for present purposes.

In *The Metaphysics of Morals*, for example, Kant asserts that

> there is no place for *reward* (*praemium*, *remuneratio gratuita*) in justice toward beings who have only duties and no rights in relation to another, but only in his [this other's] love and beneficence (*benignitas*) toward them; still less can a claim to *compensation* (*merces*) be made by such beings.[322]

My employer is obliged to pay me. My neighbour is obliged to reward me if I do something for him for which he has posted a reward (e.g., if I spend my time and money looking for his lost pet). But God is not obliged to compensate or reward me—no matter how much I exert myself and no matter how pure my motives are in

[317] This useful terminology is Garcia's. See Garcia, 'Two Concepts of Desert': 219, 222.
[318] Formally expressed: for any agent, A, let Ah stand for 'A is happy.' Let ***P*** stand for 'is permissible,' ***O*** for 'is obligatory,' and ¬ for 'is not.' Let ***O*** = ¬***P***¬. Then for any agent, A, such that ***O***¬Ah, the negation of that would be ¬***O***¬Ah, which is equivalent, by definition, to ¬¬***P***¬¬Ah and so, too, to ***P***Ah.
[319] *Rel* 6: 48-9 (93).
[320] *Rel* 6: 72 (113). See also *Rel* 6: 146 n. 3 (170 n.).
[321] *R* 6092 18: 449 (my emphasis).
[322] *MdS* 6: 489 (231).

doing so. 'We have no [claim of] merit against God,' Kant writes, 'but rather pure obligation.' If Kant does imply, in the foregoing, that there is a 'place for *reward*' in God's 'love and beneficence,' 'reward,' here, has reference to a 'happiness,' as Kant says elsewhere, 'which one does not deserve,' but which is, rather, 'from [*aus*] grace.'[323]

Elsewhere in *The Metaphysics of Morals* Kant expresses the same position in terms of conscience:

> [i]t should be noted that when conscience acquits him it can never decide on a *reward* (*praemium*), something gained that was not his before, but can bring with it only *rejoicing* at having escaped the danger of being found punishable. Hence the blessedness found in the comforting encouragement of one's conscience is not *positive* (*joy*) but merely *negative* (relief from preceding anxiety); and this alone is what can be ascribed [*beigelegt*] to virtue, as a struggle against the influence of the evil principle in a human being.[324]

This connects up with what Kant says in the *Religion* when he writes that '*[w]orthiness* has…always only negative meaning (not-unworthiness), that is, moral receptivity to such goodness,' by which 'goodness' he means, again, a happiness whose distribution is grounded in the distributor's kindness.[325] Or again, as Kant says elsewhere, 'good behaviour' (*Wohlverhalten*) does not necessitate reward, 'but rather contains only receptivity to it.'[326]

To sum up, desert is calculated in accordance with rules that govern agents' relations to one another within a system of rights and duties. The notion of desert cannot be applied to agents that have only duties in relation to another and no entitlements at all. For Kant, ethics, the part of 'morality' that cannot be externally coerced by the state, is such a system. From a moral-religious point of view, another way of looking at this is to say that the human being has only duties in relation to God, regarded as the source of the law. No matter how often I do my duty I can never accrue a *right* or *credit* in relation to the moral law (or its legislator).

Virtue as juridically meritorious

Again, on Kant's view there is no such thing as moral merit and hence no moral analogue of 'rightful' reward. Contrast this with the situation of juridical merit. The lat-

[323] *R* 7166 19: 262. See also Jacqueline Mariña, 'Kant on Grace: A Reply to His Critics,' *Religious Studies* 33, no. (1997): 382-3; Smith, 'Worthiness to Be Happy': 183.
[324] *MdS* 6: 440 (190-1).
[325] *Rel* 6: 146 n. 3 (170 n.).
[326] *R* 7114 19: 251-2.

ter arises from doing more than what is 'due'[327] and *can* necessitate reward as its consequence. In political life, as Kant says in the *Metaphysics of Morals*, 'virtue…can be said here and there to be meritorious and to deserve to be rewarded.'[328] This is because, by being moral in Kant's 'wide' sense (i.e., virtuous), it is possible to 'put others under obligation.'[329]

These 'others' are of course other members of the *polis*, which Kant does not take to include God. Obligation is, here, a strictly political notion. For example, Kant writes, '[b]y carrying out the duty of love to someone I put another under obligation; I make myself deserving from him.'[330] Simply doing what is owed (doing neither more nor less than what the law of the land demands), however, does not have this effect.[331] If something can be coerced (i.e., if its omission can be punished by the state), then doing it does not entail juridical merit.[332] Merit attaches only to what *cannot* be coerced, to those instances when 'we think,' as Johnson puts it, that '"she did not *have* to do it."'[333] This is so, too, particularly when the action in question requires the overcoming of significant obstacles. The beneficence of a wealthy person, for example, is not particularly meritorious.[334] However, to do precisely what the law demands is not even praiseworthy, whatever one's motive may be.[335]

Reward as 'rightful effect' and as kindness

Johnson argues that Kant espouses no juridical notion of (deserved, necessitated) reward either, but only of punishment—that neither juridical nor moral merit can be connected with the notion of desert in Kant.[336] As far as what is necessitated goes, Johnson argues, juridical merit entails nothing more than praiseworthiness.[337] This seems right. Admittedly, Kant refers to the '*rightful* effect…of a meritorious deed'—

[327] *MdS* 6: 390-1 (153-4).
[328] *MdS* 6: 406 (165).
[329] *MdS* 6: 448 (198).
[330] *MdS* 6: 450 (199).
[331] *MdS* 6: 448 (198).
[332] *MdS* 6: 227 (19).
[333] Johnson, 'Kant's Conception of Merit: 'Metaphysics of Morals' and Evaluating Actions': 315.
[334] *MdS* 6: 453 (202). See also ibid., 313, 325.
[335] Ibid., 331. Sverdlik makes a mistake by suggesting otherwise (see Sverdlik, 'Kant, Nonaccidentalness and the Availability of Moral Worth': 293); and Meyers also implies that an agent deserves praise if she does deeds that possess 'genuine moral worth' (Meyers, 'The Virtue of Cold-Heartedness': 233).
[336] Johnson, 'Kant's Conception of Merit: 'Metaphysics of Morals' and Evaluating Actions': 328.
[337] Ibid., 329.

a deed through which an agent does more than the law constrains him to do—as '*reward* (*praemium*).' But this is a specifically *rightful* effect, not only because the agent in question has done 'more' than she legally had to do, but also only to the extent that this 'reward' was 'promised in the law' and was her 'motive to it [i.e., to performance of that action].' Barring these conditions, meritorious deeds do not necessitate reward. In fact, Kant says, 'conduct in keeping with what is owed has no *rightful* effect at all.'[338]

By contrast with the foregoing, '[k]indly recompense (*remuneratio s. respensio*) stands in no *rightful relation* [*Rechtsverhältniss*] to a deed.'[339] Morality entails worthiness (not moral merit), as Kant says elsewhere, to receive strictly '*voluntary* goods.'[340] Happiness (or the means to it) may be regarded as a reward, but only if it is regarded as one that is voluntarily bestowed on the one that 'deserves' it and not at all as a 'rightful effect' of anything that the agent has done (i.e., such that the bestowal's omission would have been permissible). One may say to an employer, 'I will work for you only on condition that you promise to pay me'; but one may not say, 'I will only be moral if I can be certain that this will entail that I am happy.'

Unworthiness to be happy is desert of unhappiness

'Whoever does evil,' Kant avers, 'is not worthy [*nicht werth*] to be indulged or spared.'[341] This does not express a kind of option to punish, however, in the way that morality and, with it, worthiness to be happy gives rise to a kind of option to reward. Vice necessitates unhappiness. And, as we saw above, 'unworthiness to be happy' may be taken to denote a positive feature of the human being, a primary property of which 'worthiness' is the negation: 'not-unworthiness.'[342] If I am not moral, then I remain as I was: unworthy to be happy, deserving of unhappiness. *Ethically* speaking, one can deserve to be unhappy in the sense that one can deserve to be punished in some hypothetical, ultimate, eschatological scenario. As we shall see in the next chapter, however, Kant's thinking about capital punishment reveals a kind of overlapping of these political and eschatological frames of reference.

---

[338] *MdS* 6: 227-8 (19) (my emphasis).
[339] *MdS* 6: 228 (19).
[340] *R* 6998 19: 223 (456) (my emphasis).
[341] *R* 6998 19: 223 (456).
[342] *Rel* 6: 146 n. 3 (170 n.).

This chapter's main purpose, again, was simply to answer the question *how* Kant takes morality and happiness to be related, given his habitual gloss. And it has been necessary to delineate the asymmetry described above in order to answer that question clearly. The main part of my argument concerning the nature of 'negative' moral desert, however, which is unworthiness to be happy, and my further exploration of the asymmetry that cuts 'not-unworthiness' off from the political concept of 'positive desert' must wait for chapter 3.

A bad habit in the commentary

Before concluding this chapter, I will draw attention to a key difference between my approach and the approach of most of Kant's other readers—and I will elucidate its upshot. In spite of what I have claimed above, many of Kant's commentators do seem to regard '*is worthy to be happy*' and '*deserves to be happy*' (or relevantly similar constructions) as synonymous expressions. They regularly deploy 'desert' and 'deserves' as though these were synonyms for 'worthiness' and 'worthy,' confusing the concept of worthiness to be happy, which is an ethical notion, with desert of happiness, which, for Kant, can only be a juridical, political one.[343] In the main, this is

[343] See Allison, *Idealism and Freedom*, 114; Ameriks, *Kant and the Historical Turn*, 173, 177; Gerald W. Barnes, 'In Defense of Kant's Doctrine of the Highest Good,' *Philosophical Forum* 2, no. (1971): 448; Claudia Card, *The Atrocity Paradigm: A Theory of Evil* (Oxford: Oxford University Press, 2002), 86; Cicovacki, 'Illusory Fabric': 395; Cohen, 'Critique of Kant's Philosophy of Law', 280; Denis, 'Kant's Conception of Virtue', 524; Engstrom, 'Happiness and the Highest Good', 128, 132; Fendt, *For What May I Hope?*, 51, 69; M. Jamie Ferreira, 'Making Room for Faith: Possibility and Hope,' in *Kant and Kierkegaard on Religion*, ed. D. Z. Phillips and Timothy Tessin (New York: St. Martin's Press, 2000), 79; Luca Fonnessu, 'The Problem of Theodicy,' in *The Cambridge History of Eighteenth-Century Philosophy*, ed. K. Haakonssen (Cambridge: Cambridge University Press, 2005), 770; Robert Gibbs, 'Fear of Forgiveness: Kant and the Paradox of Mercy,' *Philosophy & Theology* 3, no. (1989): 325-7, 332; Guyer, *Kant on Freedom, Law, and Happiness*, 118-21, 123, 392; Hegel, *Phenomenology of Spirit*, § 624 [379], §608 [371]; Hill, 'Happiness and Human Flourishing in Kant's Ethics': 175; J. J. Howard, 'Kant and Moral Imputation: Conscience and the Riddle of the Given,' *American Catholic Philosophical Quarterly* 78, no. 4 (2004): 623; Korsgaard, 'Aristotle and Kant': 501; Leslie A. Mulholland, 'Freedom and Providence in Kant's Account of Religion: The Problem of Expiation,' in *Kant's Philosophy of Religion Reconsidered*, ed. Philip J. Rossi and Michael J. Wreen (Bloomington: Indiana University Press, 1991), 79; Thomas Nenon, 'The Highest Good and the Happiness of Others,' *Jahrbuch für Recht und Ethik* 5, no. (1997): 422; Paton, 'Kant's Idea of the Good': xxi; Pojman and McLeod, eds., 31; B. Quelquejeu, 'Ethical Autonomy and the Question of God,' in *Ethics of Liberation - the Liberation of Ethics*, ed. Dietmar Mieth and Jacques Marie Pohier (Edinburgh, Scotland: T & T Clark, 1984), 18; Philip J. Rossi, 'The Final End of All Things: The Highest Good as the Unity of Nature and Freedom,' in *Kant's Philosophy of Religion Reconsidered* (Bloomington: Indiana University Press, 1991), 151; Arthur Schopenhauer, *The World as Will and Representation*, 2 vols., vol. 1 (New York: Dover Publications, 1958), 524; Smith, 'Worthiness to Be Happy': 169, 184-5; Walter, 'Kant's Moral Theology': 288-9; Wike, *Kant on Happiness in Ethics*, 123-4; Wood, *Kant's Moral Religion*, 126; Allen Wood, *Kant's Rational Theology* (Ithica: Cornell

nothing more than a bad habit and a careless use of terminology. The conflation is not thematized and defended. In an offhand manner, it implies a view, I surmise, that few of these readers would espouse upon reflection. Nevertheless, their usage is problematic. It tends to obscure the actual significance of Kant's 'worthiness to be happy' idiom precisely in its relation to desert.[344]

As we have seen in Guyer's reading of him, admittedly, from time to time, Kant does seem to suggest the possibility of moral merit, to characterize happiness as something that moral agents deserve, and to see happiness as a reward for virtue.[345] In such cases, however, 'deserve' does not imply the categorical obligation of a third party, or the idea of practical necessitation; nor does 'reward' entail anything that is so necessitated. Kant refers to happiness, in these instances, as the product of a *kindness* that endows moral and hence worthy agents (exclusively) with a good that they are in fact *not owed*, but which it is permitted (a kind third party) to bestow on them (alone). Nothing prevents us from construing such 'reward' as a good that is grounded in the giver's kindness, a default kindness to which there is no normative impediment. There is no good reason to see this happiness as something whose bestowal is necessitated by the moral conduct of its object.[346]

University Press, 1978), 22, 24; Allen Wood, 'Rational Theology, Moral Faith, and Religion,' in *The Cambridge Companion to Kant*, ed. Paul Guyer (Cambridge: Cambridge University Press, 1992), 402. This imprecision is reflected in English translations of Kant's work. See, for example, the *Cambridge Edition*'s translation of *Rel* 6: 7 n. (60 n.), which renders '*Glückseligkeit und Würdigkeit*' as 'happiness and desert'; Wike's rendering of '*Würdigkeit*' as 'merit' (followed almost immediately, however, by '*würdig*' as 'worthy') (Wike, *Kant on Happiness in Ethics*, 117). By way of contrast, David Sussman offers an unusually incisive reading of Kant's moral theory that clearly acknowledges the asymmetry that I am discussing here, as well as the significance, for Kant's understanding/construction of the human subject, of the notions of negative desert and punishment as retribution (see David Sussman, 'Shame and Punishment in Kant's Doctrine of Right,' *The Philosophical Quarterly* 58, no. 231 (2008); David G. Sussman, *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics*, ed. Robert Nozick, Studies in Ethics: Outstanding Dissertations (London: Routledge, 2001), esp. 91-98).

[344] One sees this particularly in relation to Kant's doctrine of the *summum bonum*, which is represented (incorrectly) as a scenario in which virtuous agents are endowed with the happiness to which—given their morality—they have a right or entitlement. See, for example, Pamela Sue Anderson and Jordan Bell, *Kant and Theology*, Philosophy and Theology (London: T & T Clark, 2010), 55, 68; Denis, 'Kant's Conception of Virtue', 524; Hill, 'Is a Good Will Overrated?', 43; Packer, 'The Highest Good in Kant's Psychology of Motivation': 110; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 274-5.

[345] See, for example, *MdS* 6: 482 (225); *LMK*$_2$ 28: 766-7 (406-7); *LEV* 27: 550-1 (307); *R* 4277 17: 493 (125); *R* 1187 15: 524-5 (411); *R* 6913 19: 204 (450); *R* 6857 19: 181 (441). See also Guyer's substantive use of the latter (Guyer, *Kant on Freedom, Law, and Happiness*, 122).

[346] See *Rel* 6: 161 (182-3); *MdS* 6: 469 (215); *Vorlesungen-Religionslehre* 28: 1003, 1074, 1081 (349, 408, 414); *R* 6786 19: 160.

There is no problem then, just as such, in referring to happiness as 'reward.' I insist, however, that a distinction be made and maintained here so as not to elide the asymmetry that exists, for Kant, between the deontic *possibility* of moral agents' happiness, on the one hand, and the absolute deontic *necessity* of immoral agents' unhappiness on the other. The normative necessity that is affirmed by Kant's habit of glossing 'morality' as 'worthiness to be happy' pertains to the relationship between *immorality* and *unhappiness* alone, which the gloss represents in an indirect fashion.

As for what the gloss represents *directly*, it signifies that an agent's happiness is deontically possible (permitted her or, where it is regarded as the effect of another's action, permitted that other *qua* aim or end) *if and only if she is moral*. It follows from the claim that an agent's happiness is deontically possible (permitted) if and only if she is moral that if she is immoral then her happiness is deontically impossible (forbidden), and her unhappiness therefore deontically necessary (commanded).[347] Thus unworthiness to be happy is, for Kant, desert of unhappiness, while worthiness to be happy is not desert of anything at all. Rather it entails a certain permissibility. In short, Kant's gloss expresses the idea that, in some sense (the topic of chapters 3 and 4), the happiness of moral agents is permitted, while the happiness of immoral ones is forbidden and their unhappiness commanded. His gloss does not imply that he regards happiness as a mode of remuneration that is practically necessitated *a priori*, just in case one is moral. It does imply, however, that he regards the unhappiness of immoral agents as a mode of (just) retribution (indeed retaliation) that *is* so necessitated. Another way of putting this is to say that, in accordance with the sense of Kant's gloss, immoral agents *deserve to be* unhappy and their unhappiness is a state of affairs that *ought to be* forged extraneously, by a third party.

Kant's use of the idiom expresses what he takes to be the kind and rational being's *benevolent* interest in the happiness of moral agents. But it also expresses the rather less comfortable view that immoral agents ought simply to be unhappy and that their unhappiness is a demand of neither kindness (as it would be, for example, if unhappiness were regarded as something that could lead to insight and reform, a

[347] Formally expressed: for any agent, A, let Am stand for 'A is moral' and Ah stand for 'A is happy.' Let $\boldsymbol{P}$ stand for 'is permissible,' $\boldsymbol{O}$ for 'is obligatory,' $\leftrightarrow$ for 'if and only if,' $\rightarrow$ for 'if…then,' and $\neg$ for 'is not.' Let $\boldsymbol{P} = \neg\boldsymbol{O}\neg$. Then for any agent, A: $\boldsymbol{P}$Ah $\leftrightarrow$ Am, thus $\boldsymbol{P}$Ah $\rightarrow$ Am and $\neg$Am $\rightarrow$ $\neg\boldsymbol{P}$Ah, hence $\neg$Am $\rightarrow$ $\neg\neg\boldsymbol{O}\neg$Ah, and therefore $\neg$Am $\rightarrow$ $\boldsymbol{O}\neg$Ah.

better life, and so, too, happiness), nor of malice and resentment (as it would be if unhappiness were regarded from the point of view of an all too human spirit of vengefulness), but of pure practical reason, as such.

Conclusion

In this chapter I have executed three main tasks. First, I showed that Kant's habitual glossing of 'morality' as 'worthiness to be happy' represents morality and happiness as states of an agent's affairs that are *extrinsically* related. Next, I further specified this claim by arguing that Kant's gloss represents two states of an agent's affairs that are not only extrinsically, but also *necessarily* and *normatively* related. Third, I argued that Kant's notion of 'worthiness to be happy' points to an extrinsic, normative, necessary relation that holds, not between morality and happiness, but between unhappiness and immorality. I argued that Kant's gloss pertains to the normative necessity of the immoral agent's desert of unhappiness. I showed that Kant's use of the 'worthiness to be happy' idiom expresses a concern, then, whose primary focus is not the happiness or unhappiness of *moral* agents, but rather the unhappiness and happiness of *immoral* ones. While Kant takes morality and happiness to be normatively related, he does not take it to be the case that morality *necessitates* happiness.

Kant's idiomatic use of the notion of worthiness to be happy shows that, from the very outset of his thinking about the grounds of moral action, he takes the tension between morality and happiness to be resolvable in the actualization of a correspondence which—to indulge an anachronism—is always already enjoined by pure practical reason, as such, hence normative for all rational beings *a priori* and not merely hoped for by rational finite ones.[348]

In the next chapter I will take an even closer look at the 'ought' that is implicit in Kant's gloss, in the 'enjoinment' that it promotes. I argue that Kant understands the unhappiness of immoral agents by way of an analogy to the punishment of criminals. I show that this correlation is implicitly eschatological. The sub-set of immoral

[348] Very few commentators have taken up this particular point. But see, however, Smith, 'Worthiness to Be Happy': 173. This is not to say that Kant's readers do not notice that he sometimes represents the relationship between morality and happiness in terms of *desert* (i.e., thematically, not merely by way of the conflation of 'desert' with 'worthiness'). See, for example, Engstrom, 'Happiness and the Highest Good', 128 and Korsgaard, 'Aristotle and Kant': 499. The point is that they do not see the central importance, for Kant, of this mode of representation.

agents that are guilty of murder, in particular, ought to be put to death. This is Kant's law of punishment. I argue that, for Kant, there is an equivalent law of unhappiness. Ultimately, I will ask who is subject to it (i.e., bound to do what it commands) and how it is to be put into practice. This will involve a first step in the direction of showing that his gloss has a deeply practical significance for Kant.

CHAPTER THREE

The Law of Punishment and the 'Law' of Unhappiness

> The law of punishment is a categorical imperative.[349]

Introduction

In chapter 2 I showed that Kant's gloss expresses the idea that immoral agents are perfectly capable of happiness, but that they *do not deserve to be* happy at all. Indeed, I showed that, to the extent that the gloss expresses Kant's thinking about the relationship between immorality and unhappiness, he thinks that immoral agents *deserve to be* unhappy. I showed, moreover, that this implies that, for Kant, immoral agents *ought to be* unhappy. In this chapter, I explicate this '*ought*' more fully by showing how it is embodied in what I will refer to as Kant's 'law of unhappiness' and by exposing the latter's retributivist connections. In this way, more pointedly, I will advert to a particular mode of retributivism that is in the air whenever we encounter Kant's gloss.

In service of these ends, this chapter executes three main tasks. First, I explore Kant's thinking about punishment and affirm that when it comes to the latter's specification and justification Kant is a kind of retributivist. Second, I show that, especially in his late treatment of the topic in *The Metaphysics of Morals*, Kant's conceptions

[349] *MdS* 6: 331 (105).

of the legal and the ethical encounter one another in the practice and justification of capital punishment. Indeed, I characterize Kant's 'scaffold' as the liminal *topos* in which his thinking about law and politics extends deep into his thinking about ethics and eschatology. I argue that the unconditional, immediate necessity that Kant ascribes to capital punishment in cases of murder is key to understanding the retributivist tendencies of both his thinking about politically situated punishment and his eschatologically inflected thinking about unhappiness more generally. Third, I argue that the 'ought' that arises from Kant's implicit conviction that immoral agents deserve to be unhappy (the 'ought' implicit in his gloss) may be expressed in the form of a normative law: 'It is practically necessary that all immoral agents be unhappy.' I refer to this as Kant's 'law of unhappiness' and argue that it is the ethical and ultimately eschatological expression of his political 'law of punishment.'

The law of punishment

The law of punishment is the imperative that says, concerning 'the one that did it' (where 'it' has been judged a punishable deed): 'Punish him!' Kant's law of punishment has three main properties. The first is that it is a *categorical imperative*. The second is its applicability to *all* and *only* criminals. The third is its stipulation that criminals be punished *in exact proportion* to their crimes, in accordance, that is, with the so-called *ius talionis*. To say that the 'law of punishment' is a categorical imperative is to say that criminals ought, *unconditionally*, to be punished. It is to say that to punish is not an option, but a duty[350] and—having said that it is a duty (in Kant's sense)—it is to say that the subject that it binds (the sovereign) has a reason for doing it independently of any empirical incentives that can be adduced in favour of it. Together, the second and third properties constitute the law's core retributive character. In other words, these properties bear on the justification of particular punishments. The first property, the law's 'categorical' nature is something else again. It may be regarded as an aspect of the law's retributivism as well to the extent that, to claim that it is a categorical imperative, is to claim not only that this or that actual instance of punishment is justified, but that punishment *as such*—the having of the practice at

[350] See Garcia, 'Two Concepts of Desert': 222; D. D. Raphael, *Moral Judgment* (London: George Allen and Unwin, 1955), 71.

all—is *demanded* somehow, *a priori*. But it tokens more than this, too, for Kant—something rather more obscure.

I begin with this fundamental text:

> The law of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaemonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises, in accordance with the pharisaical saying, "It is better for *one* man to die than for an entire people to perish."' For if justice goes, there is no longer any value in human beings' living on the earth.[351]

Note what is at stake here for Kant. '[I]f justice goes,' he writes, 'there is no longer any value in human beings' living on the earth.' And what is 'justice' (*Gerechtichkeit*) here? It is to punish criminals. The 'value' (*Werth*) that Kant ascribes to 'human beings' living on the earth' is grounded, in the strongest terms, in our conformity to a law that categorically *forbids* amnesty to criminals or any action aimed at reducing the 'amount' (*Grade*) of harm due to them. Later, as we shall see, this is so above all in regard to the crime of murder. In passages filled with stirring language and imagery, at a loss for argument, as it were, Kant twice refers to the consequences of amnesty in such cases as a terrible 'blood-guilt'[352] that will cling to and contaminate the whole community where this is allowed to happen.

We will come to this imagery in due time. In this section, however, I discuss the practice that puts Kant's categorical law of punishment into effect, which enacts the 'justice' without which the earthly existence of human beings would be, for Kant, a worthless thing. First, I outline Kant's thinking about the constitution of punishments. Next I discuss Kant's retributivism, his thinking about what justifies punishments. Then I examine the significance of his claim that the law of punishment is a categorical imperative.

Punishment in connection with 'a transgression of public law'

Kant's thinking about punishment constitutes an enormous area of study and debate.[353] The various tensions that are present in this thinking have led several commentators to question whether, in Murphy's phrase, 'Kant develops anything that

[351] *MdS* 6: 331-2 (105).
[352] *MdS* 6: 333, 490 (106, 232)
[353] For a recent, detailed treatment of the historical context of Kant's thinking about punishment both in advance of and after him see Martin Reulecke, *Gleichheit Und Strafrecht Im Deutschen Naturrecht Des 18. Und 19. Jahrhunderts*, Grundlagen Der Rechtswissenschaft, vol. 9 (Tübingen: Verlag Mohr Siebeck, 2008). See also F. Zanuso, 'The Current Interest in Kant in the North American Debate on Criminal Punishment,' *History of European Ideas* 30, no. 3 (2004).

deserves to be called a *theory* of punishment at all.'[354] I neither deny nor affirm that Kant has a consolidated, systematic theory of punishment. My main claim in this and the next section is that Kant is committed—immediately, in advance of his argumentation—to a particular practice (the practice of putting murderers to death) and that the best way of describing this commitment is in the language of a retributivist theory of punishment. Before coming to *that* however, it will be necessary to say something about Kant's conception of the practice of punishment more generally.

Restricted to its core subject matter, which pertains to the justification of particular punishments, a retributivist theory of punishment tells us how the concepts of crime and punishment are related with an aim to articulating a standard that actual punishments have to meet in order to be justified. Restricted to this task, such a theory does not tell us what crime and punishment *are*—only how they are, or ought to be, *related*. Saying what crime and punishment are is an antecedent, independent task, whose results retributivism takes for granted.

Fundamentally, before being anything else then, punishment is for Kant a form of 'physical harm'[355] or 'pain,'[356] which '(rightly) offends the accused's feeling of honor, since it involves coercion that is unilateral only.'[357] Of course, as far as the brute infliction of pain goes, there is a sense in which anyone may be punished for anything at all, *ad hoc*, but not 'rightly' so. In Kant's view, of course, punishments must be preceded by crimes and the agents whose punishments are in question must naturally be the very ones that committed them. Each concept (crime, punishment) is implicated in the definition of the other. A particular crime is 'of itself punishable,' which means, as Kant says, that it (precisely 'of itself') 'forfeits happiness (at least in part).'[358] Punishments have other core features too: being unwanted,[359] of course,

---

[354] Jeffrie G. Murphy, 'Does Kant Have a Theory of Punishment,' *Columbia Law Review* 87, no. 3 (1987): 509. See also David Cooper, 'Hegel's Theory of Punishment,' in *Hegel's Political Philosophy*, ed. Z. A. Pelezynski (Cambridge: Cambridge University Press, 1971), 160; J. A. Corlett, *Responsibility and Punishment* (Dordrecht: Kluwer Academic Publishers, 2001), chapter 4; J. A. Corlett, 'Making More Sense of Retributivism: Desert as Responsibility and Proportionality,' *Philosophy* 78, no. 2 (2003): 281; Andrew Von Hirsch, 'Proportionality in the Philosophy of Punishment,' *Crime and Justice* 16, no. (1992).

[355] See *KpV* 5: 37 (34).

[356] *MdS* 6: 331 (104).

[357] *MdS* 6: 363 n. (130 n.).

[358] See *KpV* 5: 37 (35).

[359] For a clear analysis and critique of the claim that criminals will their punishment see S. Fleischacker, 'Kant's Theory of Punishment,' *Kant-Studien* 79, no. 4 (1988): 438-40. But cf. *Bem* 20: 68.

and being identifiable by way of penal law in determinate connection with a member of a class of actions that fall afoul of one or more primary imperatives, that is, identifiable as an instance of action that contravenes one of the laws of the land.

A number of questions press for an answer (here, from Kant): (1) Who is actually punishable? (Who may be punished?); (2) How should those who are authorized to punish determine the degree and kind of harm that they may inflict on those who are punishable? (What ought punishments to consist in?); (3) Why punish *at all*?; (4) Are parties that are authorized to punish also *bound* to punish? (If an agent is found to be punishable then is it necessary that she actually *be* punished? And must the punishing authority punish, on every occasion, in strict accordance with a single calculus for determining degrees and kinds of harm, or may this authority make exceptions?); (5) Who has the authority to punish? (If anyone is *bound* to punish, who would that be?).[360]

I discuss Kant's answer to (5) in chapter 4, while I lay out his answers to (1)-(4) in this chapter. This sub-section pertains to (1), (2), and (3). I discuss Kant's answer to (4) below.

Who is punishable?

In a sense, Kant's answer to this question is straightforward: a man gets to be punishable, he says, 'because he has willed a punishable action.'[361] In fact, for the purposes of this thesis, this answer very nearly suffices. Criminals, people that break the laws of the land, are punishable. We need not tarry too long with Kant's answers to the questions, 'What is crime?' and 'What (or who) is a criminal?' I will give only a very general outline.

On Kant's description, crime entails—or just is—forfeiture of one's 'civil personality' (*bürgerliche Persönlichkeit*), the loss of one's basic 'dignity' (*Würde*) as a citizen. This entails that a citizen can pass, by way of a crime, from his default state of (external) freedom, the state that he is in as long as 'he has not yet committed a crime,' to a state in which he 'forfeits his personality.' This can entail, among other things, another's 'right of ownership with regard to [the criminal],' where 'some-

[360] For a similar series of distinctions see Thomas E. Hill, Jr., 'Kant on Punishment: A Coherent Mix of Deterrence and Retribution?,' *Jahrbuch für Recht und Ethik* 5, no. (1997): 292, 294, 299, 301.
[361] *MdS* 6: 335 (108).

one…become[s] a slave through his crime.'[362] 'A transgression of public law,' Kant says, 'makes someone who commits it unfit to be a citizen.'[363] In other words, 'by his own crime,' the criminal *loses* 'the dignity of a citizen,' which is the baseline of 'dignity' that any 'human being in a state' possesses. Given this loss of dignity, 'though he is kept alive, he is made a mere tool of another's choice (either of the state or of another citizen).'[364]

In the *Metaphysics of Morals* Kant defines punishment (*poena*) as '[t]he *rightful* effect of what is culpable.'[365] In other words, punishment is not subjective and arbitrary. It is an effect, which is to say that it is a necessary consequence, given an antecedent of the right kind. But it is also 'rightful.' Punishment does not arise from innocence. It derives from culpability. Thus, no one is punishable unless, by a crime that they have committed, they have passed into a state in which they have forfeited their civil personality and lost their dignity as a citizen. But given that a human being's punishability is granted, what may or ought to be done to them?

Kant's retributivism and the '*ius talionis*'

At this point, we need a more specific definition of punishment that adverts to *what* ought to be done to criminals and *to what degree*. Kant's answer, here, makes use of the so-called *ius* (or *lex*) *talionis*—the law of (proportionate) retaliation.[366] It is important to distinguish between Kant's use of this standard and the other parts of his thinking that suggest that he is a retributivist. In fact, I suggest, in spite of Kant's sometimes identifying 'the law of retribution' with the '*ius talionis*' as such,[367] his retributivism is better characterized without fundamental reference to this notion. Just as the categorical nature of the law of punishment (its first property) is distinct from its property of being applicable to *all* and *only* criminals (its second property), the law of punishment's stipulation that criminals be punished *in exact proportion* to

[362] *MdS* 6: 283 (66).
[363] *MdS* 6: 331 (105).
[364] *MdS* 6: 329-30 (104). Of course, Kant distinguishes between the human being's inalienable 'innate personality,' which shields her from being punished as a means to an end, even one that is ostensibly good for the criminal herself, and her 'civil personality,' which she 'can be condemned to lose' (*MdS* 6: 331 [105]).
[365] *MdS* 6: 227 (19).
[366] For further analysis of the *ius talionis* see Hill, 'Kant on Punishment': 302.
[367] *MdS* 6: 332 (105).

their crimes is a distinct third property in relation to these others. And, I suggest, this is not its decisively retributivist element.

The question of how the practice of punishment is justified, in general—the primary problem to which retributivism offers a solution—is distinct from the question of how particular punishments are determined and warranted. The *ius talionis* pertains to the latter problem. As Kant puts it in *The Metaphysics of Morals:*

> only the *law of retribution* (*ius talionis*)—it being understood, of course, that this is applied by a court (not by your private judgment)—can specify definitely the quality and the quantity of punishment; all other principles are fluctuating and unsuited for a sentence of pure and strict justice because extraneous considerations are mixed into them.[368]

Just as such, in spite of his calling it 'the law of retribution,' Kant's use of the *ius talionis* here does not necessarily entail that he is a retributivist about punishment, in general. Something more is needed for that. As Hill observes, 'Kant does endorse standards of punishment *commonly associated with retributivism*.' But, he argues, 'the retributive elements in Kant's theory [here, these standards] are more firmly rooted in considerations of comparative justice and honesty in public expressions of moral judgment.'[369] This is certainly plausible and, save with respect to Kant's treatment of capital punishment, I will not dispute claims like this one.

The law of punishment stipulates not only *that* criminals be punished, then, it also specifies *how*. Kant proposes that we specify the judgment that someone ought to be punished by looking to the crime—and nowhere else. Kant argues that if a punishment were not precisely, reciprocally correlated with the crime of the punished person ('if not in terms of [the penal law's] letter at least in terms of its spirit,' he allows), if the criminal is unable to regard 'what is done to him in accordance with the penal law' as *equivalent* to 'what he has perpetrated on others,' then he would be permitted to 'complain that a wrong is done him.' This is what would happen if, in a move that 'would be literally contrary to the concept of *punitive justice*,' one were, under a general authorization to mete out punishments, '[t]o inflict *whatever* punishments *one chooses* [*willkürlich Strafen*].'[370]

---

[368] *MdS* 6: 332 (105-6). Hill offers helpful commentary on this passage in Hill, 'Kant on Punishment': 309-10.
[369] Hill, 'Punishment, Conscience, and Moral Worth', 341 (my emphasis). See also ibid., 342.
[370] *MdS* 6: 363 (130). See Hill's discussion in Hill, 'Wrongdoing, Desert, and Punishment', 332, 338.

If 'what is done to him in accordance with the penal law' is equivalent to 'what he has perpetrated on others,' then, says Kant, the criminal has no warrant for complaining. But of course, his eligibility for being punished at all (in whatever manner) is predicated upon his being, in fact, 'the one that did it' (where it is granted that 'it' is 'a transgression of public law'). Barring this condition, of course, the punished person has an even deeper warrant for complaint than in the case where he deserves to be punished, but is punished in the wrong manner or degree. Thus, as Hill aptly observes, among the 'rules commonly associated with "retributivism"' [371] that Kant includes in his account of punishment is the requirement 'that the agent had the freedom necessary to conform to the law.'[372] More precisely, for Kant, her nonconformity has to have been a matter of freedom as well—but freedom in a sense that excludes the least reference to the involvement of other agents *at any point* in the genesis of the punishable deed. The agent is only punishable if she is 'the one that did it,' the one that broke the law, *wholly* and *simply*. It is Kant's inclusion of this condition—which is included long before it is fully thematized—in tandem with his assertion that the law of punishment is a categorical imperative that makes Kant a retributivist. However, I suggest that the condition pertaining to the agent's radical freedom (which turns out to have been, all along, a radical freedom to choose between good and evil, as such) is only really clearly pertinent in the case of capital punishment and, too, in the realm into which the latter practice penetrates, the ethical and the eschatological. Hence my claims about Kant's retributivism ought not to be read as generalizations about Kant's political and legal philosophy as a whole. It is capital punishment and, too, the irrevocable, eschatological making-unhappy of immoral agents that is modeled on it, whose rational justification are at stake here.

Retributivism in general, as a doctrine that pertains to the justification of the practice of punishment, is far from monolithic. '[W]hat has been called by this name comes in many different forms and degrees,' as Hill aptly observes.[373] For the pur-

[371] Hill, 'Punishment, Conscience, and Moral Worth', 342.
[372] Ibid., 343.
[373] Hill, 'Kant on Punishment': 291. For an extremely detailed discussion of the main retributive theory-types and a strong critique of each see R. L. Christopher, 'Deterring Retributivism: The Injustice of "Just" Punishment,' *Northwestern University Law Review* 96, no. 3 (2002). See also John Cottingham, 'Varieties of Retribution,' *The Philosophical Quarterly* 29, no. (1979).

poses of this thesis, there is no need to answer the question of what retributivism, in general, 'is.' The sense and purpose of my claims concerning *Kant's* retributivism are relative to my discussion of his habit of glossing 'morality' as 'worthiness to be happy.' And my claim is simply that this gloss represents the connection between immorality and unhappiness in a manner which—when *thematized*—turns out to be best described as a variety of 'retributivism.' I will further elucidate *this* sense of the term below.

Until relatively recently, the unqualified claim that Kant is a retributivist would have carried the weight of majority consensus. This is no longer so, or at least not in a straightforward manner. Hill speaks of 'the *formerly* accepted view of Kant as a prime example of a retributivist'[374] and he is correct to do so. While some commentators proceed (or have proceeded) as though Kant's retributivism is a relatively clear-cut matter,[375] using epithets such as 'strict,'[376] 'pure,'[377] 'classical,'[378] 'bold,'[379] 'deep,'[380] and 'thoroughgoing'[381] to describe it, scholarly consensus is currently rather fractured on this score. It is now common, for example, to find affirmations of a retributivist tendency in *some* of Kant's thinking about punishment, which makes his work relevant for discussions of contemporary versions of the doctrine, but tempered by the denial that Kant is a retributivist in every relevant respect.[382] Such commentators concede no more than a 'retributive *current* in [Kant's] thought,'[383] for example, or argue that 'Kant's mature theory of legal punishment,' at least, is 'not deeply retributivist.'[384] Yet others define 'thoroughgoing retributivism' in a manner that includes features that Kant does not espouse and argue that Kant's thinking about

[374] Hill, 'Punishment, Conscience, and Moral Worth', 340 (my emphasis).
[375] See, for example, T. Brooks, 'Kantian Punishment and Retributivism: A Reply to Clark,' *Ratio* 18, no. 2 (2005): 238, 241; Cohen, 'Critique of Kant's Philosophy of Law', 285-6; Fleischacker, 'Kant's Theory of Punishment'; Gibbs, 'Fear of Forgiveness': 328; H. L. A. Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1982), 231-2; Wood, *Kantian Ethics*, 206.
[376] Douglas Lind, 'Kant on Capital Punishment,' *Journal of Philosophical Research* 19, no. (1994).
[377] David Dolinko, 'Some Thoughts About Retributivism,' *Ethics* 101, no. 3 (1991); Edmund Pincoffs, *The Rationale of Legal Punishment* (New York: Humanities Press, 1966), 6.
[378] George Sher, *In Praise of Blame* (Oxford: Oxford University Press, 2006), 68.
[379] Dolinko, 'Thoughts About Retributivism': 542.
[380] Hill, 'Wrongdoing, Desert, and Punishment', 314.
[381] S. I. Benn, 'Punishment,' in *The Encyclopedia of Philosophy*, ed. Paul Edwards (London: Macmillan, 1967), 30 (cited in Don E. Scheid, 'Kant's Retributivism,' *Ethics* 93, no. 2 (1983): 264); similarly Jeffrie G. Murphy, *Kant: The Philosophy of Right* (London: Macmillan, 1970), 142 (cited in Scheid, 'Kant's Retributivism': 265).
[382] Von Hirsch, 'Proportionality in the Philosophy of Punishment': 59-61.
[383] Hill, 'Kant on Punishment': 292 (my emphasis).
[384] Hill, 'Wrongdoing, Desert, and Punishment', 328.

punishment has reference, not to retribution (under this other definition), but to a strictly political notion of justice, or 'right,' or 'law and order.'[385]

In the main, however, commentators continue to find a retributivist *dimension*, at least, in Kant's work. The commentary evinces a variety of approaches to this dimension and its relation to the rest of Kant's thinking, but it is possible to distinguish two main strategies. The first is to argue that, while Kant does explicitly espouse retributivism, the latter position neither follows from, nor coheres with, the rest of his thinking about morality. The basic claim, here, is that he explicitly avows a position that, implicitly, he ought not to hold.[386] The second strategy is to argue that Kant's thinking about punishment more or less successfully combines retributivist and consequentialist elements. The basic idea, here, is that while Kant *is* a retributivist in some sense, his retributivism is not inherently inimical to considerations bearing, especially, on the deterrent effect of punishment and that, indeed, such considerations are present in his thinking. In short, on this view, Kant's retributivism is not nearly so 'strict' or 'thoroughgoing' as it has often appeared.[387]

Now, I said above that Kant's habit of glossing 'morality' as 'worthiness to be happy' represents the connection between immorality and unhappiness in a manner which—when *thematized*—turns out to be best described as a kind of retributivism. The sense of 'retributivism' that I have in mind is a position, more or less, that is

---

[385] See, for example, Sarah Holtman, 'Toward Social Reform: Kant's Penal Theory Reinterpreted,' *Utilitas* 9, no. 1 (1997): 4, 11. See also Thomas E. Hill, Jr., *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell University Press, 1992), 176-95; Hill, 'Is a Good Will Overrated?', 56; Hill, 'Wrongdoing, Desert, and Punishment', 310-11, 316, 329; Wood, *Kantian Ethics*, 214, 216.

[386] For particularly forceful versions of this argument see Wood, *Kantian Ethics*, 206, 219 (cf. *MdS* 6: 387-8 [151-2]); Holtman, 'Toward Social Reform': 18). See also Jean-Christophe Merle, 'A Kantian Critique of Kant's Theory of Punishment,' *Law and Philosophy* 19, no. 3 (2000); Jeffrie G. Murphy, 'Kant's Theory of Criminal Punishment,' in *Proceedings of the Third International Kant Congress*, ed. Lewis White Beck (Dordrecht: Reidel, 1972), 435.

[387] Two very influential and widely discussed arguments to this effect are B. Sharon Byrd, 'Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution,' *Law and Philosophy* 8, no. 2 (1989): 152-3; Scheid, 'Kant's Retributivism'. See Hill's appreciative, sometimes critical, but always clear elucidation of their views Hill, 'Kant on Punishment': 291, 305-9; Hill, 'Punishment, Conscience, and Moral Worth', 344-5; Hill, 'Wrongdoing, Desert, and Punishment', 336. For related 'mixed' approaches to Kant's retributivism see Christopher, 'Deterring Retributivism': 859; J. A. Corlett, 'Making Sense of Retributivism,' *Philosophy* 76, no. 1 (2001): 87; M. Tunick, 'Is Kant a Retributivist?,' *History of Political Thought* 17, no. 1 (1996): 73. For further discussion of these 'mixed' approaches see also M. Clark, 'A Non-Retributive Kantian Approach to Punishment,' *Ratio (New Series)* 17, no. 1 (2004): 13-14; Holtman, 'Toward Social Reform': 4; Merle, 'A Kantian Critique of Kant's Theory of Punishment': 316; Sussman, 'Shame and Punishment': 305; Wood, *Kantian Ethics*, 212, 216.

cognate with the one delineated by proponents of the claim that Kant is a retributivist. I want to qualify this, however, very carefully. I do not claim that Kant just *is* a retributivist; but to the extent that I affirm that he is (again, in respect, minimally, of capital punishment), then here are some useful descriptions of what I have in mind.

Murphy, for example, observes that Kant takes 'guilt,' just as such, to be not only a necessary, but also a sufficient condition 'for the legitimate infliction of punishment.'[388] Fleischacker notes 'Kant's insistence that punishment [of malefactors] is a worthy act in itself.'[389] Scheid delineates a strategy for the justification of the practice of punishment that excludes considerations of utility (here 'concern for crime control') and, at the same time, takes 'the view that whether a person may be punished and, if so, to what extent are questions to be decided solely by reference to [his or her] past legal offence.'[390] Benn defines Kant's 'thoroughgoing' retributivism as the view that 'the punishment of crime is right in itself, that it is fitting that the guilty should suffer, and that justice, or the moral order, requires the institution of punishment.' Benn makes the important point, too, that '[t]his…is not to justify punishment but, rather, to deny that it needs any justification' and to affirm that '[i]ts intrinsic value is appreciated immediately and intuitively.[391] Dolinko defines Kant's retributivism as the view that 'giving lawbreakers their just deserts is the only point or purpose of punishment.'[392] Moore avers that '[r]etributivism is the view that punishment is justified by the moral culpability of those who receive it,' so that '[a] retributivist punishes because, and only because, the offender deserves it.'[393] And Moore avers that, here, '[m]oral culpability ("desert") is…both a sufficient as well as a necessary condition of liability to punitive sanctions,' adding that this justification pertains to more than society's 'right to punish culpable offenders'; it also grounds the claim that 'the moral culpability of an offender also gives society the *duty* to punish.' This means, Moore concludes, that '[r]etributivism…is truly a theory of justice

[388] Murphy, 'Kant's Theory of Criminal Punishment', 434.
[389] Fleischacker, 'Kant's Theory of Punishment': 438.
[390] Scheid, 'Kant's Retributivism': 262. See also ibid.: 263.
[391] Benn, 'Punishment', 30. Cited in Scheid, 'Kant's Retributivism': 264.
[392] Dolinko, 'Thoughts About Retributivism': 542.
[393] Michael S. Moore, 'The Moral Worth of Retribution,' in *Punishment and Rehabilitation*, ed. Jeffrie G. Murphy (Belmont, CA: Wadsworth Publishing Co., 1973), 94 (author's emphasis removed).

such that, if it is true, we have an obligation to set up institutions so that retribution is achieved.'[394]

In their various ways, each of the foregoing expresses the retributivism that I also find in Kant. This is present in the second *Critique*, for example, when Kant avers that '[i]n every punishment as such there must first be justice and [that] this constitutes the essence of the concept.' Punishment 'must…be justified as punishment, i.e., as mere harm in itself,' he says, 'so that even the punished person, if it stopped there and he could see no glimpse of kindness behind the harshness, would yet have to admit that justice had been done and that his reward perfectly fitted his behavior.'[395]

But why would everyone, up to and including 'the punished person' have to concede this? Kant claims, simply but crucially, that 'there is *in the idea of our practical reason* something further that accompanies the transgression of a moral law, namely its deserving punishment.'[396] The one concept, the concept of 'its deserving punishment [*ihre Strafwürdigkeit*],' simply '*accompanies*' (*begleitet*) the other, the concept of 'transgression of a moral law [*Übertretung eines sittlichen Gesetzes*].' Indeed, here at least, Kant seems to suggest that the first concept (the concept of punishment), which is a notion primarily, before it is put into practice, accompanies the very 'transgression' itself, the 'it' that 'the one that did it' did—not as empirical fact, but strictly in the estimation 'of our practical reason.'

Later, in *The Metaphysics of Morals*, Kant expresses the same view in different terms.

> *Punishment by a court* (*poena forensis*)…can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*…. He must previously have been found *punishable* [*strafbar*] before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens.[397]

The only good reason for punishing a criminal is that 'he has committed a crime,' which is to say that 'he must previously have been found punishable.' Punishment can only be 'justified as punishment, i.e., as mere harm in itself,' as we read above.

---

[394] Ibid., 96.
[395] *KpV* 5: 37 (34).
[396] *KpV* 5: 37 (34) (emphasis modified).
[397] *MdS* 6: 331 (105). With respect to the warrant that Kant mentions here and in connection with his assertion that the law of punishment is a categorical imperative see Hill, 'Wrongdoing, Desert, and Punishment', 331.

And it *is*, from the outset, Kant argues, because, 'in the idea of our practical reason,' the judgment that *this* transgressor deserves to be punished is already present, preformed in accordance with a 'pure' canon for judging of such matters, which is simply, again, the concept of 'transgression of a moral law.'

In this way, however, Kant not only closes off the possibility of morally repugnant 'uses' of punishment or of the punished person; he also closes up the space between the transgression and the judgment of punishability so precipitously that we cannot remain with the transgression long enough to ask whether, in spite of our concession that the transgressor in question is 'the one that did it,' their might not be some sense in which 'it' is not—cannot be—the sort of thing whose normative consequence can be *conceived* of (let alone justified) altogether without reference to the natural and social order to which the transgressor belongs.

With respect to Kant's caveat that '[h]e must previously have been found *punishable*,' Hill observes that '[s]ome previous translations misleadingly translated "*strafbar*" as "deserving of punishment" rather than "punishable," thereby encouraging the thought that intrinsic moral desert might be the justification for inflicting suffering.'[398] Hill's basic impulse, here, is correct. The transgressor is 'punishable' (*strafbar*) if and only if she has done something that counts as 'a transgression of public law' and has thereby rendered herself 'unfit to be a citizen.'[399] This definition has no reference to 'intrinsic moral desert' at all—as long as we remain well within the boundary that separates the political from the ethical. However, the closing of the gap between the concepts of 'deserving punishment' and 'transgression of a moral law'—indeed the claim that there *is* no gap here (just a conceptual distinction)—prevents our asking or noticing whether there might not be a special case *at that boundary* in which the justification of punishment would require that intrinsic moral desert (in a special sense, modeled on the extrinsic desert entailed by crime) be demonstrably present, in addition to strict punishability, in this strict sense that has reference, not to the heart of the matter, or rather to the *character* of the man, but only to the character of the deed.

It is important to note that Kant does not claim—nor claim to be able to show in general—that this or that instance in which someone is punished is actually just. As

[398] Ibid.
[399] *MdS* 6: 331 (105).

Wood points out, it might be that the belief that the party in question is guilty is a mistaken belief in the sense that the person in question is not 'the one that did it' at all. Or it might be, even given that the person is guilty, that their punishment is either excessive or deficient in some way.[400] Retributivism does not claim to be able to avoid outcomes like these.[401]

What is at stake for Kant, here, is whether human beings are the kind of agent, in the first place, that could ever be regarded as culpable, hence punishable, in the sense that is entailed, specifically, if capital punishment, on the one hand, and the irrevocable, eschatological making-unhappy of immoral agents, on the other hand, are rationally justifiable—*given that* Kant sees (correctly) that these can only be so justified in retributivist terms (i.e., without reference to anything empirical, hence parochial). Kant takes it that the *class* of justified instances of capital punishment is not *empty in principle*, even if it might turn out to be empty in fact. He and virtually everyone in his cultural milieu takes this for granted.[402] And, given the coinciding instantiation of the law of punishment and of what I refer to as the law of unhappiness in the practice of capital punishment (which 'coinciding' I discuss in the third section of this chapter), his taking this for granted is an aspect of what is disclosed in his habit of glossing 'morality' as 'worthiness to be happy.'

The law of punishment qua *'categorical'*

So far we have seen that Kant takes it that (justified) punishments belong to a class of practices whose instantiation is always already enjoined by moral-practical rationality, to the extent that we are actually faced with instances that embody the concept of 'transgression of a moral law.' Irrespective of whether any *particular* instance of punishment has ever been fully justifiable and thus a member of that class—by being the punishment of the right person in exactly the right manner and to just the right degree—our having the practice of punishment, just as such, is justified by its objectively necessary correlation with the concept of 'transgression of public law,' whose instances we see all around us.

[400] Wood, *Kantian Ethics*, 208.
[401] See Murphy, 'Kant's Theory of Criminal Punishment', 438; Sher, *Praise of Blame*, 133.
[402] Cf. Ameriks, *Kant and the Historical Turn*, 295.

At the beginning of this chapter, I identified three main properties, which I ascribed to Kant's law of punishment. We have now examined the second and third of these: this law's property of being applicable to all and only criminals and its subsumption of the stipulation that criminals be punished in accordance with the *ius talionis*. I will now turn to this law's most basic property, the one that Kant ascribes to it directly and explicitly in his canonical formulation of it: its property of being a *categorical imperative*. This also brings us to the point of Kant's answer to the fourth of the five 'pressing questions' that I laid out earlier: Are parties that are authorized to punish also *bound* to do so?

Kant's assertion that the law of punishment is a categorical imperative is his answer to this question. It goes beyond the claim that all and only criminals be punished, which is still antecedent to the presupposition that punishments *must* take place and affirms that they must indeed. The former stipulation entails only that *if* punishments are going to take place, *then* there can be no exceptions. But this still leaves open the possibility of a perpetual state of *universal amnesty*—a politically untenable response to transgression, of course, but not so obviously an ethically/eschatologically untenable one. By asserting that the law of punishment is a categorical imperative, however, Kant blocks access to this fantastical possibility.

If an agent is found to be punishable then it is practically necessary that she actually be punished. The party that is authorized to punish such an agent is also bound to punish her. But note that this practical necessitation is not a matter of public law. To *omit* to punish—and so to transgress the law of punishment—is not itself a crime and hence punishable in turn. The authorized party's failure to punish is immoral,[403] but it is not illegal.

'By categorical imperatives,' Kant explains in *The Metaphysics of Morals*, 'certain actions are *permitted* or *forbidden*, that is, morally possible or impossible, while some of them or their opposites are morally necessary, that is, obligatory.'[404] And in the *Groundwork*, in the course of efforts aimed at showing that there are such things at all, Kant defines a categorical imperative as 'a practical law, which commands *ab-*

[403] Cohen makes a similar observation, but on different grounds (Cohen, 'Critique of Kant's Philosophy of Law', 287). Cf. Garcia's argument (contra Kant) for the moral permissibility of mercy (Garcia, 'Two Concepts of Desert': esp. 225, 230).
[404] *MdS* 6: 221 (14).

solutely of itself and without any incentives'; a law the 'observance' of which 'is *duty*.'[405]

As I said earlier, to say that the 'law of punishment' is a categorical imperative is to say that criminals ought, *unconditionally*, to be punished, that it is obligatory to punish them.[406] Wood suggests that the 'full Kantian retributivist position [is] that the ruler is *required by right* to visit the criminal with harm equal to the wrong he has inflicted and does an *injustice* if he fails to do so.'[407] Admittedly, Kant says just this. But his treatment of the topic is rather ambiguous. The 'sovereign' has 'the *right to grant clemency*,' but this is 'the slipperiest one [among his rights] for him to exercise.' Kant suggests that action that expresses this 'right' is, at the same time, 'injustice in the highest degree' and that, when it comes to 'crimes of *subjects* against one another' (in contrast to 'case[s] of a wrong done *to himself*'), to exercise this 'right,' to fail to punish, 'is the greatest wrong against his subjects.'[408] But Kant also asserts, somewhat earlier, precisely in connection with the sovereign's 'right to punish,' that the sovereign (or 'head of a state') cannot himself be punished' and that 'one can only withdraw from his dominion.'[409] On Kant's account of positive law, if an action cannot be coerced (by the threat or enactment of punishment), then it is not legally prohibited, hence not a *crime* to omit to do it. Therefore the sovereign's failure to abide by the law of punishment counts as an injustice in a 'wide,' ethical sense, but not otherwise.

If the law of punishment is a categorical imperative for the sovereign, then Wood's characterization of the nature of the 'ought' in question is at least inadequate. If the law of punishment is a categorical imperative, then any party that is bound by it is bound *ethically* to act in accordance with it—unless the failure to punish can be construed as punishable. But the only party authorized (and commanded) to act in accordance with it is unpunishable on Kant's account. If the law of punishment is a categorical imperative, then what it commands is not an action, but the having of a maxim. It is a law, conformity with which will take the latter form, first, and the form of outward action in a secondary sense only. To punish will be a 'duty of

[405] *Gr* 4: 425 (34) (my emphases).
[406] See Garcia, 'Two Concepts of Desert': 222; Raphael, *Moral Judgment*, 71.
[407] Wood, *Kantian Ethics*, 219.
[408] *MdS* 6: 337 (109-10).
[409] *MdS* 6: 331 (104-5). See also ibid. 6: 319 (95).

wide obligation.'[410] The 'ruler' will be commanded to make (crime-proportionate) 'physical harm'[411] to criminal in general, or their 'pain,'[412] or 'rightful' offense to 'the accused's feeling of honor,' or 'unilateral' coercion,[413] an end—and to seek this end without taking recourse to 'any incentives,' that is, to do this on the ground that it is her *duty* to do so.[414] On Kant's account of such ends, the ruler will always already have a *reason* for pursuing it, independently of any of his or her parochial interests. Not only is the sovereign to punish, but neither the ruler nor anyone else is to succumb to the temptation to subsume actual punishments under a description of what punishment consists in, in general, that has reference to anything that exceeds the bare requirements of retribution.

Even if it were true, in general, that 'Kant's theory of *legal* punishment is not deontological,'[415] as Tunick claims, if the law of punishment is a categorical imperative and if this imperative pertains to the *a priori* 'combination,' as ground and consequent, of the concepts of crime and authorized harm, *in general*, then this view is mistaken. I do not claim as much, here. I only claim that it is mistaken with respect to Kant's attitude towards capital punishment.[416] If, as Tunick argues, 'deterrence [is] the justification for legal (but not moral) punishment,'[417] then the law of punishment, to the extent that that it is a categorical imperative, does not pertain to legal punishment in this sense at all. But of course, this is not what Kant says. Implicit in Tunick's claims, in any case, is that retributivism *does* imply a deontological theory of punishment.[418] He simply denies that Kant is a retributivist. To the extent that punishment is regarded as retribution, however, the punishment of crimes is regarded as a *duty* that binds the sovereign unconditionally. And to the extent the punishment of crimes is regarded as a *duty* that binds the sovereign unconditionally, punishment is regarded as retribution.

---

[410] See *MdS* 6: 388-91 (152-4) for Kant's exposition of the relevant distinctions.
[411] See *KpV* 5: 37 (34).
[412] *MdS* 6: 331 (104).
[413] *MdS* 6: 363 n. (130 n.).
[414] *Gr* 4: 425 (34) (my emphases).
[415] Tunick, 'Is Kant a Retributivist?': 62.
[416] Cf., however, *MdS* 6: 335-7 (108-9); Lind, 'Kant on Capital Punishment'; Tunick, 'Is Kant a Retributivist?': 76.
[417] Tunick, 'Is Kant a Retributivist?': 64.
[418] See also Moore, 'The Moral Worth of Retribution', 97; N. T. Potter, 'Kant and Capital Punishment Today,' *Journal of Value Inquiry* 36, no. 2 (2002): 272.

Punishment all the way down: the special case of capital punishment

In this section I argue that, for Kant, the politically situated practice of capital punishment and the eschatologically situated practices associated with unhappiness (see chapter 4) are each conceived of, *univocally*, as moral (not merely legal) retribution. I argue that in each case (the political and the ethical-eschatological) the occasioning *object* of the practice is not a criminal or immoral deed, but the wicked agent *as such*. Political punishment, more generally, is addressed to the malefactor's law-transgressing *deed(s)*. The politically (legally) punishable agent is punished under a description of the practice of punishment that has reference, strictly, to these. She is punished because she has done this or that. The murderer, however, is punished *qua* 'who' or 'what' she is: a person whose murderous deed discloses the 'inner wickedness' that makes her death *good*, in addition to being right. This is so, I argue, in the same sense that the immoral agent rendered (eschatologically) unhappy—of necessity, given that her will is an evil one—is subject to unhappiness because this is an outcome that is absolutely called for and absolutely good. It is particularly clear here, then, that Kant rules out, not only the political *practicability* of mercy, but mercy's very goodness as such. I argue that in Kant's justification of capital punishment he offers us his most vivid expression of the mercy-excluding immediacy of the connection that he takes to hold between crime and punishment, on the one hand, and immorality and unhappiness, on the other.

As I mentioned earlier, some commentators concede that some of the things that Kant says about punishment suggest that he is a retributivist—even a strong one—but then deny that his retributivism is entailed by, or coheres with, the rest of his thinking about morality. This is often taken to be especially true of the things he says about the value and justification of capital punishment. One strategy in response to this situation is to concede that Kant's thinking about this topic is problematic, but that his views in this area are not central in any way and that a compelling reading of Kant's ethics can be articulated without reference to this encumbrance.[419] Another strategy is to recognize that Kant is a thinker who belongs very much to his time and place and to emphasize the improvement that his thinking represents over earlier ap-

[419] See Wood, *Kant's Ethical Thought*, 2.

proaches.[420] Another approach, which combines readily with the foregoing, is to argue that key elements of Kant's moral theory, including other parts of this thinking about punishment, imply that the death penalty is in fact either immoral or juridically incoherent on his own terms.[421] Yet another tactic is to aver that, along with Kant's retributivist proclamations, his strong claims about how murderers ought to be treated are mistakes on his part, to the extent that these are not consistent with his most fundamental commitments.[422]

I do not disagree, in principle, with any of this.[423] I claim, instead, that the particular rigidity of Kant's thinking about capital punishment is telling and that it signals a blind spot,[424] not so much *in* his thinking his thinking about law and politics, but at its boundary and so, in a sense, on its outermost surface; not so much a mistake in reasoning,[425] but the tenacious residue (at least) of an archaic practice that distorts Kant's vision from the outset. I affirm that, in Kant's thinking about punishment more generally, one finds two things going on at once: one line of thought that proceeds relatively independently of the commitments that are expressed in his habit of glossing 'morality' as 'worthiness to be happy,' the other line of thought expressing and ratifying what the habit itself sanctions and conveys. If this is so and if, as I claim, Kant's thinking about capital punishment is the 'other line of thought' in question, then his statements concerning the death of murderers are not an aberration; they make sense relative to that trajectory. Later, I will show that *this part* of Kant's thinking extends from the political context in which apparent instances of this thinking's ultimate realization find expression (on the scaffold), out into the eschatological.

---

[420] See, for example, Potter, 'Kant and Capital Punishment': 273, 277-8.
[421] For the former view see ibid., 281.; for the latter, see A. Ataner, 'Kant on Capital Punishment and Suicide,' *Kant-Studien* 97, no. 4 (2006): esp. 453, 455, 481 (but with respect to the latter's argument cf. Kant on Beccaria at *MdS* 6: 334-5 [108]).
[422] Holtman, 'Toward Social Reform': 12. The author is thinking, especially, of *MdS* 6: 333 (106). See also Scheid, 'Kant's Retributivism': 281.
[423] Sussman shows good sense by assuming no more than Kant's 'amply evidenced *moral* retributivism' (David Sussman, 'Kantian Forgiveness,' *Kant-Studien* 96, no. (2005): 88; my emphasis).
[424] For an earlier characterization of this part of Kant's thinking in these terms see Wolfgang Palaver, 'Mimesis and Scapegoating in the Works of Hobbes, Rousseau, and Kant,' *Contagion* 8, no. (2003): 140.
[425] See Potter, 'Kant and Capital Punishment': 271.

'If…he has committed murder he must die*'*

Now we come to some of what Hill calls 'Kant's most off-putting rhetorical remarks on punishment in [*The Metaphysics of Morals*],'[426] teachings that have seemed to some to be 'defective in human sympathy and understanding.'[427] Here especially, we are faced with 'rules of punishment' that are, at the very least, 'tough and inflexible.'[428]

Several pages along from his initial assertion that 'the law of punishment is a categorical imperative,' Kant refers more specifically to 'the categorical imperative of penal justice,' which he glosses as the demand that 'unlawful killing of another must be punished by death.'[429] This unique specification of Kant's 'law of punishment'—in terms, that is, of punishment, specifically, for murder—is determined, in part, by the context of his discussion. But it signifies a focus that is evident throughout this part of *The Metaphysics of Morals*: Kant has his eye on the practice of capital punishment. Some of his most concrete, urgent, and sustained *pleas* for the practice of punishment—above all for its absolute, irreducible necessity—pertain to the death penalty. Thus:

> If…he has committed murder he *must die*. Here there is no substitute that will satisfy justice. There is no similarity between life, however wretched it may be, and death, hence no likeness between the crime and the retribution unless death is judicially carried out upon the wrongdoer, although it must still be freed from any mistreatment that could make the humanity in the person suffering into something abominable.[430]

On the one hand, to be precise, the murderer must '*die*' (*sterben*). Unique among crimes murder does not admit of a variety of possible applications of the principle of *ius talionis*. On the other hand, to be equally precise, he '*must*' (*muß*) die.[431] And the sense of this 'must' is evident from a slightly later passage from the same text: 'every murderer…must suffer death; this is what justice, as the idea of judicial authority, wills in accordance with *universal laws* that are grounded *a priori*.'[432]

---

[426] Hill, 'Kant on Punishment': 291.
[427] Cohen, 'Critique of Kant's Philosophy of Law', 286 (but note that Cohen's point is a more general one about the *Rechtslehre* as a whole).
[428] Hill, 'Wrongdoing, Desert, and Punishment', 311.
[429] *MdS* 6: 336-7 (109).
[430] *MdS* 6: 333 (106) (Kant's emphasis modified). See also *MdS* 6: 463 (210); *MdS* 6: 436 (188). But cf. *MdS* 6: 441, 463-4 (191, 210); Wood, *Kant's Ethical Thought*, 134-5.
[431] See also Tunick, 'Is Kant a Retributivist?': 73.
[432] *MdS* 6: 334 (107) (emphasis added).

In a famously dramatic passage, however, interposed between the two foregoing citations, Kant spells out this same claim to necessity in a rather different manner, insisting that

> [e]ven if a civil society were to be dissolved by the consent of all its members (e.g., if a people inhabiting an island decided to separate and disperse throughout the world), the last murderer remaining in prison would first have to be executed, so that each has done to him what his deeds deserve and blood guilt [*Blutschuld*] does not cling to the people for not having insisted upon this punishment; for otherwise the people can be regarded as collaborators in this public violation of justice.[433]

One of the remarkable things about this passage is that, instead of grounding his claim that 'the last murderer…would…have to be executed' in something along the lines of his slightly later reference to 'what justice, as the idea of judicial authority, wills in accordance with *universal laws* that are grounded *a priori*,' Kant refers instead to the problem of 'blood guilt' (*Blutschuld*) and a 'public violation of justice' whose character as injustice is highly ambiguous, both because it takes place (*ex hypothesi*) on the very threshold of this society's disappearance, but also because, as we saw earlier, Kant cannot really take it to be the case the sovereign (here, presumably, 'the people' as a whole) has a 'narrow,' coercible, legal duty to punish. I will return to these matters—especially Kant's recourse to the notion of 'blood guilt'—below.

First, I want to point out that Kant's unrelenting proposal for the treatment of this 'last murderer' is a sticking point for any reading of his thinking about punishment that would claim that he is not a 'strict,' 'bold,' 'deep' (etc.) retributivist *at all*, that is, in any connection whatsoever. The passage very clearly cites the two main elements of such an approach in its justification of what it proposes. It is a problem for accounts of the kind that I referred to above, which construe Kant's theory of punishment in its political frame-work in terms of primary purposes to do with 'law and order,' the prevention of hindrances to external freedom, or deterrence. Or, in a sense, it is *not* a problem for such accounts to the extent that these readers are able to argue that, on balance, Kant's thinking is not what it appears to be here, or that he is simply being inconsistent, and that there are good reasons for prioritizing and foregrounding other tendencies in his thinking about punishment.

Hill, for example, integrates this passage into the rest of his account of Kant's rationale for the practice of punishment by arguing that, when Kant argues that fail-

[433] *MdS* 6: 333 (106).

ure to put 'the last murderer' to death would lead to a situation in which the people of the former island kingdom would be susceptible to the accusation that they had been 'collaborators in this public violation of justice,' Kant is endorsing these murderers' punishment as no more than 'a rigorous application of [his] extreme retributive *policy* that *all* of the guilty ought to be punished, following *lex talionis* so far as permissible.'[434] Hill's argument turns in significant measure, however, on an opposition between the thesis that, for Kant, particular punishments are to be justified, in general, on the basis of the 'intrinsic' or 'moral' desert of criminals—which Hill denies—and the thesis that, for Kant, particular punishments are justified to the extent that they constitute an application of the 'retributive policy,' mentioned above, 'that *all* of the guilty ought to be punished, following *lex talionis* so far as permissible.' Hill's argument is aimed, then, at readers who take the 'last murderer' passage to be evidence favouring the general claim that Kant endorses the view that punishments are justified 'because of, or even according to, intrinsic desert.'[435] I will return to this latter notion below. Suffice it to say, for the moment, that I concede Hill's claim that, generally, Kant does not take 'intrinsic desert' to be of the essence for the justification of punishment. But I do claim, however, that this is a central component of Kant's thinking, specifically, about the justification of *capital* punishment in particular.

Not all commentators are as confident as Hill about the fit between this passage and other, more evidently moderate approaches to the justification of punishment that might be found along with it in Kant. Again with respect to the 'last murderer' passage, Fleischacker, for example, notes 'that Kant here clearly shows that his reasons for the *moral worth* of punishment extend beyond even the universalizable aim of minimizing the hindrances to freedom within society.'[436] This marks a significant departure from Hill's reading. Fleischacker recognizes that Kant takes punishment—even if only here in the case of putting murderers to death—to have '*moral worth*' and not merely a political or legal value.

[434] Hill, 'Wrongdoing, Desert, and Punishment', 332. See also Tunick, 'Is Kant a Retributivist?': 74.
[435] Hill, 'Wrongdoing, Desert, and Punishment', 332.
[436] Fleischacker, 'Kant's Theory of Punishment': 438 (my emphasis).

The ultimate punishment: the political and the eschatological, together on the scaffold

If punishment has moral worth, then this is because to punish is a moral undertaking on the part of someone in particular, the fulfillment of a moral obligation, action that conforms with what duty requires of such an agent, and so forth. In the person of the sovereign (or the latter's representative) we have an agent and a possibility for action in which the *political* coincides with the (Kantian) *good*. It does not do so at all in cases where the sovereign rewards citizens in accordance with their juridical merit. To do *that* is no duty. It is possible that it does not do so, either, in the case of punishments for crimes apart from murder (if there are empirical reasons for those punishments). But desert, for Kant, where it is qualified as *desert-of-death* (given murder), is both an ethical and a juridical notion. The 'law of punishment,' in its form specifically as the 'categorical imperative of penal justice' (again, the stipulation that the 'unlawful killing of another must be punished by death') is a *moral* and no merely legal imperative. But this *requires*, as I will now show, that unique among punishments, the death penalty (for murder) be both justified with reference to the deed (death in exchange for death, as it were), but also with reference to the 'inner wickedness' of the criminal and her 'intrinsic desert' of the treatment that she is also legally judged to deserve.

'Inner wickedness,' intrinsic desert, and punishment all the way down

'Inner wickedness'

One way of distinguishing between the realm of politics or 'legality,' on the one hand, and 'morality,' on the other, is by pointing out, as Höffe does, that 'the topic of evil,' while an integral theme for thinking about 'morality,' 'is dispensable in a legal ethical theory, since legal ethics concerns merely (juridical) legality.'[437] This may be affirmed of Kant's political and legal philosophy in general, to be sure, but not in a straightforward manner at the boundary that is marked by capital punishment. This is not, however, because an extension of morality in its primary, *forward-looking* sense

[437] Otfried Höffe, *Kant's Cosmopolitan Theory of Law and Peace*, Modern European Philosophy. (Cambridge: Cambridge University Press, 2006), 73.

as action 'from duty' overtakes the legal or political with its 'strict[er] demands'[438] and gives us 'a philanthropic concept of law that aims at enforcing duties of virtue, such as beneficence,' or 'a disposition-based concept of law that is not satisfied with legality…but additionally demands inner recognition [of the moral law].'[439]

Rather, this is because an extension of the backward-looking *legal* interest in action that would tend to undermine the stability and well-being of the *polis*, overtakes and subsumes morality in *its* secondary, backward-looking sense as a point of view on human action that must contend with the problematic of human action as something both *over and done with* (not as deeds to be done, but as deeds accomplished) and, at the same time, a source of unpredictably ramifying consequences. At the boundary that is marked by the practice of capital punishment, the (Kantian) 'topic of evil' turns out *not* to be dispensable. If this topic is superfluous, for Kant, in matters pertaining to 'merely (juridical) legality,' as Höffe says, Kant does not quite retain the courage of his convictions here—and rightly so, as we shall see. He writes that:

> This fitting of punishment to the crime, which can occur only by a judge imposing the death sentence in accordance with the strict law of retribution, is shown by the fact that only by this is a sentence of death pronounced on every criminal in proportion to his *inner wickedness* (even when the crime is not murder but another crime against the state that can be paid for only by death).[440]

Kant's reference to other 'crime[s] against the state that can be paid for only by death' does not belie my claim that, on his account, the death penalty *for murder* is unique among juridically executed punishments. The high-minded rebels (Kant's example) whose sense of honour leads them to *prefer* death to 'convict labor' find, in being put to death, a fate that is proportionate to their crime—whereas forced labor would have been an unjust affront and disproportionately coercive. Here, the punishment (death) conforms to the *ius talionis*, but not *qua* death—only under a description of 'death' that has reference to the rebels' high-minded desire to preserve their honour (hence *not* in relation to their 'inner wickedness,' which is not proposed). Neither Kant, nor his imaginary rebels, interpret this punishment as some-

[438] Ibid., 86.
[439] Ibid., 3. Indeed, '[a]rmed with [his] twofold distinction' between 'legality' and 'morality,' Höffe points out, Kant avoids giving us either of these (ibid.).
[440] *MdS* 6: 333 (106-7). This is a unique passage. But cf. Wood, *Kantian Ethics*, 220.

thing that they intrinsically, that is, *morally* deserve.[441] Again, however, in the case of the *murderer* put to death, the death penalty is justifiable only to the extent that it is both conformable to the *ius talionis*, precisely *qua* death, and also proportionate to the 'inner wickedness' of the criminal. In the case of the wicked murderer it is a matter, both of what she intrinsically (morally) deserves, which is to forfeit—as far as anyone can tell—the very possibility of happiness, and also of what she is judged to deserve in a legal sense, which is to suffer an equivalent consequence for having unlawfully taken a life.[442]

I claim, again, that Kant is committed to a particular practice: the practice of putting murderers to death—under a description of the deed that causes their death that foregoes all reference to any of the social or political benefits that might arise from it. Kant's anomalous reference to 'inner wickedness' signals his understanding of what would have to be the case if particular instances of capital punishment, so understood, were actually to be justified.[443] His antecedent, remarkably rigid commitment to the practice in general, however, is readily mistaken for confidence that he has good reasons for taking the practice to be justified, in general. Kant himself mistakenly takes the *prima facie* inevitability and fittingness of the practice to which he is committed for rationally warrantable necessity.

Kant gets no further, really, than this. If the death penalty is not regarded, from the point of view of the world in which it takes place, as retribution, then there *is* something especially hideous about it. But, too, if the death penalty is retribution in this sense, then the 'sentence of death' must also be 'pronounced on every criminal in proportion to his *inner wickedness*' since only the latter's presence, which is the agent's ultimate immorality, unveiled, justifies the absolute occlusion of the human being's access to happiness. Clark observes that because, on Kant's own account, none but God knows whether the agent instantiates this wickedness, then '[p]ublic justice is (in general) deterrent in aim, but to be distributed retributively,' that is, in accordance with the principle that all and only criminals be punished and in accord-

[441] Kant does not 'forget' here, as Cohen suggests, that 'some criminals (e.g. political ones such as the Scotch Rebels) act from honorable motives' (Cohen, 'Critique of Kant's Philosophy of Law', 286-7). To the contrary, Kant bears this in mind and cleverly works his way around it.
[442] But cf. Hill, 'Wrongdoing, Desert, and Punishment', 333.
[443] Hill also acknowledges that this is, or appears to be, an 'anomaly' in Kant's thinking (ibid., 334). But cf. his apology for Kant ibid., 335-6.

ance with the *ius talionis*. He adds that, for Kant, 'divine justice,' by contrast, 'is purely retributive' and that none but 'God can punish us in accordance with our moral guilt.'[444] He asks whether, given that 'Kant believes that ideally our happiness should match our virtue,' he also holds that 'the unhappiness of criminals should match their crimes?' None but God 'can punish us for our "inner wickedness,"' Clark points out. In light of this, he asks, 'should the courts punish to secure such a match?'[445]

It is not that Kant's thinking that this is so, uniquely, in the case of capital punishment is a piece of thinking that is internal to his theory of punishment. Of course, Kant does not think that 'the courts [should] punish to secure such a match,' here. Of course Kant knows better.[446] But he also recognizes that there is at least one mode of punishment that must be regarded in precisely these terms *or else given up*. His anomalous reference to the fit that exists between the death sentence and the 'inner wickedness' of murderers expresses the grip of the practice itself, not a coherent element in Kant's theory of punishment.

Intrinsic desert

The murderer's 'inner wickedness' is the look of her immorality, as it were, the particularly dreadful look of her free acquiescence in evil, put on display in the act of murder. As I have shown, Kant's habit of glossing 'morality' as 'worthiness to be happy' expresses a pair of related thoughts about the relationship between immorality and unhappiness: immoral agents *deserve* and *ought to be* unhappy. To put a human being to death is—from the point of view of the agent of the deed and of every spectator present—an unlimited, irrevocable annihilation of whatever happiness she possessed and whatever happiness was possible for her. No one need regard this as the death penalty's aim, or the reason for its being meted out. This is, however, what it accomplishes.

---

[444] Clark, 'A Non-Retributive Kantian Approach': 18 n. 6.
[445] Ibid., 16.
[446] See, however, Brooks view (*pace* Clark) that '[i]nner wickedness is not, in fact, knowable by God alone: it is made explicitly known in acts of murder' (Brooks, 'Kantian Punishment': 241; see also ibid., 242). For the opposite view—again, however, based on the mistake of thinking that Kant's remarks about 'inner wickedness' are a general claim belonging to his theory of punishment as a whole—see Hill, 'Wrongdoing, Desert, and Punishment', 329.

Kant's reference to the murderer's unrivalled 'inner wickedness' exposes a certain discomfort in the face of this fact. It is Kant's apology for a practice whose justification must include reference to the agent's 'intrinsic desert,' which, as desert-of-death, is the political incarnation of her moral unworthiness to be happy. Normally, of course, the state may not obliterate the very possibility of happiness in cases where agents are unworthy to be happy—in the cases, say, of you and I and, presumably, Kant. However, thinks Kant, if that unworthiness to be happy, the agent's immorality, is put on display in the act of murder the state may and must do so.

'The *Tugendlehre*,' Kant's 'doctrine of virtue' (and so, too, of vice), 'teaches of internal actions of the will, opaque to perception,'[447] as Gibbs points out. In general, Kant's *Rechtslehre* does not deny this opacity. Kant's notion of 'inner wickedness,' then, is an anomaly in the context of his discussion of political life, to the extent that it has direct reference to the agent's immorality ('internal actions of the will'). An agent whose will is so qualified—given that she is a murderer—is not only unworthy to be happy and thus intrinsically deserving of unhappiness (like any other immoral agent), but intrinsically deserving of the punishment that, justified first with reference to the political *ius talionis*, also ratifies and effects this other judgment *on the spot*.

By contrast, for Kant, the desert that warrants other punishments in civil society is 'extrinsic.' In other words, particular deeds merit particular punishments, just given the class of deeds to which they belong. Capital punishment, too, is connected with a particular deed, homicide under a particular description (as vicious, unlawful killing). Here, however, the proportionate punishment's conformity to the *ius talionis* goes deeper than the deed. The punishment's fittingness encompasses not only the deed, but also the will from which the deed arises. It is this will, which is again 'in a sense…the very person himself,'[448] as Wolff observes, that deserves to die. Here, uniquely, the *person* is punished for being the kind of agent that she is.

Hill observes that 'a thorough-going *retributivist* theory of punishment' might take one of the following two forms. On the one hand, he suggests, it might propose that the 'amounts and kinds of punishment' that offenders receive be determined in terms of the '*ius talionis*, as Kant presents this,' which Hill takes to be a matter of

[447] Gibbs, 'Fear of Forgiveness': 326.
[448] Wolff, *Autonomy of Reason*, 59.

matching external, illegal *deeds* to the punishments that fit them.[449] On the other hand, argues Hill, an equally retributive standard would be to '[p]unish in the way and degree appropriate to the "inner deserts" of offenders.' And, as he goes on to observe, in order to do this we would have to 'take into account their characters, commitment to morality, obstacles, effort of will, etc.'[450] Hill objects to the latter, correctly, on the basis of the fact that, while this counts as a form of retributivism, 'it is incompatible with Kant's idea that public law should concern itself only with "external" actions.'[451]

I do not claim otherwise—in general. Nor do I claim that, in the case of capital punishment, Kant actually *argues* that the murderer must die because he is wicked, or immoral, or because he deserves absolutely to be unhappy and because putting him to death is the best way of achieving this end. Kant says quite clearly that murderers must die because death is the punishment that most perfectly fits their crimes. Kant nevertheless indicates, however, that when it comes to putting murderers to death, something *more* is present in our reflexive understanding of what we—the sovereign (or the latter's avatar, the executioner), the spectator, and the murderer herself—take ourselves to be doing on (and around) the scaffold.

Hill marks this intrusive element as well, by affirming that, in Kant's view, 'criminals *deserve*, in some sense, to be punished,' and that, at the same time, Kant also holds 'that anyone who lacks a good will is, to some degree, *unworthy to be happy*.' Kant's actual *thinking*, here, is obscure, however, as Hill also points out. Hill responds to this irritant, however, by arguing that, in any case, these positions 'do not amount to an endorsement of the intrinsic desert thesis as an action-guiding (practical) principle.'[452] Kant certainly thinks that none but virtuous agents are worthy of happiness, Hill concedes, but this does not mean that he 'endorse[s] the deep retributive idea that we ought to make the vicious suffer because they inherently deserve it' (Hill calls this 'the intrinsic desert thesis'). And, Hill suggests, 'despite ap-

[449] Hill, 'Kant on Punishment': 306.
[450] Ibid.
[451] Ibid. (author's italics removed).
[452] Hill, 'Wrongdoing, Desert, and Punishment', 324.

pearances, this thesis is not implied by Kant's official theory of punishment either.'[453]

As far as these 'appearances' go, Hill identifies two main grounds in Kant's work for thinking that he 'held that wrongdoers intrinsically deserve to suffer.' First, there are Kant's 'remarks, sprinkled throughout his works, that a good will is the condition of the *worthiness to be happy*.'[454] Second, there are Kant's 'tough-sounding remarks about punishment' in *The Metaphysics of Morals*.' None of this implies, Hill argues, 'that we are warranted, as individuals or state officials, to inflict suffering on wrongdoers or to deprive them of happiness.'[455] Kant never implies that anyone possesses 'a general warrant to interfere with the happiness of persons who are "unworthy" of it.'[456]

The main practical impediment, here, is that 'according to Kant's moral psychology we are so ignorant of the moral worth of others that we could not fairly undertake to make others happy, and unhappy, in proportion to their worthiness.'[457] The upshot of this is that, although Kant endorses the *notion* of 'intrinsic desert,' he does not allow claims about the latter—which bear in the abstract on immoral agents in general—to have any 'action-guiding' significance in regard to the punishment of criminals, or at least, Hill argues, not 'in his more mature, systematic work.'[458] By contrast, Hill does allow that the idea that immoral agents intrinsically deserve to suffer, and so to be unhappy, has a 'more anemic,' 'nonpractical,' 'faith-guiding,' 'wish-expressing,' significance.[459] Hill's general apologetic aim extends too far, however. He does not recognize, as Kant does, that there is something exceptional about capital punishment and that the latter's justification, uniquely, must have refer-

[453] Ibid., 312. See also Thomas Hurka, 'Desert: Individualistic and Holistic,' in *Desert and Justice*, ed. Serena Olsaretti (Oxford: Clarendon Press, 2007), 52; Ross, *The Right and the Good*, 56-7.
[454] Hill, 'Wrongdoing, Desert, and Punishment', 326. Hill refers, especially, to Kant's opening statements in the *Groundwork*'s first section (*Gr* 4: 393 [7]) and the first part of the 'Fragment of a Moral Catechism' at *MdS* 6: 480-81 (223-4).
[455] Ibid.
[456] Ibid., 327.
[457] Ibid., 328.
[458] Ibid., 316.
[459] Ibid., 315. See also ibid., 328. Hill is thinking, for example, of *LEC* 27: 287 (80). See also *KpV* 5: 37-8, 61, 99-100, 124-32 (34-5, 53, 83-4, 103-10); *Rel* 6: 69, 73-4, 116-17, 126 (110-11, 113-14, 147, 155); *LEV* 27: 553 (309).

ence to an 'intrinsic' mode of desert that *does* coincide with the immoral agent's unworthiness to be happy.[460]

Murphy contrasts 'the official theory presented in the *Rechtslehre*,' with its 'rather self-righteous tone'—self-righteous because it comes to the point of identifying the murderer's 'inner wickedness'—with the *Religion*'s 'insights' concerning the pervasiveness (indeed the universality) of radical evil and, as a consequence of this, the 'position of humility' which that work evinces. No one may claim, on Kant's account of 'radical evil in the *Religion*…that he has, from a morally creditable motive, restrained himself.'[461]

But this is to concede that Kant's thinking about capital punishment in *The Metaphysics of Morals* goes too far, that it really does mark a connection between the death penalty as consequence, on the one hand, and unworthiness to be happy as ground, on the other. And this *is* an anomaly within the *Rechtslehre*. This does reveal that the law of punishment, 'the categorical imperative of penal justice,' the demand that 'unlawful killing of another must be punished by death,' has become the law of *unhappiness* and rendered this law—at least upon the scaffold—an 'action-guiding' principle. Officially, the murderer's death conforms to the *ius talionis*, just given his or her deed. In this way it is like any other legal sanction. But it is nevertheless a counter-example, too, that belies all claims that Kant's retributivism *qua* adherence to the *ius talionis* and punishment of none but 'the one that did it' is warranted by an aim that falls *within* the bounds of the *polis* alone. The murderer's death *also* matches the absolute occlusion of her access to happiness with her immorality, her 'inner wickedness.' That is not a description of what the punishing agent is doing from within the discourse of politics, but rather a (re)description that has reference to an *ultimate* end.

Note that when I say, above, that the law of punishment ('the categorical imperative of penal justice') has 'become' the law of unhappiness and when I claim that the latter must be then, for Kant, as much a 'moral imperative,' as much an instance of universally practically necessitating law, as the former, it may be that my way of ex-

[460] In spite of this significant difference between Hill's reading and my own, I am adopting his definition of 'intrinsic desert.' For Hill's full discussion of this notion in relation to other senses of 'deserves' see ibid., 324-5.
[461] Murphy, 'Kant's Theory of Criminal Punishment', 440.

pressing myself will lead to misunderstanding. Note, however, that I do not claim that either the *political* law of punishment, or the *eschatological* law of unhappiness, really *does* qualify, even in Kant's own terms, as a categorical imperative. I only claim that, (i) given that Kant *takes* the law of punishment to be a categorical imperative, and (ii) given Kant's understanding of what it is that qualifies a practical principle for designation as an imperative of this kind, and (iii) given the theoretical upshot of Kant's special plea for the practice of capital punishment, it follows that (iv) *if* Kant takes immorality and unhappiness to be connected *a priori* and necessarily, but only normatively (that is, as something that is *called for*, that *ought to be so*, but is not necessarily so in fact), not physically, nor analytically, *then* (v) Kant takes it to be the case that the law of punishment has a kind of 'reach' that extends to the *eschaton* and a necessitating force that subsumes and constrains even divine action in regard to immoral agents. *At that limit*, I suggest, the law of punishment—which finds its *primary* mode of expression as an imperative whose object is the having of a particular *political end*—has 'become' the law of unhappiness, that is, a law that commands that God (and, in the special sense that I discuss, we) adopt and effect (as far as possible) an *eschatological end*: the unhappiness of immoral agents, which is, precisely, their *punishment* (but qualified now as an eschatological and not merely a political undertaking).

All the way down

In a moment I will take up Kant's own turn towards this end, which comes at the very end, fittingly, of *The Metaphysics of Morals*. First, however, I will draw attention to the most important consequence of the foregoing discussion of 'inner wickedness' and 'intrinsic desert': the demand that, if she is to be put to death, then the punishable subject must be (and always already is) regarded as 'the one that did it,' *all the way down*. The demand, here—which derives surreptitiously from a *description* of what we actually do—is that we exclude all reference to the involvement of other agents in our account of the genesis of the punishable deed. The agent is only punishable, as I said earlier, if *she*, 'the one that did it,' broke the law, *wholly* and *simply*. This demand is internal to Kant's endorsement of capital punishment, which is in effect from the outset, long before he thematizes it. It is the inchoate theory that is embodied in his habit of glossing 'morality' as 'worthiness to be happy.'

Again, the subject has to be 'the one that did it,' all the way down. This 'depth' has two aspects. As Kant says in *Theory and Practice*: 'punishments happen only to a will that is *free* but *contrary to the law*.'[462] The first aspect (mentioned here second) is the deformation of the agent's will, its evil orientation, which may be openly revealed in the obvious wickedness of murder, or remain the agent's secret, evident nowhere but in the inner experience of moral self-contempt. The second element is the absolute *ascribability* to the agent herself, to her will, all the way down, without remainder, of this very deformation (see my discussion of the imputable '*Gesinnung'* in the introduction to this thesis). This is why the human being is always already intrinsically, morally '*strafbar*.' This is why she deserves, utterly, to be cut off from happiness.

The ascribability of particular deeds is of the essence for Kant's distinction between 'punitive justice' and 'punitive prudence.' In *The Metaphysics of Morals*, he writes that

> the argument for the former is *moral*, in terms of being *punishable* (*quia peccatum est* [because a crime has been committed]) while that for the latter is *merely pragmatic* (*ne peccatur* [so that the law will not be broken]) and based on experience of what is most effective in eradicating crime.[463]

In this context, Kant is not using the epithet, 'moral' (*moralisch*), in its primary, forward-looking sense. That is 'moral,' here, which bears on the secondary, backward-looking question of whether a particular deed is actually *ascribable* to someone in particular and the question of whether the deed may be subsumed under a description (of it) that has immediate reference to the *punishability* of anyone who commits it (i.e., whether the deed may be subsumed under the concept of transgression—here, of 'public law'—which Kant takes to be immediately, analytically connected to the concept of punishment *a priori*).

For Kant, both the theory and the practice of *capital* punishment (as the actualization of the *a priori* conceptual 'combination' of the notions of murder and of the murderer's death), and the theory and the practice of the activity which (assuming that any subject were authorized and commanded to do it) would consist in the actualization of the *a priori* notional connection between unhappiness and immorality, require and deploy the notion of ascribability, on the one hand (and so, too, of 'abso-

[462] *ÜdG* 8: 288 (289) (my emphasis).
[463] *MdS* 6: 363 n. (130 n.).

lute causal spontaneity'[464]), and the notion of a freely undertaken rejection, itself unnormed and unintelligible, of a particular mode (but not every mode) of *law-governedness*, on the other. As Kant puts it in his lectures on the philosophy of religion, '[a]s soon as the human being recognizes his obligation to the good and yet does evil, then he is *worthy of punishment*, because he could have overcome his instincts.'[465]

The eschatological post-script to *The Metaphysics of Morals*

I suggested earlier that Kant's anomalous reference to the fit that exists between the death sentence and the 'inner wickedness' of murderers expresses the grip of the practice of capital punishment itself, not an integrated element of Kant's theory of punishment. The same is true with respect to his references to 'blood-guilt.'

According to its title, the conclusion of *The Metaphysics of Morals* concerns a matter that lies 'beyond the bounds of pure moral philosophy,' namely, 'religion as the doctrine of duties to God.'[466] After presenting a brief argument concerning 'religion,' 'the boundaries of the *science* to which it belongs,'[467] and the possibility 'of a "Religion *within the Bounds* of Mere Reason,"'[468] Kant offers a 'Concluding Remark' in which, among other things, he returns—not in an offhand manner, but as if returning to a theme of absolutely integral importance—to the topic of punishment. He returns to the topic of punishment and, one sentence into this recapitulation, he returns, again, to the topic of murder. The tone and language of Kant's argument is not merely religious, but also superlatively mythic and biblical,

> Punishment (according to Horace) does not let the criminal out of its sight as he strides proudly before it: rather, it keeps limping after him until it catches him.—Blood innocently shed cries out for vengeance.—Crime cannot remain unavenged; if punishment does not strike the criminal, then his descendants must suffer it, or if it does not befall him during his lifetime,

---

[464] See this thesis' conclusion for a very brief discussion of this topic. Kant's notion of 'spontaneity' and the distinction between the latter and 'autonomy' is an important and complex matter whose detailed discussion, however, lies without the limits of my current project (see Allison, *Idealism and Freedom*, 111, 140; Ameriks, *Kant and the Fate of Autonomy*, 189; Lewis White Beck, 'Five Concepts of Freedom in Kant,' in *Stephan Körner: Philosophical Analysis and Reconstruction*, ed. J. T. J. Srzednicki (Hingham: Kluwer Academic Publishers, 1987); Hill, *Dignity and Practical Reason* 93-4, 106-10).
[465] *Vorlesungen-Religionslehre* 28: 1079 (412).
[466] *MdS* 6: 486 (229). The work's 'Conclusion' is subtitled: 'Religion as the Doctrine of Duties to God lies Beyond the Bounds of Pure Moral Philosophy.'
[467] *MdS* 6: 487 (229).
[468] *MdS* 6: 488 (230).

> then *it must take place in a life after death, which is accepted and readily believed expressly so that the claim of eternal justice may be settled*.[469]

And then, in a return to language that he used in dealing with the problem of the 'last murderer,' Kant adds, immediately:

> I will not allow *blood-guilt* [*Blutschuld*] to come upon my land by granting pardon to an evil, murdering duelist for who you intercede, a wise ruler once said.—*Guilt for sins* must be expiated, even if a completely innocent person should have to offer himself to atone for it (in which case the suffering he took upon himself could not properly be called punishment, since he himself had committed no crime). All of this makes it clear that this judgment of condemnation is not attributed to a *person* administering justice (for the person could not pronounce in this way without doing others wrong), but rather that *justice* by itself, as a transcendent principle ascribed to a supersensible subject, determines the right of this being.[470]

To his earlier remark about punishment's 'tak[ing] place in a life after death' Kant adds a footnote that deploys key 'critical' distinctions in order to insure that his readers understand that (given certain of our practical interests) we *can* believe and affirm what we neither know nor perceive:

> It is not even necessary to bring the hypothesis of a future life into this, in order to present that threat of punishment as completely fulfilled. For a human being, considered in terms of his morality, is judged as a supersensible object by a supersensible judge, not under conditions of time; only this existence is relevant here. His life on earth—be it short or long or even everlasting—is only his existence in appearance, and the concept of justice does not need to be determined more closely since belief in a future life does not, properly speaking, come first, so as to let the effect of criminal justice upon it be seen; on the contrary, *it is from the necessity of punishment that the inference to a future life is drawn*.[471]

Contrast this with Kant's 'postulates of pure practical reason' in the second *Critique*, the immortality of the soul and the existence of God, 'practical' belief in which is justified with reference to the inexorability of the *a priori* claims of morality in its primary, forward-looking sense: a call to the (self-wrought) transformation of the will, the human being's vocation, and the *summum bonum* as a state of affairs causally connected to the fulfillment of that vocation.[472] Here, however, we are faced with a different *a priori*: 'the necessity of punishment.'

Earlier, in the 'last murderer' passage, Kant's emphasis seemed to fall upon the murderer's deeds. The necessity that such a one be put to death was elucidated in terms of both the general requirement 'that each [have] done to him what his deeds deserve' and the somewhat startling worry about the 'blood-guilt' that would 'cling to the people' otherwise. In that earlier passage it seemed that, having failed to punish their prisoner with death, 'the people' would be susceptible to the judgment that

[469] *MdS* 6: 489-90 (231-2) (my emphasis).
[470] *MdS* 6: 490 (232). Cf., however, *MdS* 6: 335-7 (108-9).
[471] *MdS* 6: 490 n. (231-2 n.) (my emphasis).
[472] *KpV* 5: 122-32 (102-110)

they were 'collaborators in [a] public violation of justice,' where 'justice' referred, again, to the requirement 'that each [have] done to him what his deeds deserve.' This emphasis on 'deeds' and what *they* 'deserve' allowed Hill, as we saw, to deny that this text counted as an instance in which Kant was representing 'intrinsic desert' as the justifying ground of punishment. In Kant's second reference to 'blood-guilt,' however, we encounter something rather different. The 'wise ruler's' worry about 'blood-guilt' does not have immediate reference to anyone's deeds, but rather to the consequence that would 'come upon [his] land' if he were to '[grant] pardon to an evil, murdering duelist' (by way of concession, moreover, to some unnamed party or parties interceding on the latter's behalf). It is clear that punishment's necessity—or, as here, the necessity of omitting to show mercy—is not a matter (merely) of the *ius talionis*, given instances of a certain class of deed, but a matter of what is deserved by a certain class of agent ('evil,' 'murdering').

Taken together with the earlier passages on capital punishment in *The Metaphysics of Morals* we are faced with an astonishingly rigid stance. One of its key features is the sheer immediacy of the move that progresses from the judgment that a murder has been committed to the judgment that the definitive, irrevocable occlusion of the murderer's access to happiness must take place. The move in question gives the misleading impression of being an inference from ground (the first of the foregoing two judgements) to consequent (the second). But this is not what we are presented with here. As we have seen, Kant construes this immediacy in terms of the *a priori* demand of a universal practical law. We have also witnessed his insistence that the necessity of the consequent is so absolute that it must be regarded as a judgment whose execution will be realized *no matter what*—and that this gives rise to and *warrants* 'the inference to a future life.' In this way, Kant discloses that he regards what takes place there, upon the scaffold ('criminal justice'), as uniquely eschatological in character—unique among all other political phenomena.

The unconditional immediacy, in Kant's thinking, of punishment's attachment to transgression merits closer scrutiny. This will bring us to the threshold of my claim that the categorical bond and the eschatological enactment of *the law of unhappiness* are grounded in, and give expression to, the same mode of *a priori* 'combination' that Kant insists upon in the case of murder and the murderer's own death—in the

law of punishment, that is, to the extent that it is put into effect *there*, on the scaffold, primarily and most exactly.

Kant's retributivism and the immediate connection of punishment to crime

Again, the move by which Kant's retributivism progresses from the judgment that a murder has been committed to the judgment that the definitive, irrevocable occlusion of the murderer's access to happiness *must* take place gives the impression of being a kind of inference.[473] It is not, however. Kant's retributivism moves directly (immediately) from the claim that an agent 'has committed murder' (*P*) to the claim that 'he must die' (*Q*). To be sure, Kant's thinking here does not proceed immediately from *P* to *Q*, but rather, by way of *modus ponens* from $P \rightarrow Q$, together with *P*, to *Q*. The 'immediacy' to which I am adverting characterizes Kant's commitment to the idea of a warrant for *Q*, given *P*, which does not depend upon his ever showing that the conditional, 'if he has committed murder, then he must die' is itself true.

Kant's 'must' ('he must die'), then, gives expression to the unanalyzable immediacy of his commitment to the practice of putting murderers to death, the absolute proximity, the mercy-excluding immediacy so to speak, in Kant's thinking of the notions of transgression, immorality, or crime, on the one hand, and punishment, on the other. Kant's habit of thought, here, gives the impression of endorsing a theoretical move, which is his repeated affirmation of a particular way of thinking the connection between happiness and morality.

Among commentators who either take Kant to be a retributivist, or make tempered claims to that effect, or who seek to elucidate the nature of retributivism more generally, it is common to point, momentarily at least, to the immediacy of Kant's

[473] One is reminded here, of Descartes' claim that '*cogito ergo sum*,' a piece of apparent reasoning that is generally not regarded as an inference at all. Descartes himself writes that '[w]hen someone says "I am thinking, therefore I am, or I exist," he does not deduce existence from thought by means of a syllogism, but recognizes it as something self-evident by a simple intuition of the mind (Replies 2, *AT* 7:140). Kant's assertion that 'If…he has committed murder he must die' (*MdS* 6: 333 [106]) does offer a premise that, taken together with the claim that '*x* has committed murder,' allows us to infer that '*x* must die.' The first premise (the conditional) remains undefended in Kant, however. Kant really does seem to hold that to say that 'he has committed murder, therefore he must die' is not a deduction of the normative necessity of murder from the claim that a murder has been committed, but is rather something akin (at least formally speaking) to Descartes 'simple intuition of the mind.' Kant's key claim that 'there is in the idea of our practical reason something further that accompanies the transgression of a moral law, namely its *deserving punishment*' (*KpV* 5: 37 [34]) suggests something along these lines.

procedure here (and either to affirm or to deny that Kant proceeds in this way). Thus, by calling the law of punishment a categorical imperative, Kant 'impl[ies] that…*moral* grounds [for punishment] exist,' which 'he never clearly provides.'[474] Kant takes the law of punishment to be self-evident.[475] Kant's retributivism is 'grounded on a categorical imperative which is inscrutable.'[476] Kant's exemplary, retributivist position 'maintain[s] that the punishment of crime is right in itself.' But this is 'not to justify punishment.' '[R]ather, [it is] to deny that it needs any justification' and, moreover, that '[i]ts intrinsic value is appreciated immediately and intuitively.'[477] Kant insists that 'punishment is a worthy act in itself,'[478] such that 'the reason-giving connection between wrongdoing and the infliction of harm is *immediate* and *necessary*, not indirect or contingent.'[479] And the claim that 'guilt merits punishment' may be regarded as 'a primitive and unanalyzed proposition that is morally ultimate.'[480] '[D]eep retributivism,' at least, espouses the 'fundamental principle' that 'it [is] a moral necessity, independently of the consequences, that wrongdoers *ought to be made to suffer* in proportion to their offenses'—and takes this principle to be 'in need of no further justification.'[481]

The retributivist, as Bedau points out, must either appeal to 'some good end,' over and above punishment itself, 'that is accomplished by the practice of punishment,' or he must forego appeal to 'something else' altogether. If he takes the first approach, then it may be objected that his justification for punishment is not a retributivist one at all. If he takes the second, then 'his justification…is open to the criticism that it is circular and futile.'[482] This, however, is a problem that faces 'any other deontological theory,' as Moore points out. 'Retributivism is no worse off in the modes of its possible justification than any [of those].'[483] Thus, '[o]nce the de-

[474] Fleischacker, 'Kant's Theory of Punishment': 436 (my emphasis). See also Wood, *Kantian Ethics*, 214.
[475] Wood, *Kantian Ethics*, 214. See also ibid., 219.
[476] J. G. Fichte, *Foundations of Natural Right* (Cambridge: Cambridge University Press, 2000), 245. Cited in Wood, *Kantian Ethics*, 214.
[477] Benn, 'Punishment', 30. Cited in Scheid, 'Kant's Retributivism': 264.
[478] Fleischacker, 'Kant's Theory of Punishment': 438.
[479] Wood, *Kantian Ethics*, 209.
[480] Murphy, 'Kant's Theory of Criminal Punishment', 435.
[481] Hill, 'Wrongdoing, Desert, and Punishment', 311.
[482] Bedau in Moore, 'The Moral Worth of Retribution', 97. See H. Bedau, 'Retribution and the Theory of Punishment,' *Journal of Philosophy* 75, no. (1978).
[483] Moore, 'The Moral Worth of Retribution', 97.

ontological nature of retributivism is fully appreciated, it is often concluded that such a view cannot be justified. You either believe punishment to be inherently right, or you do not, and that is all there is to be said about it.'[484]

Kant takes immoral agents to be unworthy to be happy. Unhappiness is what they deserve. In the sense of 'deserve' in play here, it is not only 'fitting' that they receive this punishment, as one commentator puts it,[485] it is *necessitated a priori*. They unconditionally *ought* to be unhappy. Recall that the main point of this chapter is to explicate this '*ought*,' to show how it is embodied in Kant's 'law of unhappiness,' to expose this law's retributivist connections, and so to clarify the mode of retributivism that is in the air whenever Kant glosses 'morality' as 'worthiness to be happy.'

The 'ought,' here, may be explicated in terms of the immediacy of Kant's commitment, its unanalyzability. Why ought immoral agents to be unhappy? They ought to be unhappy because, just given that they are immoral, it is good that they be unhappy. It is what they deserve, intrinsically. Their unhappiness is not good *for* anything—it is simply required of them. It is not that the purposes of justice are inscrutable here; it has none. There is nothing more to be said.

This means, in a sense, that retributivism is not a theory at all. It marks a refusal to theorize and so to justify a particular practice. It is a stubbornly reiterated '*because*' and a repetitive indication of the malefactor that proceeds by way of pointing at her deed ('Just look at what she did!'). The retributivism expressed in Kant's claim that the necessity of punishment grounds 'the inference to a future life' gives expression to the extent of this utter silence. The rest of Kant's thinking about the punishment of criminals may move freely around this point, flowing like water around a fixed mass. Again, I do not claim that retributivism or its 'deontological nature' constitutes the whole of Kant's thinking about punishment. But this other thinking is there, too, and its presence affects other aspects of Kant's thinking about freedom and moral accountability.

In regard to murder, Kant's 'law of punishment' is the logic that constrains the inference that moves from the judgment that 'he has committed murder' to the judg-

[484] Ibid.
[485] Feinberg, 'Justice and Personal Desert'. See also Smith, 'Worthiness to Be Happy': 187 and *KdU* 5: 443 (309-10).

ment that 'he must die.' (Note that the first judgment subsumes two moments: first, the judgment that *he* is 'the one that did it'; second, that the 'it' in question is an instance of something in particular, that is, a member of the kind, *murder*.)

Kant expresses the absolute immediacy of the relationship between transgression, immorality, crime, on the one hand, and punishment, on the other, in a variety of ways. He shows, in each of these instances, that 'punishment' (*Straf*) and the *forfeiture* of happiness are one and the same. When, in the second *Critique* Kant states that 'there is in the idea of our practical reason something further that accompanies the transgression of a moral law, namely its *deserving punishment*,'[486] nothing is required to *link* the one thing (the particular transgression) to the other ('its deserving punishment'). '[E]very crime,' Kant goes on to say, 'is of itself punishable,' which means that it—precisely 'of itself'—'forfeits happiness (at least in part).'[487]

Or again, to cite another passage that we encountered above, 'every murderer—anyone who commits murder, orders it, or is an accomplice in it—must suffer death.' Why? Kant's answer is that 'this is what justice, as the idea of judicial authority, wills in accordance with universal laws that are grounded *a priori*.'[488] Again, nothing is required to *link* the particular murderous act, on the one hand, to the death that must be suffered, on the other. By murdering (ordering murder, being an accomplice in the commission of murder) the agent forfeits happiness—and here, not 'in part,' but, *as far as we can tell*, altogether. Justice demands as much—the 'idea' of justice as punishing authority 'wills' that this be so 'in accordance with universal laws that are grounded *a priori*.' Kant does not, however, offer a deduction[489] that shows that the concepts here combined are combined *a priori*, let alone that their combination

[486] *KpV* 5: 37 (34). Reading this passage, Enderlein adverts to the possibility, at least, of arguing, from Kant, for a 'connection of moral unworthiness and the infliction of physical ills [*physischer Übelszufügung*]' (W. Enderlein, 'Die Begründung Der Strafe Bei Kant,' *Kant-Studien* 76, no. 3 (1985): 304). But the author argues, even so, that Kant does not take the *moral* unworthiness of the agent to be the basis for the infliction of punishment in a political sense (ibid., 305).
[487] See *KpV* 5: 37 (35).
[488] *MdS* 6: 334 (107).
[489] By contrast, of course, Kant does offer a 'deduction' or vindication of the claim to necessity and universality (within the bounds of experience) of our basic concepts of objects (the categories) (in the first *Critique*) and, too, of the primary, forward-looking categorical imperative's claim to practical necessity and normative universality for all rational agents (in the *Groundwork*). The *Groundwork*, in particular, is a sustained argument whose upshot is that the notion of duty, and so morality itself, is not a mere *chimaera*. The notion of a duty to put murderers to death receives no such treatment. Nor does it simply benefit from Kant's demonstration concerning the possibility of categorical imperatives in general.

accords with a universally binding norm for thinking and then *forging* their 'real connection.'

Kant makes all of this very clear in his 'Theodicy' essay of 1791, writing that 'punishment in the exercise of justice is founded in the legislating wisdom not at all as mere means but as an end: trespass is associated with ills not that some other good may result from it, but because *this connection is good in itself, i.e., morally and necessarily good*.'[490] Yet again, nothing is required to *link* 'trespasses,' on the one hand, to 'ills,' on the other. Their 'connection,' their being so connected, is unconditionally (necessarily) required because it is morally 'good in itself.' In Kant's view, then, the practice of punishment is warranted *a priori*. It is disconnected from the will of anyone in particular, from the projects or interests of any particular agent, punishing or punishable, endorsed by the general will, by pure practical reason, the perfect upshot of perfect justice (or 'judicial authority') as such.[491] And yet Kant does not take it that this immediacy—this unqueried silence about aims and consequences that is so suggestive of universal necessity *a priori*—is actually embodied in the *polis*.

This concession finds expression in the *LEC*, in a passage that shows that Kant does, in general, make a distinction between the aim or justification of punishments, to the extent that the state metes them out, and to the extent that God does. 'All punishments are either deterrent or retributive.' The first he defines as 'those which are pronounced merely to ensure that the evil shall not occur.' The latter, however, 'are those pronounced because the evil has occurred.' On the one hand, punishments are 'a means of…preventing the evil,' on the other, of 'chastising it.'[492] The *LEC* then adds that '*[a]ll punishments by authority are deterrent*, either to deter the transgressor himself, or to warn others by his example.' By contrast, 'the punishments of a being who chastises actions in accordance with morality are retributive.'[493] Punishments of the latter kind express justice; the former, however, are merely pruden-

[490] *ÜdM* 8: 257 (26) (my emphasis).
[491] See *MdS* 6: 335 (108). In this connection see also Fleischacker, 'Kant's Theory of Punishment': 442.
[492] *LEC* 27: 286 (79).
[493] *LEC* 27: 286 (79) (my emphasis).

tial.[494] Here, at least, Kant's view is that princes, governments, and other embodiments of 'authority' punish for pragmatic reasons only.

Kant expresses this distinction even more pointedly, about a decade later, in a letter to J. B. Eberhard (dated 21 December, 1792):

> In a world of moral principles governed by God, punishments would be categorically necessary (insofar as transgressions occur). But in a world governed by men, the necessity of punishments is only hypothetical, and that direct union of the concept of transgression with the idea of deserving punishment serves the ruler only as a prescription for what to do… [Punishment for crime], even if its goal is merely mechanical for the criminal and setting of an example for others, is…a *symbol* of something deserving punishment.[495]

This 'direct,' but merely *ideal* 'union of the *concept* of transgression with the *idea* of deserving punishment,' in other words, is the *telos* of a particular imperative (here 'a prescription for what to do'). The actions that it commands, particular *punishments*, do not rise to the level of this goal. They cannot do so. They are, rather, *symbolic* gestures that declares that 'something,' a particular instance of transgression, by being *that* (precisely a transgression), also instantiates 'the idea of deserving punishment.'[496] The goal itself, however, would be realized only where *every* transgression—not merely legal, but ethical—was punished, met with the forfeiture of happiness, always, everywhere.

But this is just what the law of *unhappiness* commands. The 'direct' conceptual 'union,' which 'serves the ruler only as a prescription for what to do,' by commanding that the ruler obey the injunction, also assures the ruler, along with the whole body of the *polis*, that *if* we could ever identify 'the one that did it' and if it were ever possible to confirm that *what* this one did was immoral, then this same one would *deserve*, intrinsically and morally, to be punished—even if there were no further warrant for actually punishing her (then, of course, punishment would be illegal). Reflecting on this same passage from Kant's letter to Eberhard, Murphy writes, '[h]ere, Kant seems to be admitting that human society is not the kind of society, and human criminals not the kind of individuals, corresponding to the ideals of community and personality needed to make punishment as retribution legitimate.'[497] This is not wrong, but I would add that Kant means to say that we cannot *know* whether human

[494] *LEC* 27: 286 (79).
[495] Letter to J. B. Eberhard, 21 Dec., 1792, in Arnulf Zweig, ed. *Kant: Philosophical Correspondence, 1759-99* (Chicago: 1967), 199.
[496] For a cognate reading see Fleischacker, 'Kant's Theory of Punishment': 446.
[497] Murphy, 'Kant's Theory of Criminal Punishment', 437.

society and human individuals are like this. If the question pertains to the legitimacy of punishment in a particular case, that is one thing; but Kant does not deny that, *given* the antecedent ('*A* is the one that did it' and 'it' was a criminal act), then the consequent follows necessarily ('*A* deserves to be punished').

Actual instances of 'punishment as retribution' cannot be justified as such (as instances) because the judgments upon which such punishments are based cannot be shown with any certainty to be true. But Kant balks (correctly I think) at conceding this point with respect to capital punishment. The latter cannot be justified as a symbolic gesture towards an impossible-to-ascertain moral desert, while being, in fact, 'merely mechanical' and exemplary. The death of the murderer must be regarded as an actual achievement of the retributive ideal at which other punishments gesture. The murderer's death is not justified if it does not *forge*, in the ruler's practice of punishment, in a 'real connection,' there, upon the scaffold, the *a priori* combination ('union') of the 'the concept of transgression' and 'the idea of deserving punishment.' The murderer's being put to death must be an expression of the law of punishment, but only to the extent that this law appears, here, in the eschatological guise of a law that commands that the condemned person be cut off from the very possibility of happiness.

Kant does not thematize matters in this way of course; but this, I suggest, is why he makes confused (and evidently confusing) reference to the 'inner wickedness' of murderers *in particular*. Here, at least, the standard *had better* be realized: the condemned must be 'the one that did it,' wholly and simply, all the way down, independently of all others, and the 'it' that he did must be both an utterly discrete effect of his freedom and subsumable under an unambiguous description as the transgression of laws that are universally and necessarily valid *a priori*. Barring these conditions, Kant recognizes that the agent's being put to death is an abominable mistake.

If retributivism connects crime, as ground, to equal harm, as consequence, and regards this harm (just given the crime) as an end in itself, then, politically speaking, Kant's paradigm for retribution is capital punishment. The latter is unique among punishments in that no question can be raised concerning the *fit* between crime and punishment in its instances. As we saw earlier, here uniquely 'there is no substitute

that will satisfy justice.'[498] Moreover, as Kant adds with surprising confidence a few sentences later, 'one has never heard of anyone who was sentenced to death for murder complaining that he was dealt with too severely and therefore wronged; everyone would laugh in his face if he said this.'[499] Capital punishment is unique among punishments, too, in that no question can be raised about its having been undertaken with a view to the good that it would do the one that suffers it (as deterrent vis-à-vis future crimes, or as a component in some rehabilitative strategy). The one that suffers it is to all appearances obliterated, removed entirely from the *polis*. All other punishments constitute lost freedom—in the form of lost time or lost opportunity. They 'forfeit' happiness, but only 'in part.' Capital punishment, however, is constituted by the irrevocable obstruction of any further expressions of the condemned person's freedom whatsoever.

If the rest of Kant's thinking about punishment's justification is struck through with retributivism then this is because this rigid paradigm, this recalcitrant knot of uninterrogated certainty, subsumes all other instances of punishment as though it were itself the category, 'punishment,' come down from heaven. Punishments in general must be regarded as *poenae vindicativae* so that capital punishment can be. The order of priority here, however, is obscure.

I pointed out, earlier, that the sense and purpose of my claims concerning Kant's retributivism are relative to my discussion of his habit of glossing 'morality' as 'worthiness to be happy.' And I said that I was claiming no more than that this gloss represents the connection between immorality and unhappiness in a manner which—when *thematized*—turns out to be best described as a kind of retributivism. It is important to note that this claim, because it is a claim about Kant's representation of the relationship between *immorality* and *unhappiness* pertains, first, to his ethical and, indeed, to his eschatological thinking. In other words, it does not pertain, in the first instance, to his political theory. (This is not to say that it does not apply to the latter *as well*.) As we have seen, Kant takes the connection between unhappiness and immorality to be an immediately accessible, fundamental *datum*, an *a priori* conceptual 'combination' whose actualization is enjoined in some sense by practical reason, independently of any extraneous considerations. If Kant takes the connection to be ana-

[498] *MdS* 6: 333 (106).
[499] *MdS* 6: 334 (107).

lyzable in terms of some warrant distinct from the connection's own intrinsic value he does not say so. It does not follow from this, however, that he takes the connection between *crime* and state-authorized *harm*, understood in the ultimate terms, especially, of the first part of *The Metaphysics of Morals* (the 'Metaphysical First Principles of the Doctrine of Right') to be of the same sort.

In fact, I have not (and will not) address the question whether Kant really, in general, takes the notions of state-authorized harm, on the one hand, and crime, on the other, to be a pair of 'determinations *necessarily* combined in one concept' (to borrow language that we encountered in chapter 2[500]), that is, in the empirical *concept* of punishment, and made actual by the forging of their 'real connection' in the *practice* of it. I have argued only that this is the case when it comes to Kant's thinking about capital punishment, the concept of which (I argue) does combine two elements, the concept of *murder*, on the one hand, and the notion of the murderer's *deserved death*, on the other, necessarily, and effects their 'real connection' upon the scaffold. *There* at least, I argue, the political encounters and even coincides with the ethical. My point is that even if no other part of his thinking about punishment does so, Kant's endorsement of capital punishment expresses or reflects the stubborn knot with which this thesis is concerned.[501]

The law of unhappiness

As we saw in chapter 2, the notions of merit and reward (or desert thereof) belong, for Kant, to the 'doctrine of right'; they do not belong to the 'doctrine of virtue,' unless specially qualified. The notion of worthiness to be happy, by contrast, belongs to the latter and not to the former. There is a sense, however, in which the notion of *unworthiness* to be happy belongs to both. The agent who is *unworthy* to be happy

[500] I refer here, again, to the notion of synthetic *a priori* 'combination' that I deployed in my discussion of Kant's second *Critique* account of the actualization of the 'real connection' of virtue and happiness in the highest good (i.e., by a special process of physical causation). I argued there that Kant very nearly elides the possibility that the necessary, universal ground-consequence 'connection' between virtue and happiness (two 'determinations' that are 'necessarily combined in [the] one concept [of the highest good]') is one that would have to be forged, either in adherence to, or under the limitation of, a categorical imperative. (Of course, for Kant, categorical imperatives are only objectively valid, hence, indeed, possible, if synthetic *a priori* knowledge, in general, is possible.)

[501] Thus Hill may be permitted his very detailed and articulate account of a number of possible 'extreme' positions that involve '*a deep retributivist* ground for punishing and for meting out punishments' along with his denial that 'Kant claimed…[any] of these as a comprehensive basic principle' (Hill, 'Kant on Punishment': 306).

deserves unhappiness. If omission of the action by which she discloses that she is unworthy to be happy can be coerced, then its commission is a crime and she can be punished for it. If such an action's omission cannot be coerced, then its commission, though a moral failure, cannot be punished by the state. In Kant's eschatological perspective, however, the notions of unworthiness to be happy and desert of punishment coincide. In that framework, the law of punishment, read as a command that immoral agents be made unhappy, and the law of unhappiness, understood as the injunction that they be punished, are one and the same.[502]

On Kant's account, the grounds for eternal punishment are not distinct from the grounds for empirical judicial punishment *in cases of murder*. To demonstrate the objective validity of the former mode of judgment would be, *a fortiori*, to show the validity of the latter: the judgment of condemnation declares the same thing to be true. In other words, only an agent whose eternal ejection from the human community can be a normatively necessary consequence of the use of her freedom, in general, is an agent that can be subject to the death penalty, consequent upon a particular use of that freedom.

A number of other commentators have taken note of the ambiguity that I am exploring here. In general, however, they take the ambiguity to arise from a modeling that moves from the ethical and the eschatological, as image, to the realm of the legal and the political, as reflection (to some degree, in a confused manner, etc.).[503]

Am I making the mistake of trying 'to draw a "Kantian" theory of law and politics from Kant's ethical theory'?[504] I am not. I have nothing substantive to say, really, about a 'Kantian,' or any other 'theory of law and politics.' I claim only that in his thinking about capital punishment, Kant himself touches a boundary at which the directionality of influence—from 'law and politics' to 'ethics,' or vice versa—becomes unclear. And I suggest that the immediacy of the connection between

[502] Thus, to take one illustrative example of the ambiguous merging of categories, 'sin (evil in human nature) has made penal law necessary (as if for slaves)' (*Streit* 7: 43 [268]).
[503] See, for example, Alain Badiou, *Ethics: An Essay on the Understanding of Evil* (London: Verso, 2001), 8; Fleischacker, 'Kant's Theory of Punishment': 446-7; Gibbs, 'Fear of Forgiveness'.
[504] An approach deplored by Wood and perpetrated, he claims, by much Anglophone Kant commentary on this topic. See Wood, *Kantian Ethics*, 213. For a good overview of the place of *The Metaphysics of Morals* relative to the rest of Kant's critical period thinking, the relationship between the two parts of that work (on the Doctrines of Right and Virtue, respectively), and so, too, of the relationship between ethics and politics in Kant see Allen Wood, 'The Final Form of Kant's Practical Philosophy,' in *Kant's Metaphysics of Morals*, ed. Mark Timmons (Oxford: Oxford University Press, 2003).

ground and consequence in the case of capital punishment then functions as a paradigm for the justificatory style that Kant adopts in regard to punishments more generally, and that this is the reason why it is hard to get a clear answer from him on this very topic.

'Sometimes Kant asserts retributivism in general moral or religious contexts,' Wood observes. But, he adds, '[t]hese passages cannot possibly be read as statements from within a legal practice.'[505] This is fair enough. I do not take the retributivist orientation that is implicit in Kant's habit of glossing 'morality' as 'worthiness to be happy,' to be an orientation that is structured by an interest in 'legal practice.' Nevertheless, I claim that Kant's habit expresses the same commitment that is expressed by his remarks about capital punishment in *The Metaphysics of Morals*. Wood makes his point (Kant's retributivism as expressed in 'general moral or religious contexts' cannot to be seen 'as statements from within a legal practice') in the course of a critique of commentators, however, who take Kant to be a retributivist *only* when he is 'writing and lecturing…from within the practice of punishment.' And Wood denies that Kant ever speaks from this perspective.[506] My claim, however, is that Kant's remarks about capital punishment in *The Metaphysics of Morals* are not instances in which he is 'writing and lecturing…from within the practice of punishment.' Kant's remarks about the necessity of putting murderers to death are not a *theoretical* activity that takes place 'from within' that practice at all. They are *habitual*, just like his gloss, and evince the same commitment (they are marked by the practice, continuous with it, also practical).

Wood concedes, in any case, that 'Kant does try to link the idea that the good will is a condition of worthiness to be happy with his retributivism about punishment.' But this, Wood argues, 'is chiefly (or even exclusively) when he is thinking about God as a judge of the world.' And, adds Wood, 'Kant even sometimes represents God's proportioning of human happiness to worthiness as the doing of punitive justice.'[507] This, I suggest, is just what the law of unhappiness commands. I will take up the question of *who* it commands—and so, too, its 'action-guiding' significance, in chapter 4.

---

[505] Wood, *Kantian Ethics*, 213.
[506] Ibid.
[507] Ibid., 221.

The idea of eternal punishment, the practice of capital punishment

One of my main claims is that Kant's juridical notion of (negative, objective, intrinsic) desert of punishment coincides at a certain point with his ethical notion of unworthiness to be happy. I also claim that both notions supervene on the practice that they rationalize and that Kant's commitment to this practice constrains him to some of his key theoretical moves, up to and including moves belonging to his theory of radical evil. Kant places morality and so, too, worthiness to be happy, at the centre of our attention. But he does so in a manner that obscures our view of the thing that is at the centre of *his* attention: the (today still-ongoing) problematization of the notion that it is possible for a human being to deserve to die at the hands of her community (or one of the latter's avatars).

The priority of Kant's commitment to the idea of an unconditionally punishable agent (i.e., an agent that is punishable irrespective of the fact that nothing is gained by punishing her) is embodied in his commitment to the idea of an agent that is able to be worthy of happiness, not because worthiness is equivalent to deservingness (it is not), but because, in one special case, the unworthiness to be happy of every radically evil one of us is disclosed as desert of the ultimate punishment. We are not all murderers, but we are all agents who, just in case we commit murder, will have put one and the same thing on display. Cohen thinks that, when he 'speak[s]…of making the punishment proportional to the internal wickedness of the criminal' Kant 'forget[s]…that no human being can determine the internal wickedness of another.'[508] It is true, here, that Kant 'forgets,' but again only with respect to the scenario in connection with which he speaks of this 'internal wickedness' in the first place. This is not a general pattern in Kant's thinking about punishment.

Divine *poenae vindicativae* and unworthiness to be happy as intrinsic desert of unhappiness

We have already witnessed the transition from the political to the eschatological that occurs at the end of *The Metaphysics of Morals*. Again, I do not claim that Kant's thinking about politics is modeled on his thinking about eschatology; but nor do I claim, exactly, that his thinking about the latter is modeled on his thinking about po-

[508] Cohen, 'Critique of Kant's Philosophy of Law', 286-7.

litical life. The model for each (in its backward-looking sense) is, in a sense, the practice of capital punishment: the obliteration of 'the one that did it.' The imperative to do *that* is the law of punishment, in one frame of reference, and the law of unhappiness in the other.

If 'the necessity of punishment' warrants 'the inference to a future life,' it also warrants a particular set of theological claims. Thus, in the first *Critique*, Kant observes that 'everyone also regards the moral laws as *commands*.' Everyone *does* this, Kant says, and then he says that the moral laws could not even *be* commands, in his sense, 'if they did not connect appropriate consequences with their rule *a priori*, and thus carry with them *promises* and *threats*.'[509] Next, he adds a condition on the foregoing: the moral laws could not 'connect appropriate consequences' with their rule *a priori* 'if they [these laws] did not lie in a necessary being, as the highest good, which alone can make possible such a purposive unity,' i.e., by *distributing* happiness and unhappiness and so carrying out these threats and fulfilling these promises.[510] More specifically, as the later *LEV* (1793-4) puts it:

> [W]hen once we acknowledge our act to be worthy of punishment, we straightway think of someone who has the authority to punish us, and for this reason, and because we cannot punish ourselves for our offences, there naturally follows the idea that we think of God as the moral judge, who will deal out evils appropriate to our own unlawful actions, as to those of other men.[511]

The *Opus Postumum* adds a significant dimension here, too, by adverting to moral self-contempt as the affective counterpart of the agent's acknowledgement that she is unworthy to be happy. Her unworthiness to be happy, Kant writes, is 'the transgressor['s] own reprehensibility.' It is disclosed to the subject 'in' the categorical imperative's 'rigorous command of *duty*,' by way of her failure to conform to the latter. Then, Kant says, 'if abstraction is made from sensible appearance, not only is the transgressor's worthiness of being happy [*diese Würdigkeit*] denied him, but he himself [is] condemned through an irrevocable verdict (*dictamen rationis*). Not technical-practical but moral-practical reason absolves or condemns.'[512]

---

[509] *KrV* A811/B839. See also the more or less contemporaneous *LMM* 29: 777 (133). Cf. Suzanne M. Uniacke, 'Responsibility and Obligation: Some Kantian Directions,' *International Journal of Philosophical Studies* 13, no. 4 (2005): 468.
[510] *KrV* A812/B840. See also Fleischacker, 'Kant's Theory of Punishment': 434.
[511] *LEV* 27: 555 (310).
[512] *OP* 21: 13 (221).

This late reference to 'moral-practical reason' as the source of absolution and condemnation is a key component of Kant's thinking about the rules in accordance with which God judges. Kant is being inconsistent on this score when, in the first *Critique*, he claims that '*promises*' (of reward) are 'connected' with the laws *a priori*.[513] The 'threats' are; the 'promises,' however, are not. In his lectures on the philosophy of religion Kant is more careful—at least some of the time. There he is criticritical of what he takes to be the typical view of God's justice, where the latter is 'divided into *justitiam remunerativum et punitivam*, according as God punishes evil and rewards good.' He denies that divinely ordained 'rewards' proceed from God's justice at all. These are expressions of God's benevolence and we have no 'right to demand them.' God is not 'bound to give them to us.' They do not express justice to the extent that '[j]ustice gives nothing gratuitously,' but rather 'gives to each only the *merited* reward.' Kant could not be clearer: 'Human beings may certainly merit things of *one another* and demand rewards based on their mutual justice; but we can give nothing to God, and so we can never have any right to rewards from him.'[514] In short, God owes 'no *justitiam remunerativam* toward us' and 'all the rewards he shows us must be ascribed to his benevolence.'

God's 'justice,' however, '*is* concerned…with punishments.'[515] Moreover, by 'punishment' Kant means '*poenae vindicativae*' or retributive punishments, grounded in the mere fact *that* an agent broke the law (punishment given '*quia peccatum est*'). '[W]e see that there must be *poenae vindicativae*,' Kant writes, 'because they alone constitute what is proper to justice.'[516] Deterrent punishments (punishments given '*ne peccatur*'), whether corrective ('*poenae correctivae*') or exemplary ('*poenae exemplares*'), must be 'grounded on *poenae vindicativae*.' Kant affirms here, again, that 'an innocent human being may never be punished as an example for others.' But neither, however, may a guilty one be punished just in order to set such an example. An agent may be punished only on condition that 'he deserves the punishment himself.'[517]

---

[513] Cf. *R* 6317a 18: 632 (373-4); *Vorlesungen-Religionslehre* 28: 1098-9 (427).
[514] *Vorlesungen-Religionslehre* 28: 1085 (417).
[515] *Vorlesungen-Religionslehre* 28: 1086 (417) (my emphasis). But Kant is inconsistent here, too (see *Vorlesungen-Religionslehre* 28: 1084 [416]).
[516] *Vorlesungen-Religionslehre* 28: 1086 (418).
[517] *Vorlesungen-Religionslehre* 28: 1086 (418).

Even if benefits accrue to punishment, 'all [divine] corrective punishments' must be regarded as '*avenging* punishments.' Kant is aware, however, that the idea of vengeance 'always presupposes a feeling of pain impelling one to [reciprocal acts of revenge].' If 'vengeance' is associated with God, this absolutely may not have reference to 'feeling,' 'pain,' or 'impulse.' It might be safer, Kant says, to regard 'the punishments inflicted by divine justice on sins' as a mode of distributive justice ('*justitiae distributivae*'). God is that rational being that distributes rewards and punishments, but the criterion for the latter is such that not rewards, but only punishments are distributed *necessarily* in accordance with justice. Here, as elsewhere, Kant sees this ultimate distribution of punishment in negative terms. Distributive justice 'limit[s] the apportionment of benevolence' in accordance with '*the laws of holiness*.' But even here Kant is compelled to return to the earlier, more basic idiom. He does this so as not to betray its ostensibly rational core. Again, he says, 'we see that there must be *poenae vindicativae*, because they alone constitute what is proper to justice.'[518]

Ethical-eschatological punishment

Earlier in these same lectures, a certain enthusiasm and good nature takes the upper hand in Kant's thinking, when he paints a picture of

> the human race [as] a class of creatures which through there own nature are someday to be released and set free from their instincts [and so too, in this context, from evil]…. The whole is *someday to win through to a glorious outcome*, though perhaps only after enduring many punishments for their deviation.[519]

Obviously, upon reflection, Kant cannot take this 'whole' in such a way that it includes every one of us. The subject(s) of the 'glorious outcome' and the object(s) of the 'many punishments' that he mentions can hardly turn out to be one and the same. Only a few pages later, Kant's thinking evinces a much less hopeful perspective.

> Moral perfection in this life will be followed by moral growth in the next, just as moral deterioration in this life will bring a still greater decline of morality in that life…. [One] has no reason to believe that a sudden reversal will occur in the next life. Rather, the experience of his state in the world and in the order of nature in general gives him clear proofs that his moral

[518] *Vorlesungen-Religionslehre* 28: 1086 (418). For Kant's official, carefully qualified, endorsement of '*Rache*' see *MdS* 6: 460 (207-8). See also *LEV* 27: 688 (417). Cf. Smith, 'Worthiness to Be Happy': 183).
[519] *Vorlesungen-Religionslehre* 28: 1079 (412).

> deterioration, and the punishments essentially necessary with it, will last indefinitely or eternally, just as will moral perfection and the well-being inseparable from it.[520]

Add to this Kant's insistence that there is *no* sense in which these divine (or any other justified) punishments are grounded in benevolence. Divine justice does 'not ordain punishments…in order to teach.' It has no other aim than 'to punish the offense by which [an agent] has violated the law,' to punish a violation, altogether imputable to him, through which the agent *himself* has 'made himself unworthy of happiness.'[521]

Because of their evil disposition, Kant says in the *Religion*, 'every human being has to expect *infinite* punishment and exclusion from the Kingdom of God.'[522] In that text, Kant describes punishment as 'satisfaction [that] must be rendered to Supreme Justice, in whose sight no one deserving of punishment can go unpunished.' The fact that an agent has freely opted for the changed disposition that is consequent up a radical, self-effected conversion of the will, does not change the 'moral' fact of her still intact 'punishability.'[523]

The 'Concluding Remark' of the 'Conclusion' of *The Metaphysics of Morals* seems to set forth a *caveat* here, however, in relation to any possible discourse concerning God's relation to human beings. '[I]t is clear that in ethics,' Kant writes,

> as pure practical philosophy of internal lawgiving, only the moral relations of *human beings to human beings* are comprehensible by us. The question of what sort of moral relation holds between God and human beings goes completely beyond the bounds of ethics and is altogether incomprehensible for us. This, then, confirms what was maintained above: that ethics cannot extend beyond the limits of human being's duties to one another.[524]

Questions about the kind of 'moral relation' that 'holds between God and human beings' go, as Kant says, 'beyond the bounds of ethics.' But it is not entirely clear what limit he means to effect by claiming that this 'question' is 'altogether incom-

---

[520] *Vorlesungen-Religionslehre* 28: 1084 (416). For a more detailed, 'critical' working out of the idea, specifically, of 'the immutability of one's disposition in progress toward the good' see *KpV* 5: 123 n. (103 n.).
[521] *Vorlesungen-Religionslehre* 28: 1086-7 (418).
[522] *Rel* 6: 72 (113).
[523] *Rel* 6: 73-4 (113-14). Kant's rather unsatisfying solution, which need not delay us here, is an account of the convert's 'punishment' that is 'adequately executed in the situation of conversion itself.' Kant 'think[s] [i.e., conceives of] that situation as entailing such ills as the new human being, whose disposition is good, can regard as having been incurred by himself (in a different context) and, [therefore], as *punishment* whereby satisfaction is rendered to divine justice' (ibid.). Gibbs points out that 'Kant himself is not satisfied with this solution [the *Religion*'s rational doctrine of self-atonement], because while this atonement argument appears in the *Religion*,' the *Tugendlehre,* which was a later publication, expresses the view that 'past sins cannot be erased through continuing improvement' (Gibbs, 'Fear of Forgiveness': 330).
[524] *MdS* 6: 491 (232).

prehensible for us.' Beyond the 'critical' limitation that turns such questions back at the boundary separating the theoretical from the practical, Kant certainly does not mean to say that nothing can be *said* about the 'moral relation' in question. In any case, if the 'question' takes us 'beyond the bounds of ethics' it does so by tarrying at the differential moment of transition and creating an opening in this boundary. *There*, as I will show in chapter 4, it turns out, *not* that God and human beings stand in a 'moral relation' analogous to 'the moral relations of *human beings to human beings*' in any sense of 'moral' that pertains to morality's primary, forward-looking concerns, but rather that human beings, together with God, must submit *together* to the secondary, backward-looking, categorical law of punishment/unhappiness and that, to this extent, they are members in common of the single community for which this law is constitutive—called to submit, in common, and tempted, together, by the allure of mercy.

The law of unhappiness is a categorical imperative

On Dews reading, Kant holds that 'glaring discrepancies between virtue and happiness…mar our world,' and that 'morality demands [that these] should be overcome.'[525] As I have shown, this is not quite right. The pair, 'virtue and happiness,' does not name a 'demand' that is set by 'moral-practical reason,' just as such. First the latter commands. And then, as we saw above, it 'absolves or condemns.'[526] But then, after that, it commands *again* concerning immorality and unhappiness. And the law of unhappiness is a categorical imperative. If the law of punishment is a categorical imperative, then punishment's condition, which is crime, must necessitate it internally. The same is true of the law of unhappiness. Kant says that punishment is connected with transgression *necessarily* as the latter's inexorable consequence; any other kind of connection is contestable. The same is true of unhappiness and immorality.

Now, while I use the term, 'law,' in my expression, 'the law of unhappiness,' in a manner that is meant to be understood, univocally, as a term that signifies the same notion that is put forth in Kant's expression, 'the law of punishment,' Kant does not *himself* ever speak of a law of unhappiness in my sense of an eschatological, or pro-

[525] Dews, *Idea of Evil*, 22.
[526] *OP* 21: 13 (221).

spective ‘extension’ of his law of punishment. This fact—and an associated worry about being misunderstood—led me to place the word ‘law’ in this thesis’ title’s reference to ‘the “law” of happiness’ in scare quotes. I would have preferred, in fact, *not* to take this approach (and do not generally do so in the body of this work), since this tends to dilute the force of my claim that some of Kant’s basic commitments really do suggest that, in a mostly unthematized manner, he is committed to the existence, coherence, and bindingness of such a law.

Nevertheless, I do not think that there is, or could be, such a thing as a genuinely Kantian (i.e., categorical, moral) ‘*law* of unhappiness.’ I do not think that the ‘law’ of unhappiness could qualify as such a law, anymore than Kant’s so-called ‘law of punishment’ could, since what the supposed ‘law’ commands in both cases is not that an empirical state of affairs (being-punished, being-made-unhappy) ‘*per se*’ (*qua* empirical) be brought to pass, but that a particular *notional* relationship, the one that supposedly holds, *a priori*, between immorality and unhappiness, or crime and punishment, be actualized or instantiated. It is important to note that I would make the same claim (regarding each one’s possible failure to really *be* a ‘law’ in Kant’s sense) concern both the ‘laws’ of *punishment* and *unhappiness*. If I signal my doubt in this regard by referring to ‘the “law” of unhappiness’ (with the scare quotes in place), I could do the same by referring to ‘the “law” of punishment’—and for the same reason. In neither case, is it obvious that the (supposedly normative) association in question (crime-ought-to-be-punished, immorality-ought-to-be-met-with-unhappiness) is given to us *a priori*, rather than *a posteriori*, by way of nature, or in the form of habits of judging socially acquired.

Indeed, it is my view, more generally (a view that I will confess to holding, but will not defend here), that there are no such things *at all* as the kind of imperatives that Kant calls ‘categorical’ and that—as Kant fears—‘morality’ really is a ‘*chaemera*’ (but who needs Kant’s ‘morality’ anyway?—only retributivists like Kant, as far as I can see). Again, I do not need to press this point, however, since arguing that the law of punishment is not really a law in Kant’s sense, but rather a practical ‘rule’ lies outside the scope of my thesis. All that I am claiming, here, is that Kant *asserts* as much.

But this is a really bald assertion on Kant's part. As far as I can see, the punishing, or making-unhappy, of criminals and/or immoral agents is an undertaking whose justification is possible only *a posteriori*, with reference to the empirical ends served by the practice, rather than with reference to any supposedly pure interests of practical reason. And yet Kant's use of his 'worthiness to be happy' idiom, which gives expression to certain of his deep practical commitments, has the theoretical upshot, never thematized by *him*, that punishment is an undertaking whose justification is possible *a priori* (that is what Kant means when he asserts, baldly, that the law of punishment is a categorical imperative). This upshot remains implicit, as far as its eschatological context goes, but comes out into the open in his comments about 'the law of punishment,' to the extent that these comments belong, specifically, to his thinking about *capital* punishment for murder.

That is all that I claim. There is a danger that I have caused confusion by calling this law 'the law of unhappiness,' which might sound to some readers too much like the converse of a kind of 'law of happiness,' which would be obviously an entirely un-Kantian construct. Perhaps, as I said earlier in this thesis, 'the law of the-making-unhappy-of-immoral-agents' would be a clearer rendering of what I intend. Nevertheless, trusting that the foregoing (along with my discussion of this matter in this thesis' introduction) clarifies matters adequately, I will maintain my current idiom—and proceed.

Now, the context of the law of unhappiness has an enormous breadth, for Kant. In his essay on 'The End of All Things,' Kant remarks that Jesus Christ 'announces *punishments*,'

> as a loving warning, arising out of the beneficence of the lawgiver, of preventing the harm that would have to arise inevitably from the transgression of the law (for: *lex est res surda et inexorabilis* [the law is deaf and inexorable]. Livy.); because it is not Christianity as a freely assumed maxim of life but the law which threatens here; and the law, as an unchanging order lying in the nature of things, is not to be left up to even the creator's arbitrary will [*Willkür*] to decide its consequences thus or otherwise.[527]

The law of unhappiness is binding for a universe of beings that subsumes even God. In general, as Kant explains in *The Metaphysics of Morals*, '[t]hat action is *permitted* (*licitum*) which is not contrary to obligation; and this freedom, which is not limited

[527] *Ende* 8: 338-9 (230-31). No one—not even God—may 'declare [vice or sin] less punishable than they are' (*Vorlesungen-Religionslehre* 28: 1074 [408]). Moreover, however, God may not refrain from punishing what pure reason declares punishable (see chapter 4).

by any opposing imperative, is called an authorization (*facultas moralis*). Hence it is obvious what is meant by *forbidden* (*illicitum*).'[528] It is clear enough what is forbidden here. Under the limiting law of punishment, the sovereign is permitted beneficently to reward a subject for the latter's (juridically) meritorious deeds, commanded categorically (hence morally) to put him to death in case he turns out to be, for example, an 'evil, murdering duelist', and (morally, but not juridically) forbidden to show mercy. Under the limiting law of unhappiness, the 'creator' is granted, in effect, the same authorization and made to submit to the same command and proscription—with respect to each creature's happiness or unhappiness, as such. If the law of punishment is a categorical imperative, I argued, then to punish is a 'duty of wide obligation'[529] for the 'ruler.' It is not illegal for him to show mercy, but it is unethical. What of the law of unhappiness?

Kant takes the judgment that all immoral agents ought to be unhappy to be grounded, *a priori*, in a principle of pure practical reason to which God is also subject (as we shall see in chapter 4, Kant cannot regard God as 'holy' in this connection at all). Here again, the necessity that immoral agents be unhappy must have immediate and exclusive reference to each one's being 'the one that did it.' Ultimately, the 'it' in question, for Kant, will turn out to be the unique deed/fact/law that, in the *Religion*, he identifies with the agent's irreducibly radical nature, the disposition (*Gesinnung*) of her power of choice (*Willkür*).

For the moment, it is important to see that Kant does not hold merely that immoral agents *deserve* to be unhappy. He needs to block access to any account of the 'ought' that is implicit in this claim which, while including this concession, would nevertheless *deny* that this desert (i.e., unworthiness), just as such, entails that such an agent's unhappiness is objectively *good*, or deny that this desert absolutely demands that God secure the unhappiness of human agents (i.e. that mercy is *evil*), or deny that human beings ought, unconditionally, to contemplate its hypothetical actualization with an attitude of prospective approbation (see chapter 4).

[528] *MdS* 6: 222 (15).
[529] See *MdS* 6: 388-91 (152-4).

Kant says in the first *Critique* that 'the true original' of virtue (or morality, or moral goodness) is '[un]changeable with time and circumstances.'[530] Kant thinks that the same is true of the idea of worthiness to be happy. We know that the representation of the relationship between morality and happiness encapsulated in it is valid, even though we never encounter it, as such, in experience: its content 'rest[s] on mere ideas of pure reason and [can] be cognized *a priori*.'[531]

Kant never shows, however, as he attempts to do, at least, with respect to the concept of duty or moral obligation, that his understanding of the relationship between morality and happiness—as encapsulated in his habit of glossing 'morality' as 'worthiness' to be happy—is 'cognized through reason.' As I pointed out in chapter two, he does not demonstrate, in other words, that reason, rather than nature, or habits of judging socially acquired, is the source of this idea.

As I said earlier, Kant offers no 'deduction' on this score. And, I suggest (asserting, I think, what is obvious), the claim that the relationship in question holds is *not* 'an a priori synthetic practical proposition'[532] whose validity, whose demand on our thinking and acting, can be established *a priori*. It is a claim whose *appearance* of validity is forged and secured in the course of human practices whose inevitability—from the point of view, precisely, of mercy—is far from obvious.

For Kant, however, the products of the human being's free agency (i.e., her deeds, the radical disposition of the will) and her wellbeing or ill (i.e., as states of the affairs of a being whose wellbeing must be conceived under the empirical rubric of 'happiness') are always to be regarded as constituents of a *single* system, not logically (analytically), but normatively put together: a deferred, eschatological possibility, on the one hand, but also realized, in exceptional cases, upon the scaffold (on the assumption, of course, that the relevant judgments of action-ascription and guilt are true).

---

[530] *KrV* A315/B371-2.
[531] *KrV* A806/B834.
[532] *Gr* 4: 420 (30). Kant is speaking here, of course, of the 'categorical imperative or law of morality.' See also ibid., 4: 440, 454.

Conclusion

In this chapter I executed three main tasks. First, I explored Kant's thinking about punishment and affirmed that, at least in a qualified sense, when it comes to the specification and justification of punishments, Kant is a retributivist. Next, I showed, with particular reference to his treatment of the topic in *The Metaphysics of Morals*, that Kant's thinking about the 'narrow,' 'external' domain of 'public law,' on the one hand, and his thinking about the 'wide' domain of ethics, on the other, interpenetrate one another in his treatment of capital punishment.

I argued, moreover, that the two domains themselves coincide in the practice of it. Indeed, I characterized Kant's 'scaffold' as the liminal *topos* in which his thinking about law and politics punctures and extends deep into his thinking about ethics and eschatology. I argued, in particular, that the unconditional, immediate necessity that Kant ascribes to capital punishment in cases of murder is key to understanding, not only the retributivist tendencies of his thinking about politically situated punishment, in particular, but the retributivism of his eschatological notion of unhappiness, more generally.

Finally, I argued that the 'ought' arising from Kant's implicit conviction that immoral agents deserve intrinsically to be unhappy may be expressed in the form of an imperative that is binding, not only on the agent tasked with temporal punishments (the sovereign), but on God himself. I characterized this imperative as the ethical and ultimately eschatological expression of Kant's political 'law of punishment' and referred to it as his 'law of unhappiness.'

This is the retributivism, then, that is in the air whenever we encounter Kant's gloss. It remains to be shown, however, that Kant's law of unhappiness, construed in terms, strictly, of its reference to *unhappiness*—without referring it, that is, to the exceptional case of capital punishment for murder—has 'action-guiding' significance. The plausibility of my claim that Kant's gloss evinces a deep, antecedent, durable commitment to a particular set of practices depends upon my showing that the law of unhappiness as such, and so, too, Kant's mundane uses of the worthiness to be happy idiom, have such significance.

CHAPTER FOUR

Kant's God and the Practical Significance of the Law of Unhappiness

> [A] judge who pardons is not to be thought of![533]

Introduction

As we have seen, Kant's habit of glossing 'morality' as 'worthiness to be happy' signifies his assent to the thesis that immoral agents deserve intrinsically to be unhappy. We have also seen how this converts to the view that immoral agents *ought to be* unhappy. In the last chapter I showed that this 'ought' may be expressed in the form of a categorical imperative. I referred to the latter as Kant's 'law of unhappiness.' As I showed in chapter 2, however, it is in the nature of the relationship between morality and happiness, as represented by Kant's gloss, that the unhappiness of immoral agents is only *assured* to the extent that their access to happiness is blocked extraneously, by circumstances or by a third party. I characterized the latter mode of occlusion as the 'forging' of a 'real connection' between immorality and unhappiness that conforms to their ostensibly *a priori*, normative, conceptual 'combination.'

[533] *Vorlesungen-Religionslehre* 28: 1086 (418).

In this chapter, I discuss the 'action-guiding' significance of Kant's law of unhappiness, the ostensibly *a priori* legislation that commands that this forging take place. I concede that the latter has such significance only to the extent that Kant also takes it to be binding on some agents or other. I identify the latter and show what their enactment of the imperative consists in.

To these ends, this chapter executes four main tasks. First, I open with a brief review of the practical significance of Kant's law of *punishment*. Second, I demonstrate that Kant's commitment to the thesis that immoral agents ought to be unhappy is a fundamentally practical commitment as well. Third, I argue more specifically that Kant's law of unhappiness is binding on two kinds of agent, human and divine, in two distinct contexts: one mundane, the other eschatological. I argue that these two contexts correspond to two distinct ways in which the law of unhappiness is put into practice. Fourth, I argue that these two contexts of the law's realization *coincide* to the extent that, under the law of unhappiness (but not under the primary moral law), we, together with God, are members in common of a single community, a community for which the law of unhappiness is uniquely constitutive. And I argue, finally, with respect to this community, that for Kant it must be an absolutely mercy-free zone.

The action-guiding significance of the law of punishment

In this section, I review the sense in which Kant's commitment to the thesis that all murderers ought to die (just given that they are murderers) may be regarded as a fundamentally practical one. My discussion here sets up a framework for the main inquiry of this chapter's subsequent sections.

The subject of the law of punishment

In chapter 3, although I did not construe matters in precisely these terms, I already identified the subject of Kant's 'law of punishment.' This was a relatively straightforward matter: Kant's categorical 'law of punishment' has action-guiding significance for the sovereign. The latter is its subject, the one that it binds. I will make this point again now, briefly, in a more programmatic manner. But I will also argue that there is a special sense, too, in which Kant's law of punishment has action-guiding significance for the sovereign's subjects.

In the face of a transgression of 'public law,' that is, action that obstructs or interferes with the 'external' freedom of other members of the *polis*, (proportionate) punishment is unconditionally necessary. The laws of the land are correlated with particular sanctions. However, that criminals be (proportionately) punished, *in general*, is the object of Kant's categorical 'law of punishment.' The subject of this law is the sovereign (or the state regarded as a 'moral person') and its object(s) is (are) the sovereign's subject(s).[534] With respect to the sovereign, as such, no punishment is possible.[535] It is the sovereign's unique right to punish,[536] but also—given Kant's law of punishment—his moral duty. Admittedly, the deeds that actualize what the law of punishment demands are undertaken by the sovereign's representatives (jailers and executioners), executive figures who act legitimately, not in their capacity as individual citizens, but in their authorized capacity as executor's of the sovereign's will.[537] In short, the law of punishment is not an imperative that is binding upon concrete, specific persons, *qua* individual citizens—at least, it does not command that they *do* anything.

It does, however, command a response from them. There is a sense, in other words, in which the law of punishment does have a mode of expression, for Kant, such that its subjects include the criminal's fellows. As Hill points out, one of punishment's 'constitutive element[s]…is expression of public disapproval of wrong-doing.'[538] The 'deliberate infliction of undesirable consequences in response to overt injuries'[539] is not an undertaking of any private citizen in particular, to be sure; but, to the extent that this 'public disapproval' is embodied in the affective makeup of individual citizens, it is a sign of the individual's judgment that (given crime) punishment is a good, a sign of their (impotent) will that punishment be carried out. The 'citizen's disapproval of wrong-doing' converts to an approval of particular punishments, which, as affective endorsement of them, signifies the individual citizen's will that something be done which, however, he defers to the subject actually authorized to carry it out.

---

[534] See *MdS* 6: 362 (130).
[535] See *MdS* 6: 331, 347 (104-5, 117).
[536] *MdS* 6: 328, 331 (102, 104).
[537] *MdS* 6: 460 (207-8).
[538] Hill, 'Kant on Punishment': 310.
[539] Hill's terse definition of punishment (ibid., 310-11).

Thus, while the law of punishment is normative for the realm of sovereign action, it is also normative (but of course not coercible) for the citizen's *judging* and a certain mode of *feeling* about crime and punishment. When, after asserting that '[t]he law of punishment is a categorical imperative,' Kant goes on to write, ' *woe to him* who crawls through the windings of eudaemonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises,'[540] he is addressing *anyone* who might be tempted to crawl about in these 'windings.' Both sovereign and subject are warned ('woe to him') of the consequence that threatens if each fails to endorse, to will, to approve of the criminal's full, adequate punishment: 'if justice goes, there is no longer any value in human beings' living on the earth.'[541]

Does the law of unhappiness have action-guiding significance?

In his article on 'Worthiness to be Happy and Kant's Concept of the Highest Good,' Steven Smith points to a possible 'objection…against construing virtue as worthiness to be happy, and consequently against the alleged *a priori* connection between virtue and happiness,' namely, that this construal lacks any 'practical (action-guiding) significance' and that it 'answers only to the concern of hope.' [542] Smith is onto something important here. Kant's idiomatic construal of 'virtue as worthiness to be happy' does not express the idea, merely, that there is an '*a priori* connection between virtue and happiness.' It indicates that this ideal connection, which obviously does not hold now, *ought* to hold. So understood, the problem with Kant's construing morality and happiness in this way is that it expresses an imperative in a situation where the only agents around are *incapable* of realizing the end that it prescribes. In Kantian terms, an imperative that commands what cannot be done is no imperative at all, a contradiction in terms.

In this section, I argue that, by implying that there is a categorical 'law of unhappiness,' Kant's habit of glossing 'morality' as 'worthiness to be happy' suggests indirectly, too, that someone or other is bound by it and that there is (or are) in fact a

[540] *MdS* 6: 331 (105) (my emphasis).
[541] *MdS* 6: 332 (105).
[542] Smith, 'Worthiness to Be Happy': 184.

practice (or practices) that would count as enactments of it. In other words, I argue, the 'action-guiding' significance of Kant's gloss is, at the same time, the action-guiding significance of the law of unhappiness to which it adverts. And whenever Kant glosses 'morality' as 'worthiness to be happy,' I argue, he signals his commitment to the practice (or practices) that execute(s) and conform(s) to this law.

Smith's response to the 'objection' that Kant's construal of the relationship between morality and happiness lacks action-guiding significance is to point to an ethical 'interest' that is directly correlated with the primary, forward-looking 'problem of the determining ground of the will.' For Smith, this other interest finds expression, not in a new articulation of the latter problem—in relation, now, to the question whether anyone ought to undertake to actualize the '*a priori* connection between virtue and happiness'—but instead in the 'irreducible dimension of moral experience' that consists in our 'approving or disapproving contemplation of real states of affairs.' He characterizes the latter as 'a direct corollary' of the 'formal rule of willing' and 'an exhibition and confirmation' of the latter.[543] This is not so much to affirm that Kant's construal of matters has action-guiding significance after all, however, but rather to show that the backward-looking, affective 'dimension of moral experience' in which it finds expression is directly connected to the forward-looking one. This is to affirm that Kant's mysterious construal of virtue as worthiness to be happy has a *kind* of significance for us as moral agents, to be sure, but it is also to concede the original objection. Kant is simply pointing to something that we inevitably and in some sense rightly *feel* in connection with (moral, free) success and failure in matters pertaining to the 'formal rule of willing.' But on Smith's reading, when Kant glosses 'morality' as 'worthiness to be happy' (or vice versa) he is not expressing the idea of a call to *do* something.

Like Smith, Hill notices that there is an objection to be made here. He affirms that when Kant construes virtue as worthiness to be happy he expresses the view that 'anyone who lacks a good will is, to some degree, *unworthy to be happy*,'[544] that there is some sense in which 'we are liable to suffer for our wrongdoing' or, more generally, 'that wrongdoers *deserve* to suffer.'[545] Kant appears to employ a notion of

[543] Ibid., 184-5..
[544] Hill, 'Wrongdoing, Desert, and Punishment', 324.
[545] Ibid.

desert here, however, 'that floats free from systems of law and systems of informal social sanctions.'[546] Hill, as we saw in chapter 3, refers to this as 'intrinsic moral desert.' He affirms that Kant's construal of virtue as worthiness to be happy expresses the view that human beings are, or can be, deserving in this way. But then he denies that, for Kant, 'wrongdoers *deserve* to suffer in any practical sense that entitles others to contribute to their suffering,'[547] or that Kant's use of the worthiness to be happy idiom 'amounts to an endorsement of the intrinsic desert thesis as an action-guiding (practical) principle.'[548] In essence, Hill ends up where Smith does, but without connecting Kant's idiom to the affective dimension of moral feeling. Implicitly, Smith concedes the objection's corollary claim that Kant's gloss 'answers only to the concern of hope'; and Hill allows, as we saw in chapter 3, that Kant's idea that immoral agents deserve to be unhappy has an 'anemic,' 'nonpractical,' 'faith-guiding,' 'wish-expressing,' significance.[549]

But I claim that Kant's habitual construal of virtue as worthiness to be happy *does* have action-guiding significance. It has such significance because it is tantamount to Kant's assertion that there is a *law of unhappiness* that, just like his political law of punishment, is a categorical imperative. As I showed in the last chapter, it has such significance, at least, in the special case where the sovereign is called upon to forego mercy and to put murderers to death. But it has a broader significance than that.

'That it is fitting for God to take up this task [of ensuring that wrongdoers get what they deserve],' Hill allows, 'is a faith-guiding idea that Kant sometimes seems to endorse, but it is not what grounds or determines our responsibilities.'[550] I claim however, first, that there is a special sense in which the 'idea that Kant sometimes seems to endorse' *does* determine a particular kind of activity for us. And I claim, second, that Kant's habit of glossing 'morality' as 'worthiness to be happy' also expresses the idea that it is more than merely '*fitting*' that God 'take up this task.' It is *demanded* of him. At first glance, this would seem to be out of the question, since

[546] Ibid., 325.
[547] Ibid., 324. See also ibid., 311.
[548] Ibid., 324. See also Wood, *Kantian Ethics*, 220-21 and *MdS* 6: 460-1 (207-8).
[549] Hill, 'Wrongdoing, Desert, and Punishment', 315.
[550] Ibid., 328. Cf. Beck's even more deflationary approach in Beck, *A Commentary on Kant's Critique of Practical Reason*, 245.

God, on Kant's account, is not subject to moral imperatives at all. Nevertheless, as I will show below, although he is bound by no forward-looking moral imperatives (the forward-looking laws that are imperatives for us are, for God, laws of the latter's very nature), Kant's God *is* among the agents that are bound by the backward-looking law of unhappiness.[551]

Indirectly, in spite of his concession, Smith also points to the imperative that is encapsulated in Kant's gloss. 'The criterion of worthiness to be happy,' he writes, 'does not command the creation of happiness *ex nihilo*; rather, it controls the apportionment of the well-being that happens to be possible.'[552] The agent that is subject to the 'command' is not commanded to create happiness out of nothing, but rather to ensure that happiness (whatever 'well-being happens to be possible') is distributed in a particular manner. Whatever else this entails, it entails that *immoral* agents are to have no access to happiness (or the means to it) at all. The 'criterion of worthiness to be happy' *commands* that.

Before taking this matter up, however, and before turning to the sense in which human beings are also called upon to act in conformity with the law of unhappiness, we need to be clear about Kant's thinking, more generally, about imperatives, their subjects and their objects. The first thing to note is that to affirm that the law of unhappiness is an imperative is to say that what it demands (the choice, the course of action, the end) is not merely natural for the being that must put it into effect. The law takes the form of an imperative for such a being because the latter is endowed with inclinations upon which the law itself puts 'pressure.'[553] By saying that it is a *categorical* imperative I am affirming that the agent, called unconditionally to act against the pull of these inclinations, is called to do what is objectively good.

All of this follows directly from Kant's definition of an imperative:

> The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an *imperative*.
>
> All imperatives are expressed by an *ought* and indicate by this the real relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation). They say that to do or to omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be

[551] Cf. Smith, 'Worthiness to Be Happy': 185.
[552] Ibid., 189.
[553] I owe this very evocative mode of expression to Ameriks observation that, for Kant, if 'the basic laws of value' exert 'an imperative "pressure" on us [this] presupposes a context of sensibility' (Ameriks, *Kant and the Fate of Autonomy*, 137).

> good to do that thing. Practical good, however, is that which determines the will by means of representations of reason, hence not by subjective causes but objectively, that is, from grounds that are valid for every rational being as such.[554]

'Ought,' says Kant, 'expresses a possible *action*, the ground of which is nothing other than a mere concept.'[555] In the present instance, the concept is the notion of a state of affairs in which unhappiness and immorality are perfectly correlated. '*[A]ll* imperatives,' as Paton puts it, 'tells us that something would be good to do or to leave undone,'[556] but, as Kant puts, it '[imperatives] say [this] to a will that does not always do something just because it is represented to it that it would be good to do that thing.' Where the law of unhappiness is pronounced, the object of the rational will, or 'the state of affairs which [the rational agent's] action is intended to produce,'[557] is a state of affairs whose production requires a kind of asceticism on the agent's part.

It is good that immoral agents be unhappy, thinks Kant; and, as Engstrom observes, '[Kant] identifies the good in general with the practically necessary' (in contrast, that is, to the merely 'agreeable').[558] 'The good…according to Kant,' as Paton puts it, is 'a necessary object of a rational will in accordance with a principle of reason.'[559] In the present instance, this means that, by affirming the objective goodness of the unhappiness of immoral agents, Kant indicates the practical necessity of the demand that this good be realized. But the demand that this conjunction be realized (as too the proposed—but never demanded—conjunction of morality and happiness) is, at the same time, a sign that Kant is thinking of something that is 'not yet actual,' that is first 'postulated' as practically necessary and only then achieved.[560] Kant defines a 'moral world' as 'the world,' hence presumably *this* world, 'as it would be if it were in conformity with all moral laws.' Kant adds parenthetically that this conformity is possible, on the one hand, 'in accordance with the *freedom* of rational beings' and also practically necessary, on the other hand, that is, *called for* 'in

---

[554] *Gr* 4: 413 (24-5).
[555] *KrV* A547-8/B575-6.
[556] Paton, 'Kant's Idea of the Good': iv.
[557] Ibid., iii.
[558] Engstrom, 'Happiness and the Highest Good', 112. See *Gr* 413-14 (24-6) and *KpV* 5: 58 (51).
[559] Paton, 'Kant's Idea of the Good': ii.
[560] See Hegel, *Phenomenology of Spirit*, § 602 (367). But note that, as Kant's readers generally do, he takes the 'demand' to pertain to 'the harmony of morality and happiness.'

accordance with the necessary laws of *morality*.'[561] The world as it *ought* to be, then, is an idea of reason. And, as Neiman observes of such ideas more generally, it 'is linked to reality by its claim that it ought to be realized in the future.'[562]

The subjects of the law of unhappiness and its modes of enactment

In this section I answer the questions, 'Who is the subject of Kant's law of unhappiness?' and 'How is it enacted?' I argue that Kant's God is a subject of the law of unhappiness and that human beings are too. I argue that, for the former, conformity to the law takes the form of a suspension of kindness and an occlusion of access to happiness or the means to it (irrespective of whether or not this entails a further *production* of unhappiness). For human beings, I argue, the law of unhappiness entails a class of practically necessary dispositions of the will, which find expression in attitudes (but not outer actions) that bear on the happiness and unhappiness of immoral agents (in general).

I argue, to be more precise, that on Kant's gloss the judgment that a particular agent is *moral* implies, maximally, that her being rendered happy would be deontically possible (i.e., permissible) for any being capable of acting in her behalf to actually secure her happiness and, minimally, that the (hypothetical, prospective) *approbation* of such action is deontically possible for beings such as ourselves, who are capable of willing or intending that moral agents (in general) be happy, but who are not endowed with power or insight adequate to the task of securing this end. On the other hand, the judgment that an agent is *immoral* implies maximally, for Kant, that her being rendered happy is deontically impossible (i.e., forbidden) and that her being rendered unhappy is deontically necessary (i.e., commanded) for any being capable of actually effecting or occluding her happiness and, minimally, that the (hypothetical, prospective) *approbation* of such action is either deontically impossible or necessary (as the case may be) for beings such as ourselves, who are capable of

[561] *KrV* A808/B836.
[562] Neiman, *The Unity of Reason*, 113. See also Cicovacki, 'Illusory Fabric': 394. Cf. Friedman's confused reference to 'a *causal* relationship' that, as he puts it, 'ought to prevail' (Friedman, 'Virtue and Happiness: Kant and Three Critics': 110).

willing or intending that immoral agents (in general) be unhappy, but who are not endowed with power or insight adequate to the task of securing this end.

Human beings as subjects of the law of unhappiness

In the *Metaphysics of Morals*, Kant argues that the human being is morally obligated to promote the happiness of others (to make their happiness an end)[563] and, indeed, without confusing this with the '*meek toleration* of wrongs,' to be 'forgiving.'[564] If, like the political 'law of punishment,' the ethical 'law of unhappiness' is a categorical imperative, then in what sense does this law necessitate, precisely for human beings, a determinate course of action *a priori*?

In this sub-section I argue that the first context in which the law of unhappiness is enacted is the mundane context in which human beings are called to live in accordance with the demands of morality in Kant's primary, forward-looking sense and the demands of the political law of punishment. This mundane context has both an external aspect, in so far as it is the spatio-temporal context in which the deeds of moral or immoral agents are enacted, and an internal one, which is the individual human being's conscience.

I argue that in the case of human beings (apart from the sovereign, that is, at least in the case of empirical punishment) the law of unhappiness is enacted in this 'inner' world only. The free activity that conforms to it is a particular class of judgment that finds expression in an attitude of (prospective) disapprobation of the happiness of immoral agents (in general) and in the (prospective) approbation of their unhappiness. But it is at the same time an impotent or 'virtual' *will* that immoral agents (in general) be unhappy, a kind of 'need of reason' that this turn out, somehow, to be always and everywhere the case. Deploying Kant's notion of 'respect' (*Achtung*), I argue that these attitudes of approbation and disapprobation, far from being merely 'reactive' ones, are in Kant's view the empirical manifestations of *a priori* judgments of practical reason concerning the badness of happiness and the goodness of unhappiness—given any agent's immorality.

[563] *MdS* 6: 387-8 (151-2).
[564] *MdS* 6: 461 (208).

Moral self-contempt and the disapprobation of the deeds and character of immoral agents are the affective correlates of objectively valid judgments of pure practical reason, as such (i.e., irrespective of the deity or humanity of reason's bearers). The mere representation of immorality and unhappiness as necessarily combined, the 'act' of taking an interest in the actualization of this representation, the act of representing this actualization as practically necessary—all of this is already an expression of freedom. Allison is helpful on this score, explaining that 'Kant…takes seriously the conception of practical spontaneity and therefore distinguishes between having a desire, which is a matter of nature, and being interested, which is (at least partly) a matter of freedom.'[565] To 'have' an interest in something is to spontaneously '*take*' an interest in it: interests 'are products of practical reason' for Kant. If they were not, then they would be indistinguishable from inclinations. This spontaneous 'taking,' Allison argues, 'necessarily involves the projection of some end as in some sense desirable,' whether in a moral sense or any other.[566] To be sure, as Kant recognizes, only the 'causality of the will' is able to 'bring about *the existence* of its object.'[567] Nevertheless, the law of unhappiness is normative for this antecedent activity of *taking an interest* in seeing this realized. It 'practically' necessitates the production of this representation and forbids the production of one that is formed along the contours of the 'windings of eudaemonism.'

Even if this practice and the (divine) agent capable of undertaken it must remain 'problematic,' or hypothetical for us, a practice that would gratify reason, as embodied in a subject whose justified contempt and disapprobation are stirred by the very thought of a happy but immoral agent, serves an interest, not of the kind of being for whom such contempt and disapprobation happen to be natural possibilities (i.e., the affectively modifiable human one), but an interest of all rational beings as such (including, if there were such a thing, any rational being that was capable of undertaking the practice in question). Kant's commitment to this thesis is the theoretical counterpart of his more basic commitment to these affects and these practices, of his

[565] Allison, *Theory of Freedom*, 196
[566] Ibid., 89.
[567] See *KrV* B159. See also *KpV* 5: 89 (76); Gerold Prauss, 'Theory as Praxis in Kant,' in *Kant's Practical Philosophy Reconsidered*, ed. Yirmiahu Yovel (London: Kluwer Academic Publishers, 1989), 94.

unwillingness or inability to forego *them*, to do without them—or even to imagine what it would be like to do so in a decisive manner.

Affect

Kant is struck by the fact, noticed classically by Plato's Socrates, that the immoral agent, when she suffers empirical ill, is constituted in such a way that she is able, in an important sense, to appreciate, value, and approve of the latter, precisely to the extent that she regards herself as immoral.[568] '[B]y deviating from [virtue],' as Kant points out in *Theory and Practice*, 'a human being can…bring upon himself reproach and purely moral self-censure and hence dissatisfaction.'[569] Indeed, Kant adds a few pages later, 'someone's transgression of [duty], even without his considering the disadvantages to himself resulting from it, works immediately upon his mind and makes him reprehensible and punishable in his own eyes.'[570] For Kant, the lack of fit between my happiness and my morality instigates an immediate itch in reason, as it were. Reason relieves this irritation, not by way of (unwarranted[571]) action aimed at bringing about the suffering that the agent evidently deserves, but by way of a particular mode of representation of the moral law, which gives rise to self-contempt and which, as a painful *feeling*, undermines my happiness.

Kant holds that 'observance or transgression of [one's duty] is indeed connected with a pleasure or displeasure of a distinctive kind (moral *feeling*).'[572] This even means that 'an aesthetic of morals' corresponds, as a 'subjective presentation' of its content, to the objective metaphysics of morals. This 'aesthetic' pertains to 'the feelings that accompany the constraining power of the moral law (e.g., disgust, horror, etc., which make moral aversion sensible)' and 'make its efficacy felt' (albeit without being the absolute *ground* of this efficacy).[573] The primary '*a priori* feeling'[574] is

---

[568] See, for example, Plato, 'Crito,' in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns (New York: Bollingen Foundation, 1961), 31-3, 46d-48d. Cf. Friedman, 'Virtue and Happiness: Kant and Three Critics': 101. See also Howard, 'Kant and Moral Imputation: Conscience and the Riddle of the Given': 611; Sussman, *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics*, 91-8).
[569] *ÜdG* 8: 283 (285).
[570] *ÜdG* 8: 288 (289).
[571] See, for example, Hill, 'Wrongdoing, Desert, and Punishment', 311; Wood, *Kantian Ethics*, 220-21.
[572] *MdS* 6: 221 (15).
[573] *MdS* 6: 406 (165).

respect for the moral law. It is an immediate, felt 'effect' of the latter upon finite agents. It is a product of the subject's own spontaneity, a matter of 'what we ourselves do,'[575] a mode of paradoxical self-affection and not 'something that we merely passively feel.'[576] But it may also be regarded as 'the effect of pure practical reason upon our sensuous nature.'[577] As such it allows the moral law to serve as an incentive for agents, like us, that are motivated by sensible enticements.[578]

There is a close relationship, for Kant, between respect for the moral law, on the one hand, and moral self-contentment, morally motivated self-contempt, and other-directed moral approbation and disapprobation, on the other.[579] Each of these is an affective state that relates to 'the representation of the law.'[580] Of course, respect contrasts as *a priori* and prospective with the feelings that are correlated with judgments about action and character, which are *a posteriori* and retrospective. Nevertheless, Kant's affirmation of these retrospective feelings is a further inflection of his concession of a role to affect in his view that reflection on the notion of duty modi-

---

[574] See Hinman, 'On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty': 262; Louden, 'Kant's Virtue Ethics': 486.
[575] Cf. *KrV* B152-3 and see W. W. Sokoloff, 'Kant and the Paradox of Respect,' *American Journal of Political Science* 45, no. 4 (2001): 770.
[576] *Gr* 5: 117; 98.
[577] Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue': 525-6. See *KpV* 5: 74-6, 78-9 (64-5, 67-8). See also Hinman, 'On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty': 264; Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', 9-10, 12, 22-3; Stratton-Lake, *Kant, Duty and Moral Worth*, 4.
[578] *KpV* 5: 72 (62-3, 65). Kant needs to make his claim that the moral law, or respect for the latter, is 'the sole and also undoubted moral incentive' (ibid., 67) fit with his thesis that the moral law is 'of itself and *immediately* a determining ground of the will' (*KpV* 5: 72 [62] [my emphasis]). See also *Gr* 5: 117 (98); *MdS* 6: 376-7 (142); Alexander Broadie and Elizabeth M. Pybus, 'Kant's Concept of 'Respect',' *Kant-Studien* 66, no. 1 (1975); Richard McCarty, 'Kantian Moral Motivation and the Feeling of Respect,' *Journal of the History of Philosophy* 31, no. 3 (1993); Richard McCarty, 'Motivation and Moral Choice in Kant's Theory of Rational Agency,' *Kant-Studien* 85, no. 1 (1994); R. D. Miller, *Schiller and the Ideal of Freedom: A Study of Schiller's Philosophical Works with Chapters on Kant* (Oxford: Clarendon Press, 1970), 17; Paton, *The Categorial Imperative: A Study in Kant's Moral Philosophy*, 50. But cf. Beck, *A Commentary on Kant's Critique of Practical Reason*, 215; Hegel, *Phenomenology of Spirit*, § 655 [404]; Johnson, 'Kant's Conception of Merit: 'Metaphysics of Morals' and Evaluating Actions': 330).
[579] Cf. Allison, *Theory of Freedom*, 125; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 281; Wike, 'Kant on Happiness': 83.
[580] *MdS* 6: 399 (160). For a compact overview of Kant's thinking about respect see Walker, 'Achtung in the *Grundlegung*'. See also Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*, 181; Dieter Schönecker, *Grundlegung III: Die Deduktion Des Kategorischen Imperativs*, Alber Symposion (Freiburg: Alber, 1999), 81; Jens Timmermann, *Sittengesetz Und Freiheit: Untersuchungen Zu Immanuel Kants Theorie Des Freien Willens* (Berlin: Walter de Gruyter, 2003), 17.

fies the latter.[581] That which respect is, in prospect, the feelings of self-approbation and moral self-contempt are, in retrospect. The former evinces an awareness that one has done one's duty, the latter an awareness that one has failed in this respect. One responds 'to one's own moral failure,' in particular, with 'misery.'[582]

The distinctively moral feeling of 'displeasure' take two main forms, first as 'awe,' which is 'respect coupled with fear'[583] at the mere thought of running afoul of the moral law. First, pure practical reason '*strikes down* self-conceit altogether.' This is entailed by the fact that 'all claims to esteem for oneself that precede accord with the moral law are null and quite unwarranted.'[584] This is not pleasurable at all, as Kant explains in the second *Critique*. The 'intimidating respect' that we have for the moral law is directly connected with the agent's consciousness of her own moral inadequacy before it. The human being 'want[s] to be free' from having to feel this respect because it reveals this inadequacy 'with such severity.'[585] This painful moral 'sensation' is not connected with transgression, however, but rather warns against it.

A second form of moral 'displeasure' arises when the agent 'forgets himself so far as to act' immorally. Then, Kant writes, his 'own reason bears witness against him…and makes himself despicable and abominable in his own eyes.'[586] 'Contempt' (*Verachtung*), in general, is the affective correlate of the judgment that something is 'worthless.'[587] Moral self-contempt is correlated with the reflexive judgment that one is morally worthless. 'Humiliation' in the shadow of the moral law is not equivalent to moral self-contempt. The latter, but not the former, presupposes that an agent has disobeyed the law. Moral self-satisfaction, on the other hand, is connected with consciousness of conformity to this law. And, as Kant says in the second *Cri-*

---

[581] See I. Goy, 'Immanuel Kant Über Das Moralische Gefühl Der Achtung,' *Zeitschrift für philosophische Forschung* 61, no. 3 (2007): 337; Frank Schalow, *The Renewal of the Heidegger-Kant Dialogue: Action, Thought, and Responsibility*, Suny Series in Contemporary Continental Philosophy. (Albany: State University of New York Press, 1992), 273; Sokoloff, 'Kant and the Paradox of Respect': 768. Cf. Acton, *Kant's Moral Philosophy*, 14.
[582] Langton, 'Duty and Desolation': 484
[583] *MdS* 6: 438 (189).
[584] *KpV* 5: 73 (63). See also *MdS* 6: 435 (187). See Reath's excellent discussion in Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', esp. 15-16. See also Gauthier, 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue': 528.
[585] *KpV* 5: 77 (67).
[586] *Vorlesungen-Religionslehre* 28: 1011 (356).
[587] *MdS* 6: 462 (209). See also *KdU* 5: 443 (309); *ÜdG* 8: 283 (285); *R* 7202 19: 278 (466); *R* 7315 19: 312.

tique, the human being is capable, in principle, of 'certainty of a disposition in accord with [the moral] law' and so capable of consciousness that she has met 'the first condition of any worth of a person.'[588] 'Moral happiness' is 'the assurance of the reality and *constancy* of a disposition that always advances in goodness (and never falters from it),'[589] or is 'inherent in the consciousness of his progress in the good.'[590]

Of course, for Kant, the context in which the relevant judgments and the associated affective states take shape is an internal one. Conscience hosts a scene in which the one 'I,' 'the same *human being* (*numero idem*),' as Kant puts it, is both accused and accuser (the latter being prosecutor and judge together). The two roles, however, are completely distinct.[591] The 'human being who accuses and judges himself in conscience must think of a dual personality in himself,' Kant argues, 'a doubled self.'[592] We accuse ourselves, Kant says, and we find ourselves guilty, punishable, and finally condemned.[593] But of course, as Wood observes, this implies no 'duty to punish ourselves for our misdeeds (as by depriving ourselves of the happiness of which we judge ourselves unworthy).' Rather, 'Kant…insists that our happiness or misery is left for the ruler of the world to decide.'[594]

When it comes to questions about the unhappiness of other agents, however, there is a danger that the feeling of resentment and a desire for revenge will come into the foreground. Not only Kant, but some of his main predecessors had already noticed this and attempted to offer a general warrant for the having of feelings along these lines. For Hume, moral feeling, whether as 'indignation' or 'resentment,' on the one hand, or as 'approbation' or 'approval,' on the other, is not grounded in an antecedent *thinking*, not a matter of judgment by means of moral concepts. Rather, it arises immediately through exposure to certain social states of affairs (even by way of hearsay): 'I feel an immediate indignation arise in me against such violence and injury'[595] and 'our breasts are affected with the liveliest resentment against the author

---

[588] *KpV* 5: 73 (63).
[589] *Rel* 6: 67 (109).
[590] *Rel* 6: 75 n. (115 n.).
[591] *MdS* 6: 438-9 (189).
[592] *MdS* 6: 439 n. (189 n.).
[593] The *OP* is particularly vivid: 'if abstraction is made from sensible appearance, not only is the transgressor's worthiness of being happy denied him, but he himself [is] condemned through an irrevocable verdict (*dictamen rationis*)' (*OP* 21: 13 [221]).
[594] Wood, *Kantian Ethics*, 187. See *MdS* 6: 439 n., 440, 460-1 (189 n., 190-1, 207-8).
[595] Hume, *An Equiry Concerning the Principles of Morals*, 5.21 (110)

of these calamities.'[596] If the feeling itself is warranted, then this is because it gives expression to interests that human beings actually have and which they have, moreover, *in common*. As Hume puts it

> [i]n all determinations of morality, this circumstance of public utility is ever principally in view; and wherever disputes arise, either in philosophy or common life, concerning the bounds of duty, the question cannot, by any means, be decided with greater certainty, than by ascertaining, on any side, the true interests of mankind.[597]

In this same connection, Adam Smith goes to great lengths to show that, when it comes to the treatment of malefactors, action arising from sympathy with the latter (mercy) ought never to be allowed to take priority over action grounded in resentment.[598] Like Hume and Smith, Kant takes *sympathy* (no matter how broadly construed) to be a kind of partiality that is grounded in sensibility or inclination. But unlike these predecessors, he wants to *deny* this of his analogue of *resentment*—which is moral contempt (*Verachtung*). For Kant, the latter—as expressed, for example, by the thief's judgment that 'I am a worthless man although I have filled my purse'[599]—signifies a normative judgment that is not only categorically true of a whole class of deeds (in a way that subsumes even that deed-like thing, the evil character from which they spring), but enjoys a universal validity whose 'universe' extends beyond the merely anthropological one.

When we are faced with the idea of punishment as divine '*justitiam remunerativum*,' Kant takes it that an agent's self-contempt and, with the latter, her consciousness of deserving to be unhappy, implicitly concedes the objective (practical) reality of this idea. But the concession does not pertain to her merely. It is a universally valid judgment to affirm that *all* immoral agents deserve this treatment. Kant secures the normativity of both legal and theological notions of retributive punishment within this 'virtual' format: that is, as both a free interest-taking judgment and its correlative affect. In our actual practice, with respect to others—even given the apparent 'pervasiveness' of radical evil—we are to be sympathetic and humane. We are to be 'stern with ourselves in the pursuit of moral self-perfection,' but we are not warranted in

[596] Ibid., 5.27 (111).
[597] Ibid., 2.17 (81).
[598] Adam Smith, *The Theory of Moral Sentiments*, Cambridge Texts in the History of Philosophy (Cambridge: Cambridge University Press, 2004), 190-1. On main contrasts between Kant's views and those of 'the British sentimentalists whom he admired' see Korsgaard, 'Morality as Freedom', 186 n. 21.
[599] *KpV* 5: 37 (34).

seeking to undermine the happiness of our immoral (but non-criminal) fellows.[600] Nevertheless, this humaneness aside, Kant takes it that the human being's endorsement of the unhappiness of immoral agents *in general* (her own included), is grounded in an interest of the same faculty (pure practical reason) that is concerned with answering the forward-looking question, 'What ought I to do?'—the *pure* faculty to which Kant adverts when he solves the 'problem of the determining ground of the will' by identifying the 'formal rule of willing.'[601]

But Kant never demonstrates this thesis. He never shows that this interest is more than a merely anthropological *datum*, a property of our 'common life,' a habit that we share. Kant takes it that the feeling of respect plays a mediating role between pure practical reason and sensibility. Likewise, pure practical reason mediates its backward-looking demand—the law of unhappiness—by way of the feeling of moral self-contempt, which is then generalized affectively as the prospective disapprobation of the happiness of immoral agents in general.

Indeed, in the fascinating *R* 7202 Kant claims to 'find in [himself] a principle of disapprobation [*Missbilligung*] and of inextinguishable inner aversion [*unauslöschlichen innern Abscheu*].' In a sense, he stacks the deck by referring to this as 'a principle,' from the outset. Nevertheless, Kant goes on to ask what 'this disapprobation rest[s]' upon. He cites various possibilities, including an 'immediate feeling of shamefulness,' 'hidden reflection on harmfulness,' and 'fear of an invisible judge.' He does not offer a positive answer to the question that he has posed, but insists, in any case, that *whatever* the ground of this disapprobation turns out to be '*it cannot be habit*'—an assertion that he bases in the claim that the disapprobation in question is 'universal and unconquerable.' Habit, he implies, is not 'universal and unconquerable' and so moral disapprobation cannot be founded in that.[602]

Even if we concede this, however, in relation to judgments about the moral goodness and badness of prospective or already undertaken courses of action—what of the approbation of the unhappiness of immoral agents and the disapprobation of their happiness? For Kant, nothing short of the purity and rationality of their ground

[600] P. Formosa, 'Kant on the Radical Evil of Human Nature,' *Philosophical Forum* 38, no. 3 (2007): 245. The author refers to *LEC* 27: 295-6.
[601] Cf. Smith, 'Worthiness to Be Happy': 184-5.
[602] *R* 7202 19: 280-1 (468) (my emphasis). See also *R* 7217 19: 288 (473).

vindicates 'moral sentiments.'[603] Pure and rational, then, this ground must be. Barring this originary purity, the impartial spectator's prospective disapprobation of 'the uninterrupted prosperity of a being graced with no feature of a pure and good will'[604] may readily be called into question. So too practical reason's supposedly pure, anticipatory judgment that the eschatologically deferred, divine practice of punishment, conceived in strictly retributivist terms, is absolutely *good*.[605] And so too, more proximately, practical reason's ostensibly *pure* judgment that the sovereign's merciful cancellation of the death penalty is so morally repugnant that were it made *policy*, there would 'no longer [be] any value in human beings' living on the earth.[606]

Kant is no revolutionary in these matters.[607] We are faced here, to be sure, with one tendency among several, but Kant's thinking with respect to mercy and punishment (capital, divine) is highly conservative. Even if it is true, in general, that Kant's 'narrative of progress and universal history constitutes a *delegitimation* of practices and institutions that might otherwise be taken for granted,'[608] it is not true here. And if, as Taylor puts it, a 'revolutionary project' is one that is 'put forward' as something that 'ought to supersede the status quo,[609] then Kant is no radical here. Ameriks is right to describe Kant's project in terms of a 'realiz[ation] that what he needed was…a good apology, a story of how the best examination of all the latest options of metaphysics and science…shows that there is still room for (what he took to be) our most important common beliefs.'[610] But in a sense not intended by Ameriks[611] it also turns out that the importance of these 'most important common beliefs' turns to some significant degree, at least, upon the importance of certain cherished, evidently universal, apparently inevitable *practices*.[612]

---

[603] See Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 274; Smith, 'Worthiness to Be Happy': 168-9.
[604] *Gr* 4: 393 (7).
[605] See, for example, *Vorlesungen-Religionslehre* 28: 1086 (417-18).
[606] *MdS* 6: 332 (105).
[607] Hill makes the same point, at least implicitly, in Hill, 'Kant on Punishment': 310-11.
[608] Sankar Muthu, *Enlightenment against Empire* (Princeton: Princeton University Press, 2003), 167
[609] Taylor, *Sources of the Self: The Making of the Modern Identity*, 204-5.
[610] Ameriks, *Kant and the Fate of Autonomy*, 67.
[611] Ameriks argues, in fact, that we can 'extract from [Kant's] work an attractive apologetic strategy that gives philosophy the modest negative role of primarily defending modern agents simply against philosophy itself and its ever-growing alienating effects, including its challenges to the very notion of science as a crucial and distinctive from of knowing' (ibid.).
[612] See Taylor's claim that things like 'moral ideals, understandings of the human predicament, concepts of the self' are 'embedded in practices' and 'for the most part exist in our lives' in no other way

Here, even in this 'virtual' format, we witness the priority of the practical in Kant's thinking—by which I do not mean so much the priority of practical *reason* over theoretical, but, to deploy Cicovacki's useful distinction, of the 'practical *realm'* over the theoretical one.[613] Just as Kant 'offered a detailed reconstruction of our cognitive experience,' he 'sought to reconstruct the underlying principles of our moral practice.'[614] But this does not apply to the practice with which this chapter is concerned—the 'virtual' and 'deferred' practice (see below) that conforms to the law of unhappiness. If these practices were to turn out *not* to be 'reconstructable' in the way that Kant requires, then this would be equivalent to the discovery that we have no objectively, readily shareable *good reasons* for keeping on with them (in Kant's sense of 'good reasons'). In a sense, here, the practical outstrips Kant's efforts at 're-construction' and persists more or less 'as is.'

But Kant clearly takes the *feelings* in question to be 'principled' ones—and not merely *ex post facto*, to the extent that they might turn out to have given *rise* to the very principles that he thinks they express. Kant observes, rather, that it is a special 'propensity of *reason*' in them that leads (in his example) young people 'to enter with pleasure upon even the most subtle examination of the practical questions put to them.'[615] Kant allows that

> frequent practice in knowing good conduct in all its purity and approving it and, on the other hand, marking with regret or contempt the least deviation from it, even though it is carried on only as a game of judgment in which children can compete with one another, yet will leave behind a lasting impression of esteem on the one hand and disgust on the other, which by mere habituation, repeatedly looking on such actions as deserving approval or censure, would make a good foundation for uprightness in the future conduct of life.[616]

But Kant does not conclude that these judgments of 'approval or censure' and the feelings of 'regret or contempt' and 'esteem' or 'disgust' reflect interests that human beings have independently of pure reason. And yet both morality, on the one hand, and the happiness or unhappiness of immoral agents, on the other, are practical possibilities in whose realization human beings always already take an interest, without

---

(Taylor, *Sources of the Self: The Making of the Modern Identity*, 204). His definition of 'practice' is apposite for my purposes as well: 'By "practice", I mean something extremely vague and general: more or less any stable configuration of shared activity, whose shape is defined by a certain pattern of dos and don'ts…. [I]deas frequently arise from attempts to formulate and bring to some conscious expression the underlying rationale of the patterns.' (ibid.).

[613] Cicovacki, 'Illusory Fabric': 396

[614] Ibid., 385.

[615] *KpV* 5: 154 (127) (my emphasis).

[616] *KpV* 5: 154-5 (127).

the benefit of, and well in advance of, philosophical reflection. Kant does not have to persuade us to take an interest, here, or teach us what to attend to. At the same time, however, the consistency and frequency of Kant's deployment of the notion of 'worthiness to be happy' show that he takes the practical impulse to which it refers to have become eroded in some way. Kant's notion of 'worthiness to be happy,' then, may be regarded as a tool for reconstituting and stabilizing a 'virtual' practice (as free judgment and correlated affect) whose end is the universal connection of unhappiness and moral unworthiness, a practical possibility concerning which human beings are or have become ambivalent.

Interests

Kant's habit of glossing 'morality' as 'worthiness to be happy' indicates the presence of an implicit interest, which pertains to a particular practice, a form of willing, and an affective attitude, each of which takes as its object the state of affairs in which the correspondence between immorality and unhappiness is, or would be, realized. Kant holds that it is as rational beings, as such, and not as specifically human and so also earthly ones, that we are interested in the proportionate distribution of happiness in accordance with morality. We notice that this proportionality is lacking and this fact bothers us. But Kant thinks that our irritation in this respect expresses the frustration of a need of pure reason.

This is the thrust of Kant's claim, in the *Groundwork*, that 'an impartial spectator can take no delight in seeing the uninterrupted prosperity of a being graced with no feature of a pure and good will.'[617] And in his philosophy of religion lectures, Kant puts the same point in theological terms, asserting that 'God could [never] take pleasure in seeing other beings happy without their being worthy of it.'[618] This does not get us quite as far as saying that reason is positively *pained* at the sight of such a thing, nor need it.[619] In his essay on 'Orientation in Thinking' Kant writes, in any case, that

> '[r]eason does not feel; it has insight into its lack and through the *drive for cognition* it effects the feeling of a need. It is the same way with moral feeling, which does not cause any moral

[617] *Gr* 4: 393 (7). See also *R* 6090 18: 450.
[618] *Vorlesungen-Religionslehre* 28: 1102 (430).
[619] See, however, *R* 6871 19: 187 (445) and *R* 7196 19: 270 (461).

> law, for this arises wholly from reason; rather, it is caused or effected by moral laws, hence by reason, because the active yet free will needs determinate grounds.'[620]

Generally, for Kant, there is a direct relationship between the feelings of pleasure and displeasure, on the one hand, and what it means to take an interest in some course of action or in the effect that it aims at, on the other. In *The Metaphysics of Morals*, he cites our 'susceptibility to feel pleasure or displeasure merely from being aware that our actions are consistent with or contrary to the law of duty.' And he goes on to argue that '[e]very determination of choice proceeds *from* the representation of a possible action *to* the deed through the feeling of pleasure or displeasure, taking an interest in the action or its effect.'[621] In another context, Kant avers that 'well-pleasedness is pleasure in an object,' irrespective of whether or not the object exists (e.g., being 'well-pleased with a house, even if I can see only the plans'), while an '*interest*' (*Interesse*) is 'well-pleasedness in the *existence* of an object.'[622] This brings us to the threshold of practice since 'a necessary practical postulate is the same thing in regard to our practical interest as an axiom is in regard to our speculative interest.'[623] In each case, movement *from* the axiom/postulate to the conclusion/effect serves the interest in question by establishing its object.

It is generally accepted that Kant's notion of rational interest may be specified in two distinct ways. Reason has theoretical interests, on the one hand, and it has practical ones, on the other.[624] Kant avers, however, that 'all interest is ultimately practical.'[625] The theoretical interests of reason are served by the practice of explanation, at least in part, while reason's practical interests are served, in the first instance, by action that conforms to 'the law of duty.' When it comes to the possibility of achieving the latter, pure reason's needs with respect to 'doing or acting'[626] render the claim that 'everything is mere nature' (the Third Antinomy's antithetical position[627]) untenable. When it comes to the human being's freedom, her immortality, and the existence of God, reason's practical interest demands that the reality of these things

[620] *Orient* 8: 140 n. (12 n.).
[621] *MdS* 6: 399 (160).
[622] *Vorlesungen-Religionslehre* 28: 1065 (400).
[623] *Vorlesungen-Religionslehre* 28: 1083 (415).
[624] For a typical formulation see P. Rossi, 'Kant's Doctrine of Hope: Reason's Interest and the Things of Faith,' *New Scholasticism* 56, no. 2 (1982): 229.
[625] *KpV* 5: 121 (102).
[626] *KrV* A475-6/B503-4.
[627] *KrV* A475-6/B503-4.

be postulated 'as an hypothesis,' even though it is understood that the postulates in question can be neither proven nor disproven.[628] These ideas 'have not been thought up arbitrarily,' Kant argues. They answer directly to this need or interest of reason. The projection of 'thinking' beyond the bounds of nature leads reason to these ideas necessarily. And this projection is warranted to the extent that the interests that it serves reflect a demand of pure reason itself.[629]

What, however, of the thesis 'that wrongdoers deserve to suffer'? 'To many,' Berman observes, this 'seems truer [even] than that they deserve particular treatment from a contingent entity, the state.'[630] Hegel notes critically that

> [t]he designation of an individual as immoral *necessarily* falls away when morality in general is imperfect, and has therefore only an arbitrary basis. Therefore, the sense and content of the judgment of experience is solely this, that happiness simply as such should not have been the lot of some individuals, i.e. the judgment is an expression of *envy* which covers itself with the cloak of morality.[631]

Hegel's 'judgment of experience' and Berman's reference to what 'seems truer,' here, are apposite, not because I am setting out specifically to demonstrate that something sinister is going on in Kant's thinking at this point. Instead, these observations raise the stakes in connection with my proposal that Kant does not, in any case, *show* that, in addition to its forward-looking interest in the agent's ultimate conformity to 'the law of duty,' pure practical reason, just as such, has an interest in the happiness or unhappiness of immoral agents. This is not to question whether reason is involved here, at all, but only to question reason's 'purity' on this score. But this is what Kant claims when, in the third *Critique*, he proposes that we

> [c]onsider a person at those moments in which his mind is disposed to moral sensation…. Cleverly to dig for incentives behind these feelings would be in vain, for they are immediately connected with the purest moral disposition, since *thankfulness*, *obedience* and *humiliation* (subjection to deserved chastisement) are particular dispositions of the mind toward duty.[632]

The fact that the human being freely 'takes an interest' in some end or other does not mean, however, that her having that end in the first place is a function of anything to do with reason (anymore than her taking an incentive to be a good reason for acting implies that the incentive itself is given to her by reason). For Kant, particular affective states are connected with judgments concerning the goodness or otherwise of

---

[628] *MdS* 6: 354 (123).
[629] *KrV* A462/B490.
[630] Mitchell N. Berman, 'Punishment and Justification,' *Ethics* 118, no. 2 (2008).
[631] Hegel, *Phenomenology of Spirit*, § 625 (379).
[632] *KdU* 5: 445-6 (311-12).

prospective courses of action. But it does not follow from the claim that this forward-looking approbation and disapprobation expresses an interest of *reason*, that this is so in connection with judgments concerning the goodness or badness of the happiness of the members of a particular class of agent. Even if we concede Kant's point with respect to these 'particular dispositions of the mind toward duty,' does this really tell us anything about the incentives that one might find (if one were to 'dig') 'behind these feelings,' behind 'moral sensation,' in the case of the prospective approbation of the unhappiness of immoral agents—or, say, in cool reflection on the prospect of letting 'the last murderer' go free?

Virtual and deferred practice

I propose that, when they anticipate and endorse the eschatological forging of the connection between immorality and unhappiness whose objective goodness is represented by Kant's notion of worthiness to be happy, the affects of approbation and disapprobation may be regarded as virtual or deferred forms of that ultimate practice. The disapprobation or approbation of the deeds of morally worthless or upright men is a corollary of the judgment that such men or deeds possess or lack a certain kind of goodness. It is not a deferred practice. But the disapprobation that one feels at the very prospect of an immoral, but happy agent—including a 'passively healed' one (see below) that has not repaid her debts—*is* practical. It does not merely approve of a particular deed—it adopts and (impotently) wills a particular end.

The attitudes that Kant describes, these moral 'sensations,' 'feelings,' and so forth are virtual practices. On Kant's view, the interests that these practices serve are *pure* ones. Although both 'reactive attitudes'[633] and the attitudes that Kant describes 'attribute responsibility to others,'[634] Kant's moral feelings are no mere *reactions*. Rather they give expression to moral judgments, made freely, concerning interests freely 'taken.' That is 'virtual,' in the sense that I intend, then, which *would* be the case if pure reason had a power that was commensurate with that which interests it, with what it wills, in this regard.

---

[633] The idea of 'reactive attitudes' originates in Strawson, 'Freedom and Resentment'). See also Bernard Williams, *Ethics and the Limits of Philosophy* (London: Routledge, 2006 [1985]), 36-8.
[634] See Korsgaard, 'Creating the Kingdom of Ends', 196; Langton, 'Duty and Desolation': 486; Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame*, 188-9, 274.

'[N]ot everyone can or should sustain every complaint,' notes Williams. But '[i]t is another consequence of the fiction of the moral law that this truth does not occur to us. It is as if every member of the notional republic were empowered to make a citizen's arrest.'[635] This 'arrest,' however, is deferred, by which I mean that, while it is willed and practiced virtually, it is left to another to execute. Analogously, the actual, political practice of punishment is an outworking of that which remains virtual in the spectators' approbation of it and which is deferred to the sovereign. If our disapprobation of the happiness of immoral agents were *efficient* then we would be beings who could, at least, occlude it. But our disapproval is not efficient and we are not such beings—or we are, with respect to the manner of our judging and willing, but no further than that. As finite agents operating under the conditions of space and time, we lack the *insight* and *power* that would be required to insure that no immoral agent was happy.[636] As immoral agents ourselves, too, we would also have to regard the end of our action as a form of self-punishment, which, for Kant, is a contradiction in terms.[637] Nevertheless, in spite of the impossibility of our doing what we will here, Kant thinks that we have insight *a priori* into the *principle* in accordance with which, if we possessed adequate power and insight and were ourselves morally superior to the objects of the practice (immoral agents), we would be *constrained* to proceed. This is part of what it means to say that the law of unhappiness is a categorical imperative.

Contrast Kant's implicit view, here, with his explicit treatment of the theoretical '*a priori*' of the first *Critique*. Both of these modes of aprioricity constitute a horizon within which the service of particular interests unfolds. There is no going beyond this limit, nor is their any sense in which the horizon offers clues as to its origin or the origin of these interests. On the one hand, there is that which lies *a priori* and so 'pure' in reason, as its ideals, in the understanding, as its categories, in sensibility, as the formal properties (space and time) of any possible human receptivity to affection by independently existing things. This marks out a horizon within which, for Kant, it is possible to find out, and to think about, what merely happens to be the case and

[635] Williams, *Ethics and the Limits of Philosophy*, 192.
[636] In fact, Kant allows that, empirically speaking, we cannot really establish the presence of immorality at all (see A551 n/B579 n.). See also Jacob Rogozinski, 'It Makes Us Wrong: Kant and Radical Evil,' in *Radical Evil*, ed. Joan Copjec (London: Verso, 1996), 35.
[637] *MdS* 6: 335, 485 (108, 227).

what is necessarily the case. On the other hand, there is that which is always-already of practical interest to the human being. This marks a horizon for thinking about norms, about what *ought* to be the case, about what ought to be done or left undone.

In the juxtaposition of 'theory' and 'practice,' theory is not merely a matter of what is 'intellectual,' but of what is *normative*—for both cognition and practice. But to the extent that it is normed, there is a sense in which thought may be regarded as a mode of practice.[638] The normative may be regarded as theoretical in the sense that it inflects a kind of judgment. The judgment that '*x* ought to be the case' is, after all, a knowledge claim proposing that, as a matter of fact, something (*x*) has a particular *value*.

'Normative,' here, refers to something that is related to the thinking, hoping, planning, and so forth of a particular kind of agent. The putting-into-practice of what has been judged normative does not, in this case, lie within our physical capacities. Nevertheless, we do empirically follow the law in a virtual format. 'Ought,' here, implies not 'can *achieve*,' but rather 'can *intend*,' or 'can *will*.' The same is true of the moral law. This is why the *good will* and not the good deed is the primary object of moral assessment. The moral law does not command, primarily, that one do this or that, but rather that one will a particular *class* of deeds, in a particular *manner*. In respect of the law of unhappiness, human beings must *will* that which none but 'God' can do; they must have this same leaning, but without being able to actualize it, a 'leaning' that is embodied in a certain mode of judging and feeling.

Hill remarks that 'Kant did not focus much on issues of moral praise and blame.'[639] This is a surprising contention, given what we have seen so far. But if Hill misses the point, this is in part because the point is an eschatological one. Kant does not think that, *in addition* to the concept of worthiness to be happy, we are equipped with the necessary insight for making supra-legal judgments that would ground specifically *moral* (and not merely legal) praise and blame. He only thinks that our possession of the concept of worthiness to be happy—and its interchangeability with a particular concept of 'morality'—shows that there is nothing incoherent

[638] See Prauss, 'Theory as Praxis in Kant', 94, 102. But cf. Adorno, *Problems of Moral Philosophy*, 7; Allison, *Idealism and Freedom*, 131; Dieter Henrich, 'Die Deduktion Des Sittengesetzes,' in *Darmstadt*, ed. A. Schwan (Denken im Schatten des Nihilismus: 1975), 64-70.
[639] Hill, 'Is a Good Will Overrated?', 55.

about claiming that, given such insight, or given an agent that had it, the *class* of such judgments is possible.

God as subject of the law of unhappiness

With respect to the second context in which the law of unhappiness is enacted, my task is to make explicit a point of view that is generally implicit in Kant's thinking. In this sub-section I argue that this second context is the eschatological scenario in which God is called upon to omit the mercy to which he is universally inclined (as benevolent) and to act in strict accordance with the categorically imperative law of unhappiness. I argue that the virtual and deferred practice by which human beings conform to the demand of this law, the end that is impotently *willed* by them and deferred to another more powerful being—a deferral that leaves, as its sign, the prospective disapprobation of the prospective happiness of immoral, but happy agents—can only be put into full effect by a special kind of rational agent. Kant refers to this agent as 'God.'

The term 'God' is not univocal in Kant's thinking and there are several uses to which he puts it. From among these, however, I focus on contexts in which 'God' refers to a being that, though unlike any of the human community's members in some key respects (i.e., vis-à-vis cognition and morality in its forward-looking sense), is nevertheless regarded, from the point of view of the law of unhappiness, as a member of that community and the ultimate guarantor of that community's interest in this regard.

As we saw above, when it comes to the law of unhappiness, human beings are its subjects, putting it into effect in a format that constitutes it as both virtual and deferred, as an impotent will, expressed in an affective attitude of approbation, which also surrenders the task itself to another. We are commanded by the law of unhappiness to disapprove (prospectively) of the happiness of immoral agents, forbidden therefore to approve of (wish or hope for) the latter, and permitted (but not commanded) to approve of the happiness of moral ones. 'God,' on the other hand, is Kant's name for the unique agent who—assuming there is such a being—is bound to *actualize* the state of affairs that we endorse from within the structure of our capacity for 'pure feeling.'

In his philosophy of religion lectures of the mid 1780s, Kant says that '[m]orality alone…gives me a *determinate* concept of God.'[640] Later, in the second *Critique*, Kant will argue that the concerns of morality are the source of the postulate of God's *existence*.[641] In the lectures, however, Kant emphasises morality's role as the ground of specific claims concerning God's nature, rather than the latter's existence.[642] And there is no need to argue with Kant about whether morality really *does* give you or me or anyone else 'a determinate concept of God.' It is sufficient simply to allow that *Kant's* 'determinate concept of God' is implicit in Kant's understanding of morality—which is what he claims. But recall that Kant regularly glosses 'morality' as 'worthiness to be happy.' Is Kant's 'determinate' concept of God grounded, then, in 'morality,' precisely in its secondary, backward-looking sense, as encapsulated in the notion of worthiness to be happy as well? If so, it should be possible to work backwards, reverse engineering Kant's description of this agent, to get at his thinking about 'morality' in this connection, particularly in relation to its action-guiding significance.

The priority of Kant's 'morality' over his 'determinate concept of God' is, at the same time, the dominion of pure practical reason over this concept's formation. This is so whether 'the practical' is specified, further, in terms of the absolutely universal requirement of duty, or the inexorable demand that immoral agents be allowed no access to happiness. Kant describes a 'faith in God' that is 'as certain as a mathematical demonstration.' The 'foundation' of this faith 'is *morals*, the whole system of duties, which is cognized *a priori* with apodictic certainty through pure reason.'[643] The great teacher, here, is reason. And what the latter 'has taught us about God is faultless and free from error.'[644] This priority entails that 'the mark of [any revelation's] divinity (at least as the *conditio sine qua non*) is its harmony with what reason pronounces worthy of God.'[645]

---

[640] *Vorlesungen-Religionslehre* 28: 1073 (407).
[641] See, for example, *KpV* 5: 125 (104).
[642] The experience of the demand of the moral law, along with the affective consequences that accrue to disobedience, do not, in general, depend for Kant on the human being's conceding that God exists. On this score see, especially, *KdU* 5: 452 (317).
[643] *Vorlesungen-Religionslehre* 28: 1011 (356). See also *R* 8090 19: 634.
[644] *Vorlesungen-Religionslehre* 28: 1047 (386).
[645] *Streit* 7: 46 (270).

Indeed, in the *Religion*, Kant 'ask[s] whether morality must be interpreted in accordance with the Bible, or the Bible, on the contrary, in accordance with morality.'[646] Morality and so pure practical reason takes priority here again. Witness this strange parthenogenesis:[647] 'the teaching of the Bible' is 'a father which our reason can develop out of itself.'[648] In a related vein, Kant's 'pure moral religion' is 'the euthanasia of Judaism,' religion 'freed from all the ancient statutory teachings, some of which were bound to be retained in Christianity (as a messianic faith).'[649] To say that the foundation of this faith is 'morals,' in Kant's ultimate sense, means that the associated sense of 'religion' cannot entail a faith that is, to be precise, 'faith that we can obtain God's favor or pardon by anything other than a pure moral attitude of will.'[650] Indeed , in this sense of the terms 'moral' and 'religion,' 'the only thing that matters in religion is *deeds*.'[651]

Omniscience, omnipotence, and eternity: God's non-moral attributes from the point of view of morality

What morality 'gives' us, here, in addition to the postulate that God exists, is the notion of a being who is entirely like us, *qua* rational being,[652] denuded of the properties that are specific to our earthliness ('the human being according to his species [is] an earthly being endowed with reason'[653]), but enhanced by way of the addition of omniscience, omnipotent, and eternity. Firstly, God is all-knowing. This accords with morality's demand that someone 'be acquainted…with the most secret stirrings of my heart.' Kant thinks that morality presupposes, at least, the possibility of a point of view on the human being whose subject, if there were such a thing, would be competent to judge 'according to the principles of morality, whether I am worthy of happiness.' Secondly, God is all-powerful. Although morality does not put happi-

[646] *Rel* 6: 110 (142).
[647] See *KrV* A763/B791: 'all the concepts, indeed all the questions that pure reason lays before us [the ideas of God, the simple and imperishable soul, and freedom], lie not in experience but themselves in turn only in reason…. [R]eason has given birth to these ideas from its own womb alone.'
[648] *Streit* 7: 59 (280).
[649] *Streit* 7: 53 (276).
[650] *Streit* 7: 52 (275).
[651] *Streit* 7: 41-2 (267).
[652] For a particulary strong expression of this view see M. Häyry, 'The Tension between Self Governance and Absolute Inner Worth in Kant's Moral Philosophy,' *Journal of Medical Ethics* 31, no. 11 (2005): 647.
[653] *Anthro* 7: 119 (3).

ness forward as an incentive, I cannot coherently make worthiness to *be* happy (or my being a decisively moral agent, under precisely this description) an end, thinks Kant, unless I take it to be the case that an agent exists who, just in case I turn out to be worthy of it, 'must also make me actually participate in happiness.' Therefore morality postulates a being who, when it comes to 'matters of morality,' judges *just as I do* (to the extent that I judge impartially, as I ought to do), but who has 'the whole of nature under his power.'[654] Thirdly, morality proposes that God is eternal. To the extent that the latter is regarded as a being that will ultimately 'arrange and direct the consequences of the different states of my existence,' God must be regarded, too, as unconstrained by the condition of time.[655]

God's 'moral perfections'

God's non-moral attributes answer to the requirement that the systematic correlation of happiness with morality be thinkable as a physical possibility. But the 'objective reality [of] moral duties,' where the latter's fulfillment is regarded as a ground of the moral agent's worthiness to be happy, also presupposes God's unlimited 'moral perfections': *holiness, benevolence, and justice*. God's moral perfections are the elements of Kant's determinate concept of God that most clearly have their 'source' in morality. But this is predominantly a matter of morality in its secondary, backward-looking sense. Kant gives a careful account of the relation between these attributes, which is a consequence of the priority of 'morality,' regarded as worthiness to be happy, over 'ecclesiastical' religion. Holiness is primary, benevolence secondary. Justice is a property of the *relation* of these other attributes to one another. Holiness is an attribute of 'the laws,' or the 'supreme principle of legislation.' The latter's demand is unconditional: 'strictly good conduct or the highest virtue.' Benevolence, on the other hand, 'is a special idea whose object is happiness,'[656] or 'an immediate well-pleasedness with the welfare of others.'[657] Its object is not an unconditional one, however. Rather, 'a restrictive condition always precedes God's benevolence,' namely, that 'human beings are to become worthy of the happiness flowing to them.' Kant says that, like God's other attributes, 'in and for itself benevolence is without

[654] *Vorlesungen-Religionslehre* 28: 1073 (407).
[655] *Vorlesungen-Religionslehre* 28: 1073 (407-8).
[656] *Vorlesungen-Religionslehre* 28: 1074 (408).
[657] *Vorlesungen-Religionslehre* 28: 1076 (409).

limit.'[658] However, we read a little later, 'the application of [God's] benevolence is limited *in concreto* through the constitution of the subject in which it is to be shown.'[659] Kant does not explain how it is that *this* aspect of his 'determinate concept of God' is put forward by 'morality.' Benevolence is only *necessary* in order to explain the happiness of the moral agent on the counterfactual assumption that it is actualized. But this actualization is not in itself necessary. God is not *constrained*, at least not immediately, to make the moral agent happy. As Kant defines it in another context, worthiness is a property of the human agent that 'must in God's decision be the condition of his benevolence.'[660] But Kant adds that 'under divine rule even the best of human beings cannot found his wish to fare well on divine justice but must found it on God's beneficence, for one who only does what he owes [*seine Schuldigkeit*] can have no rightful claim on God's benevolence.'[661]

If God is not regarded as benevolent, then the human being has no grounds for hoping that she will be made happy, even on condition that she is moral. If God is not regarded as benevolent, but as merely just and holy, then the only being capable of making human beings happy would have no interest in doing so, no reason for acting to this end. Morality in its primary, forward-looking sense does not 'give' us this notion of God at all. Morality in *that* sense gives us the idea that God is holy—the idea of a being for whom the categorical imperative is no imperative, but rather a law of this being's very nature.[662] Holiness demands 'strictly good conduct or the highest virtue,' but it does so without promising anything in return. Holiness has no direct reference to happiness at all. And the Kantian notion of divine *justice* is directly connected with morality only to the extent that the latter is regarded as a condition without which action aiming at the happiness of the human being is *forbidden*. The fact that this leaves action aiming at the happiness of moral agents merely *permissible* demands the addition of kindness to the roster of divine attributes. As one commentator puts it, '[f]or actions…which are morally acceptable but neither prescribed nor

[658] *Vorlesungen-Religionslehre* 28: 1074 (408).
[659] *Vorlesungen-Religionslehre* 28: 1076 (409-10).
[660] *ÜdM* 8: 257 (26).
[661] *ÜdM* 8: 258 (26).
[662] See *KpV* 5: 32-3, 47 (29-30, 42).

forbidden, we need other (nonmoral) reasons for action.'[663] This holds true here, for Kant, of God as well.

Our reverse engineering of Kant's notion of God shows that his concept of benevolence is not an element of the 'determinate concept of God' that Kant's 'morality' offers us. It would only have been present there if worthiness to be happy were intrinsic moral desert of happiness. However, as we have seen, it is not. God's kindness would have to be regarded either as *ex pacto*, or as something that exceeds morality. Either way, to affirm the presence and expression of this divine attribute is to go beyond what Kant's 'morality,' or pure practical reason, offers us.

There is an instructive tension here. On the one hand, Kant seems to want to say that the *limit* that is applied to God's benevolence is in no way ascribable to God. This limitation '*in concreto*' arrives 'through the constitution of the subject in which it is to be shown.' Neither this constitution, nor this limitation is 'in' God, up to God, God's work, or from God. The only possible obstacle to God's benevolence is set up by the very beings who hope to be happy through God's omnipotent arrangement and direction of 'the consequences of the different states of [their] existence.' This occlusion is their work; they are its authors; it is imputable to them without remainder.[664] As Kant puts it in another context, God's kindness is 'in itself infinite [*an sich unendlich*].' God's goodness is limited only by the human agent's own 'unworthiness [*Unwürdigkeit*],' which is ascribable to her. In the medium of God's holiness, the transgressor limits God's benevolence absolutely, which limiting-relation is justice.[665] But then, on the other hand, Kant wants (and needs) to show that there is something in God that *necessitates* this limitation of what is supposed to be unlimited, otherwise, to the extent that it is a property that inheres *in God*: thus, he says, it is an expression of justice, God's third moral perfection. Justice is the '*limitation of benevolence by holiness* in apportioning happiness.' In short, justice demands that benevolence 'express itself in the apportionment of happiness *according to the proportion of worthiness in the subject*.'[666] It is a matter of kindness, not justice, that

[663] O'Connor, 'Kant's Conception of Happiness': 202. See also Herman, 'On the Value of Acting from the Motive of Duty': 375.
[664] As Kant puts it in the first *Critique*, 'complete' happiness 'knows no other limitation before reason except that which is derived from our own immoral conduct' (*KrV* A813-14/B841-2).
[665] *R* 6686 19: 132.
[666] *Vorlesungen-Religionslehre* 28: 1073-4 (408). See also *R* 6113 18: 459 (337).

happiness be distributed *at all*, but it is a matter of justice that it be distributed in this particular way: that is, 'only *according to the subject's worthiness*.'[667] Since Kant is a rigorist about morality, this reference to 'proportion' is *de trop*. Finite rational agents (or their wills) are either wholly good or they are wholly evil. Human beings, in particular, are evil, even if, as Kant will argue in the *Religion*, an original, ineffaceable disposition to good remains in them.[668] The will is innocent or corrupt, even if the arbitrary freedom of *Willkür* holds open the possibility of change; agents are therefore destined to be either happy or unhappy rather than to enjoy some measure of each.

God's attributes and the idea that the unhappiness of immoral agents is good in itself

Kant's discussion of 'counterpurposiveness' in his 'Theodicy' essay of 1791 further elucidates his commitment to a concept of God that is in thrall, as it were, to the antecedent thesis that the *a priori* combination of the notions of immorality and unhappiness is utterly inexorable.

After defining 'theodicy' as 'the defense of the highest wisdom of the creator against the charge which reason brings against it for whatever is *counterpurposive* [*das Zweckwidrige*] in the world,'[669] Kant describes three kinds of counterpurposiveness. First, there is 'absolute' or 'moral' counterpurposiveness, which is 'evil [*Böse*]' or 'sin.' Second, there is 'conditional' or 'physical' counterpurposiveness, which is 'ill [*Übel*]' or 'pain.' Significantly, however, Kant insists that 'the proportion of *ill* to *moral evil*, if the latter is once there' is not a matter of counterpurposiveness, but, to the contrary, an instance of '*purposiveness* [*Zweckmäßigkeit*].' As such 'the conjunction of ills and pains, as penalties, with evil, as crime' neither 'can nor should be prevented.' And this implies, finally, that 'the disproportion between *crimes* and *penalties* in the world'—but not, let us note, between virtue and reward—is a *third* kind of counterpurposiveness.[670]

Kant now relates these three modes of counterpurposiveness to the three divine attributes to which they 'stand out as objections.' These are God's holiness, good-

[667] *Vorlesungen-Religionslehre* 28: 1087 (418).
[668] *Rel* 6: 26, 43 (74, 88).
[669] *ÜdM* 8: 255 (24).
[670] *ÜdM* 8: 255 (25).

ness, and justice, respectively. Kant's understanding of divine justice, here, is particularly significant. Although he denies that there can be a 'philosophical' (i.e., theoretical or speculative) resolution of these matters, he affirms that, from the point of view of practical reason's pure interest in morality, 'the disproportion between the impunity of the depraved and their crimes' is indicative of a 'bad state…in the world.' This means that, for Kant, the objection to the claim that God is *just* has a special status. Remarkably, it also means that, for Kant, this problem does not bear at all on the problem of the suffering of *moral* agents.[671]

Now, Kant thinks that these three attributes (holiness, goodness, and justice) are basic and irreducible; he also thinks that their 'rank' reflects what is always already established in the human being's 'own pure (hence practical) reason.' This ranking implies that the 'dignity' of moral legislation and the 'firm concept of duties' precludes any compromise on the condition that the human being must meet if God is to regard her happiness as a good. Echoing the second *Critique*, Kant says that 'the human being wishes to be happy first,' but understands and accepts ('though reluctantly') 'that the worthiness to be happy, i.e., the conformity of the employment of his freedom with the holy law, must in God's decision be the condition of his benevolence.' Why? Because 'the wish that has the subjective end (self-love) for foundation cannot determine the objective end (of wisdom) prescribed by the law that unconditionally gives the will its rule.'[672]

This opposition of subjective and objective ends, however, applies to the human being, not God. In God's case, for Kant, the subjective end is a matter of his kindness or goodness vis-à-vis his creatures and the objective end is a matter of what is prescribed, not by the moral law—which for God is no imperative at all—but by the law of unhappiness. Thus there is this *other* objective end: the resolution of 'the disproportion between the impunity of the depraved and their crimes' and repair of this 'bad state…in the world.' This end is the unhappiness of evildoers, which is regarded as absolutely good in itself and an inexorable demand that is placed on God as a being that is not entirely inclined to execute it—to the extent that this rubs God's benevolence the wrong way.

---

[671] *ÜdM* 8: 257 (25-6).
[672] *ÜdM* 8: 257 n. (26 n.).

As we saw in chapter 3, Kant tends predominantly towards the view that God owes 'no *justitiam remunerativam* toward us,' that 'all the rewards he shows us must be ascribed to his benevolence,' and that God's 'justice is concerned…with punishments.'[673] It is *good* that moral agents be happy. And it is *good* that immoral ones be unhappy. In these two affirmations, however, 'good' is not univocal. In the first case 'good' has reference to a state of affairs of which a rational and *kind* being would approve. In the second, 'good' has reference to a state of affairs of which all rational beings, just as such, would approve, however much it might pain them to see it—whether because of their inclination to 'self-love' or their other-directed kindness.

As we saw in chapter 2, an impartial and merely rational observer, one that lacked the attribute of *kindness*, would not be disposed even to notice the unhappiness of moral agents, let alone to do anything about it. In the first *Critique*, Kant makes it clear that the 'wish' for happiness is an inclination of the being that harbors it, whether the happiness wished for is one's own or another's (the happiness, say, of a loved one). For Kant's God, too, kindness is a mode of inclination. Reason approves or disapproves of happiness (and unhappiness), but this is not a matter of inclination for Kant, as we have seen.[674] So, too, with respect to the 'hope to partake of [happiness].' If, '[i]n order to complete [happiness], he who has not conducted himself so as to be unworthy of happiness must be able to hope to partake of it,'[675] this hope cannot be grounded, entirely, in considerations about what God will do *qua* rational being, but must add to God's strictly rational interests a benevolent interest in the happiness of his creatures.[676]

The impartial, rational, and *kind* observer's disapprobation of the unhappiness of *moral* agents, and so too a positive demand for their happiness, can only be traced back as far, then, as the inclination to see them happy. What, however, of such a being's disapproval of the happiness of *immoral* agents and the positive demand that

---

[673] *Vorlesungen-Religionslehre* 28: 1086 (417-18). Kant is not always consistent on this score, a point that I conceded in chapter 3.
[674] *KrV* A813/B841.
[675] *KrV* A813/B841.
[676] Kant seems to think that his treatment of hope constructs a concept of the latter that is based in pure practical reason alone. The necessary thesis that God is kind, however, can only be a matter of *a posteriori* revelation. As far as I am aware, Kant never addresses this contradiction. Cf. Sidney Axinn, 'Kant on Possible Hope: The Critique of Pure Hope,' in *The Proceedings of the Twentieth World Congress of Philosophy: Modern Philosophy*, ed. M. D. Gedney (Charlottesville: Philosophy Document Center, 2000), esp. 79-80.

they be unhappy? Is this grounded in nothing more profound than an inclination to see them suffer, given that they have failed to measure up morally? Is it a mere habit that mistakes for something necessitated the regular contiguity of the judgment that an agent is immoral and the disapprobation of her (actual or prospective) happiness? In Kant's view the answer to these questions is clearly negative. Kant does not bother to justify the implicit claim that, along with respect for the moral law, the 'humiliating' abasement of self-conceit before it, and moral self-contempt in the wake of disobedience of it, disapprobation of the happiness of immoral agents and the demand that they be unhappy are a feeling and a demand that are grounded in pure practical reason.

God's 'virtue'

Far from being a law of God's very nature (as Kant takes the primary, forward-looking moral law to be), the law of unhappiness is an imperative that puts 'pressure' on that nature. This is so to the extent that Kant's God is a kind being whose kindness is limited by and so subordinated to his justice. With respect to the end that this commands, however, I construe Kant's God not as 'holy' (which, in a special sense, Kant takes him to be vis-à-vis the primary moral law) but 'virtuous.'[677] In order to 'obey' the law here, as I make explicit, he must overcome his ineliminable desire to make and see his creatures happy. Relative to the happiness of immoral agents, God's desire to show mercy has the form of a constant temptation to disobedience.

God's holiness consists in the fact that the moral law, for us an imperative, is for God an aspect of God's nature. This just means, in effect, that God is not endowed with any attribute whose practical upshot resists that law. The secondary, backward-looking 'law of unhappiness,' however, must be regarded as an imperative even for God, just given the fact that he is endowed with an attribute, *kindness*, whose practical upshot does resist it. Conversely, the law of unhappiness is an imperative for God because it puts 'pressure' on God's kindness, commanding that his tendency towards limitless benevolence be suppressed. This is God's duty and so, as Kant says in an-

[677] Virtue, for human beings, is moral 'strength' (*MdS* 6: 392, 405, 477 [155, 164, 221]), or 'self-overcoming' (*Vorlesungen-Religionslehre* 28: 1075 [409]; see also *MdS* 6: 383 [148]). The same applies implicitly to Kant's God, here, in connection with the latter's conformity to the law of unhappiness.

other connection, an 'objective constraint' and 'a moral imperative limiting our [here God's] freedom [here God's freedom to act benevolently].'[678]

The distinction between the moral law, in its primary, forward-looking sense, and the 'law of unhappiness,' as a kind of secondary moral law, sets up a 'third man' problem. Kant's account of finite rational agency and the distinction between the latter and the agency of a 'holy' being implies the following. For every class of being that is subject to laws of its 'nature,' on the one hand, but also subject, on the other hand, to imperatives whose expression is not necessitated by the latter laws at all, it is possible to conceive of a distinct class of being for whom these imperatives are laws of nature, but to whom the laws that govern the nature of the first class are not applicable at all. A problem emerges, however, when one goes on, as I suggest that Kant does, to posit a law that is distinct from both those laws that are, for you and me, laws of nature, and those laws that are, for us, on the one hand, imperatives while being, for a 'holy' subject, on the other hand, laws of such a being's ('supersensible') nature. This is precisely what the 'law of unhappiness' consists in; and its postulation opens up the possibility of an infinite regress that render Kant's moral notion of 'holiness' merely relative.

God and human beings together under a single constitution

The human community's practice of expelling, from its midst, individual members belonging to it, by way of death, is a practice that, whether conceived under the rubric of criminal punishment or not, forms and has always already formed a specifically human, practical horizon. Kant's God falls *within* this horizon to the extent that he is subject to the law that Kant takes to command this. Like the earthly sovereign, judge, or executioner, Kant's God is disbarred from coming to the defense of, or providing refuge for, the one that is 'rightly' accused by the members of this unanimous community.

As I argued above, the 'determinate concept of God' that Kant's morality offers up is the notion of a being who, as far as the rules for judging concerning what it is good to do or to omit to do go, is entirely like us, a rational being like any other. As Häyry aptly puts it, 'the universality of practical reason [makes] God and all his ra-

[678] See *MdS* 6: 437-8 (188).

tional creatures equal in the domain of morality.'[679] The difference between God and us does not pertain to these rules, but to God's way of knowing, his power over nature, his freedom from the limitations of time and space, his absolute conformity to the moral law, and his unlimited resolve in the face of the temptation to be kind or merciful.

Steven Smith's 'discover[y] [of] the true basis of the idea of virtue as "worthiness to be happy"' affirms, first, that 'worthiness implies desert' (but of course Smith means primarily the moral agent's desert of *happiness*) and, second, that 'desert implies either a previous stipulation—a contract in the mechanical sense—or a solicited concurrence of free will' with respect to the actualization of whatever it is that is deserved. He goes on to argue that '[t]he moral contract offered by the universalizing will is to be taken up and fulfilled by other members of the moral community' and avers that 'it remains problematic whether God, the author and controller of nature as well as the giver of the moral law, is a member of that community.'[680] On my reading, however, given Kant's revelation that *his* 'determinate concept of God,' at least, derives from his understanding of 'morality,' there is no good reason to doubt that God—*Kant's* 'God' to be precise—is *regarded*, at least, as a member of the community that is constituted by the law of unhappiness.

With respect to 'the ushering in of the highest good,' Kant says in the *Religion* that God's role is simply to make up for what is lacking by way of 'human capacity' (*Menschenvermögen*) when it comes to the causal connection that is supposed to hold between 'the strictest observance of the moral laws,' on the one hand, and this ultimate end, on the other. '[A]n omnipotent *moral* being must be assumed as ruler of the world, under whose care this would come about.'[681] God is not a transcendent other, here, but a powerful ally in a matter of shared interest. But God's and moral agents' shared interest in the latter's happiness runs in parallel with their and all rational agents' shared interest in the conjunction of immorality and unhappiness. Kant refers to '[t]he verdict of conscience upon the human being,' which '*acquit[s]* or *condemn[s]* him with rightful force.'[682] This 'rightful force' is a kind of impetus of

[679] Häyry, 'The Tension between Self Governance and Absolute Inner Worth in Kant's Moral Philosophy': 646.
[680] Smith, 'Worthiness to Be Happy': 189-90.
[681] *Rel* 6: 7-8 n. (60 n.) (my emphasis).
[682] *MdS* 6: 440 (190).

the condemnation that extends, virtually, into the consequence that is attached to it: punishment, unhappiness. '[T]he proceedings are concluded,' Kant writes, and 'the internal judge, as a person *having power*, pronounces the sentence of happiness or misery, as the moral results of the deed.' And, he continues, '[o]ur reason cannot pursue further his power (as ruler of the world) in this function; we can only revere his unconditional *iubeo* or *veto* ["I command" or "I forbid"].'[683]

In connection with this shared interest Michalson writes, '[t]he issue is not one of morality, but of proportionality.'[684] He takes there to be a contradiction here. But this is to miss the point, to evince a typical problem. It is true that

> on the grounds of Kant's own moral theory, to be worthy of happiness (i.e., to be virtuous) and yet to be without it probably does not confront us with a genuinely moral issue. The spectre of an unhappy person of great virtue may be distressing and deeply poignant, but it hardly threatens the moral status of the person in question.[685]

It is also true that '[p]ostulating God's existence hinges on the idea of happiness; yet happiness…satisfies a rational yet non-moral need.'[686] Michalson does not notice that the conjunction of *immorality* and *happiness*, however, does pose a problem that Kant clearly regards as a moral one: a problem that finds expression in a judgment of 'counterpurposiveness' and an attitude of disapprobation that are grounded *a priori*, for Kant, in reason.

God and human beings are not merely bound by the same law in terms of what is to be *done* about happiness, given that human beings are immoral. Kant also identifies the Good with the moral in *both* senses—as an interest in doing what is good and right and an interest in seeing happiness distributed in a *good* manner. Because of this unqualified sameness of the rules in accordance with which both we and God judge concerning *that*, it is possible to read Kant's 'Fragment of a Moral Catechism' as a query whose interlocutors are the human being and God himself.

The catechist, drawing his student forward into an awareness of what he always already knows, asks for more than a clear answer in the matter of what one ought to do, what duty consists in. He asks his student to imagine a counterfactual scenario—really, an eschatological one. '[I]f it were up to you to dispose of all happiness (possible in the world),' he asks, 'would you keep it all for yourself or would you share it

---

[683] *MdS* 6: 439 n. (189 n.).
[684] Michalson, *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*, 23.
[685] Ibid., 24.
[686] Ibid.

with your fellow human beings?'[687] The pupil responds as Kant's benevolent God would do: 'I would share it with others and make them happy and satisfied too.' In response, the teacher observes 'that [this] proves that you have a good enough *heart*.' He then proposes to test the student, to 'see whether you have a good *head* to go along with it.'[688]

This contrast of 'heart' and 'head' signals a transition, of course, from kindness to reason—which is to say, to the determination and doing of duty (or, for God, simply holiness), on the one hand, and the determination and doing of justice, on the other. The teacher then describes a series of types, 'a lazy fellow,' 'a drunkard,' 'a swindler,' and 'a violent man' and—here trafficking, apparently, in a notion of happiness that reduces the latter to nothing more than the pleasure of the moment[689]—asks the student whether he would 'really' hand over the various things that these types of people most want (as a means to happiness, each 'in his own way'): 'soft cushions,' 'wine and whatever else [one] needs to get drunk,' 'a charming air and manner to dupe other people,' and 'audacity and strong fists.'[690]

In a sense, the catechist makes it too easy for the pupil to answer as he does: 'No, I would not.'[691] It is as though he is being asked, simply, to agree that he would refrain from aiding and abetting immoral agents in their immorality. Happiness is identical, here, with persistence in immorality and, indeed, a deepening of it. Kant's point, however, is rather broader. So too is the teacher's point when he observes, building on the student's answer, that 'you…would not give [happiness] without consideration to anyone who put out his hand for it; instead you would first try to find out to what extent each was worthy of happiness.' Kant distracts from the generality of this 'each' by describing situations in which obviously deleterious consequences accrue to providing the *means* to (idiosyncratically determined) happiness, where there are good reasons, distinct from their unworthiness to be happy, for not

[687] *MdS* 6: 480 (223-4). Other instances in which Kant refers to happiness' principled disposition, distribution, granting, or apportionment (Kant's vocabulary is similarly varied) appear, for example, in *KrV* A806/B834; *Gr* 4: 450 (55); *ÜdG* 8: 278 n. (281 n.); *Orientierung* 8: 139 (12); *Vorarbeiten-MdS* 23: 404, 413; *R* 7049 (19: 235 [457-8]); *R* 6975 (19: 218 [454]; *R* 4277 (17: 493 [125]).
[688] *MdS* 6: 480 (224).
[689] Note that I adverted to this peculiar text in chapter 2 in my discussion of Kant's thinking about happiness. Kant's main point here, however, is not to give his thoughts about what happiness consists in, but to show that the distribution of happiness ought to be grounded in the morality of its beneficiaries.
[690] *MdS* 6: 480-81 (224).
[691] *MdS* 6: 481 (224).

providing access to happiness 'without consideration to anyone who put out his hand for it.'

The generality of the matter comes to the fore, however, in the very next question, when the teacher asks: 'But doesn't it occur to you to ask, again, whether *you yourself* are worthy of happiness?' and the pupil answers, '[o]f course.'[692] Naturally, posed as a question to God, this would make no sense. As Kant says a page later, 'a human being's observance of his duty is the universal and sole condition of his worthiness to be happy, and his worthiness to be happy is identical with his observance of duty.'[693] This does not apply to God. Nevertheless, the basic structure of the human being as a being that is called upon to do her duty, mirrors the elements of Kant's 'determinate concept of God.' The teacher declares that

> the force in you that strives only toward happiness is *inclination*; but that which limits your inclinations to the condition of your first being worthy of happiness is your *reason*; and your capacity to restrain and overcome your inclinations by your reason is the freedom of your will.

In God, of course, in relation to duty, 'freedom' is just holiness—and so, in a sense, not freedom at all (rather the divine nature). The 'capacity to restrain and overcome your inclinations by your reason' is, precisely, the capacity of the agent to do her duty—and to do her duty for reasons that are entirely independent of any inclinations she might have to do so (i.e., so that she *can* do so even if such inclinations are lacking, or others militate against it). In God, 'the force…that strives only toward happiness' strives towards the happiness of God's creatures. God is inclined, without internal limitation, to make them happy. God is, in short, benevolent. But note that reason makes two appearances here: first, in connection with 'freedom of the will' and the ability to do one's duty independently of inclination, not identical with that capacity, but a feature of it; second, simply *as* 'that which limits your inclinations to the condition of your first being worthy of happiness.' This is rule-making reason as the source of the idea of *justice* and of the demand that justice be done. It is not nature, even in God, but freedom. It meets resistance in the form of the human inclination that strives after happiness for oneself. And it meets resistance in the form of the divine inclination that seeks the happiness of creatures. Reason in *this* sense, 'a reason,' as Kant says at one point, that both judges and 'distributes happiness universal-

[692] *MdS* 6: 481 (224).
[693] *MdS* 6: 482 (225).

ly,'[694] 'the morally disposed reason [that] cannot conceive of happiness without good conduct,'[695] is predicated univocally of both God and human beings.

In this respect, God is called to follow the same rules as human beings. That which God (setting aside the warrant for taking God to exist) would both judge to be good to do and put into effect, is equivalent to that which human beings *qua* rational also judge to be objectively good and *would* put into effect if only they had sufficient insight and power. As Kant's catechetical pupil puts it, at the close of the fragmentary 'moral catechism,' 'with regard to the moral order, which is the highest adornment of the world, we have reason to expect a…wise regime, such that if we do not make ourselves *unworthy of happiness*, by violating our duty, we can also hope to *share* in happiness.'[696] And if we *do* violate our duty? Then, of course, we have reason for dread.

In his first *Critique* Kant uncovers the limitations under which reason operates, at least in human beings, in order to show that there can be no warrant for metaphysical (i.e., cosmological, psychological, or theological) dogmatism. By showing these limits, however, Kant also sets forth the regulative ideal of a perspective from which all of our claims about the way things 'really' are might yet be subject to a mode of judgment that is not constrained by the conditions of possibility of our specifically human kind of knowledge.

Kant's critique of pure reason's theoretical pretensions shows that claims pertaining to specifically human experience or cognition are subject to a kind of 'last judgment' whose specific content no finite agent could ever determine, but whose possibility such agents can nevertheless take into account. Kant shows that our knowledge claims do not have the status of ultimate judgments and that they lack this status, not as a matter of happenstance (as though they *could* be grounded in some unconditioned insight, but happen not to be), but inevitably, because of the way in which they are constituted. The human word, in short, where this is a 'theoretical' one, is never the 'last' word—irrespective of whether any such word is forthcoming from some non-empirical 'elsewhere' or not.

---

[694] *R* 7242 19: 293 (474).
[695] *R* 6317a 18: 632 (373).
[696] *MdS* 6: 482 (225).

If this is 'Kantian humility,' to borrow Rae Langton's apt epithet,[697] one finds something very different when one turns from Kant's treatment of the theoretical pretensions of pure reason to his thinking about the interests that reason has in its character as practice-guiding. One is confronted with something else when one turns to interests that are associated with pure reason in its character as both practice-guiding and action-motivating. Here, too, judgment is at issue. This is not assertoric, but rather normative judgment; not judgment concerning what exists and what existing things are like, but rather judgment concerning what ought to be case even if, in fact, nothing is as it should be.

Of course, Kant does not take all normative judgments to have the same status. Judgments concerning what ought to be done, given an agent's interest in this or that end, are normative, certainly, but these rest on fallible judgments concerning the means that would actually be effective relative to particular ends. It is possible to be mistaken, here, since the truth of such judgments depends upon what happens empirically to be the case. By contrast, judgments about what human beings ought or ought not to do, *unconditionally*, do not depend for their truth upon anything that merely happens to be the case. Such judgments are not shadowed by the ideal of a 'last judgment,' as metaphysical, or even empirical knowledge claims are. To the contrary, Kant thinks that we are equipped to make such judgments, *a priori*. He thinks that, like other judgments of this kind, our judgments about what human beings ought, or ought not, to do can be 'apodictically' certain.

If we lack access to that non-discursive mode of understanding that Kant contrasts with our own, that 'divine' intellectual intuition in comparison with which our own knowing (combining elements both intellectual and sensible) is infinitely more humble, there is nevertheless no such distinction when it comes to the judgments of pure practical reason. God and human beings judge in accordance with the same rules.

The mercy-free community

In the foregoing, I argued that the law of unhappiness is constitutive for a community of which we, together with God, are members in common. I will now argue that the

[697] See Rae Langton, *Kantian Humility: Our Ignorance of Things in Themselves* (Oxford: Clarendon Press, 1998).

upshot of this community's constitution is that, like the merely human *polis*, this larger community is a mercy-free zone—here, however, with a view to the broad scope of immorality (regarded, in a manner of speaking, as crime against morality), as such, rather than the narrow scope, merely, of external crimes against 'public law.' I characterize this coinciding of the two contexts (in which the law of unhappiness is put into practice) in terms of a puncturing of the ethical, by which it is susceptible to incursions of an antecedently, politically constituted notion of the Good, or a conflation of good and right.[698] More precisely, I argue, this is so to the extent that Kant regards murder, uniquely, as a political manifestation of the 'inner wickedness' of immoral agents and converts the (political) necessity of systems of criminal sanctions, in general, into the objective badness of mercy. This means that, for Kant, the enactment and the approbation of particular instances of capital punishment are practices (one actual and the other 'virtual') that conform, not merely to the law of punishment, but to the law of unhappiness as well. Capital punishment (conceived as retribution), as I argued in chapter 3, has a uniquely political and eschatological character for Kant. I argue, here, that long before he ever gets to the point of the theoretical moves with which this thesis is concerned, he is already committed to this long extant, really archaic practice—and to the 'virtual' practice that consists in the desire and the will (albeit a generally, but not universally, impotent one) that murderers be removed from the world.

Two key notions and two distinctive Kantian claims bear directly on the matter at hand, expressing universally applicable claims about punishment and giving especially clear expression to the thinking that is implicit in Kant's gloss. The first of these is Kant's key assertion, which we discussed in the last chapter, that '[t]he law of punishment is a categorical imperative.' The second is a view that the latter encapsulates, implicitly, and which Kant expresses with particular clarity in his lectures on the philosophy of religion, when he declares that 'a judge who pardons is not to be thought of.'[699]

---

[698] For a classic argument detailing the deleterious consequences of these notions' conflation see Ross, *The Right and the Good*.

[699] *Vorlesungen-Religionslehre* 28: 1086 (418). See Kant's equivalent, but less dramatic declaration in *Rel* 6: 146 n. (171 n.). These pronouncements sum up the canonical point of view on mercy that we encountered earlier in the 'woe to him' (*wehe dem*) of *MdS* 6: 331-2 (105), Kant's prescription con-

In its context, Kant's assertion concerning the 'law of punishment' refers to punishment in its political articulation as state-authorized harm. But Kant specifies 'the categorical imperative of penal justice' as the demand that 'unlawful killing of another must be punished by death,'[700] which restricts this imperative's scope, politically, but, at the same time, extends the law to the point where the practice that conconforms to its demand coincides with action that conforms, too, to Kant's law of unhappiness.

In *its* context, the second assertion refers a general elucidation of what it means to be a judge back to its immediate (contextual) subject: God. Kant's law of punishment and Kant's law, as it were, of judging (Thou shalt not pardon!) apply across the board; the first within an ethical as much as a political frame of reference, the second within a political as much as an eschatological (and so, too, an ethical) one. I have avoided claiming that Kant's thinking about the relationship between crime and punishment, *in general*, is modeled on his thinking about murder and capital punishment, in particular. I do make such a claim, however, with respect to Kant's representation of the normative 'combination' of immorality and unhappiness. The direction of the 'modeling' moves from the political to the eschatological. Moreover, the theoretical dimension is subordinated to its antecedently (indeed archaically) practical one: Kant's understanding of the relationship between unhappiness and immorality is informed by his antecedent attitude towards the practice of capital punishment—an attitude of approbation whose object is not this or that (maybe hideous) instance of the thing, but the practice as such. This attitude, or the commitment that it expresses, is not thematized, in Kant, in retributivist terms from the outset. Rather, its being a retributivist commitment is, as I argued in chapter 3, a function of its stubborn, philosophically unanalyzable immediacy.

Mercy as immorality for God

In this sub-section I argue that the main place where Kant's commitment to capital punishment, on the one hand, and to the (prospective) approbation of the unhappiness of immoral agents, on the other, is evident, is in his denigration of mercy, his

---

cerning treatment of 'the last murderer' (6: 333 [106]), and the wise ruler's refusal to 'allow *blood-guilt* to come upon [his] land' (6: 490 [232]).

[700] *MdS* 6: 336-7 (109).

view that mercy is absolutely and unmitigatedly bad, an expression of kindness that remains eternally out of reach to both politics and ethics, forbidden to both the human judge and to God. Again, Kant is unable to articulate a notion of the Good such that while justice (or 'right') is good, mercy is good in some sense as well.

Once again, I connect Kant's strong bias against mercy and his habitual manner of representing the relationship between immorality and unworthiness to be happy with his retributivism. And I claim that, when it comes to 'matters of morality,' Kant is fundamentally opposed to any other idea of God and of God's action in behalf of human beings. Thus he excludes, from the outset, the merciful God of scripture, God the *Paraklete*, God as shelter for 'the one that did it' in the face of the executioner and the mob who—in accordance with justice—would effect and approve of his or her death. For the same reason, I argue, when it comes to the question whether it is *good* for immoral agents to be happy, Kant is opposed to any answer that would tend to weaken our resolve for the mundane (deferred, impotent, 'virtual') forms of the practices that his God (God the merciless judge, God the executioner, God the mob) and earthly judges and executioners actually (in God's case eschatologically) enact.

'Kant has loaded upon us the guilt of the past, which Kantian practical reason will not allow God to forgive.'[701] In short, a merciful God is not to be thought of.[702]

---

[701] Gibbs, 'Fear of Forgiveness': 332. Indeed, Gibbs writes, '[Kant] fears forgiveness because he believes that the promise of forgiveness will destroy duty' (ibid., 331). See also O. Dekens, 'Initiation À La Vie Malheureuse: De L'impossibilité Du Pardon Chez Kant Et Kierkegaard,' *Revue Philosophique de Louvain* 96, no. 4 (1998); John R. Silber, 'The Ethical Significance of Kant's *Religion*,' in *Religion within the Limits of Reason Alone*, ed. Theodore M. Greene and Hoyt H. Hudson (New York: Harper and Row, 1960), cxxxii (but cf. Wood's critical assessment of the latter in Wood, *Kant's Moral Religion*, 242-3).

[702] This ought to give pause to the host of recent commentators who find, in Kant, an able, fruitful, more-or-less orthodox partner for work in theology, that is, for the doctrine of God (R. Arp, 'Vindicating Kant's Morality,' *International Philosophical Quarterly* 47, no. 1 (2007); R. F. Brown, 'The Transcendental Fall in Kant and Schelling,' *Idealistic Studies* 14, no. 1 (1984): 49; Chris L. Firestone, 'Making Sense out of Tradition: Theology and Conflict in Kant's Philosophy of Religion,' in *Kant and the New Philosophy of Religion*, ed. Chris L. Firestone and Stephen Palmquist, Indiana Series in the Philosophy of Religion (Bloomington: Indiana University Press, 2006); Chris L. Firestone and Nathan Jacobs, *In Defense of Kant's Religion*, Indiana Series in the Philosophy of Religion (Bloomington, IN: Indiana University Press, 2008); Bernd Hildebrandt, 'Kant Als Philosoph Des Protestantismus,' in *Was Ist Und Was Sein Soll*, ed. Udo Kern (Berlin ; New York: Walter de Gruyter, 2007); P. Rossi, 'Kant as a Christian Philosopher: Hope and the Symbols of Christian Faith,' *Philosophy Today* 25, no. 1 (1981); Rossi, 'Kant's Doctrine of Hope: Reason's Interest and the Things of Faith'; P. Rossi, 'Autonomy and Community: The Social Character of Kant's Moral Faith,' *Modern Schoolman* 61, no. 3 (1984); Rossi, 'The Final End of All Things: The Highest Good as the Unity of Nature and Freedom'; Philip J. Rossi, *The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind*, Suny Series in Philosophy (Albany: State University of New York Press, 2005); Philip J. Rossi, 'Reading Kant through Theological Spectacles,' in *Kant and the New*

And this goes beyond the notion that God, being just, must punish sinners, occlude their happiness, make them unhappy. This means that the hope that breaks into the desperate experience of moral inadequacy, the hope that what one can never achieve on one's own can be done for and to one—from sheer love—is a delusion. Divine 'indulgence' or a 'dispensation' that regards an agent as 'fully adequate to God's will,' when she clearly is not, 'do[es] not harmonize with justice.'[703] Forgiveness is out of the question for agents that have not themselves retrieved what was lost through their own fault. 'The justice of [moral] judgment must be unexceptionable and unrelenting.'[704] But this pertains directly, not merely to the question whether or not agents have done well, or failed, morally. It pertains to the consequences accruing necessarily to their failures. When Kant teaches that 'a judge who pardons is not the be thought of!' he means, specifically, that the object of God's judgment must be

---

Philosophy of Religion, ed. Chris L. Firestone and Stephen Palmquist, Indiana Series in the Philosophy of Religion (Bloomington: Indiana University Press, 2006)), for thinking about the relationship between human and divine agency (Anderson and Bell, *Kant and Theology*, 33; Mariña, 'Kant on Grace: A Reply to His Critics'), for thinking theologically about hope (Elizabeth C. Galbraith, 'Kant and Richard Schaeffler's Catholic Theology of Hope,' *Philosophy & Theology* 9, no. 3-4 (1996): 333; Richard Schaeffler, 'Kant Als Philosoph Der Hoffnung,' *Theologie und Philosophie* 56, no. (1981)), or love (J. Mariña, 'Making Sense of Kant's Highest Good,' *Kant-Studien* 91, no. 3 (2000): 196); or who see Kant as a competent theologian in his own right (see J. P. Lawrence, 'Radical Evil and Kant's Turn to Religion,' *Journal of Value Inquiry* 36, no. 2-3 (2002): 319); or who see him as a source for improving in key ways upon orthodoxy itself (see C. Cope, 'Freedom, Responsibility, and the Concept of Anxiety,' *International Philosophical Quarterly* 44, no. 4 (2004): 549; Franz Gniffke, 'Auf Der Suche Nach Dem Verantwortlichen Subjekt: Eine Hinfuhrung Zu Kants Grundlegung Der Ethik,' in *Zur Geschichtlichkeit Der Beziehungen Von Glaube, Kunst Und Umweltgestaltung* (Frankfurt: Haag & Herchen Verlag, 1977); but cf. Morris Stockhammer, 'Responsibility and Freedom: The Kantian Solution,' *Judaism* 14, no. (1965)); who welcome Kant's notion of autonomy as a tool for thinking about Christian faith (E. Gaziaux, 'Trois Modeles D'autonomie En Morale: Kant, Steinbuechel, Auer,' *Revue Theologique De Louvain* 28, no. 3 (1997); E. Gaziaux, 'L'autonomie En Morale: Entre L'affirmation De L'homme at La Quête De Dieu,' *Revue Theologique De Louvain* 30, no. 3 (1999); Johannes Schwartländer, 'Sittliche Autonomie Als Idee Der Endlichen Freiheit: Bemerkungen Zum Prinzip Der Autonomie Im Kritischen Idealismus Kants,' *Theologische Quartalschrift* 161, no. 1 (1981)); or who endorse Kant's denigration of 'ecclesiastical' Christianity and his proposals for a purely moral, rational religion (Lawrence, 'Radical Evil and Kant's Turn to Religion': 333; Friedrich Paulsen, *Immanuel Kant, His Life and Doctrine*, trans., J. E. Creighton and A. Lefevre (London: J. C. Nimmo, 1902), 339. Cf., however, the critical approaches of J. DiCenso, 'Kant, Freud, and the Ethical Critique of Religion,' *International Journal for Philosophy of Religion* 61, no. (2007); Christopher J. Insole, *The Realist Hope: A Critique of Anti-Realist Approaches in Contemporary Philosophical Theology*, Heythrop Studies in Contemporary Philosophy, Religion, & Theology (Aldershot: Ashgate, 2006); G. P. Schner, 'Moral Ontology in Kant: At What Cost Freedom and Perpetual Peace?,' *Toronto Journal of Theology* 18, no. 1 (2002): 154; Bernard Wand, 'Religious Concepts and Moral Theory: Luther and Kant,' *Journal of the History of Philosophy* 9, no. 3 (1971)) and the more nuanced and limited appreciation of Paul D. Janz, *God, the Mind's Desire: Reference, Reason, and Christian Thinking*, Cambridge Studies in Christian Doctrine (Cambridge: Cambridge University Press, 2004).

[703] *KpV* 5: 123-4 (103).

[704] *Vorlesungen-Religionslehre* 28: 1074 (408).

allowed no more than 'that measure of happiness which is proportionate to his worthiness.'[705]

The immediacy that is embodied in the claim that the law of punishment is a categorical imperative closes off the possibility of imagining mercy as a good. '[T]he concept of duty includes unconditional necessitation,' Kant says in the *Religion*, 'to which gracefulness stands in direct contradiction.'[706] To punish is God's duty, God's grace notwithstanding.[707] And this punishment is conceived of in retributivist terms as *proportionate* to the offense—which is total (immorality all the way down)—as a consequence that is addressed, wholly and simply, to *none* but the entirely discrete 'one that did it,' and as a consequence that *must* be addressed to the latter. If the grace of mercy or forgiveness is possible, then this cannot be understood in any sense that compromises this retributivism. In general, commentators have not appreciated this specific problem, but frame matters in terms of the general importance that Kant places on freedom.[708]

It is not merely that Kant's 'determinate idea of God' is offered up by his thinking about morality. His thinking about religion, faith, revelation, grace, original sin, his Christology, his soteriology—all of this may be reverse engineered in a manner that gets back to 'morality'—but often in the sense with which I have been concerned in this thesis. Detailing all of this would take us far beyond the scope of this project. For now, the main thing—in connection with the present task of establishing as far as possible that Kant's law of unhappiness has action-guiding significance for the agent that he calls 'God' (which is what gives content to my claim that *our* law-conforming practice is both virtual and an act of deferral)—is to show that mercy is occluded, not just in the sense that it is shown to be injustice, but in the sense that it cannot be regarded as good at all.

---

[705] *Vorlesungen-Religionslehre* 28: 1087 (418).
[706] *Rel* 6: 23 (72).
[707] For arguments to the contrary see Mulholland, 'Freedom and Providence in Kant's Account of Religion: The Problem of Expiation', 100; Sussman, 'Kantian Forgiveness'. Sussman's argument turns, in particular, on a very narrowly defined notion of forgiveness that makes antecedent 'repentance, apology, and supplication' centrally important for its 'morally acceptable instances' (ibid., 86).
[708] Dews, *Idea of Evil*, 34; Mariña, 'Kant on Grace: A Reply to His Critics': 388; Paton, 'Kant's Idea of the Good': xxi; Silber, 'The Ethical Significance of Kant's *Religion*', cxxxii; Nicholas P. Wolterstorff, 'Conundrums in Kant's Rational Religion,' in *Kant's Philosophy of Religion Reconsidered*, ed. Philip J. Rossi and Michael J. Wreen (Bloomington: Indiana University Press, 1991), 45. Sussman, however, notes as I do that 'Kant's retributivism' is pronounced enough to constrain his thinking about 'true forgiveness' (Sussman, 'Kantian Forgiveness': 88).

For Kant, nothing can be good that is not a product of freedom. Thus, as we saw before, 'religion itself' or 'the faith in question cannot be faith that we can obtain God's favor or pardon by anything other than a pure moral attitude of will.'[709] Or again, 'true religion is not to be placed in the knowledge or profession of what God does or has done for our salvation, but in what we must do to become worthy of it.'[710] God is 'the loving one,' but, as Kant adds parenthetically, he is the one 'whose love is that of moral *approbation* of human beings so far as they conform to his holy laws.'[711] Kant decries 'the lazy and timid cast of mind (in morality and religion), which has not the least trust in itself and waits for external help.' This attitude, he writes, 'unharnesses all the forces of a human being and renders him unworthy even of this help.'[712] As for faith, Kant maintains that 'the improvement of [one's] life [is] the supreme condition under which alone a saving faith can occur.'[713] In short, '[i]t depends only on us whether we will become objects of [God's] grace or of his punitive justice.'[714] And Kant really means this in the most radically individualistic manner possible: 'if the teaching of the church were directed straight to morality, the judgment of his conscience would be…that *he must answer to a future judge for any evil he has done that he cannot repair*, and that no ecclesiastical means, no faith or prayer extorted by dread, can avert this fate.'[715] But instead, the (ecclesiastically) religious subject thinks of God on the model of the merciful human sovereign and God's wisdom on the model of 'the fallible wisdom of the human will' so as to separate out and bypass God's justice 'by appealing exclusively to his *grace*.'[716]

---

[709] *Streit* 7: 52 (275).
[710] *Rel* 6: 133 (160). For a rather discouraging expression of this view see *Rel* 6: 196 n. (211 n.).
[711] *Rel* 6: 145 (170).
[712] *Rel* 6: 57 (101).
[713] *Rel* 6: 118 (149). Not surprisingly, Kant's *Religion*, in particular, is shot through with passages in which Kant makes grace, faith, God's love, or salvation, depend upon antecedent, utterly unassisted effort. Among others and in addition to those instances cited above, see, for example, *Rel* 6: 53, 72, 75, 75 n. †, 118, 120, 159-61, 174, 184, 189, 191-92, 200-202 (97, 112-13, 115, 115 n. †, 148, 150, 181-82, 193, 201-2, 205, 207, 214-15).
[714] *Vorlesungen-Religionslehre* 28: 1039 (379). Thus, too, '[i]t is impertinent to pray for happiness or even for release from punishment if one is not a better human being' (*R* 7093 19: 247 [460]). The force of Kant's 'it depends only on us' is evident, too, in his claim that we 'have only our trespasses before our eyes, together with the consciousness of our freedom and of the violation of our duty for which we are wholly to be blamed, and hence have no ground for assuming generosity in the judgment passed on us' (*Rel* 6: 141n. [166 n.]).
[715] *Streit* 7: 60 n. (281n.).
[716] *Rel* 6: 200 (214).

The rigorous, universal requirement that human beings prove themselves worthy of happiness by a self-transformation and an ultimate, irrevocable turn towards the moral law—the perfection of perfect virtue—is not susceptible to exceptions. The universality of the primary, forward-looking law's demand, on the one hand, and the universality of the law of unhappiness, on the other, find expression in Kant's declaration that 'God's impartiality consists in the fact that God has no favorites.' God, Kant argues, is not like a parent who has 'a special love for a child which has not especially distinguished itself.' To the contrary, says Kant, 'it cannot be thought of God that he would choose some individual subject over others as his favorite with no regard to the subject's worthiness.'[717] But this is very telling for what it *excludes*: not to have favorites, to be impartial, is perfectly consistent with loving *each one* of one's children without regard for the degree to which each or any of them has 'distinguished' himself or herself. Kant cannot say, however, that 'it cannot be thought of God that he would choose some individual subject over others as his favorite'—and leave it at that. The supplementary qualification with regard to 'the subject's worthiness' is a necessary addendum. Or again:

> It is arduous to be a good *servant* (here one always hears only talk of duties); hence the human being would rather be a *favorite*, for much is then forgiven him, or, where duty has been too grossly offended against, everything is again made good through the intercession of some one else who is favored in the highest degree, while he still remains the undisciplined servant [*der lose Knecht*] he always was.[718]

It is not a question of God's favouring this or that 'individual' without making this favour turn on worthiness to be happy. The issue is not the condition that has to be met for God's favour—it is rather the individuation that has to be presupposed if God's disfavor expresses nothing *in* or *about* God, but is rather an effective 'pressure' that comes to bear on God in view of the agent's unworthiness to be happy, and if God's occlusion of access to happiness is regarded as good in itself, independently of anything that either God or the subject derives from it. The unhappiness of immoral agents is unconditionally good; their happiness bad.

The mercy of forgiveness does not *follow* upon anything. It is not a consequence of anything. It makes no connections. Like punishment, it puts a stop to consequenc-

[717] *Vorlesungen-Religionslehre* 28: 1087-8 (419).
[718] *Rel* 6: 200 (214).

es,[719] but unlike punishment it is not *called for*. It is something 'new,' unprecedented. There is no need to dispute Kant's point: a merciful judge is a contradiction in terms. Mercy is a good whose goodness cannot be explained from within any system of penal law that presupposes that 'the one that did it,' did 'it,' wholly and simply, all the way down. Kant is only able to imagine the good in terms of his two candidates: in relative terms, as the happiness of the individual finite rational agent, or, in absolute terms, as the goodness of the will. But there is a third option that he does not imagine: reconciliation between enemies, between rival neighbours, brothers, parents and children, siblings, husbands and wives, in a situation where the task of identifying 'the one that did it' is as easy as ever, while the task of saying what 'it' is that this one did is virtually impossible.[720]

This is why, again, when it comes to the sovereign's '*right to grant clemency* to a criminal,' Kant must insist that this is 'the slipperiest one [of his rights] for him to exercise.' Kant thinks that it must be regarded as 'doing injustice in the highest degree.' I argued in chapter 3, however, that this has to be qualified; that the merciful sovereign is doing what is immoral, on Kant's own terms, not what is illegal.[721] The sovereign's mercy cannot be punished, in turn; but it is absolutely *bad*.

There is no higher good than 'morality,' for Kant. But if morality is not to be had, then it is also good, within a new frame of reference, that happiness not be 'the lot' of certain agents. Wolff points out, with respect to Kant's opening moves in the *Groundwork*, that '[i]t is noteworthy that the philosopher most completely identified with the doctrine of stern duty should begin, not with a statement about what we ought to do, but rather with a judgment of what is unqualifiedly *good*.'[722] But of course this is Kant's strategy: to work from the notion of absolute goodness to an answer to the question what we ought to do—by way of the ascription of this kind of goodness to, precisely, a *will*, a 'doer' of deeds.

The rational, impartial spectator's attitude of disapprobation at the very thought of a happy, but immoral agent, expresses a value that Kant takes to be universally

---

[719] See Hannah Arendt, *The Human Condition*, ed. Margaret Canovan (Chicago: University of Chicago Press, 1958), 241.
[720] Hegel is highly suggestive on this point. See Hegel, *Phenomenology of Spirit*, § 669 (407). See also Gibbs, 'Fear of Forgiveness': 329-30.
[721] *MdS* 6: 337 (109-10).
[722] Wolff, *Autonomy of Reason*, 57.

shared—if only we will judge impartially; if only God will suffer the aim of his kindness to be thwarted by evildoers. In contrast, Kant cannot conceive of an act of mercy to which *everyone* would agree since, in effect, he always regards pure practical reason itself as both the injured party and as the legislator of the (rightly) retaliatory sanction. '[T]he verdict of the highest *judge* (*supremi iudicis*) [i.e., pure practi-practical reason] is *irreversible* (cannot be appealed).'[723]

Wood indicates that Kant's problem with forgiveness, which finds expression especially in Kant's strong claims about divine justice in the *Religion*, is that it appears to be 'something "outside" morality, or something "higher than" it.'[724] Likewise, Wood points to Kant's worry that grace is not 'something *rational*.'[725] Kant wants to be able to 'distinguish forgiveness from simple immorality and inhumanity, from avarice, murder, or deceit,'[726] or even just 'immoral leniency.'[727] His solution is to describe grace and mercy in terms that accord with 'a definite moral standard or rule'[728] and so, really, to put these within the limits of what pure practical reason is able to warrant *a priori*. But this is just to evacuate the notions of grace and mercy and to replace them with something else.

One consequence, here, is that the representation of God as an agent who, uniquely, stands alongside the accused (even the 'rightly' accused) against the multitude of her accusers, out of love for that one, altogether without reference to anything that the accused has done or will yet do is, for Kant, repugnant, if not incoherent. This is not to say that Kant would take the unanimity of human communities with respect to those of their members that they call to account and expel from their midst as a guarantee that justice and reason are on the side of the 'many,' rather than on the side of 'the one that did it.' But, for Kant, the unanimity of the whole class of rational beings, or of all human beings just *qua* rational, does ground such an inference. It is *this* 'many,' or the human 'many' so conceived, with whom the agent that Kant calls 'God' would inevitably side—if such an agent existed. In short, for Kant,

[723] *MdS* 6: 316 (93).
[724] Wood, *Kant's Moral Religion*, 240.
[725] Ibid.
[726] Ibid., 241.
[727] Ibid.
[728] Ibid.

the practically-rational, even if not the theoretically-rational, 'many' includes both this 'God' and human beings among its members.

The good that is aimed at receptivity, the *given* and *received*, is happiness, flourishing, eternal life. The good that is produced by spontaneity is the *created*. No good that is *merely* received is good enough, for Kant. Again, Kant thinks there is no higher good than morality, that nothing is really good that is not grounded in freedom. Contrast the 'unconditionally good' with the *unconditionally given*. Kant presupposes that the *given* is valueless, inert. As Korsgaard puts it, Kant's 'argument [in *Groundwork* I] is essentially that only human reason is in a position to confer value on the objects of human choice.'[729] Or again,

> [n]othing else justifies our ends and actions; it is our rational autonomy itself that does so.… [W]e regard them as good whenever they are chosen with full rational autonomy; so full rational autonomy itself is the source of their value. Since this holds for other rational beings as well as myself…so it turns out to be a good will that is the source of all value.[730]

'God' is the avatar of this 'all' that includes 'myself.' With respect to 'a good will that is the source of all value,' Korsgaard is right. Only agents that have been forgiven see the value of mercy. It is still possible to choose the standards of justice and of the good will instead of mercy. On this account the community is always split.

Indeed Kant *does* say in the second *Critique* that '[w]hat we are to call good must be an object of the faculty of desire in the judgment of every human being.'[731] Note that Kant does not say: 'in the judgment of every *kind* being.' Surely, then, this applies to the unhappiness of the immoral agent. And if, as Korsgaard argues, 'the reasons for "calling" a thing good must be universalizable,' then this unhappiness cannot be such an object unless it can be 'called' good for such universally shareable reasons. There are such reasons in the case of the happiness of moral agents. In the case of the happiness of an immoral agent, however, there are none: 'rational beings [alone]…determine what is good; rational nature confers value on the object of its choices and is itself the source of all value.'[732] Kant's God embodies this 'rational nature.' This leaves the immoral agent defenseless against her accusers and deprived of all refuge in the face of the unhappiness that she deserves.

---

[729] Korsgaard, 'Aristotle and Kant': 499.
[730] Ibid., 500.
[731] *KpV* 5: 61 (53).
[732] Korsgaard, 'Aristotle and Kant': 500.

The problem of the immoral, but happy agent: the special case of 'passive healing'

In this sub-section I show that the main consequence of Kant's denigration of mercy is that his God is barred, not merely from rendering immoral agents happy, but from intervening in their immorality, as it were, by healing their evil wills, rendering them good, and adding happiness to this healing as well. I show that, for Kant, an agent whose will's goodness is a consequence—in *any* way or degree—of another agent's intervention, such that, without this intervention, this transformation would never have been achieved, remains just as unworthy to be happy as she was when her will was an evil one. In short, I argue, the notion of unworthiness to be happy to which Kant's gloss adverts is connected to a notion of immorality that *subsumes* even the goodness of a good will whose goodness is a consequence of such healing.

Immoral and happy?

As I argued in chapter 2, it hardly seems likely—assuming that Kant does not mean by 'happiness' (*Glückseligkeit*) the mere pleasure of the moment—that the happiness of which Kant takes immoral agents to be unworthy is a state of affairs that could subsist for long in the presence of moral self-contempt. Kant describes this in particularly vivid language in his philosophy of religion lectures:

> We must not be blinded by the outward glitter that frequently surrounds the vicious person. If we look within, we read constantly…his reason's admission: *You are nevertheless a villain*. The restlessness of his conscience torments him constantly, agonizing reproaches torture him continually, and all his apparent good fortune is really only self-deceit and deception.[733]

Either Kant intends by 'happiness' something that is incompatible with the presence of self-contempt (not immorality, but this *affect* would be the problem here), or else he intends something weaker, to the point of that which inheres in his description of the fraud, the drunkard, and so forth.

But if such an agent's happiness, *qua* empirical, as satisfaction of desire or gratification of inclinations, were attended by her consciousness that, though happy, she was alienated from humanity, not worthy to be happy after all or, indeed, that she positively deserved to be unhappy, then what kind of happiness would that be? If immoral agents can be happy and happiness is impossible in the absence of (genuine) moral self-satisfaction, then there must be a sense of 'immoral' that is compatible

[733] *Vorlesungen-Religionslehre* 28: 1081 (413-14).

with being morally self-satisfied. If happiness is impossible in the presence of self-contempt and immoral agents can be happy, then either there is sense of 'immoral' that is compatible with the absence of self-contempt due to the presence of moral self-satisfaction *or* the only immoral and happy agents are ones that have no pangs of conscience even when they have reason to have them.

Here, as elsewhere, Kant evinces more than one tendency. Sometimes the two tendencies appear together. Just as Kant's causal and normative accounts of the connection between morality and happiness, or his distributive and retributive representations of justice, are not always clearly distinguished, representations of the 'capacity' for happiness are mixed up with talk of 'worthiness.'[734] Kant's main tendency, however, is to regard the relationship between morality and happiness in the very way that he appears to do when he deploys the idea of 'worthiness to be happy.' His main tendency is to regard morality as a necessary, *normative* condition bearing on the distribution, or the approval, of a mode of empirical well-being that may be achieved without reference to morality, through skill, prudence, 'rational self-love,' the cooperation of nature, of other human beings, or of God—an illegitimately procurable happiness.

The question of the possibility, for Kant, of an immoral, but happy agent is seldom raised as such.[735] As we have seen Kant's habitual glossing of 'morality' as 'worthiness to be happy' implicitly affirms this possibility. The famous opening lines of the *Groundwork*'s first section also express the view that such a thing is both thinkable and unconditionally, objectively bad. Evidently, here, Kant does not take '[t]he uninterrupted prosperity of a creature graced with no feature of a pure and good will' to be *unthinkable*.[736] If he did, he would not problematize this (perhaps counterfactual) prosperity as he does. To do so would be *de trop*; there would be no need. If Kant holds such 'prosperity' to be possible, then the happiness that is possi-

[734] See, for example, *LMM* 29: 907 (272-3).
[735] See, however, Hill, 'Happiness and Human Flourishing in Kant's Ethics': 144; Römpp, 'Kant's Ethics as a Philosophy of Happiness: Reflections on the "Reflexionen"': 275. More typically, Kant's readers implicitly affirm this possibility by referring to such things as happiness that lacks objective value or worth (Hill, 'Happiness and Human Flourishing in Kant's Ethics': 158; Korsgaard, 'Aristotle and Kant': 499; Sikka, 'On the Value of Happiness: Herder Contra Kant': 531; Wood, *Kant's Ethical Thought*, 128, 312); happiness that is morally 'bad' (Paton, 'Kant's Idea of the Good': xx; Wood, *Kant's Ethical Thought*, 24); or happiness, the means to, or the pursuit of, which conflicts with what morality requires (Acton, *Kant's Moral Philosophy*; Hills, 'Kant on Happiness and Reason': 243-4).
[736] *Gr* 4: 393 (7). See also *KrV* A813/B841; *Idee* 8: 26 (116); *R* 6317a (18: 632 [373-4]).

ble without morality cannot be a state of the agent's empirical affairs for which her morality is a transcendental or any other kind of necessary internal or causal condition.

When he thinks the relationship between morality and happiness by way of the notions of worthiness and unworthiness to be happy Kant does not think that the wish for happiness-without-morality is an incoherent one, based upon a misunderstanding. It is not like the case of an agent that aims at mutually exclusive gratifications (seeing the sunrise, sleeping in) or disregards the ways in which her own well-being shares some of its integral grounds with the well-being of other members of her community. Rather, he holds only that a fundamental, unwavering, irreversible commitment to morality is a condition independently of which it is wrong to *seek* happiness for oneself, or to approve of its (ultimate) *distribution* to others.

Kant's recourse, in *The Metaphysics of Morals*, to the very idea of an agent that can 'dispose of *all* happiness,' along with the possibility of a principle in accordance with which this would (perhaps counterfactually) be undertaken, implies that none of the necessary conditions for happiness are irreducibly and solely in the power of the kind of agent whose happiness is in question. This is implicit in what Kant says. We may conclude, again, that Kant does not take the agent's *morality*—the goodness of her will or of its character—nor the ascribability of this goodness to her, wholly and simply, to be transcendental, constitutive conditions without which her happiness is simply impossible.

If the notion of worthiness to be happy is not conceptually *de trop*, then, if we take this Kantian idiom seriously, affirming that it is more than a mere manner of speaking, then we have two options. Either we follow Wike and others and take 'happiness' in Kant to refer to the concept of a *system* of empirical ends, with reference to a logically and prudentially-practically integrated will (or such a will's 'willing'), but without logically intrinsic or analytic reference to Kant's critical notion of *morality*; or we take Kant not to have had *any* account of happiness that is compatible with the conception of the relationship between morality and happiness that is represented by the notion of worthiness to be happy. In short, happiness must not demand anything that pertains to *pure* practical reason; it must be compatible with

the immorality of the subject (in Kant's sense). *This* must be the happiness of which moral agents are worthy and immoral ones unworthy.

Kant's observations about the cheat and the 'upright man' show that it is in the nature of the satisfaction of specifically *human* beings (agents that are both 'earthly' and 'endowed with reason') that a lack of *fit* between the character in accordance with which an agent acts and the character of the empirical satisfaction that she enjoys tends to undermine this satisfaction. This is so irrespective of whether this is in consequence of her acting, or through some more haphazard natural or social arrangement. The gratification of the desires of a particular human animal has too shallow a reference to the specifically *human* good, in a sense, for it to interest the human being, *per se*, in the kind of sustained manner implicit in the framing of a whole life. This conclusion is based upon empirical observations that stand in need of further clarification and reinforcement, but seems quite plausible. Kant, however, thinks that the experience of moral self-judgment and moral self-contempt give expression, not to principles of empirical anthropology and psychology, but to *a priori* principles of pure, practical reason, principles that are aligned with what he takes to be the pure idea of '*nomos'* itself.

Kant does not demonstrate that this is so. I want to affirm the general claim, however, that the value of the human being's empirical 'well-being' or 'gratification' is not unconditional, neither to her, nor to others. It lacks objective value wherever there is a lack of *fit* between the empirical state of her affairs and the state of those transcendental, deed-grounding and deed-governing features of her agency that also figure in the way that she comes to find her life significant in the first place; her life regarded as a whole that is greater than the mere concatenation of her gratifications and frustrations, a life qualified as happy or unhappy.

But imagine, against the backdrop of this concession, that one agent says to another: 'I am neither happy, nor "fit" to be happy; but you (perhaps through my cooperative entanglement with you and others) can make me both happy and fit to be so.' Imagine, for example, the Christian believer who says: '*Domine, non sum dignus, ut intres sub tectum meum, sed tantum dic verbo, et sanabitur anima mea*.'[737] To say this is not to presuppose that I *could* have made myself worthy at some earlier time,

[737] 'Lord, I am not worthy that Thou shouldst enter under my roof: but only say the word and my soul shall be healed.'

but failed to do so. An agent may well say that she is not worthy to be happy if she has recourse to something like Kant's *notions* of worthiness and morality, perhaps because she has acquired these concepts from other members of her community, even if she is nevertheless, all the same, irremediably an agent that could *never* instantiate the kind of agency that is thought in this concept—and not through moral frailty or through the self-corruption of which she stands accused, but just because it is not in the nature of her agency as such.

Here, the sense of 'worthiness' implicated in the Christian believer's assertion that she is not '*dignus*' in no way resembles what Kant has in mind. This is nowhere more obvious than in the petition that follows the prayerful concession of unworthiness: 'but only say the word and my soul shall be healed.' Whatever else Kant means by 'worthiness to be happy' and 'morality,' he does not take the latter to be a qualification of the agent that could ever be achieved in her by *another*. In that case neither her happiness, nor the agent's 'dignity' would be unambiguously ascribable to the individual human agent as her product. The very possibility of individuation by way of action-ascription is, for Kant, of the essence of morality.

Kant thinks that the non-empirical condition of the deontic possibility of a 'grant' of happiness is an endogenous, or 'autogeneous' transformation of the agent. It cannot be, say, an exogenous transformation of the cooperating 'many' with whom she is entangled. Her morality is something that can be traced back to her, to the agent herself, without any social or natural remainder. The approach that is here implicitly rejected takes 'worthiness,' now an excessive designation, to be an exogenous, or heterogeneous outcome: a transformation of the agent—perhaps in and through the transformation of her social and natural situation as a whole—that is undertaken by another, or others, perhaps, or by the agent, but only by way of her cooperation with one or more simply indispensable partners. To be 'healed' and so to be made fit for happiness (prepared, as Blake says, to 'bear the beams of love'[738]) leaves one unworthy to be happy—and indeed immoral, in Kant's sense.

[738] William Blake, 'The Little Black Boy,' in *Songs of Innocence & Songs of Experience* (Mineola, NY: Dover Publications, 1992 [1789]), 10.

In the *Opus Postumum*, Kant asks 'Whether God could…give man a good will?' The answer: 'No,' he replies, 'rather, that requires freedom.'[739] The immoral agent is in a fundamental predicament, then, which cannot be resolved by way of God's mercy. How, then, is it to be resolved? In his draft 'catechism' Kant has the teacher ask, '[c]an another…make you worthy' and has the pupil answer, 'I must do it myself.'[740] Guyer observes that this question in particular

> makes it plain that worthiness is only earned or merited by one's own action, not by the action of another…. [And Kant] seems to assume that what makes you worthy of happiness is that you choose to act in a certain way even when you could have done otherwise. Merit depends on one's own action, and on one's free action; so *worthiness* to be happy can stem only from one's own free action.[741]

Leaving aside Guyer's reference to 'merit,' here, this must be generalized from action-on-an-occasion to the goodness of the will's character itself, its disposition towards unconditional obedience. As Kant puts it in the *Religion*, 'whatever does not originate from himself and his own freedom provides no remedy [*Ersatz*] for a lack in his morality.'[742] And for Kant, as Sikka points out, 'only acts of the will are truly one's own.' Indeed, '[t]he rest of the human psyche…[is] outside the will.'[743] Obviously, third parties are even more decisively alien than this. As Kant puts it in his *Conflict of the Faculties*, '[a]ction must be represented as issuing from the human being's own use of his moral powers, not as an effect [resulting] from the influence of an external, higher cause by whose activity the human being is passively healed.'[744] In a similar vein Kant says that

> [w]hen it is said that it is in itself a duty for a human being to make his end the perfection belonging to a human being as such (properly speaking, to humanity), this perfection must be put in what can result from his *deeds*, not in mere *gifts* for which he must be indebted to nature; for otherwise it would not be a duty.[745]

But note that the 'lack' (*Mangel*) that Kant describes is the particular problem that it is precisely to the extent that it is imputed to the agent with the ethical analogue of

---

[739] *OP* 21: 34 (237). In fact, the question concerns the human being's original *creation*, but the question—or, especially, its answer—is relevant subsequently as well.
[740] *R* 7315 19: 312.
[741] Guyer, *Kant on Freedom, Law, and Happiness*, 121.
[742] *Rel* 6: 3 (57). For an illuminating contrast with the approach of 'the ancients,' which demonstrates the radical character of Kant's innovation, see Hume, *An Equiry Concerning the Principles of Morals*, Appx. 4.20 (182-3).
[743] Sikka, 'On the Value of Happiness: Herder Contra Kant': 521.
[744] *Streit* 7: 42-3 (267). See also *ibid.* 7: 55-6 (277-8). See also Anderson and Bell, *Kant and Theology*, 68.
[745] *MdS* 6: 386 (150).

'rightful force.'[746] If the human being is to be transformed, if she is to find herself whole, saved from the shadow of desert in relation to her own unhappiness, her transformation must be imputable *to her*. It cannot be a matter of 'healing' or a gift of 'nature.' Otherwise, nothing will have changed, really, with respect to what is demanded by the law of unhappiness. The 'healed' agent, as much as the agent with an evil will, would still owe a debt—minimally for having ever chosen to have an evil will in the first place—and so still instantiate an unremedied gap with respect to the *ius talionis*. It is in the nature of 'worthiness to be happy' that God's 'external' influence on the human being could never make her worthy of anything that she was unworthy of before.[747] That is obvious enough. The issue, however, is whether the happiness of such an agent could count as good and whether God would be permitted to offer it to her. Kant's answer is that both the mercy of this transformation and the mercy of this gift of happiness are barred. Kant's 'woe' concerning 'the windings of eudaemonism' is addressed to these eschatological mercies as much as it is to the temptation to forego capital punishment in cases of murder.

For Kant, then, gifts bestowed on immoral agents are not good and the healing of their wills is without value. As Korsgaard puts it, Kant's critical view is that '[t]he only value there is is that which human beings give to their own lives.'[748] This is true of both morality and empirical well-being. The former is a function of freedom, all the way down, but it must be possible, too, to regard the latter as, in some essential respect, a state of affairs that is her very own unassisted production. Her goodness cannot be regarded as a gift, a gift that entails a transformation that also renders her both 'fit' for happiness and actually happy. On Kant's view, there can be no subject who, like Alastair Sims' Ebenezer Scrooge, rejoices in his happiness, while averring both that he does not 'deserve to be so happy' and that he 'cannot help it.'[749] There can be no such thing as a human subject whose fitness-for-happiness was

[746] See *MdS* 6: 227, 440 (19, 190). I should like to take this matter up in greater detail, but cannot, given the restrictions of this thesis. The basic idea is that imputation 'with rightful force' is a judgment that both imputes a deed to the agent and simultaneously sentences her to be punished.
[747] See Anderson and Bell, *Kant and Theology*, 69.
[748] Korsgaard, 'Aristotle and Kant': 505.
[749] Brian Desmond Hurst, "A Christmas Carol," (UK: United Artists, 1951). One can easily imagine what Kant would have to say about a subject, transformed in spite of himself, but also driven to conversion *and* happiness along the channels of his own past resistance, a subject who exclaims, 'I am as light as a feather, I am as happy as an angel, I am as merry as a schoolboy. I am as giddy as a drunken man…. I don't know what day of the month it is...I don't know how long I've been among the Spirits. I don't know anything. I'm quite a baby. Never mind. I don't care. I'd rather be a baby' (ibid.).

achieved through cooperation with, or surrender to, another (or others). The Christian subject who says, '*Domine, non sum dignus*' (etc.), not only remains, on the present account, unworthy of happiness (regarded, say, as intimate reception of such a subject's 'Lord'), 'healed' though she be—but her happiness is objectively bad, evil, forbidden. What is missing in these instances is not the transformation of the subject, the subsequent goodness of her will (ultimately, say), but rather this transformation's being her own unassisted achievement, so that the order in her life, the integrity of her will, is an effect of her own spontaneous activity. Nothing strictly given one, received by one, can be regarded as fundamentally, or 'originally' good.[750] Only the free deliverances of a will constrained by duty, on the one hand, and justice, on the other are good.

Think of an agent whose heart was so transformed that she had been rendered unfailingly kind-hearted, unfailingly merciful, who loved others as she did herself, but whose entire motivation was her loving care for others, her identification with them in their joys and sorrows, and so on: the happiness of this agent would be repugnant to reason. In other words, you and I, in Kant's ultimate view, ought to disapprove of the prospective happiness of such an agent. The problem, however, is simply this: the agent's transformation is not an independent achievement; she had help and the help that she had was (*ex hypothesi*) absolutely integral for her being changed.

Suppose that one's empirical interests were fully served, every inclination that one happens to have gratified, and every source of empirical distress occluded. It is important to be clear that Kant is not saying that if this gratification or occlusion were secured through immorality, then a more profound moral suffering would persist in the agent's consciousness, right alongside the enjoyment associated with the gratification of her inclinations. This is merely to claim that the agent's full satisfaction in its empirical and moral registers requires that she be free of any sense that she has secured the gratification of her desires through immorality, or at the expense of others, or in a manner that blocks others' access to happiness. Hill makes a point along these lines, writing that, for Kant, 'none of us…could live without inner conflict and self-disapproval if we pursued personal happiness by plainly immoral

[750] See *R* 7196 19: 270 (461).

means.'[751] But this is not precise enough. It is not just that one's achieving happiness by immoral means undermines happiness; nor, in any case, does becoming happy by moral means, just as such, render happiness secure. The point is that no matter *how* one becomes happy—whether by fortune, or by the intervention of a superhuman agent, or by effort that conforms perfectly to the conditions of the possibility of human beings (not just oneself) being happy in general—an 'I am moral' must be able to accompany one's empirical qualification as 'happy.' And here, for Kant, 'morality' references not merely the goodness of the will's character, but the radical ascribability *to her* of the sheer fact that her will is a good one. The *moral* goodness of the good will is not its property of being a will that wills what is good, just as such. Rather, it is this property under the limiting condition that the will's having this property at all be ascribable to the agent herself. This property of being-ascribable is something that the good and the evil of good and evil wills have in common, on Kant's account. Good is not good without it, evil not evil.[752]

To be healed is thus an intolerable shortcut. Indeed, the hope for healing is immoral and forbidden. The rejoicing of a passively healed agent, or of one that gets a glimpse of such healing, or of one that takes herself to be caught up in a process of being healed, is incompatible with Kant's disapprobation of her happiness. If the rejoicing were justified, then this disapprobation could not be. Kant holds, however, that the objective badness of the happiness of such an agent is obvious. And there is a sense in which he is right about this—if we take mercy-excluding penal justice to be good in itself. But if the objective badness of the happiness of the 'healed' agent is conceded, then the value inherent in the objectivity of the judgment that deems it so is not a value whose endorsement is consistent with a Christian understanding of divine mercy. The issue is not that Kant says that happiness is impossible without a special receptivity to the gift of happiness, nor that he identifies morality with efforts aimed at becoming thus receptive. The problem is that Kant does not allow—and takes himself to have a rational warrant for disallowing—that in the event that an agent fails to make herself adequately receptive to happiness, God might be regarded as an agent that would willingly assist her in this, *irrespective* of what she has done so far and of how hard *she* has tried.

[751] Hill, 'Happiness and Human Flourishing in Kant's Ethics': 174-5.
[752] See Guyer, *Kant on Freedom, Law, and Happiness*, 121.

Conclusion

In this chapter, I executed four main tasks. First, I reviewed the practical significance of Kant's law of punishment. Next, I demonstrated that Kant's commitment to the thesis that immoral agents ought to be unhappy is a fundamentally practical one as well. Third, I further specified this claim by arguing that Kant's law of unhappiness is binding on two kinds of agent, human and divine, and that it is binding in two distinct contexts. I characterized the first of these as mundane, the other as eschatological and I argued that these two contexts correspond to two distinct ways in which the law of unhappiness is put into practice. Finally, I argued that these two contexts *coincide* to the extent that, under the law of unhappiness (but not under the primary moral law), we, together with God, are members of a single community. I described this community as an order for which the law of unhappiness is uniquely constitutive and I showed that, for Kant, this community must be an absolutely mercy-free zone.

I reject this representation of the human community. By doing so, however, I am not claiming that Kant is wrong about the goodness of justice in some general sense. I am only suggesting that his taking mercy to be immoral (and so, in his terms, rooted in evil) shows that his notion of the good is not a Christian one. The Christian notion of divine mercy is a threat to the practice of punishment, as Kant conceives of it, and, theoretically speaking, a threat to the justification of punishment in retributivist terms. This, I suggest, is as it should be.

Kant insists that God's involvement at the origin of the human being *qua* moral agent 'is not thinkable,' irrespective of whether this involvement be direct and 'miraculous,' or mediated by the social and natural order in which the human creature is embedded. The radical involvement, at the source of specifically human action as such, of 'something' not identical to, but nevertheless incorporating, 'the one that did it,' would render human action 'mechanism and not freedom,' Kant thinks. He emphasises the central point, here, by saying that '[m]an is himself regarded as cause of his actions that take place in the world.'[753] To repeat an earlier cited formulation of this integral focus: '[t]he human being must make or have made *himself* into whatever he is or should become in a moral sense, good or evil';[754] and 'what is to be im-

[753] *Fort* 20: 336 (414).
[754] *Rel* 6: 44 (89).

puted to us as morally good conduct must take place not through foreign influence but through the use of our own powers.'[755] Moves like these block access to the tools for thinking and hoping that inhere in the Christian notions of original sin, grace, creator-creature cooperation (*divine concursus*), redemption, and conversion—tools for thinking the possibility of mercy and for hoping that this possibility can never be limited by mere human trespass.

[755] *Rel* 6: 191 (207).

CONCLUSION

In this thesis I have argued that Kant's habit of glossing 'morality' as 'worthiness to be happy' discloses the presence of an implicit, durable cluster of unjustified commitments. I showed that these are, at bottom, an immediate devotion to a particular set of practices: on the one hand, that is, the practice of capital punishment and, on the other, the practice—'virtual' and 'deferred' for us, but taken up eschatologically by Kant's 'God'—that occludes immoral agents' access to happiness. I demonstrated that Kant's steadfastness in relation to these commitments (or to what remains, at least, unexamined *in* them) both antedates and survives his 'critical period.'

The law of unhappiness is practically significant if there is a subject that it binds and a practice that it prescribes. I have shown that, for Kant, there is such a subject and such a practice. What, however, of the object of this practice? Is the human being really an agent such that the goodness and badness (or evil, in Kant's sense) of her will may be ascribed to her as the consequence of an utterly unassisted undertaking on her part? In the *Religion*, as we have seen, Kant claims that the human being is such an agent. In this way, he affirms that the practice that puts the law of unhappiness into effect has an object.

Space does not permit me to take the next step here, to orient the results of my discussion, so far, to the ultimate standard that such an object must meet: a particular conception of her susceptibility to the imputation of deeds, a standard that is equivalent to a particular way of being free. An adequate discussion of the latter topic, beginning with the question, 'What does the imputability of deeds consist in for Kant?' lies without the scope of this thesis. The latter merely sets the scene for this inquiry, as well as for an investigation of other questions concerning Kant's thinking about freedom.

In my introduction I pointed to the fundamental motivation for the task executed in the foregoing chapters. I claimed that there are good reasons for thinking that the

commitments that I describe condition the trajectory by which Kant's 'critical' thinking about freedom and moral accountability passes from its initial state in about the 'Third Antinomy' of the first *Critique* to its final state in the *Religion*'s theory of radical evil. I cannot offer, here, a robust demonstration of the claim that the commitments expressed by Kant's habit warp the conceptual space within which his thinking about moral agency develops in the critical period. I cannot show now, decisively, that the Kantian commitments that I have identified operate like an enormous, but mostly invisible, mass that distorts the theoretical space within which his thinking develops—that the final disposition of Kant's theory of moral accountability, as embodied in his theory of radical evil, takes the form that it does as a consequence of the constant proximity of this 'black star.'

All of this remains to be shown. This thesis sets the scene, merely, for a story of the development of Kant's thinking about freedom and moral accountability that remains to be told—a story whose climactic moment would be Kant's *Groundwork* (re)description of 'matters of morality' (*Sittlichen*) as a *topos* in which 'imitation' (*Nachahmung*) simply does not 'take place' (*findet…gar nicht statt*) and whose denouement would be the developments leading to the *Religion*'s description of the human being as an agent whose radical 'disposition' (*Gesinnung*), the very character from which she acts, is imputable to her alone, wholly and simply. Indeed, these two moments mark the implicit horizon within which the argument of this thesis has unfolded, fundamental concerns that I have had to exclude almost entirely from my presentation.

By way of conclusion, however, I will offer a sketch of the account that builds upon the point of departure that I have established here. This sketch will situate the contribution that this thesis makes (in itself, in terms of the treatment of Kant's 'gloss,' the significance of his retributivism, the consequences of the latter for his thinking about divine mercy) relative to the further avenues of research to which this project is already oriented.

The first of these pertains to the distinction (and relationship) between what I have referred to in this thesis as 'forward-looking' (or 'primary') and 'backward-looking' (or 'secondary') senses of Kantian 'morality': on the one hand, that is, the *ab origine* open-ended, forward-looking inquiry that is inspired by the primary ques-

tion of ethics, 'What ought I to do?'; on the other hand, the backward-looking discipline inspired by Kant's always already settled *answer* to the question, 'How are happiness and morality related?' Together, these two perspectives express an important bifurcation in Kant's thinking.

In this thesis, I unpacked the first component of the implicit cluster of commitments that I claim are present whenever Kant glosses 'morality' as 'worthiness to be happy.' I showed that these commitments are most fundamentally speaking commitments to particular practices or, at least, to a set of 'virtually' practical attitudes. There is more to be said on this score, however—in relation, that is, to Kant's deep practical commitments. The practice of capital punishment, on the one hand, and the practical attitude of approbation with respect to such punishment, along with the human being's deferral to God's ultimate treatment of immoral agents, on the other, do not constitute the most *primary* practical objective to which Kant is committed. The primary, most immediately proximate practice—the practical threshold across which the human community must pass in order to get to the point, say, of regarding the putting-to-death of one of its members as an undertaking that is justified with reference exclusively to *that one* and *what she did* (hence retributivistically)—is a particular inflection of the practice of *imputation*. This is the second component of the core of practical commitments that Kant brings along, by way of his habitual gloss, into his critical period thinking about human agency.

Kant expresses the decisiveness of this focus relatively late in his career when, in a draft version of his contest essay on 'Progress in Metaphysics,' he identifies the '[o]rigin of the critical philosophy' as 'morality, in regard to the imputability of actions [*der Zurechnungsfähigkeit der Handlungen*].'[756] In spite of the rudimentary form of the source here, this concession is very illuminating: Kant retrospectively (ca. 1791) characterizes the 'origin' (*Ursprung*) of the 'critical philosophy,' in terms of an antecedent orientation towards 'morals' (*Moral*) whose primary focus is not the objective *goodness*, but rather the '*imputability*' (*Zurechnungsfähigkeit*) of actions.

Generally speaking, accounts of the course of developments leading into and through Kant's 'critical' thinking about morality regard his interest in the latter in terms amenable to the tradition of enquiry and critique embodied in the projects of

[756] *Fort* 20: 335 (413-14) (translation modified).

influential predecessors such as Hume, Shaftesbury, Hutcheson, and Smith.[757] Here, Kant's main focus may be described in terms, for example, of 'the foundations of morals,'[758] 'the "roots of oughtness,"'[759] or the 'source' of 'our obligations.'[760]

However, if we take seriously Kant's revelation concerning the '*Ursprung*' of the critical philosophy, as specified above, we find 'morals' construed as a science focused, not only on the question of what finite rational agents ought unconditionally to do (or to have done), but 'morals' regarded as a disciplined thinking about practical agency that is always already oriented to the practice of imputation. I do not want to over-stress Kant's concession, here. It strikes me, however, as extremely compelling. And there is more to my claim than this, something else that warrants my foregrounding this retrospective assessment from Kant's own pen. For this is the sense of 'morals' or 'morality' that Kant deploys whenever he glosses the latter as 'worthiness to be happy.'

Kant's retrospective assessment confirms that the durable backward-looking perspective has theoretical priority over the open-ended forward-looking one. And it is incumbent on Kant's readers, I suggest, to give the backward-looking perspective its full and distinctive due when reflecting on the development of Kant's thinking about freedom and moral accountability. The path that leads to the *Religion*'s account of the imputable '*Gesinnung*' is not simply an extension of the path that passes through the *Groundwork*'s 'invention of autonomy,'[761] for example. In a sense, I propose, both paths—the paths along which Kant pursues the interests of forward- and backward-looking morality respectively—pass together through this and a number of other key points (I catalogue some of these below), overdetermining them. Indeed, I would suggest that the path that Kant pursues in the interest of his forward-looking inquiry *terminates* (roughly speaking) with his theory of autonomy.

---

[757] See, for example, Lewis White Beck, *Early German Philosophy: Kant and His Predecessors* (Cambridge, MA: Thoemmes Press, 1996); Darwall, 'Norm and Normativity'; Norton and Kuehn, 'The Foundations of Morality'; J. B. Schneewind, 'Autonomy, Obligation, and Virtue,' in *The Cambridge Companion to Kant*, ed. P. Guyer (New York: Cambridge University Press, 1992); J. B. Schneewind, *The Invention of Autonomy: A History of Modern Moral Philosophy* (Cambridge: Cambridge University Press, 1997).

[758] Norton and Kuehn, 'The Foundations of Morality', 976. With respect to the relationship between Kant and his British and German predecessors see ibid., 976ff.

[759] Darwall, 'Norm and Normativity', 988.

[760] Neiman, *The Unity of Reason*, 106.

[761] The phrase is Schneewind's (see Schneewind, *The Invention of Autonomy: A History of Modern Moral Philosophy*).

An important task for the future, then, would be to show how Kant's worries about the backward-looking problems of imputation and punishment orient his thinking in the direction of the *Religion*'s notion of the imputable '*Gesinnung*.' In both its forward- and backward-looking senses, Kant is aware that 'morality' could turn out to be a *chimaera*—something to which we only *seem* to be called, given that it entails our doing something that we are not in fact able to do (and given that, for Kant, 'ought implies can').[762] This worry is common to both kinds of moral inquiry, but evinces a distinct peril in each case.

The topic of imputation is a central element in discussions of law and its application, particularly as this relates to questions of desert, punishment and reward. As we saw in chapters 3 and 4, the influence and contemporary significance of Kant's thinking is evident in recent work in this area. It is at least atypical, however, to argue that, for Kant, concerns bearing on the problem of imputation and imputability play a decisive, fundamental role in the development of his thinking about morality, *in general*, as I am suggesting now, and not merely in his thinking about 'rightful,' 'external' relations in civil society.[763] It is more typical to hold, for example, that 'Kant did not focus much on issues of moral praise and blame.'[764]

I have shown that this is false: Kant's thinking about morality, when this has primary reference to the 'internal' scene of 'ethics' and not merely to the 'external' one of 'rights,' has constant reference, too, to the categorically imperative 'law of unhappiness,' which is directly analogous to the civil-political 'law of punishment.' Indeed, we saw that in (non-dogmatic, practically oriented) theological-eschatological contexts, Kant uses the language of punishment in relation to immorality as such—i.e., where the latter is regarded as unworthiness to be happy, or intrinsic desert of unhappiness. I argued that here, at least, Kant's thinking about the ethical is based on his understanding of the political. *Pace* Hill, not only is their 'room for moralistic praising and blaming'[765] in Kant's ethics, but the stability of the political practice of blaming and punishing murderers (at *least*) for their crimes is of

[762] See *Gr* 4: 402, 445 (15, 51); *ZeF* 8: 368 (337).
[763] See, however, Dean Moyar, 'Practical Apperception: Self-Imputation and Moral Judgment,' in *Law and Peace in Kant's Philosophy / Recht Und Frieden in Der Philosophie Kants (Akten Des X. Internationalen Kant-Kongresses)*, ed. Valerio Rohden et al. (Berlin: Walter de Gruyter, 2008), 281-2.
[764] Hill, 'Is a Good Will Overrated?', 55. See also ibid., 56; Hill, *Dignity and Practical Reason* 176-95; Korsgaard, 'Creating the Kingdom of Ends', 189.
[765] Hill, *Dignity and Practical Reason* 176-95.

fundamental interest to Kant. Kant thinks that the wicked *ought* to suffer, that they ought not to be happy—but he thinks this is so in a specifically moral, hence, intrinsic and categorical sense, as something that is good in itself.

Hill, for example, does not see the eschatological character of praise and blame, as Kant conceives of these, and the way that this eschatology arises from a pressure, as it were, a pushing forward, in Kant's thinking, from his political into his ethical theory, which shapes the latter's development. Hill's view (tempered, admittedly, by his concession of 'surprise' in this regard) is simply mistaken. The bare possibility of 'moralistic praising and blaming' is very important for Kant. This means, too, that the conditions of the possibility of imputation are also a key object of inquiry.

Again, the practice of imputation, rather than the practice of (ultimate) punishment, which is mostly 'virtual' and 'deferred,' is the most immediate or proximate focus for Kant's thinking about freedom. Imputation is a major focus for him. But his commitment to the practice of capital punishment and to the practical attitude of approbation with respect to such punishment, along with the human being's deferral of the ultimate treatment of immoral agents to God, are the source of the *urgency* that characterizes Kant's late claims concerning, especially, the imputability of the human being's underlying moral disposition to *her* alone.

Building upon the work done in this thesis, then, the next step would be to explore Kant's account of imputation and to connect the latter both with these other commitments and with his thinking about freedom. I foresee three main foci here. First, I would rehearse Kant's analysis of imputation into its two constitutive moments: *action-ascription*, which is also agent-identification, and *action-qualification*, or the subsumption of particular actions under the concepts or categories that they are taken to instantiate.[766] Next, I would discuss Kant's descriptions of our actual practice of imputation, along with his unearthing of the 'thoughts' that he takes to be implicit in it.[767] Finally, I would establish that, as 'imputation with rightful force,'[768] imputation is the *avant garde*, as it were, of the movement by which Kant's thinking about the political punctures and extends into his thinking about the ethical and the

[766] See, for example, *MdS* 6: 223, 227, 438 (16, 19, 189); *Vorarbeiten-MdS* 23: 245; *LEC* 27: 288 (80-1).
[767] See, for example, *KrV* A554-6/B582-4; *KpV* 5: 99-100 (83-4).
[768] See *MdS* 6: 227, 438, 440 (19, 189-190).

eschatological. Here, the relevant *topos* would turn out to circumscribe, not merely the practice of capital punishment, but the antecedent practice of imputation.

With this reading of Kant on imputation in hand, a further step would be to advert to one of the main consequences of his analysis of this notion and its practice. I call this Kant's '*casus datae legis* problem.' This problem bears, as its name suggests, on the requirement that, in addition to being *ascribable* to an agent (*qua* their author or cause), imputable actions must always be regarded as cases that instantiate (or fall foul of) a given law, that fall within or without a given legal or ethical category, or as instances of a juridically or ethically salient kind, falling under some description or other, and so on. Because these two elements of the practice of imputation place distinct, and at points conflicting, demands on Kant's thinking about freedom, the *casus datae legis* problem is a key aspect of the material that fuels the thinking that ends in his theory of radical evil. I suggest that, in the *casus datae legis* problem, we have a key that will help us to identify the conditions under which Kant would have to concede that his backward-looking notion of morality, in particular, had turned out to be a *chimaera*. I contend, that is, that Kant would have to concede that his backward-looking morality (in the sense that we get, again, whenever he glosses 'morality' as 'worthiness to be happy') was a *chimaera* if he could not show that the human agent was susceptible to the imputation of her actions in a sense that goes *all the way down* with respect to *both* of imputation's moments (action-ascription, action-qualification). I would suggest, indeed, that Kant never does manage to show that human beings are like this. He simply insists upon it, particularly and ultimately, in his theory of the imputable '*Gesinnung*.'

Ameriks speaks of 'the conception of the human as free being that is the basis of Kant's moral philosophy.'[769] And he affirms 'the central role of the notion of freedom in the development of Kant's entire system' and 'Kant's attachment to absolute freedom.'[770] But which 'freedom' constitutes this 'basis'? Is it Kant's notion of 'absolute causal spontaneity' (which he calls the 'the real ground of [any action's] imputability'[771]) or is it 'autonomy'? And if 'moral philosophy' is at issue, then 'moral' with which orientation and emphasis? I propose that the terminus towards which

[769] Neiman, *The Unity of Reason*, 120.
[770] Karl Ameriks, *Interpreting Kant's Critiques* (Oxford: Clarendon Press, 2003), 162.
[771] *KrV* A448/B476. See also *Versuch* 2: 202-3, 239-40; *LMH* 28: 12; *LMV* 28: 404.

Kant's thinking tends here (the theory of radical evil) is not a consequence of his commitment to the thesis that 'freedom is autonomy.' Rather, it derives from his commitment to the thesis that '[f]reedom is imputability [*Freiheit ist die Zurechnungsfähigkeit*].'[772]

Kant sees 'absolute causal spontaneity' as a necessary condition for action-*ascription*, for the absolutely unambiguous identification of 'the one that did it.' And he recognizes too, from the outset, that this gives rise to a problem for action-*qualification*, to the extent that the individual deliverances of a merely lawless will cannot be regarded as instances of any kind in particular.[773] Kant's notion of absolute causal spontaneity refers to the absolutely unconditioned originality of a being that is able to bring certain phenomena, namely her actions, into existence *ex nihilo*. An agent that can do *this* is one to whom actions may be ascribed wholly and simply, without remainder. An immediate problem arises, however, with respect to the realization of this condition. I call this Kant's 'nomological problem' since it pertains to the question whether the kind of freedom that is required for action-*ascription* is a law-governed, hence intelligible, form of causality, or whether it is merely a kind of unfettered, lawless surging. I suggest that Kant's answer to this question bears directly on his solution to the *casus datae legis* problem.

This is because Kant recognizes that if the unhappiness of immoral agents is regarded as retribution—that is, as good in itself, as perfectly proportionate to the conduct that grounds it, and as addressed, wholly and simply, to the unique subject of that conduct—then, when it comes to the *qualification* of individual human actions in moral terms, reference to any nomological constraints not of the agent's own leg-

---

[772] *Vorarbeiten-MdS* 23: 245.

[773] In the Dialectic section of the first *Critique* Kant evinces two tendencies with respect to the question whether 'absolute causal spontaneity' entails the sheer lawlessness of the will endowed with it. The argument for the antithesis of the Third Antinomy seems to presuppose that the parties to the conflict (the proponents of the two sides of the antinomy) simply agree that transcendental freedom, or spontaneity, is a matter of unqualified *lawlessness* (*KrV* A447/B475; see also Michelle Kosch, *Freedom and Reason in Kant, Schelling, and Kierkegaard* (Oxford: Oxford University Press, 2006), 33). The basic opposition that emerges in this context is a dichotomy between a causality that is governed by laws (*simpliciter*) and one that is not. In his account of empirical and intelligible 'character,' later in the Dialectic, however, Kant experiments with a different approach (KrV A538-47, 557/B566-75ff, 585; see also *Fort* 20: 336 (413); Henry E. Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, Rev. and enl. ed. (New Haven: Yale University Press, 2004), 10). I cannot go deeper into this here. Kant's position at this point is unstable in any case, I suggest, and he is actively working on a problem whose solution does not come onto the scene until his 'invention of autonomy' in the *Groundwork*.

islating (physical, social, psychological, spiritual, religious, or any other kinds of heteronomous legislation) belies the radical individualism that his eschatologically inflected retributivism demands.[774] In other words, although Kant's commitment to the practice of imputation is a function of his antecedent commitment to the practice of punishment, as expressed (indirectly) in his use of the 'worthiness to be happy' idiom, his understanding of imputation (given its two distinct moments, action-ascription and action-qualification) raises a problem for the practice of punishment or making-unhappy—precisely, again, given Kant's specifically *retributivist* understanding and justification of it. I suggest that the *casus datae legis* problem is the problem (for punishment *qua* retribution) to which Kant's theory of imputable '*Gesinnung*' is supposed to be a solution.

One of the upshots of this solution, I suggest, is that, to the extent that punishment and unhappiness are regarded as retribution, the claim that human beings are ineluctably *heteronomous* agents must be problematic for Kant, from the outset—independently of his invention and elevation of the notion of autonomy. I would sketch the generative consequences of the problem of heteronomy (as the problem, say, of nomological plurality, cooperation, collectivity, mutual participation, imitation, and other forms of mimetic entanglement) in terms of the moves that Kant makes in order to remain one step ahead of it. To this end, I would argue that the analysis of imputation into its two moments, action-ascription and action-qualification, follows the contours of a distinction between two aspects of human agency, which might be described under the rubric of '*nomos*' and '*dynamis*.'

In the next phase of the account that I am outlining here, then, I would argue that Kant's retribution-sensitive understanding of imputation demands a truly radical solution, as follows. Among the habits, patterns, rules, and principles (in sum, the *'nomoi*') relative to which an agent's action (regarded, *dynamically*, as something of which she is the unique author) can be said to be a determinate instance of some kind or other, are the habits, patterns, rules, and principles that constitute her character. Thus Kant is compelled, having made a variety of other moves in this direction, to say that the human being is endowed with an underlying disposition whose instaura-

[774] Kant expresses his basic presupposition concerning this radical discreteness of the human being when, in his *Anthropology*, he declares that, 'taken collectively,' the human race is fundamentally 'a multitude of persons, existing successively and side by side' (*Anthro* 7: 331 [236]).

tion, configuration, and decisively rigorous orientation towards good or evil, is *her* doing, one of her deeds, and hence an originary, radical constraint whose functioning allows us to say in a perfectly decisive manner that not merely each deed, but the very being-moral or being-immoral of every one of her particular deeds, is entirely her fault.

There are two ways then, relative to '*nomos*' and '*dynamis*,' respectively, in which a human being might turn out *not* to be 'the one that did it.' And, I contend, Kant has to show that neither of these characterizes human agency across the board—else backward-looking morality, in his sense, would turn out to be a *chimaera*. One way is not to have done 'it' (*dynamically* speaking)—whatever it is—at all. The other way is to have done 'it'—so that this action *can* be unambiguously ascribed to you as its author—but to have done this thing too, whatever it is, in accordance with principles ('*nomoi*,' principles relative to which it is an instance of the kind that it is) that cannot be taken to have been legislated by you *qua* individual human being. Then, while you will have done 'it,' you cannot be taken to have done 'it' on your own, alone—radically and utterly alone—under any description of what you have done that has integral reference to those principles. Kant sees this possibility and it worries him.

What Kant requires then, I suggest, is an account of human agency such that, when it comes to '*Sittlichen*'—that is, to her morality or to her immorality—the human being is unassisted *in every respect*. She acts without any dynamical assistance whatsoever (without, say, anything like a push or a pull) and she legislates, originarily and decisively, that her actions will be exactly and *necessarily* whatever kind of deeds they turn out to be, morally speaking—whether good or evil. As I said in my introduction, we might refer to this ideal as the originary and terminal unity of the acting and suffering subject: 'the one that did it' and 'the one that ought to be unhappy' are an utterly consolidated unit, one *dynamis*, one *nomos*, without remainder.

Kant's account of '*Gesinnung*' may be regarded as a solution that, by excluding all reference to other agents, human or divine, is tailored to a particular line of questioning that might be articulated in defense of 'the accused.' This line of questioning might run as follows. In what sense and on what grounds may the individual human agent be, first, held accountable for her actions and then, second, judged deserving of

unhappiness in light of them, if the specific form of her practical subjectivity (her fundamental character, or the complex of reasons, laws, loves, or principles in accordance with which she is motivated to act) is radically 'unoriginal'? In other words, what if she is constituted from the ground up through her mimetic entanglement with various models—not themselves chosen by her—from the moment of her birth; models, moreover, whose own practical activity is governed antecedently in the same manner? Then, even if she may be regarded, *dynamically speaking*, as the independent source of her own deeds, so that these are ascribable to her in some sense uniquely, does her motivational dependence upon others for shared 'reasons for acting' not mean, nevertheless, that there is always a legislative remainder, a logic that is prior to her particular undertakings, whose origin cannot be so ascribed?

The tasks that I have executed in this thesis set the scene for an account that would show why and how, when it comes to the problem of heteronomous agency, Kant's decisive theoretical elimination of imitation (*Nachahmung*) is the climactic, the most characteristic, and the most troubling of his moves in service of that problem's resolution. In this particular regard, this thesis opens onto a number of additional tasks. There is the task of saying what the specifically intersubjective phenomenon of '*Nachahmung*' consists in, for Kant, and showing how his treatment of the latter relates to his thinking about other, related phenomena: '*Nachmachung*,' '*Nachäffung*,' and, especially, '*Nachfolge*.' There is the task of showing how Kant's resolution of the problem posed by imitation has specific, pernicious theoretical and practical consequences for theology, anthropology, politics, and ethics. And there is the task of connecting the anthropological *datum* that human beings are born imitators (which Kant readily concedes) with an account of imputation and 'punishability' that comes to the defense of the human subject by declaring that mercy is a good.

Kant's work contributes to the occlusion of earlier resources for thinking about the consequences of this fact for our understanding of imputation and punishment. In particular, the significance, sources, and consequences of Kant's trivialization of imitation cannot be adequately identified or addressed without recourse to theological

tools with which many (but by no means all[775]) serious readers of Kant have dispensed: the doctrines, especially, of grace and original sin.

Instead, commentators often focus on Kant's view of imitation in relation to the themes of genius and fine art, particularly as expressed in the third *Critique*,[776] or in relation to his thinking about the special *a priori* status of philosophy's subject matter.[777] Few attend to Kant's thinking about imitation as a specifically intersubjective phenomenon,[778] while those who do omit to connect the latter with the practical problems of imputation and punishment.[779] A number of others, however, provide rich resources for work in this area, albeit without immediate reference to Kant in every case. Interdisciplinary work aimed at repairing Kant's extremely influential thinking about human agency might make use, especially, of key figures such as René Girard[780] and Horkheimer and Adorno,[781] as well as a growing body of re-

[775] Rowan Williams provides a probing example of the alternative when he observes that 'human activity is misunderstood if it is seen as a sequence of "responsible" decisions taken by conscious and self-aware persons in control of their lives. More often it is a confused, partly conscious, partly instinctive response to the givenness of a world we do not dominate, a world of histories and ideas, languages and societies, structures we have not built' (Rowan Williams, 'An Enemy Hath Done This,' in *A Ray of Darkness* (Cambridge, MA: Cowley Publications, 1995), 76).

[776] Henry E. Allison, *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment*, Modern European Philosophy (Cambridge: Cambridge University Press, 2001); Andrew Bowie, *Aesthetics and Subjectivity: From Kant to Nietzsche*, 2nd ed. (Manchester: Manchester University Press, 2003); Martin Gammon, '"Exemplary Originality": Kant on Genius and Imitation,' *Journal of the History of Philosophy* 35, no. 4 (1997); Timothy Gould, 'The Audience of Originality,' in *Essays in Kant's Aesthetics*, ed. Ted Cohen and P. Guyer (Chicago: University of Chicago Press, 1982); Stephen Halliwell, *The Aesthetics of Mimesis: Ancient Texts and Modern Problems* (Princeton, NJ: Princeton University Press, 2002); F. Hughes, 'Taste as Productive Mimesis,' *Journal of the British Society for Phenomenology* 37, no. 3 (2006); Tom Huhn, *Imitation and Society: The Persistence of Mimesis in the Aesthetics of Burke, Hogarth, and Kant*, Literature and Philosophy (University Park, PA: Pennsylvania State University Press, 2006); Rudolf Makkreel, 'Reflection, Reflective Judgment, and Aesthetic Exemplarity,' in *Aesthetics and Cognition in Kant's Critical Philosophy*, ed. Rebecca Kukla (Cambridge: Cambridge University Press, 2006).

[777] R. Piercey, 'Active Mimesis and the Art of History of Philosophy,' *International Philosophical Quarterly* 43, no. 1 (2003).

[778] See, however, Günther Buck, 'Kants Lehre Vom Exempel,' *Archiv für Begriffgeschichte* 11, no. (1967); Jacques Derrida, 'Economimesis,' *Diacritics* 11, no. 2 (1981); Gammon, '"Exemplary Originality": Kant on Genius and Imitation'; Huhn, *Imitation and Society: The Persistence of Mimesis in the Aesthetics of Burke, Hogarth, and Kant*; Robert B. Louden, 'Go-Carts of Judgment: Exemplars in Kantian Moral Education,' *Archiv für Geschichte der Philosophie* 74, no. 3 (1992); Daniel Whistler, 'Kant's *Imitatio Christi*,' *International Journal for Philosophy of Religion* 67, no. (2010); Dieter Witschen, 'Nicht Nachahmung, Sondern Nachfolge: Kants Reflexionen Zum Ethischen Exempel,' *Zeitschrift für katholische Theologie* 130, no. 3 (2008).

[779] For an exception in this regard see Palaver, 'Mimesis and Scapegoating '.

[780] See, for example, René Girard, *Deceit, Desire, and the Novel: Self and Other in Literary Structure*, trans., Yvonne Freccero (Baltimore: John Hopkins Press, 1965); René Girard, *Violence and the Sacred*, trans., Patrick Gregory (Baltimore: Johns Hopkins University Press, 1977); René Girard, Jean-Michel Oughourlian, and Guy Lefort, *Things Hidden since the Foundation of the World:*

search on imitation and cognate phenomena (e.g., 'joint attention') in psychology, neuroscience, and related disciplines.[782]

This thesis opens onto further work, then, that would foreground the phenomenon of imitation as a mode of deeply social interaction and, more precisely, as the ground of an especially social mode of entangled, heteronomous agency. Future research along these lines would offer an alternative to Kant's assessment of 'imitation' in '*Sittlichen*' that is both cognizant of his central importance for thinking about human agency, but that also holds open the possibility that the latter's moral dimension is not merely compatible with intersubjective imitation, but, for better and for worse, profoundly conditioned and sustained through a radical, mutual 'motivational' entanglement of agents that takes place by means of it.

To the extent that this future work would seek, too, to build upon and deploy Kant's 'critical' insights into the structure and limits of human cognition, it would not entail a wholesale repudiation of the key distinctions to which Kant's mature philosophy points his readers. Rather, the extended account that I envision would clarify the development of Kant's thinking by distinguishing between his epistemological concerns and his worries about morality in the latter's primary, forward-looking sense, on the one hand, and the thinking to which Kant is constrained by the commitments that I have detailed in this thesis, on the other. The account that I propose, then, would situate some of Kant's major moves in relation, specifically, to the

---

Research Undertaken in Collaboration with Jean-Michel Oughourlian and Guy Lefort (Stanford, CA: Stanford University Press, 1987).

[781] Theodor Adorno, *Minima Moralia: Reflections from Damaged Life*, trans., E. F. N. Jephcott (New York: Verso, 1974); Theodor W. Adorno, *Aesthetic Theory*, trans., Robert Hullot-Kentor, Continuum Impacts (London: Continuum, 2004); Max Horkheimer and Theodor Adorno, *Dialectic of Enlightenment: Philosophical Fragments*, trans., Edmund Jephcott (Stanford, CA: Stanford University Press, 2002).

[782] See, for example, Susan Hurley and Nick Chater, 'Introduction: The Importance of Imitation,' in *Perspectives on Imitation: From Neuroscience to Social Science: Vol. 1: Mechanisms of Imitation and Imitation in Animals*, ed. Susan Hurley and Nick Chater (Cambridge, MA: MIT Press, 2005); C. Moore, 'Intentional Relations and Triadic Interactions,' in *Developing Theories of Intention: Social Understanding and Self-Control*, ed. Janet W. Astington, David R. Olson, and Philip David Zelazo (Mahwah, NJ: Lawrence Erlbaum Associates, 1999); Chris Moore and Philip J. Dunham, *Joint Attention: Its Origins and Role in Development* (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1995); Giacomo Rizzolatti and others, 'From Mirror Neurons to Imitation: Facts and Speculations,' in *The Imitative Mind: Development, Evolution, and Brain Bases*, ed. Andrew N. Meltzoff and Wolfgang Prinz (Cambridge: Cambridge University Press, 2002); Michael Tomasello, *The Cultural Origins of Human Cognition* (Cambridge, MA.: Harvard University Press, 1999); Michael Tomasello and Malinda Carpenter, 'Intention Reading and Imitative Learning,' in *Perspectives on Imitation: From Neuroscience to Social Science: Vol. 2: Imitation, Human Development, and Culture*, ed. Susan Hurley and Nick Chater (Cambridge, MA: MIT Press, 2005).

Groundwork's imitation-excluding (re)description of 'matters of morality,' on the one hand, and his account of the imputable '*Gesinnung*,' on the other. These moves would include Kant's distinction between the practices of 'explanation' and 'imputation,' practices whose conditions he discusses in his treatment of the thesis and antithesis positions of the 'Third Antinomy'; the first *Critique*'s distinction between 'intelligible' and 'empirical' character; that work's discussion of 'absolute causal spontaneity' in relation to the unique ascribability of actions; the first *Critique*'s worries about the lawlessness of mere spontaneity; the eclipse of heteronomy, in general, in the context of Kant's *Groundwork* 'invention of autonomy'; the problem, attendant on Kant's earliest treatment of autonomy, of the imputability of evil deeds (not *qua* actions, as such, but precisely *qua* evil); Kant's distinction between *Wille* and *Willkür*; and, in relation to the latter, his articulation of the so-called 'incorporation thesis' along with his claim, in the *Religion*, that the human being's *Willkür* ('power of choice') is endowed with an ineliminable predisposition to moral goodness.

I have claimed that Kant's ultimate description of the human being, as an agent whose deepest moral properties are imputable to her alone, is a key and ultimate move in a project aimed at rationalizing and securing the practice of imputation, under a description of the latter that has integral reference to punishment, where punishment is understood in strictly retributivist terms. Kant works hard to get to his description of the ultimately, retributively 'punishable' subject. This thesis does not traverse the path by which Kant reaches this description in his theory of radical evil. The more detailed story of these rather desperate measures remains to be told. Always anterior to the stages along this path, in any case—an obscure presence that conditions his thought—Kant's antecedent commitment to (retributive) punishment and, with it, to radical, retribution-sensitive imputation, remains in force. In another sense, however, it is not really in the background at all: it is right there, under our noses, whenever Kant glosses 'morality' as 'worthiness to be happy.'

BIBLIOGRAPHY

Acton, H. B. *Kant's Moral Philosophy* New Studies in Ethics. London: Macmillan, 1970.

Adorno, Theodor. *Minima Moralia: Reflections from Damaged Life*. Translated by E. F. N. Jephcott. New York: Verso, 1974.

Adorno, Theodor W. *Problems of Moral Philosophy*. Stanford, CA: Stanford University Press, 2001.

_____. *Aesthetic Theory*. Translated by Robert Hullot-Kentor Continuum Impacts. London: Continuum, 2004.

Allison, Henry E. 'The Concept of Freedom in Kant's Semi-Critical Ethics.' *Archiv Fur Geschichte Der Philosophie* 68, no. 1 (1986): 96-114.

_____. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.

_____. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press, 1996.

_____. *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment* Modern European Philosophy. Cambridge: Cambridge University Press, 2001.

_____. 'On the Very Idea of a Propensity to Evil.' *Journal of Value Inquiry* 36, no. 2-3 (2002): 369-382.

_____. *Kant's Transcendental Idealism: An Interpretation and Defense*. Rev. and enl. ed. New Haven: Yale University Press, 2004.

Ameriks, Karl. 'Kant on the Good Will.' In *Grundlegung Zur Metaphysik Der Sitten: Ein Kooperativer Kommentar*, edited by Ottfried Höffe, 45-65. Frankfurt am Main: Vittorio Klostermann, 1989.

_____. *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy* Modern European Philosophy. Cambridge: Cambridge University Press, 2000.

_____. *Interpreting Kant's Critiques*. Oxford: Clarendon Press, 2003.

_____. *Kant and the Historical Turn: Philosophy as Critical Interpretation*. Oxford: Clarendon Press, 2006.

Anderson, Pamela Sue, and Jordan Bell. *Kant and Theology* Philosophy and Theology. London: T & T Clark, 2010.

Arendt, Hannah. *The Human Condition*, Edited by Margaret Canovan. Chicago: University of Chicago Press, 1958.

Arp, R. 'Vindicating Kant's Morality.' *International Philosophical Quarterly* 47, no. 1 (2007): 5-22.

Ataner, A. 'Kant on Capital Punishment and Suicide.' *Kant-Studien* 97, no. 4 (2006): 452-482.

Atwell, John. *Ends and Principles in Kant's Moral Thought*. Dordrecht: Martin Nijhoff, 1986.

Axinn, Sidney. 'Kant on Possible Hope: The Critique of Pure Hope.' In *The Proceedings of the Twentieth World Congress of Philosophy: Modern Philosophy*, edited by M. D. Gedney, 79-87. Charlottesville: Philosophy Document Center, 2000.

Badiou, Alain. *Ethics: An Essay on the Understanding of Evil*. London: Verso, 2001.

Baier, Annette C. 'Moralism and Cruelty: Reflections on Hume and Kant.' In *Moral Prejudices*, 268-93. Cambridge, MA: Harvard University Press, 1994.

Baker, Judith. 'Do One's Motives Have to Be Pure?' In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, edited by Richard E. Grandy and Richard Warner. Oxford: Oxford University Press, 1986.

Barnes, Gerald W. 'In Defense of Kant's Doctrine of the Highest Good.' *Philosophical Forum* 2 (1971): 446-58.

Baron, Marcia. *Kantian Ethics Almost without Apology*. Ithaca: Cornell University Press, 1995.

Baumgarten, Alexander Gottlieb. *Initia Philosophiae Practicae Primae*. Halle, 1760.

Baxley, Anne Margaret. 'The Practical Significance of Taste in Kant's Critique of Judgment: Love of Natural Beauty as a Mark of Moral Character.' *Journal of Aesthetics and Art Criticism* 63, no. 1 (2005): 33-45.

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: University of Chicago Press, 1960.

_____. 'Five Concepts of Freedom in Kant.' In *Stephan Körner: Philosophical Analysis and Reconstruction*, edited by J. T. J. Srzednicki. Hingham: Kluwer Academic Publishers, 1987.

_____. *Early German Philosophy: Kant and His Predecessors*. Cambridge, MA: Thoemmes Press, 1996.

Bedau, H. 'Retribution and the Theory of Punishment.' *Journal of Philosophy* 75 (1978): 601-20.

Benn, S. I. 'Punishment.' In *The Encyclopedia of Philosophy*, edited by Paul Edwards, 7. London: Macmillan, 1967.

Benson, Paul. 'Moral Worth.' *Philosophical Studies* 51, no. 3 (1987): 365-82.

Berman, Mitchell N. 'Punishment and Justification.' *Ethics* 118, no. 2 (2008): 258-90.

Bernstein, Richard J. *Radical Evil: A Philosophical Interrogation*. Cambridge: Polity Press, 2002.

Blake, William. 'The Little Black Boy.' In *Songs of Innocence & Songs of Experience*, 10-11. Mineola, NY: Dover Publications, 1992 [1789].

Bowie, Andrew. *Aesthetics and Subjectivity: From Kant to Nietzsche*. 2nd ed. Manchester: Manchester University Press, 2003.

Broad, C. D. *Five Types of Ethical Theory* International Library of Psychology, Philosophy, and Scientific Method. London: Routledge & K. Paul, 1930.

Broadie, Alexander, and Elizabeth M. Pybus. 'Kant's Concept of 'Respect'.' *Kant-Studien* 66, no. 1 (1975): 58-64.

Brooks, T. 'Kantian Punishment and Retributivism: A Reply to Clark.' *Ratio* 18, no. 2 (2005): 237-45.

Brown, R. F. 'The Transcendental Fall in Kant and Schelling.' *Idealistic Studies* 14, no. 1 (1984): 49-66.

Buck, Günther. 'Kants Lehre Vom Exempel.' *Archiv für Begriffgeschichte* 11 (1967): 148-83.

Burgh, Richard. 'Do the Guilty Deserve Punishment?' *The Journal of Philosophy* 79 (1982): 193–210.

Byrd, B. Sharon. 'Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution.' *Law and Philosophy* 8, no. 2 (1989): 151-200.

Card, Claudia. *The Atrocity Paradigm: A Theory of Evil*. Oxford: Oxford University Press, 2002.

Carnois, Bernard. *The Coherence of Kant's Doctrine of Freedom*. Chicago: University of Chicago Press, 1987.

Caygill, Howard. *A Kant Dictionary* The Blackwell Philosopher Dictionaries. Oxford: Blackwell Reference, 1995.

Christopher, R. L. 'Deterring Retributivism: The Injustice of "Just" Punishment.' *Northwestern University Law Review* 96, no. 3 (2002): 843-976.

Cicovacki, P. 'The Illusory Fabric of Kant's True Morality.' *Journal of Value Inquiry* 36, no. 2-3 (2002): 383-399.

Clark, M. 'A Non-Retributive Kantian Approach to Punishment.' *Ratio (New Series)* 17, no. 1 (2004): 12-27.

Cohen, Morris. 'A Critique of Kant's Philosophy of Law.' In *The Heritage of Kant*, edited by George Tapley Whitney and David Frederick Bowers, 279-302. Princeton University Press: Princeton, 1939.

Cooper, David. 'Hegel's Theory of Punishment.' In *Hegel's Political Philosophy*, edited by Z. A. Pelezynski. Cambridge: Cambridge University Press, 1971.

Cope, C. 'Freedom, Responsibility, and the Concept of Anxiety.' *International Philosophical Quarterly* 44, no. 4 (2004): 549-566.

Corlett, J. A. 'Making Sense of Retributivism.' *Philosophy* 76, no. 1 (2001): 77-110.

_____. *Responsibility and Punishment*. Dordrecht: Kluwer Academic Publishers, 2001.

_____. 'Making More Sense of Retributivism: Desert as Responsibility and Proportionality.' *Philosophy* 78, no. 2 (2003): 279-287.

Cottingham, John. 'Varieties of Retribution.' *The Philosophical Quarterly* 29 (1979): 238-46.

Cupitt, Geoffrey. 'Desert and Responsibility.' *Canadian Journal of Philosophy* 26 (1996): 83–100.

Curzer, H. J. 'From Duty, Moral Worth, Good Will.' *Dialogue* 36, no. 2 (1997): 287-322.

Dahl, Norman O. 'Obligation and Moral Worth: Reflections on Prichard and Kant.' *Philosophical Studies* 50 (1986): 369-399.

Darwall, Stephen. 'Norm and Normativity.' In *The Cambridge History of Eighteenth-Century Philosophy*, edited by K. Haakonssen, 2, 987-1025. Cambridge: Cambridge University Press, 2005.

Dekens, O. ‘Initiation À La Vie Malheureuse: De L'impossibilité Du Pardon Chez Kant Et Kierkegaard.’ *Revue Philosophique de Louvain* 96, no. 4 (1998): 581-597.

Denis, Lara. ‘Kant's Conception of Virtue.’ In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer, 505-37. Cambridge: Cambridge University Press, 2006.

Derrida, Jacques. ‘Economimesis.’ *Diacritics* 11, no. 2 (1981): 2-25.

Dews, Peter. *The Idea of Evil*. Malden, MA: Blackwell Pub., 2007.

Di Giovanni, George. ‘Freedom and Religion in Kant and His Immediate Successors: The Vocation of Humankind, 1774-1800.’ (2005): xvi, 373 p.

DiCenso, J. ‘Kant, Freud, and the Ethical Critique of Religion.’ *International Journal for Philosophy of Religion* 61 (2007): 161-179.

Dolinko, David. ‘Some Thoughts About Retributivism.’ *Ethics* 101, no. 3 (1991).

Enderlein, W. ‘Die Begründung Der Strafe Bei Kant.’ *Kant-Studien* 76, no. 3 (1985): 303-327.

Engstrom, S. ‘Happiness and the Highest Good in Aristotle and Kant.’ In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by S. P. Engstrom and J. Whiting, 102-138. Cambridge: Cambridge, 1994.

Engstrom, Stephen. ‘The Inner Freedom of Virtue.’ In *Kant's Metaphysics of Morals*, edited by Mark Timmons, 289-315. Oxford: Oxford University Press, 2003.

Ewing, A. C. ‘Paradoxes of Kant's Ethics.’ *Philosophy* 13, no. 49 (1938): 40-56.

Feinberg, Joel. ‘Justice and Personal Desert.’ In *Doing and Deserving*, 55-94. Princeton: Princeton University Press, 1970.

Fendt, Gene. *For What May I Hope?: Thinking with Kant and Kierkegaard* American University Studies. New York: P. Lang, 1990.

Ferreira, M. Jamie. ‘Making Room for Faith: Possibility and Hope.’ In *Kant and Kierkegaard on Religion*, edited by D. Z. Phillips and Timothy Tessin, 73-88. New York: St. Martin's Press, 2000.

Fichte, J. G. *Foundations of Natural Right*. Cambridge: Cambridge University Press, 2000.

Firestone, Chris L. ‘Making Sense out of Tradition: Theology and Conflict in Kant's Philosophy of Religion.’ In *Kant and the New Philosophy of Religion*, edited

by Chris L. Firestone and Stephen Palmquist, 141-56. Bloomington: Indiana University Press, 2006.

Firestone, Chris L., and Nathan Jacobs. *In Defense of Kant's Religion* Indiana Series in the Philosophy of Religion. Bloomington, IN: Indiana University Press, 2008.

Fischer, Norbert. 'Tugend Und Gluckseligkeit: Zu Ihrem Verhaltnis Bei Aristoteles Und Kant.' *Kant-Studien* 74, no. 1 (1983): 1-21.

Fleischacker, S. 'Kant's Theory of Punishment.' *Kant-Studien* 79, no. 4 (1988): 434-449.

Fonnessu, Luca. 'The Problem of Theodicy.' In *The Cambridge History of Eighteenth-Century Philosophy*, edited by K. Haakonssen, 2, 749-78. Cambridge: Cambridge University Press, 2005.

Formosa, P. 'Is Radical Evil Banal? Is Banal Evil Radical?' *Philosophy & Social Criticism* 33, no. 6 (2007): 717-35.

_____. 'Kant on the Radical Evil of Human Nature.' *Philosophical Forum* 38, no. 3 (2007): 221-245.

Forschner, M. 'Moralität Und Glückseligkeit in Kants Reflexionen.' *Zeitschrift für philosophische Forschung* 42, no. 3 (1988): 351-370.

Friedman, R. Z. 'Virtue and Happiness: Kant and Three Critics.' *Canadian Journal of Philosophy* 11, no. 1 (1981): 95-110.

Frierson, Patrick R. *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge: Cambridge University Press, 2003.

Galbraith, Elizabeth C. 'Kant and Richard Schaeffler's Catholic Theology of Hope.' *Philosophy & Theology* 9, no. 3-4 (1996): 333-350.

Gammon, Martin. '"Exemplary Originality": Kant on Genius and Imitation.' *Journal of the History of Philosophy* 35, no. 4 (1997): 563-592.

Garcia, J. L. A. 'Two Concepts of Desert.' *Law and Philosophy* 5, no. 2 (1986): 219-35.

Gauthier, J. A. 'Schiller's Critique of Kant's Moral Psychology: Reconciling Practical Reason and an Ethics of Virtue.' *Canadian Journal of Philosophy* 27, no. 4 (1997): 513-544.

Gaziaux, E. 'Trois Modeles D'autonomie En Morale: Kant, Steinbuechel, Auer.' *Revue Theologique De Louvain* 28, no. 3 (1997): 338-358.

_____. ‘L'autonomie En Morale: Entre L'affirmation De L'homme at La Quête De Dieu.’ *Revue Theologique De Louvain* 30, no. 3 (1999): 315-335.

Gibbs, Robert. ‘Fear of Forgiveness: Kant and the Paradox of Mercy.’ *Philosophy & Theology* 3 (1989): 323-334.

Girard, René. *Deceit, Desire, and the Novel: Self and Other in Literary Structure*. Translated by Yvonne Freccero. Baltimore: John Hopkins Press, 1965.

_____. *Violence and the Sacred*. Translated by Patrick Gregory. Baltimore: Johns Hopkins University Press, 1977.

Girard, René, Jean-Michel Oughourlian, and Guy Lefort. *Things Hidden since the Foundation of the World: Research Undertaken in Collaboration with Jean-Michel Oughourlian and Guy Lefort*. Stanford, CA: Stanford University Press, 1987.

Gniffke, Franz. ‘Auf Der Suche Nach Dem Verantwortlichen Subjekt: Eine Hinfuhrung Zu Kants Grundlegung Der Ethik.’ In *Zur Geschichtlichkeit Der Beziehungen Von Glaube, Kunst Und Umweltgestaltung*, 6-44. Frankfurt: Haag & Herchen Verlag, 1977.

Gould, Timothy. ‘The Audience of Originality.’ In *Essays in Kant's Aesthetics*, edited by Ted Cohen and P. Guyer, 179-193. Chicago: University of Chicago Press, 1982.

Goy, I. ‘Immanuel Kant Über Das Moralische Gefühl Der Achtung.’ *Zeitschrift für philosophische Forschung* 61, no. 3 (2007): 337-360.

Gregor, Mary J. *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik Der Sitten*. Oxford: Blackwell, 1963.

Gressis, Robert. ‘How to Be Evil: Kant's Moral Psychology of Immorality.’ In *Rethinking Kant*, edited by P. Muchnik, 1, 191-216. Newcastle upon Tyne: Cambridge Scholars Publishing, 2008.

Gross, Hyman. *A Theory of Criminal Justice*. Oxford: Oxford University Press, 1979.

Gunkel, Andreas. *Spontaneität Und Moralische Autonomie: Kants Philosophie Der Freiheit* Berner Reihe Philosophischer Studien. Bern: Verlag Paul Haupt, 1989.

Guyer, Paul. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000.

_____. *Kant* Routledge Philosophers. London: Routledge, 2006.

Halliwell, Stephen. *The Aesthetics of Mimesis: Ancient Texts and Modern Problems*. Princeton, NJ: Princeton University Press, 2002.

Hart, H. L. A. *Punishment and Responsibility*. Oxford: Oxford University Press, 1982.

Häyry, M. 'The Tension between Self Governance and Absolute Inner Worth in Kant's Moral Philosophy.' *Journal of Medical Ethics* 31, no. 11 (2005): 645-647.

Hegel, G. W. F. *Phenomenology of Spirit*. Translated by A. V. Miller. New ed. Oxford: Clarendon Press: Oxford University Press, 1979.

Henrich, Dieter. 'Die Deduktion Des Sittengesetzes.' In *Darmstadt*, edited by A. Schwan. Denken im Schatten des Nihilismus, 1975.

_____. 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason.' In *The Unity of Reason: Essays on Kant's Philosophy*, edited by R. Velkley, 55-87. London: Harvard University Press, 1994.

_____. 'Ethics of Autonomy.' In *The Unity of Reason*, edited by R. Velkley, 89-121. London: Harvard University Press, 1994.

Henson, Richard G. 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action.' *Philosophical Review* 88, no. 1 (1979): 39-54.

Herman, Barbara. 'On the Value of Acting from the Motive of Duty.' *Philosophical Review* 90, no. 3 (1981): 359-82.

Hildebrandt, Bernd. 'Kant Als Philosoph Des Protestantismus.' In *Was Ist Und Was Sein Soll*, edited by Udo Kern, 477-494. Berlin ; New York: Walter de Gruyter, 2007.

Hill, Thomas E., Jr. *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, NY: Cornell University Press, 1992.

_____. 'Kant on Punishment: A Coherent Mix of Deterrence and Retribution?' *Jahrbuch für Recht und Ethik* 5 (1997): 291-34.

_____. 'Happiness and Human Flourishing in Kant's Ethics.' *Social Philosophy & Policy* 16, no. 1 (1999): 143-175.

_____. 'Is a Good Will Overrated?' In *Human Welfare and Moral Worth: Kantian Perspectives*, edited by Thomas E. Hill, Jr., 37-60. Oxford: Clarendon Press, 2002.

_____. ‘Punishment, Conscience, and Moral Worth.’ In *Human Welfare and Moral Worth: Kantian Perspectives*, edited by Thomas E. Hill, Jr., 340-61. Oxford: Clarendon Press, 2002.

_____. ‘Wrongdoing, Desert, and Punishment.’ In *Human Welfare and Moral Worth: Kantian Perspectives*, edited by Thomas E. Hill, Jr., 310-39. Oxford: Clarendon Press, 2002.

Hills, A. ‘Kant on Happiness and Reason.’ *History of Philosophy Quarterly* 23, no. 3 (2006): 243-262.

Hinman, L. M. ‘On the Purity of Our Moral Motives: A Critique of Kant's Account of the Emotions and Acting for the Sake of Duty.’ *Monist* 66, no. 2 (1983): 251-267.

Höffe, Otfried. *Kant's Cosmopolitan Theory of Law and Peace* Modern European Philosophy. Cambridge: Cambridge University Press, 2006.

Holtman, Sarah. ‘Toward Social Reform: Kant's Penal Theory Reinterpreted.’ *Utilitas* 9, no. 1 (1997): 3-21.

Horkheimer, Max, and Theodor Adorno. *Dialectic of Enlightenment: Philosophical Fragments*. Translated by Edmund Jephcott. Stanford, CA: Stanford University Press, 2002.

Horty, John Francis. *Agency and Deontic Logic*. Oxford: Oxford University Press, 2001.

Howard, J. J. ‘Kant and Moral Imputation: Conscience and the Riddle of the Given.’ *American Catholic Philosophical Quarterly* 78, no. 4 (2004): 609-627.

Hughes, F. ‘Taste as Productive Mimesis.’ *Journal of the British Society for Phenomenology* 37, no. 3 (2006): 308-326.

Huhn, Tom. *Imitation and Society: The Persistence of Mimesis in the Aesthetics of Burke, Hogarth, and Kant* Literature and Philosophy. University Park, PA: Pennsylvania State University Press, 2006.

Hume, David. *An Equiry Concerning the Principles of Morals* Oxford Philosophical Texts. Oxford: Oxford University Press, 1998 [1751].

Hurka, Thomas. ‘Desert: Individualistic and Holistic.’ In *Desert and Justice*, edited by Serena Olsaretti, 45-68. Oxford: Clarendon Press, 2007.

Hurley, Susan, and Nick Chater. ‘Introduction: The Importance of Imitation.’ In *Perspectives on Imitation: From Neuroscience to Social Science: Vol. 1: Mechanisms of Imitation and Imitation in Animals*, edited by Susan Hurley and Nick Chater, 1-52. Cambridge, MA: MIT Press, 2005.

Hurst, Brian Desmond. ‘A Christmas Carol.’ 86 min. UK: United Artists, 1951.

Insole, Christopher J. *The Realist Hope: A Critique of Anti-Realist Approaches in Contemporary Philosophical Theology* Heythrop Studies in Contemporary Philosophy, Religion, & Theology. Aldershot: Ashgate, 2006.

Irwin, T. H. ‘Kant's Criticisms of Eudaemonism.’ In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by S. P. Engstrom and J. Whiting, 102-138. Cambridge: Cambridge, 1994.

Jacobs, Brian, and Patrick Kain, eds. *Essays on Kant's Anthropology*. Cambridge: Cambridge University Press, 2003.

Janz, Paul D. *God, the Mind's Desire: Reference, Reason, and Christian Thinking* Cambridge Studies in Christian Doctrine. Cambridge: Cambridge University Press, 2004.

Jensen, H. ‘Kant on Overdetermination, Indirect Duties, and Moral Worth.’ In *Proceedings of the Sixth International Kant Congress*, edited by G. Funke and T. M. Seebohm, 161-170. Washington, D.C.: University Press of America, 1989.

Johnson, Robert N. ‘Kant's Conception of Merit: ‘Metaphysics of Morals’ and Evaluating Actions.’ *Pacific Philosophical Quarterly* 77, no. 4 (1996): 310-334.

_____. ‘Happiness as a Natural End.’ In *Kant's Metaphysics of Morals*, edited by Mark Timmons, 317-30. Oxford: Oxford University Press, 2003.

Kelley, A. ‘Kant on Freedom, Happiness, and Peace.’ In *Spiritual and Political Dimensions of Nonviolence and Peace*, edited by D. Boersema and K. G. Brown, 169-178. Amsterdam: Editions Rodopi B. V., 2006.

Keyworth, Donald R. ‘Kant's Concept of Happiness in the Moral Argument.’ *Personalist* 43 (1962): 21-33.

Kienzle, Bertram. ‘Macht Das Sittengesetz Unglücklich?’ In *Was Ist Und Was Sein Soll*, edited by Udo Kern, 267-84. Berlin: Walter de Gruyter, 2007.

Korsgaard, C. M. ‘Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations.’ In *Creating the Kingdom of Ends*, 188-221. Cambridge: Cambridge University Press, 1996.

_____. ‘Kant's Formula of Humanity.’ In *Creating the Kingdom of Ends*, 106-32. Cambridge: Cambridge University Press, 1996.

_____. 'Morality as Freedom.' In *Creating the Kingdom of Ends*, 159-87. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. 'Aristotle and Kant on the Source of Value.' *Ethics* 96, no. 3 (1986): 486-505.

_____. 'The Right to Lie: Kant on Dealing with Evil.' *Philosophy & Public Affairs* 15, no. 4 (1986): 325-349.

Kosch, Michelle. *Freedom and Reason in Kant, Schelling, and Kierkegaard*. Oxford: Oxford University Press, 2006.

Langton, Rae. 'Duty and Desolation.' *Philosophy* 67, no. 262 (1992): 481-505.

_____. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press, 1998.

Laska, P. 'Kant on Moral Worth: A Reply to Murphy.' *Kant-Studien* 59 (1968): 374-383.

Lawrence, J. P. 'Radical Evil and Kant's Turn to Religion.' *Journal of Value Inquiry* 36, no. 2-3 (2002): 319-335.

Lind, Douglas. 'Kant on Capital Punishment.' *Journal of Philosophical Research* 19 (1994): 61-74.

Louden, Robert B. 'Kant's Virtue Ethics.' *Philosophy* 61, no. 238 (1986): 473-489.

_____. 'Go-Carts of Judgment: Exemplars in Kantian Moral Education.' *Archiv für Geschichte der Philosophie* 74, no. 3 (1992): 303-22.

Makkreel, Rudolf. 'Reflection, Reflective Judgment, and Aesthetic Exemplarity.' In *Aesthetics and Cognition in Kant's Critical Philosophy*, edited by Rebecca Kukla, 223-244. Cambridge: Cambridge University Press, 2006.

Mannion, Gerard. 'Kant and the Defeat of Egoism: Schopenhauerian Concerns and Some Reappraisals and Rejoinders.' *Kant-Studien* 99, no. 2 (2008): 220-228.

Mariña, J. 'Making Sense of Kant's Highest Good.' *Kant-Studien* 91, no. 3 (2000): 329-355.

Mariña, Jacqueline. 'Kant on Grace: A Reply to His Critics.' *Religious Studies* 33 (1997): 379-400.

McCarty, Richard. 'Kantian Moral Motivation and the Feeling of Respect.' *Journal of the History of Philosophy* 31, no. 3 (1993): 421-435.

_____. ‘Motivation and Moral Choice in Kant's Theory of Rational Agency.’ *Kant-Studien* 85, no. 1 (1994): 15-31.

Merle, Jean-Christophe. ‘A Kantian Critique of Kant's Theory of Punishment.’ *Law and Philosophy* 19, no. 3 (2000): 311-338.

Meyers, C. D. ‘The Virtue of Cold-Heartedness.’ *Philosophical Studies* 138, no. 2 (2008): 233-244.

Michalson, Gordon E. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge: Cambridge University Press, 1990.

Miller, R. D. *Schiller and the Ideal of Freedom: A Study of Schiller's Philosophical Works with Chapters on Kant*. Oxford: Clarendon Press, 1970.

Moore, C. ‘Intentional Relations and Triadic Interactions.’ In *Developing Theories of Intention: Social Understanding and Self-Control*, edited by Janet W. Astington, David R. Olson and Philip David Zelazo, 43-62. Mahwah, NJ: Lawrence Erlbaum Associates, 1999.

Moore, Chris, and Philip J. Dunham. *Joint Attention: Its Origins and Role in Development*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1995.

Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903.

_____. *Ethics*. Oxford: Oxford University Press, 1912.

Moore, Michael S. ‘The Moral Worth of Retribution.’ In *Punishment and Rehabilitation*, edited by Jeffrie G. Murphy, 94-130. Belmont, CA: Wadsworth Publishing Co., 1973.

Moyar, Dean. ‘Practical Apperception: Self-Imputation and Moral Judgment.’ In *Law and Peace in Kant’s Philosophy / Recht Und Frieden in Der Philosophie Kants (Akten Des X. Internationalen Kant-Kongresses)*, edited by Valerio Rohden, Ricardo R. Terra, Guido A. de Almeida and Margit Ruffing, 3, 281–290. Berlin: Walter de Gruyter, 2008.

Muchnik, Pablo. *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Aprioricity of History*. Lanham: Lexington Books, 2009.

Mulholland, Leslie A. ‘Freedom and Providence in Kant's Account of Religion: The Problem of Expiation.’ In *Kant's Philosophy of Religion Reconsidered*, edited by Philip J. Rossi and Michael J. Wreen, 77-102. Bloomington: Indiana University Press, 1991.

Murphy, J. G. ‘Kant's Concept of a Right Action.’ *Monist* 51, no. 4 (1967): 574-98.

Murphy, Jeffrie G. *Kant: The Philosophy of Right*. London: Macmillan, 1970.

_____. ‘Kant's Theory of Criminal Punishment.’ In *Proceedings of the Third International Kant Congress*, edited by Lewis White Beck. Dordrecht: Reidel, 1972.

_____. ‘Does Kant Have a Theory of Punishment.’ *Columbia Law Review* 87, no. 3 (1987): 509-532.

Muthu, Sankar. *Enlightenment against Empire*. Princeton: Princeton University Press, 2003.

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

Nenon, Thomas. ‘The Highest Good and the Happiness of Others.’ *Jahrbuch für Recht und Ethik* 5 (1997): 419-35.

Nietzsche, Friedrich. *On the Genealogy of Morality: A Polemic*. Translated by Alan J. Swensen and Maudemarie Clark. Indianapolis, IN: Hackett Publishing Co., 1998.

Norton, David Fate, and Manfred Kuehn. ‘The Foundations of Morality.’ In *The Cambridge History of Eighteenth-Century Philosophy*, edited by K. Haakonssen, 2, 941-85. Cambridge: Cambridge University Press, 2005.

Nugen, A. T. ‘Just Desert.’ *Journal of Value Inquiry* 31 (1997): 221-30.

O'Neill, Onora. ‘Kant after Virtue.’ *Inquiry* 26, no. 4 (1984): 387-405.

O’Connor, D. ‘Kant's Conception of Happiness.’ *The Journal of Value Inquiry* 16, no. 3 (1982): 189-205.

Olsaretti, Serena, ed. *Desert and Justice*. Oxford: Clarendon Press, 2007.

Packer, M. ‘The Highest Good in Kant's Psychology of Motivation.’ *Idealistic Studies* 13, no. 2 (1983): 110-119.

Palaver, Wolfgang. ‘Mimesis and Scapegoating in the Works of Hobbes, Rousseau, and Kant.’ *Contagion* 8 (2003): 126-48.

Paton, H. J. ‘Kant's Idea of the Good.’ *Proceedings of the Aristotelian Society* 45 (1944-5): ii-xxv.

_____. *The Moral Law: Or Kant's Groundwork of the Metaphysic of Morals* The Senior Series. London: Hutchinson's University Library, 1947.

_____. *The Categorial Imperative: A Study in Kant's Moral Philosophy*. Philadelphia: University of Pennsylvania Press, 1971 [1947].

Paulsen, Friedrich. *Immanuel Kant, His Life and Doctrine*. Translated by J. E. Creighton and A. Lefevre. London: J. C. Nimmo, 1902.

Piercey, R. 'Active Mimesis and the Art of History of Philosophy.' *International Philosophical Quarterly* 43, no. 1 (2003): 29-42.

Pincoffs, Edmund. *The Rationale of Legal Punishment*. New York: Humanities Press, 1966.

Pinkard, Terry. *Hegel's Phenomenology: The Sociality of Reason*. Cambridge: Cambridge University Press, 1996.

Pippin, R. B. 'Idealism and Agency in Kant and Hegel.' *Journal of Philosophy* 88, no. 10 (1991): 532-541.

Pippin, Robert B. 'Kant's Theory of Value: On Allen Wood's Kant's Ethical Thought.' *Inquiry* 43, no. 2 (2000): 239-66.

Plato. 'Crito.' In *The Collected Dialogues of Plato*, edited by E. Hamilton and H. Cairns. New York: Bollingen Foundation, 1961.

Pojman, Louis P., and Owen McLeod, eds. *What Do We Deserve?: A Reader on Justice and Desert*. Oxford: Oxford University Press, 1999.

Potter, N. T. 'Kant and Capital Punishment Today.' *Journal of Value Inquiry* 36, no. 2 (2002): 267 - 282.

Potter, Nelson. 'Kant and the Moral Worth of Actions.' *Southern Journal of Philosophy* 34, no. 2 (1996): 225-242.

Prauss, Gerold. 'Theory as Praxis in Kant.' In *Kant's Practical Philosophy Reconsidered*, edited by Yirmiahu Yovel, 93-105. London: Kluwer Academic Publishers, 1989.

Quelquejeu, B. 'Ethical Autonomy and the Question of God.' In *Ethics of Liberation - the Liberation of Ethics*, edited by Dietmar Mieth and Jacques Marie Pohier, 16-23. Edinburgh, Scotland: T & T Clark, 1984.

Rachels, James. 'What People Deserve.' In *Justice and Economic Distribution*, edited by J. Arthur and W. H. Shaw, 150-63. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Ranasinghe, N. 'Ethics for the Little Man: Kant, Eichmann, and the Banality of Evil.' *Journal of Value Inquiry* 36, no. 2-3 (2002): 299-317.

Raphael, D. D. *Moral Judgment*. London: George Allen and Unwin, 1955.

Reath, Andrews. ‘Hedonism, Heteronomy, and Kant's Principle of Happiness.’ In *Agency and Autonomy in Kant’s Moral Theory*, edited by Andrews Reath, 33-67. Oxford: Oxford University Press, 2006.

_____. ‘Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination.’ In *Agency and Autonomy in Kant's Moral Theory: Selected Essays*, 284-302. Oxford: Oxford University Press, 2006.

Reid, James. ‘Morality and Sensibility in Kant: Toward a Theory of Virtue.’ *Kantian Review* 8 (2004): 89-114.

Reulecke, Martin. *Gleichheit Und Strafrecht Im Deutschen Naturrecht Des 18. Und 19. Jahrhunderts*. Vol. 9 Grundlagen Der Rechtswissenschaft. Tübingen: Verlag Mohr Siebeck, 2008.

Rizzolatti, Giacomo, Luciano Fadiga, Leonardo Fogassi, and Vittorio Gallese. ‘From Mirror Neurons to Imitation: Facts and Speculations.’ In *The Imitative Mind: Development, Evolution, and Brain Bases*, edited by Andrew N. Meltzoff and Wolfgang Prinz, 247-66. Cambridge: Cambridge University Press, 2002.

Rogozinski, Jacob. ‘It Makes Us Wrong: Kant and Radical Evil.’ In *Radical Evil*, edited by Joan Copjec, 30-45. London: Verso, 1996.

Römpp, G. ‘Kant’s Ethics as a Philosophy of Happiness: Reflections on the “Reflexionen”.’ *Modern Schoolman* 71, no. 4 (1994): 271-284.

Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930.

Rossi, P. ‘Kant as a Christian Philosopher: Hope and the Symbols of Christian Faith.’ *Philosophy Today* 25, no. 1 (1981): 24-33.

_____. ‘Kant's Doctrine of Hope: Reason's Interest and the Things of Faith.’ *New Scholasticism* 56, no. 2 (1982): 228-238.

_____. ‘Autonomy and Community: The Social Character of Kant's Moral Faith.’ *Modern Schoolman* 61, no. 3 (1984): 169-186.

Rossi, Philip J. ‘The Final End of All Things: The Highest Good as the Unity of Nature and Freedom.’ In *Kant's Philosophy of Religion Reconsidered*, 132-164. Bloomington: Indiana University Press, 1991.

_____. *The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind* Suny Series in Philosophy. Albany: State University of New York Press, 2005.

_____. ‘Reading Kant through Theological Spectacles.’ In *Kant and the New Philosophy of Religion*, edited by Chris L. Firestone and Stephen Palmquist, 107-23. Bloomington: Indiana University Press, 2006.

Rossvaer, Viggo. *Kant's Moral Philosophy*. Oslo: Universitetsforlaget, 1979.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. London: Belknap Press, 2008.

Schaeffler, Richard. 'Kant Als Philosoph Der Hoffnung.' *Theologie und Philosophie* 56 (1981): 244-57.

Schaller, W. E. 'Kant on Virtue and Moral Worth.' *Southern Journal of Philosophy* 25, no. 4 (1987): 559-573.

Schalow, Frank. *The Renewal of the Heidegger-Kant Dialogue: Action, Thought, and Responsibility* Suny Series in Contemporary Continental Philosophy. Albany: State University of New York Press, 1992.

Scheid, Don E. 'Kant's Retributivism.' *Ethics* 93, no. 2 (1983): 262-282.

Schilpp, Paul Arthur. *Kant's Pre-Critical Ethics* Northwestern University Studies in the Humanities. Evanston: Northwestern University, 1938.

Schmucker, Josef. *Die Ursprünge Der Ethik Kants in Seinen Vorkritischen Schriften Und Reflektionen*. Meisenheim am Glan: Anton Hain, 1961.

Schneewind, J. B. 'Autonomy, Obligation, and Virtue.' In *The Cambridge Companion to Kant*, edited by P. Guyer. New York: Cambridge University Press, 1992.

_____. 'Kant and Stoic Ethics.' In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by S. P. Engstrom and J. Whiting, 285-302. Cambridge: Cambridge, 1994.

_____. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press, 1997.

_____. 'Active Powers.' In *The Cambridge History of Eighteenth-Century Philosophy*, edited by Knud Haakonssen. Cambridge: Cambridge University Press, 2005.

Schner, G. P. 'Moral Ontology in Kant: At What Cost Freedom and Perpetual Peace?' *Toronto Journal of Theology* 18, no. 1 (2002): 153-166.

Schönecker, Dieter. *Grundlegung III: Die Deduktion Des Kategorischen Imperativs* Alber Symposion. Freiburg: Alber, 1999.

Schopenhauer, Arthur. *The World as Will and Representation*. Vol. 1. 2 vols. New York: Dover Publications, 1958.

_____. *On the Basis of Morality*. Translated by Eric F. J. Payne. Indianapolis: Bobbs-Merill, 1965.

Schroeder, H. H. 'Some Common Misinterpretations of the Kantian Ethics.' *The Philosophical Review* 49 (1940): 424-46.

Schwartländer, Johannes. 'Sittliche Autonomie Als Idee Der Endlichen Freiheit: Bemerkungen Zum Prinzip Der Autonomie Im Kritischen Idealismus Kants.' *Theologische Quartalschrift* 161, no. 1 (1981): 20-33.

Sedgwick, S. 'Hegel, Mcdowell, and Recent Defenses of Kant.' *Journal of the British Society for Phenomenology* 31, no. 3 (2000): 229-247.

Sher, George. *Desert*. Princeton: Princeton University Press, 1987.

_____. *In Praise of Blame*. Oxford: Oxford University Press, 2006.

Sidgwick, Henry. *Outlines of the History of Ethics for English Readers*. Boston: Beacon Press, 1960.

Sikka, Sonia. 'On the Value of Happiness: Herder Contra Kant.' *Canadian Journal of Philosophy* 37, no. 4 (2007): 515-546.

Silber, John R. 'The Ethical Significance of Kant's *Religion*.' In *Religion within the Limits of Reason Alone*, edited by Theodore M. Greene and Hoyt H. Hudson. New York: Harper and Row, 1960.

Simmons, Keith. 'Kant on Moral Worth.' *History of Philosophy Quarterly* 6, no. 1 (1989): 85-100.

Smith, Adam. *The Theory of Moral Sentiments* Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2004.

Smith, Steven G. 'Worthiness to Be Happy and Kant's Concept of the Highest Good.' *Kant-Studien* 75, no. 2 (1984): 168-190.

Sokoloff, W. W. 'Kant and the Paradox of Respect.' *American Journal of Political Science* 45, no. 4 (2001): 768-779.

Sorell, Tom. 'Kant's Good Will and Our Good Nature: Second Thoughts About Henson and Herman.' *Kant-Studien* 78, no. 1 (1987): 87-101.

Sterba, James. 'Justice and the Concept of Desert.' *The Personalist* 57 (1976): 188-97.

Stockhammer, Morris. 'Responsibility and Freedom: The Kantian Solution.' *Judaism* 14 (1965): 72-80.

Stratton-Lake, Philip. *Kant, Duty and Moral Worth*. New York: Routledge, 2000.

Strawson, P. F. 'Freedom and Resentment.' In *Freedom and Resentment and Other Essays*, 1-25. London: Methuen, 1974.

Sullivan, Roger J. 'The Categorical Imperative and the Natural Law.' In *Proceedings of the Sixth International Kant Congress*, edited by G. Funke and T. M. Seebohm, 219-28. Washington, D.C.: University Press of America, 1989.

Sussman, David. 'Kantian Forgiveness.' *Kant-Studien* 96 (2005): 85-107.

_____. 'Shame and Punishment in Kant's Doctrine of Right.' *The Philosophical Quarterly* 58, no. 231 (2008): 299-317.

Sussman, David G. *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* Studies in Ethics: Outstanding Dissertations, Edited by Robert Nozick. London: Routledge, 2001.

Sverdlik, S. 'Kant, Nonaccidentalness and the Availability of Moral Worth.' *Journal of Ethics* 5, no. 4 (2001): 293-313.

Taylor, Charles. *Hegel*. Cambridge: Cambridge University Press, 1975.

_____. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Cambridge University Press, 1989.

Taylor, Richard. *Good and Evil*. New York: Macmillan, 1970.

Timmermann, Jens. *Sittengesetz Und Freiheit: Untersuchungen Zu Immanuel Kants Theorie Des Freien Willens*. Berlin: Walter de Gruyter, 2003.

Timmons, Mark. 'Motive and Rightness in Kant's Ethical System.' In *Kant's Metaphysics of Morals*, edited by Mark Timmons, 255-88. Oxford: Oxford University Press, 2003.

Tomasello, Michael. *The Cultural Origins of Human Cognition*. Cambridge, MA.: Harvard University Press, 1999.

Tomasello, Michael, and Malinda Carpenter. 'Intention Reading and Imitative Learning.' In *Perspectives on Imitation: From Neuroscience to Social Science: Vol. 2: Imitation, Human Development, and Culture*, edited by Susan Hurley and Nick Chater, 133-48. Cambridge, MA: MIT Press, 2005.

Tunick, M. 'Is Kant a Retributivist?' *History of Political Thought* 17, no. 1 (1996): 60-78.

Uniacke, Suzanne M. 'Responsibility and Obligation: Some Kantian Directions.' *International Journal of Philosophical Studies* 13, no. 4 (2005): 461-75.

Velkley, Richard L. *Freedom and the End of Reason: On the Moral Foundation of Kant's Critical Philosophy*. Chicago: University of Chicago Press, 1989.

Vleeschauwer, J. H. *The Development of Kantian Thought: The History of a Doctrine*. Translated by A. R. C. Duncan. London: T. Nelson, 1962.

Von Hirsch, Andrew. 'Proportionality in the Philosophy of Punishment.' *Crime and Justice* 16 (1992): 55-98.

Walker, Ralph C. S. 'Achtung in the *Grundlegung*.' In *Grundlegung Zur Metaphysik Der Sitten: Ein Kooperativer Kommentar*, edited by Ottfried Höffe, 97-116. Frankfurt am Main: Vittorio Klostermann, 1989.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Walter, J. E. 'Kant's Moral Theology.' *Harvard Theological Review* 10, no. 3 (1910): 272-295.

Wand, Bernard. 'Religious Concepts and Moral Theory: Luther and Kant.' *Journal of the History of Philosophy* 9, no. 3 (1971): 329-48.

Ward, Keith. 'Kant's Teleological Ethics.' *Philosophical Quarterly* 21, no. 85 (1971): 337-51.

_____. *The Development of Kant's View of Ethics*. Oxford: Blackwell, 1972.

Watson, Gary. 'Kant on Happiness in the Moral Life.' *Philosophy Research Archives* 9 (1983): 79-108.

Whistler, Daniel. 'Kant's *Imitatio Christi*.' *International Journal for Philosophy of Religion* 67 (2010): 17-36.

Wike, Victoria S. 'The Role of Happiness in Kant's *Groundwork*.' *Journal of Value Inquiry* 21, no. 1 (1987): 73-78.

_____. 'Kant on Happiness.' *Philosophy Research Archives* 13 (1988): 79-90.

_____. 'Does Kant's Ethics Require That the Moral Law Be the Sole Determining Ground of the Will.' *Journal of Value Inquiry* 27, no. 1 (1993): 85-92.

_____. *Kant on Happiness in Ethics* Suny Series in Ethical Theory. Albany: State University of New York Press, 1994.

Williams, Bernard. *Ethics and the Limits of Philosophy*. London: Routledge, 2006 [1985].

Williams, Rowan. ‘An Enemy Hath Done This.’ In *A Ray of Darkness*, 75-9. Cambridge, MA: Cowley Publications, 1995.

Witschen, Dieter. ‘Nicht Nachahmung, Sondern Nachfolge: Kants Reflexionen Zum Ethischen Exempel.’ *Zeitschrift für katholische Theologie* 130, no. 3 (2008): 323-33.

Wolff, Robert Paul. *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysic of Morals* Harper Torchbooks. New York: Harper & Row, 1973.

Wolterstorff, Nicholas P. ‘Conundrums in Kant's Rational Religion.’ In *Kant's Philosophy of Religion Reconsidered*, edited by Philip J. Rossi and Michael J. Wreen. Bloomington: Indiana University Press, 1991.

Wood, Allen. *Kant's Moral Religion*. Ithaca: Cornell University Press, 1970.

_____. *Kant's Rational Theology*. Ithica: Cornell University Press, 1978.

_____. ‘Rational Theology, Moral Faith, and Religion.’ In *The Cambridge Companion to Kant*, edited by Paul Guyer. Cambridge: Cambridge University Press, 1992.

_____. *Kant's Ethical Thought*. New York: Cambridge University Press, 1999.

_____. ‘The Final Form of Kant's Practical Philosophy.’ In *Kant's Metaphysics of Morals*, edited by Mark Timmons, 1-22. Oxford: Oxford University Press, 2003.

_____. ‘The Supreme Principle of Morality.’ In *The Cambridge Companion to Kant and Modern Philosophy*, edited by Paul Guyer, 342-80. Cambridge: Cambridge University Press, 2006.

_____. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.

Zanuso, F. ‘The Current Interest in Kant in the North American Debate on Criminal Punishment.’ *History of European Ideas* 30, no. 3 (2004): 329-348.

Zweig, Arnulf, ed. *Kant: Philosophical Correspondence, 1759-99*. Chicago, 1967.