Kant's Anatomy of Evil

EDITED BY
Sharon Anderson-Gold
and Pablo Muchnik

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521514323

This page intentionally left blank

Kant's Anatomy of Evil

Kant infamously claimed that all human beings, without exception, are evil by nature. This collection of essays critically examines and elucidates what he must have meant by this indictment. It shows the role which evil plays in his overall philosophical project and analyzes its relation to individual autonomy. Furthermore, it explores the relevance of Kant's views for understanding contemporary issues such as crimes against humanity and moral reconstruction. Leading scholars in the field engage a wide range of sources from which a distinctly Kantian theory of evil emerges, both subtle and robust, and capable of shedding light on the complex dynamics of human immorality.

Sharon Anderson-Gold is Professor and Chair at the Department of Science and Technology Studies in the Rensselaer Polytechnic Institute. Her previous publications include *Cosmopolitanism and Human Rights* (2001) and *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (2001), which was nominated for the North American Society for Social Philosophy Book Prize.

Pablo Muchnik is an Associate Professor of Philosophy at Siena College. He is the author of *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Aprioricity of History* (2009), and editor of *Rethinking Kant* (vol. I, 2008; vol. II, forthcoming).

Kant's Anatomy of Evil

Edited by

SHARON ANDERSON-GOLD

PABLO MUCHNIK

CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521514323

First published in print format 2009

ISBN-13 978-0-511-69145-4 eBook (NetLibrary)

ISBN-13 978-0-521-51432-3 Hardback

Contents

Contributors

SHARON ANDERSON-GOLD is Professor of Philosophy at Rensselaer Polytechnic Institute. She is the author of *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (State University of New York Press, 2001) and *Cosmopolitanism and Human Rights* (University of Wales Press, 2001). She has written numerous articles on Kant's moral, social, political philosophy, and philosophy of history. She is the outgoing President of the North American Kant Society.

CLAUDIA CARD is the Emma Goldman Professor of Philosophy at the University of Wisconsin. Her publications include: *The Atrocity Paradigm: A Theory of Evil* (Oxford University Press, 2002), *The Unnatural Lottery: Character and Moral Luck* (Temple University Press, 1996), and *Lesbian Choices* (Columbia University Press, 1995). She is currently in the last semester of a five-year appointment as Senior Fellow at the Institute for Research in the Humanities (Madison, WI), where she is at work on another book on evil and an introduction to feminist philosophy.

PATRICK FRIERSON is an Associate Professor at Whitman College in Walla Walla, WA. Frierson is author of *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge University Press, 2003). In addition, he is co-editor of *Kant: Observations on the Beautiful and Sublime and other writings* (Cambridge University Press, 2008) and has published articles on Kant in journals such as *The Journal of the History of Philosophy*, *Philosopher's Imprint*, and *Kantian Review*. Frierson also serves on the editorial board of *Studies in the History of Ethics*.

JEANINE M. GRENBERG is an Associate Professor at St. Olaf College in Northfield, MN. Her specialties include Kant, ethics, and the history of modern philosophy. She is the author of *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005). She has also published in the *Journal of the History of Philosophy* and *Kant-Studien*. She has received major writing grants from the American Council of Learned Societies, the American Association of University Women, and the Woodrow Wilson Foundation. She is currently working on a book about freedom and human limits in the seventeenth and eighteenth centuries.

ROBERT B. LOUDEN is Professor of Philosophy at the University of Southern Maine. He is the author of *The World We Want: How and Why the Ideals of the Enlightenment Still Elude Us* (Oxford University Press, 2007), *Kant's Impure Ethics: From Rational Beings to Human Beings* (Oxford University Press, 2000), and *Morality and Moral Theory: A Reappraisal and Reaffirmation* (Oxford University Press, 1992). Louden has also translated Kant's *Anthropology from a Pragmatic Point of View* (Cambridge University Press, 2006), and is co-editor and translator of two volumes in the Cambridge Edition of the Works of Immanuel Kant – *Anthropology, History, and Education* (Cambridge University Press, 2007) and *Lectures on Anthropology* (Cambridge University Press, forthcoming).

GORDON E. MICHALSON Jr. has served as President of New College of Florida since 2001. He is the current President of the North American Kant Society. He is a specialist in modern Western philosophy of religion and the author of *Kant and the Problem of God* (Blackwell, 1999), *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge University Press, 1990), *The Historical Dimensions of a Rational Faith: The Role of History in Kant's Religious Thought* (University Press of America, 1977), as well as numerous articles on Kant's ethical and religious philosophy. He has served as American consulting editor of *The Blackwell Encyclopedia of Modern Christian Thought*.

PABLO MUCHNIK is an Associate Professor of Philosophy at Siena College (Albany, NY). He specializes in Kant, early modern philosophy, ethics, and political philosophy. He is the author of *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Aprioricity of History* (Lexington Books, 2009), editor of the first two volumes of *Rethinking Kant* (Cambridge Scholar Publishers, vol. I, 2008; vol. II,

forthcoming), and director of the series *Kantian Questions* (Cambridge Scholar Publishers). He is the recipient of various national and international scholarships and awards, and is currently Vice-President of the North American Kant Society.

PHILIP J. ROSSI, S. J. is Professor of Theology at Marquette University and specializes in the philosophy of religion and Christian ethics. He has published extensively on the theological import of the work of Immanuel Kant, and is author of *The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind* (State University of New York Press, 2005), *Together Toward Hope: A Journey to Moral Theology* (University of Notre Dame Press, 1983), co-editor (with Michael J. Wreen) of *Kant's Philosophy of Religion Reconsidered* (Indiana University Press, 1992), and co-editor (with Paul Soukup, S. J.) of *Mass Media and the Moral Imagination* (Sheed and Ward, 1994).

DAVID SUSSMAN is an Associate Professor at the University of Illinois at Urbana-Champaign. He is the author of *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* (Routledge, 2001). He has published numerous articles on Kant including "Kantian Forgiveness" (*Kant-Studien*, 96, 1, 2005), "The Authority of Humanity" (*Ethics*, 113, 2, January 2003), and "What's Wrong with Torture?" (*Philosophy and Public Affairs*, 33, 1, Winter 2005). His current work focuses on emotional expression, practical irrationality, and the limits of moral responsibility.

ALLEN W. WOOD is Ward W. and Priscilla B. Woods Professor at Stanford University. Before assuming his present position at Stanford, he taught at Cornell University (1968–1996) and Yale University (1996–2000). He has also been Professor of Philosophy at Indiana University, and held visiting positions at the University of Michigan, University of California at San Diego, and Oxford University. His interests include modern philosophy, especially Kant and the German idealist tradition, and also moral and political philosophy and the philosophy of religion. His publications on Kant include *Kantian Ethics* (Cambridge University Press, 2008), *Kant* (Blackwell, 2004), *Kant's Ethical Thought* (Cambridge University Press, 1999), *Kant's Rational Theology* (Cornell University Press, 1978), and *Kant's Moral Religion* (Cornell University Press, 1970). He has edited several collections on Kant and published numerous articles. He is co-general editor of the Cambridge Edition of the Works of Immanuel Kant.

Abbreviations

Writings of Immanuel Kant are cited by volume and page number of the Akademie Edition: *Immanuel Kants gesammelte Schriften* [Ak.], Ausgabe der königlich preussischen Akademie der Wissenschaften (Berlin: W. de Gruyter, 1902–). Unless otherwise indicated, the English translations are from the Cambridge Edition of the Works of Immanuel Kant (Ca.) (New York: Cambridge University Press, 1992–). The following abbreviations are used throughout the book:

EF	*Zum ewigen Frieden: Ein philosophischer Entwurf* (1795), Ak. 8 *Toward Perpetual Peace: A Philosophical Project*, Ca. Practical Philosophy
G	*Grundlegung zur Metaphysik der Sitten* (1785), Ak. 4 *Groundwork of the Metaphysics of Morals*, Ca. Practical Philosophy
I	*Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht* (1784), Ak. 8 *Idea for a Universal History with a Cosmopolitan Aim*, Ca. Anthropology, History, and Education
KpV	*Kritik der praktischen Vernunft* (1788), Ak. 5 *Critique of Practical Reason*, Ca. Practical Philosophy
KrV	*Kritik der reinen Vernunft* (1781, 1787); cited by A/B pagination *Critique of Pure Reason*, Ca. Critique of Pure Reason
KU	*Kritik der Urteilskraft* (1790), Ak. 5 *Critique of the Power of Judgment*, Ca. Critique of the Power of Judgment
MA	*Mutmaßlicher Anfang der Menschengeschichte* (1786), Ak. 8

Conjectural Beginning of Human History, Ca. Anthropology, History, and Education

MS *Metaphysik der Sitten* (1797–8), Ak. 6
Metaphysics of Morals, Ca. Practical Philosophy

MVT *Über das Mißlingen aller philosophischen Versuche in der Theodicee* (1791), Ak. 8
On the Miscarriage of All Philosophical Trials in Theodicy, Ca. Religion and Rational Theology

O *Was heißt: Sich im Denken orientieren?* (1786), Ak. 8
What Does It Mean to Orient Oneself in Thinking?, Ca. Religion and Rational Theology

P *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können* (1783), Ak. 4
Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science, Ca. Theoretical Philosophy after 1781

R *Religion innerhalb der Grenzen der bloßen Vernunft* (1793–4), Ak. 6
Religion within the Boundaries of Mere Reason, Ca. Religion and Rational Theology

SF *Streit der Fakultäten* (1798), Ak. 7
Conflict of the Faculties, Ca. Religion and Rational Theology

TP *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis* (1793), Ak. 8
On the Common Saying: That May Be Correct in Theory But It Is of No Use in Practice, Ca. Practical Philosophy

VA *Anthropologie in pragmatischer Hinsicht* (1798), Ak. 7
Anthropology from a Pragmatic Point of View, Ca. Anthropology, History, and Education

Vorlesungen über Anthropologie, Ak. 25
Lectures on Anthropology, Ca. Lectures on Anthropology

VE *Vorlesungen über Ethik*, Ak. 27
Lectures on Ethics, Ca. Lectures on Ethics

VpR *Vorlesungen über die philosophische Religionslehre*, Ak. 28
Lectures on the Philosophical Doctrine of Religion, Ca. Rational Religion and Theology

WA *Beantwortung der Frage: Was ist Aufklärung?* (1784), Ak. 8
An Answer to the Question: What is Enlightenment?, Ca. Practical Philosophy

Introduction

Sharon Anderson-Gold and Pablo Muchnik

Contemporary debates in moral philosophy have primarily been focused on meta-ethical questions about the justification of morality, disregarding the ease with which perfectly justified norms are displaced by non-moral considerations.[1] Given the scope, magnitude, and inventiveness of human wrongdoing, this philosophical trend seems utterly misguided. The challenge does not lie so much in how to justify morality, but in understanding how perfectly justified judgments are so easily disregarded by self-serving calculations.[2]

Kant's doctrine of radical evil has much to tell us about this. Against the widespread tendency to explain evil in terms of the pernicious power of natural inclinations, Kant believed that evil represented "an invisible enemy, one who hides behind reason and hence [is] all the more dangerous" (R 6: 57). The enemy is invisible, for "no matter how far back we direct our attention to our moral state, we find that this state is no longer *res integra*" (R 6: 58n.). And it is exceptionally dangerous, for the corruption in question is self-imposed: "genuine evil consists in our *will* not to resist the inclinations when they invite transgression" (ibid.). Since this type of volition rests on a maxim, and maxim formation in Kant always takes place under the constraints of

[1] See Otfried Höffe, "Ein Thema wiedergewinnen: Kant über das Böse," in O. Höffe und A. Pieper (eds.), *Über das Wesen der menschlichen Freiheit* (Berlin: Akademie Verlag, 1995), pp. 11–34.

[2] See Pablo Muchnik, "Kant on the Sources of Evil," in *Proceedings of the Tenth International Kant Congress* (Berlin: Walter de Gruyter, 2009), pp. 287–97.

the categorical imperative, evil hides at the heart of practical reason: it is the deliberate attempt to subordinate what we ought to do in favor of what pleases us. This subordination entails a reversal of the moral order of priority between the incentives in the human will: "self-love and their inclinations [become] the condition of compliance with the moral law – whereas it is the latter that, as the *supreme condition* of the satisfaction of the former, should have been incorporated into the universal maxim of the power of choice as the sole incentive" (R 6: 36).

As a result of the excessive influence of the *Groundwork* in the Anglo-American reception of Kant, however, Kant's reflections on evil have been largely ignored in the secondary literature. Kant's optimistic thesis about the analyticity of freedom and morality, by which the autonomous will (*Wille*) is equated to practical reason, has been mistakenly taken as Kant's last word regarding human freedom.[3] This view overlooks Kant's gloomier reflections about the inextirpable propensity to evil in human nature, for which we are nonetheless responsible.

This collection of essays is an effort to set the record straight. Its primary goal is to explore the intellectual resources available in Kant for dealing with the question of evil. It places Kant's views in the context of the critical system, interprets some of Kant's most controversial assumptions, and extends his conception in novel ways to deal with urgent contemporary issues. There is more at stake, however, than settling a family dispute among Kantians here: acknowledging the promptness with which human beings are willing to neglect the claims of morality invites an account of human motivation and agency in which a robust conception of evil plays a central role. This is an invitation contemporary moral philosophers should not refuse. By making Kant's conception of evil more available, we hope to contribute (if only indirectly) to an overdue shift in philosophical attention.

I

The anthology opens with Philip Rossi's essay, "Kant's 'Metaphysics of Permanent Rupture': Radical Evil and the Unity of Reason." Following

[3] For the seeds of this common misunderstanding, see, e.g., M. Kosch, *Freedom and Reason in Kant, Schelling, and Kierkegaard* (Oxford: Clarendon Press, 2006), H.E. Allison, *Kant's Theory of Freedom* (Cambridge University Press, 1990), and G. Prauss, *Kant über Freiheit als Autonomie* (Frankfurt am Main: Vittorio Klostermann, 1983).

Susan Neiman, Rossi argues that Kant's philosophy is not merely a response to certain epistemological and metaphysical questions (i.e., how are a priori synthetic judgments possible).[4] More importantly, it is a response to the presence of evil, which threatens the very intelligibility of the world and our need to feel at home in it. Evil makes manifest a rift between the world as *it is* and the world as it *ought to be*, inciting us to find unity and overcome the fracture. According to this reading, the key to that unity lies in the rationalist principle of sufficient reason, which introduces the regulative demand that *is* and *ought* should coincide. Thus, an aspect largely ignored by mainstream Kantian interpretation comes to the fore: perplexity about evil is the impetus behind Kant's unification of theoretical and practical reason. The bafflement and threat of futility that overtake us when evil breaks the nexus of intelligibility drive the Kantian philosophical enterprise. For, as Rossi indicates, the most effective line of defense against evil is human solidarity, the promotion of which requires a drastic transformation of current social practices. Kant's philosophical ingenuity resides, then, in having channeled our metaphysical perplexity in the face of evil into productive practical uses. Critical philosophy is ultimately a kind of "anthropodicy," an immanent attempt at humanizing the world that makes transcendent flights into theodicy look outmoded and unwarranted.

Radical evil, "the foul stain of our species" (R 6: 38), it would seem, presents the most formidable obstacle against this project of human vindication. In "Kantian Moral Pessimism," however, Patrick Frierson shows how Kant's unflinching awareness of our moral deficiencies is not only compatible with moral progress, but also preferable to the anthropological optimism prevalent in contemporary moral theorizing. According to the latter, the main failings of human beings are explained by non-moral factors (knowledge, competence, social conditions, non-culpable negligence, etc.), which have little to do with "evil." This optimism pervades, for example, recent work in empirical social psychology (the situationism of Gibert Harman and John Doris), and even the best normative ethics of Kantian extraction. As case in point, Frierson interprets central themes in Barbara Herman.

[4] S. Neiman, *Evil in Modern Thought: An Alternative History of Philosophy* (Princeton University Press, 2002).

Her rules of moral salience, analysis of non-moral motivation, and discussion of the impact of morality in our identity come under Frierson's fire. For they operate "under morally optimistic background assumptions" (p. 38). The problem is that these assumptions lead Herman to interpret our misdeeds in terms of factors for which we do not acknowledge full responsibility, and this interpretation legitimizes strategies Kant would consider self-ingratiating and self-deceptive. Although Kant's anthropological pessimism stymies these strategies, it does not let us fall into despair. On the contrary, Kant offers an inspiring vision of moral hope, "of endless progress [toward] complete conformity with the moral law" (KpV 5: 122). This hope, however, comes at a price: since the corruption of our moral character is radical, and at the same time it is our own fault, evil cannot be extirpated "through human forces" (R 6: 37) and requires the supernatural cooperation of God's enabling grace.

Kant's leap into transcendence is filled with tensions. In "Kant, the Bible, and the Recovery from Radical Evil," Gordon Michalson questions the feasibility of Kant's strategy to reduce the Bible to a rational/ethical core independent from theology. Michalson argues that Kant's appeal to the religious language of a "new man" and a "rebirth" to capture the temporal character of moral conversion does not work as it is supposed to, i.e., as a mere illustration of a self-standing moral argument. Rather, biblical references "serve as a *substitute* [for an argument nowhere to be found], as pictorial filler for a conceptual lacuna" (p. 58). Without this "filler," the moral community would lapse into apathy, for it would have no representation of what it is aspiring to. Yet, biblical references transcend the boundaries of applicability of Kantian concepts and are meant to account for a noumenal change that eludes rational explanation. Michalson detects, then, a fundamental aporia in Kant's *Religion*: on the one hand, it is necessary for us to *imagine* moral change in order to bring it about; yet, on the other, without violating the critical strictures, it is impossible to provide a conceptual account of such a change. Here is where biblical narrative comes to Kant's rescue: religious imagery "conveys the incommensurability between moral change and temporality while still offering language that helps us to represent the change" (p. 64). Although biblical language is not conceptual, it occupies a space whose void would otherwise be intolerable. "Biblical allusion thus becomes a kind of placeholder – an apparently indispensable placeholder – for the

narrative element that Kant's philosophical position requires but cannot provide" (p. 65). Michalson's analysis shows that religious narratives are not mere "parerga," as Kant used to believe, but have a function similar to the schematism in the first *Critique*. In both cases, something entirely rational (moral change, the categories) can be "represented" to the senses without erasing their respective boundaries.

II

Reduced to its bare essentials, Kant's conception of evil rests on three assumptions: (1) evil constitutes the underlying disposition of the human will (and hence is "radical"); (2) evil consists in the motivational primacy of the principle of self-love; and (3) there is a universal propensity to evil in all human beings, even the best.

All these assumptions are ripe for dispute. In "Kant's Moral Excluded Middle," Claudia Card argues that Kant's conception is flawed in at least two fundamental ways. First, Kant's theory of the will is "rigorist" and thus excludes all moral conditions that might be called intermediary, i.e., "neither good nor evil." Motivating Card's concern is the suspicion that the human will may not be a unitary, uniform, and internally consistent decision-making mechanism, as Kant presumed it to be. The best evidence we have to discover the nature of our will consists in the patterns of choice we observe over time. Here, Card notices, phenomena overwhelmingly point at the presence of conflicting volitional patterns, which suggest ambivalence and pluralism not the monolithic picture Kant favors. Furthermore, Card maintains that not all moral wrongs are evils: "culpability increases, other things equal, with increase in the harm the perpetrator is wrongfully willing to inflict" (p. 75). According to Card, Kant's harm-insensitivity sets him at odds with ordinary moral judgments: Kant's exclusive concern with culpability not only leads him to conflate serious and minor transgressions, but also to overlook the widespread phenomenon of having "moral scruples" and "making concessions" to morality, even among those who are committed to the principled pursuit of self-love. Kant can be spared from these blunders and remain true to himself, Card suggests, by incorporating a harm-sensitive dimension to his theory. "Radical harm," then, would complement Kant's "radical culpability," bringing radical evil in line with our ordinary judgments.

In "Evil Everywhere: The Ordinariness of Kantian Radical Evil," Robert B. Louden mounts a sustained defense of Kant's position against the most frequent objections in the literature. Most criticisms, Louden argues, rest on misunderstandings – once they are cleared away, the alleged shortcomings prove to be "in fact a strength" (p. 95). To begin with, Louden dismisses the objection of explanatory impotence, most thoroughly developed by Richard Bernstein.[5] This criticism is off target: Kant never sets out to explain *why* human beings use freedom the way they do. Due to our epistemological limitations such explanations would be self-defeating: the source of free acts and the nature of our motives are inscrutable in principle. This does not mean, of course, that evil must be passed over in silence. Kant unambiguously identifies self-love as "the source of all evil" (R 6: 45). But, again, this identification seems naïve and disappointing to many interpreters. As H. Arendt famously argued, horrendous crimes cannot be explained "by comprehensible motives" such as "self-interest, greed, covetousness, resentment, lust for power, and cowardice."[6] All these motives fall under the rubric of self-love, and this principle seems too shallow to account for the totalitarian rendering of "all men … equally superfluous," a crime that "breaks down all standards we know." Although at one time Louden was sympathetic to this line of thought, he now maintains that self-love is a broad motivational notion and should not to be confused with selfishness. For Kant, the problem with self-love is that it refuses to recognize moral restrictions.[7] Moral incorrigibility, not egotism or a trivial concern for happiness, is what makes self-love a candidate for "evil." Thus interpreted, self-love is a motivational source capable of encompassing a variety of distinct types of desires and inclinations, and is even compatible with a great deal of unselfishness. It is not necessary, then, to invoke a diabolical will to account for egregious moral transgressions. Kant's rejection of diabolical evil has nothing to do with the limitations of his moral

[5] R. J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity, 2002).

[6] Hannah Arendt, *The Origins of Totalitarianism*, new edition with added Prefaces (San Diego: Harcourt, 1994), p. 459.

[7] Louden follows Andrews Reath here. See A. Reath, "Kant's Theory of Moral Sensibility," in *Agency and Autonomy in Kant's Moral Theory* (Oxford: Clarendon Press, 2006), p. 16.

psychology, as John Silber used to argue.[8] It rests on the grounds that moral accountability requires the capacity to consciously judge one's actions as being contrary to the moral law. The outright rejection of morality would turn the agent into a wanton, incapable of making moral discriminations, and thus unanswerable for the havoc she wreaks.

In "An Alternative Proof of the Universal Propensity to Evil," Pablo Muchnik develops an argument to justify the synthetic a priori character of Kant's claim "man is evil by nature." His strategy is to draw a systematic distinction between the seemingly identical concepts of "disposition" (*böse Gesinnung*) and "propensity" (*Hang zum Bösen*). While the notion of "disposition" indicates the fundamental moral outlook of an *individual agent*, the notion of "propensity" is meant to refer to the moral character of the *whole species*. The single appellative "evil," therefore, ranges over two different types of moral failure: an "evil disposition" is a failure to realize the *good* (i.e., to give duty motivational priority), whereas an "evil propensity" is a failure to realize the *highest good* (i.e., to engage in the collective project of shaping nature according to the demands of freedom). The correlation between units of moral analysis and types of obligation, Muchnik contends, clears the path for a philosophical justification of Kant's infamous claim: the attribution of radical evil to the species hinges on the same anthropological limitations that give rise to the doctrine of the highest good. According to this reading, Kant's proof is not really missing, as many interpreters have argued, but misplaced and buried where no one expects to find it, namely, in the Preface to the first edition of the *Religion*. Kant's coveted proof, Muchnik acknowledges, will probably disappoint the purists, since it falls short of the strict demonstrative standards of the first *Critique*. There is no denying it: the "transcendental" argument Kant advances in the *Religion* incorporates elements of his moral psychology arrived at by experience and is unabashedly "impure." Yet, it goes a long way to justify the subjective

[8] J. R. Silber, "The Ethical Significance of Kant's *Religion*," in T. M. Grene and H. H. Hudson (eds.), *Religion within the Limits of Reason Alone* (La Salle: Open Court, 1934; 2nd edn, New York: Harper & Row, 1960), and "Kant at Auschwitz," in G. Funke and T. M. Seebohm (eds.), *Proceedings of the Sixth International Kant Congress* (Washington, D.C.: Center for Advanced Research in Phenomenology and University Press of America, 1991), pp. 177–211.

necessity, universality, and a priori character of the propensity to evil. "[Its] *hybrid* nature … is in line with the general thrust of the *Religion*, a book whose moral anthropology has also a quasi-transcendental ring, neither reducible to empirical observation nor totally severed from it" (p. 118). By striking a middle ground, Muchnik's alternative proof is intended to solve "an unfortunate dilemma Kant poses to the interpreter: either to emphasize the widespread social/empirical dimensions of evil at the expense of its noumenal origin (the path Wood follows), or to stress its noumenal origin at the expense of its social/empirical dimension (Allison's alternative)" (pp. 127–28).

III

Even if the reader were convinced by Kant's controversial assumptions regarding rigorism, self-love, and the infamous claim that "all human beings are evil by nature," the problem of how best to interpret evil still remains. In "Kant and the Intelligibility of Evil," Allen Wood argues that a *sine qua non* for taking evil seriously is to regard it as "intelligible" – that is, as an objective phenomenon we have decisive reasons for *not* doing. But, if an evil action is one there are decisive reasons *not* to do, then evil is a species of motivated irrationality, a coherent description of which is notoriously difficult. According to Wood, Kant tackles this problem in two stages: first, he identifies "the fundamental maxim of evil," which allows him to conceptualize "evil choices as following a highly general pattern" (p. 150); secondly, he interprets "this general pattern … as fitting into human nature as it shows itself under the conditions in which human life has developed on earth" (ibid.). Wood calls these two explanatory stages "the maxim problem" and "the propensity problem," respectively. We need the second, broader sense of intelligibility, because without understanding why evil is such a persistent feature of the human condition, we would not know how to struggle against it. This becomes clear if one relates the *Religion* with Kant's essays on history, where he identifies radical evil with the dynamics of "unsocial sociability." According to Wood, "*the human propensity to evil arises in the social condition, and develops along with the processes of cultivation and civilization that belong to it*" (p. 159). These processes bring about a situation of mutual dependency tied up with an anxiety "to gain worth in the opinion of others" (R 6: 37). Although originally a desire for equality, this anxiety gradually (though ineluctably, given the development of civilization)

becomes a striving for ascendancy, i.e., "an unjust desire to acquire superiority for oneself over others [upon which] can be grafted the greatest vices of secret or open hostility to all whom we consider alien to us" (ibid.). Linking the moral excesses of individual and collective competitiveness with the development of social organization, Kant renders evil as intelligible as it can be. As a consequence, institutional arrangements become the battleground for moral progress, because it is at this level that the competitive tendencies associated with radical evil can be better controlled. The nub of Wood's interpretation, then, is that evil is "a mechanism employed by natural purposiveness in developing our species's predispositions in history" (p. 163).

In "Social Dimensions of Immanuel Kant's Conception of Radical Evil," Jeanine Grenberg finds three basic difficulties with Wood's account: (1) it tends to undermine the individual's responsibility and autonomy; (2) it obliterates the transcendental origin Kant attributes to the propensity to evil; and (3) it overlooks the fact that, unfortunately, evil takes many forms. Although Wood clearly is an individualist when it comes to moral responsibility, Grenberg finds a troubling ambivalence in the explanatory role he attributes to society in the genesis of evil. There is a trivial sense in which the presence of others provides a materially necessary condition for injuring them. But Wood, Grenberg contends, is claiming more than that: he endorses the Rousseauian view that in solitude the individual is good and tranquil, and it is people that "mutually corrupt each other's disposition" (R 6: 93). Undoubtedly, the social setting provides the most notorious example of our competitive/comparative frame of mind. Yet, in the Kantian account, the propensity to evil must pre-exist our social engagements. Blaming others for my own moral corruption is a form of self-deception – a symptom of the inversion of the ethical order of priority, not an explanation of how it came about. Grenberg's complaint, then, is that Wood confuses the cause with the symptom, and this confusion tends to dilute our individual responsibility. Furthermore, Grenberg takes issue with the problematic empirical status of "unsocial sociability," the cornerstone of Wood's interpretation: "reducing evil to a tendency in our interactions with other persons, Wood seems to have forgotten both that choice of this propensity is 'prior to every use of freedom' … and that evil is a tendency to place concerns for self over 'morality' or 'the moral law' (R 6: 36), not simply over 'others'" (pp. 178–79). To support this last point, Grenberg develops an account of the "social" in Kant, which she identifies with "shared purposes."

Not all moral transgressions can be reduced to this sphere: suicide, for instance, contravenes the duty of self-preservation (associated with the predisposition to animality), but does not necessarily undermine "shared purposes." Grenberg's point is that the possibilities for evil exceed the limits of the predisposition to humanity and the dynamics of "unsocial sociability." Morality does not simply overlap with what we share with one another. Regrettably, evil has a polymorphic character and is irreducible to a single form.

A reply to this type of criticism can be found in the last section of Wood's essay. There Wood argues that the social dynamics of evil are compatible with Kant's commitment to transcendental freedom. Furthermore, to the extent that the propensity to evil is meant to elucidate "why we have a propensity to give the rationally weaker incentives of inclination or self-love priority over the rationally stronger incentives of morality" (p. 167), and that it is in the social condition that we come to value our status in the eyes of others more than our dignity as moral persons, Wood contends that the propensity to evil should not be limited to the violations of duties toward others, but also includes the condition for the possibility of violating duties to oneself. At the end of the day, in Wood's reading, Kant's appeal to the social condition "provides the necessary context for developing our radical propensity," but does not entail that "society forces us to choose evil maxims, removing or diminishing our responsibility for these choices" (pp. 168–69). According to Kant, good or evil is always up to us, and those who blame society for their corrupt disposition are already "morally bankrupt" (p. 169).

IV

To give a taste of the relevance of Kant's view to contemporary moral discussions, we conclude our *Anatomy of Evil* with reflections on genocide and moral reconstruction.

Because of its collective nature, extraordinary moral gravity and scope, genocide seems to mock our hopes for moral progress. Although no philosopher has championed the value of humanity more forcefully than Kant, genocide represents a form of "radical harm" of the type Claudia Card holds Kant is not prepared to accommodate. In "Kant, Radical Evil, and Crimes against Humanity," Sharon Anderson-Gold challenges this conclusion. She argues that self-love, as it operates in individuals, is not limited to the "interests of the physical self"

(p. 196), but can be extended to collective identities and goals. Since for Kant "individual identities arise in a social context where self-love shapes itself in accordance with the interests of those with whom we identify" (ibid.), there is no reason to endorse an individualistic reading of the Kantian self. To the extent that group identity also exists in a comparative/competitive social environment, "extended self-love" can become the basis for the infliction of grave harms on those who do not share the identity favored by the hegemonic group. Given that moral character is always formed in a social and cultural context, evil may come to expresses the internalization of social norms embodying morally corrupt objectives. Drawing on the work of several genocide scholars, Anderson-Gold describes the process whereby the identity of a devalued group becomes gradually represented as a "threat," preparing the way for a program of extermination. This process, however, does not abolish personal responsibility: "While individuals may be differently situated with respect to the enactment of specific harms and thereby hold different degrees of guilt, individuals nonetheless share responsibility for the identities that they mutually construct. Members of social groups are responsible for the attitudes that they hold and which provide support for the actions of other group members" (p. 206). Shared responsibility is thus compatible with individual freedom and accountability. Although genocide is an extreme manifestation of culturally based conflict, its explanation does not require a special incentive structure, different from regular forms of evildoing. The social character of the Kantian self can explain how people are capable of committing extraordinary crimes out of ordinary self-love. Radical harm does not call for moral monsters.

David Sussman's essay, "Revolution and Reconciliation: Toward a Kantian Account," tackles the problem of "moral reconstruction" of political communities which have undergone traumatic experiences of injustice against some of their own members. Sussman notes that Kant's contractualist commitments lead him to draw a stark dichotomy between the state of nature and the civil condition. This dichotomy produces a deadlock when it comes to punishing persons who have committed grave crimes under the old regime. If the authority under the former regime is to be considered legitimate, then the perpetrators of injustice are not punishable; yet, if there was no legitimate authority, neither was there any morality to transgress in the previous condition, because individuals were in a state of nature. Bereft of a unified perspective from which we can require a public accounting

on the part of the perpetrators, it seems, the new regime must wipe the slate clean in order to count as legitimate. But this wholesale exculpation is unacceptable and morally offensive: no just polity can ask its citizens to accept on equal terms those who had severely abused human rights among its members. Although Kant's political philosophy is unable to resolve this conceptual deadlock, Sussman argues, Kant's model of individual "moral revolution" provides a blueprint with which to reconstruct the moral fabric of communities torn asunder by historical injustices. Drawing on the notion of "suffering" necessary for repentance, Sussman maintains that the new polity and its citizens, including those who had been treated unjustly, must bear the burden of accepting malefactors back into the community. Yet, malefactors, though legally immune from punishment, must be willing to forfeit this immunity and openly confess their culpability, i.e., they must voluntarily undergo public accounting for their crimes.

Sussman's artful analogy between "moral revolution" and "political reconstruction" underlines a common theme in these pages: the nature of evil forces us to think of ways to connect, in a single explanatory framework, the individual and the whole – the micro- and the macro-levels of moral analysis. Crimes against humanity are a good example of this interconnection, and Kant's theory shows how they can be made intelligible. Without ignoring evil's phenomenological complexity, Kant's identification of principled self-love as its fundamental source can account for the most insidious moral failures; and, when interpreted in the context of human historical development, can also account for why such failures are so persistent and pervasive.

Kant's most lasting contribution to human welfare has been to turn awareness of the "foul stain of our species" into a spur for cleansing it: "for as long as we do not remove it, [evil] hinders the germ of the good from developing as it otherwise would" (R 6: 38). By rooting evil in human freedom, Kant placed solidarity at the forefront of the moral struggle, pointing us towards the ethical community. If nothing else, the Kantian project of transforming the world according to the demands of morality teaches us to supersede moral despair with moral hope. If, as Kant says, "morality inevitably leads to religion" (R 6: 6), Kant's is a religion where redemption must be brought to earth by our hands. It is moral activism that makes us worthy of grace, anthropodicy that leads to a new form of theodicy.

1

Kant's "Metaphysics of Permanent Rupture"

Radical Evil and the Unity of Reason

Philip J. Rossi, S.J.

Introduction

In *Evil in Modern Thought: An Alternative History of Philosophy*, Susan Neiman traces the history of modern philosophy – and of Kant's pivotal role in that history – along a trajectory shaped by the problem of evil rather than by the problems of knowledge, certainty, and doubt that have been the staple of standard readings of that history. She characterizes Kant's account of our human circumstances as a "metaphysic of permanent rupture" in which

> [t]he gap between nature and freedom, *is* and *ought*, conditions all human existence … Integrity requires affirming the dissonance and conflict at the heart of experience. It means recognizing that we are never, metaphysically, at home in the world. This affirmation requires us to live with the mixture of longing and outrage that few will want to bear.[1]

I would like to thank Aaron Smith for helpful comments on an early draft of this essay and Michael Cumings for assistance in preparation of the final copy.

[1] S. Neiman, *Evil in Modern Thought: An Alternative History of Philosophy* (Princeton University Press, 2002), p. 80. For a similar reading of the challenge that Kant takes reason to face, see O. O'Neill, "Reason and Autonomy in *Grundlegung* III," in *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge University Press, 1989), p. 61: "From the first sentence of the first *Critique* we are warned of the predicament of a reason that aspires to a task that it cannot achieve. Reason's failure is that it cannot give a unified account of nature and freedom. *The metaphor of the intelligible world signals the finitude, not the transcendence of human reason.*"

In this essay I plan to show how the duality that Neiman marks out as "the dissonance and conflict at the heart of experience" functions to outline the contours of a philosophical anthropology that is embedded in Kant's critical project. The spatio-temporally embodied freedom of finite human reason stands at the conceptual center of this anthropology and serves as locus for Kant's account of evil.[2] That account exhibits evil as marking a fissure that lies athwart human efforts to render fully intelligible the world that presents itself to us, in our embodied freedom, both as nature – an object for reason's theoretical inquiry – and as freedom – a field for human action shaped by reason's moral exercise.

My presentation of the path that leads to this anthropology, as well as the brief sketch of it offered in the final section, builds upon Neiman's reading of the central role that the question of evil plays in Kant's thought, but it will not directly attempt to supplement the case that she makes in favor of the faithfulness of that reading to Kant's thought. This essay is thus offered not primarily as an exercise in analyzing and reconstructing particular arguments about evil in Kant's texts but as an interpretative exploration of a central question that Neiman's reading of Kant's treatment of evil raises: Why, in the face of the intractable resistance evil presents to human efforts to render it intelligible, is it important – indeed, even necessary – for the integrity of our humanity to persist in those efforts?[3] The reason that Neiman proposes to justify such persistence – "To abandon the attempt to comprehend evil is to abandon every basis for confronting it, in thought as in practice"[4] – is more than an expression of a moral concern that, if we cease to engage in intellectual efforts to make sense of evil, we eventually will falter and ultimately fail in our moral efforts to resist and overcome it. Neiman's remark also expresses an incisive understanding

[2] Kant's discussions of "incentives," though not explicitly framed in terms of embodiment, nonetheless articulate a central dimension of the embodied character of human freedom: Our freedom is such that we can incorporate into maxims determining our action incentives both from reason and from inclination, which, in view of our embodiment, functions under spatio-temporal determinants. See KpV 5:71–8; R 6: 36–7, 44–52.

[3] See R 6: 47–51; KU 5: 450–3 for two important texts in which Kant underscores how the sustaining of moral effort is a function of a hope originating in the recognition of the "moral vocation" we have in virtue of our freedom.

[4] Neiman, *Evil in Modern Thought*, p. 325.

that at stake in the question of evil for Kant is nothing more nor less than a principle that lies at the heart of his critical project: the unity of the theoretical and the practical uses of our finite human reason that is necessary for our efforts to render intelligible the world that we engage both in thought and in action.[5] The unity of reason provides our most fundamental human recourse against the power that evil has – as unintelligible surd, adamantly resistant to efforts to exact sense from it – to shatter our efforts to make sense of the world and to fracture into disarray whatever hope we may have to give meaning to our human lives. So the question needs to be posed: How is it possible for us to hold together as one – as Kant affirms we can and must for the very integrity of our humanity – these fragile powers of our reason in the face of the metaphysical rupture that evil presents?

What I thus also hope to show in this essay is how we may understand Kant's affirmation of the unity of reason as an integral feature of his account of evil and our human possibilities for overcoming it. Within that account, the unity of reason is not given beforehand but rather *enacted by the exercise of our finite freedom in resistance to evil*.[6] In the absence of that resistance, evil otherwise presents itself as thoroughly intractable to our human efforts to make sense of it as a factor in the world in which we think and act. Affirming the unity of our human uses of reason, in the face of a "metaphysical rupture" that runs both through the world and through the very makeup of our humanity, is thus recognition that reason gives us power to stand against evil: the only way to "make sense" of evil is to commit oneself to the project of resisting it. In affirming the unity of reason we affirm the power reason provides us to envision – and to act upon – ways to stand against evil by bringing together the fractured pieces of the world and of our own humanity that lie along the fissure that evil drives through our

[5] Cf. KrV "The Canon of Pure Reason," esp. A795–819/B823–47, for Kant's articulation of the unity of reason at the outset of his critical project. As is the case for many of the key aspects of that project, Kant revisits, refines, and reformulates his account at a number of later points. See S. Neiman, *The Unity of Reason: Rereading Kant* (New York: Oxford University Press, 1994) for an account of the trajectory along which Kant's account moves.

[6] That the unity of reason is enacted rather than fully given beforehand should not be surprising in view of the primacy that Kant assigns to the practical use of reason, i.e., the use of reason through which "the highest good" is to be effected. See KpV 5: 134–6.

attempts to make coherent sense of our experience of the world as, at once, nature and freedom. The unity of finite human reason is thus not simply given, nor can it be taken for granted as unproblematically attainable; it is a unity that is forged and constantly re-forged in and through human resistance to evil.

Evil and the Relentless "Why" of Reason

Kant used a variety of coordinate terms to characterize the duality which, on Neiman's account, constitutes a "dissonance and conflict at the heart of experience" that renders problematic the unity of human reason's effort to resolve it. These terms have vexed generations of sympathetic and hostile commentators alike – perhaps most famously and problematically, the distinction between "phenomenon" and "noumenon."[7] It has rarely been the case, however, that the question of radical evil that Kant articulates in *Religion within the Boundaries of Mere Reason* has been pressed into service as a key interpretative guide to the contours of the fissure that he sees running through our human engagement with the world. That discussion of the moral structure of evil seems to offer little promise for interpretative purchase upon distinctions fundamental to the critical project so long as Kant's affirmation of a duality of nature and freedom is understood – as it has often been – as a response to epistemic and metaphysical issues that are taken to stand in isolation from moral and anthropological ones.[8] In

[7] See KpV 5: 5–8 for Kant's affirmation of the importance of the distinction between phenomenon and noumenon with respect to differentiating the theoretical and the practical uses of one and the same reason.

[8] The reasons for such interpretative separation are multiple. Some arise from tensions within Kant's texts about how these forms of inquiry stand to one another within the critical project, such as his claims about the primacy of the practical use of reason. Others stem from larger anti-metaphysical and a-metaphysical trajectories taken in the philosophical discourses into which Kant's work was received for much of the twentieth century, particularly among English-language commentators. Within these trajectories Kant's ethics can be read as unproblematically detachable from the metaphysical and epistemic context in which the critical project locates human moral activity; or, conversely, the metaphysical and epistemic context of the uses of reason can be understood to stand in independence from the moral character Kant attributes to the full range of human reason's engagement with the world. Still other reasons for the separation lie in the fact that Kant's most explicit and extensive treatment of evil occurs quite late in his articulation of the critical project; this suggests it might be merely a codicil to that enterprise rather than a fundamental interpretative locus.

consequence, his explicit engagement of the question of human evil in the later phases of the critical enterprise has often been considered marginal to the main conceptual and argumentative strands of his monumental endeavor to delimit the scope of human reason's engagement with the cosmos of which it is a part, in which it functions, and beyond which it drives itself to aspire.

This section will thus dispute such relegation of Kant's treatment of evil to a minor role in his critical philosophy. It takes its cue from Neiman's re-reading of the history of modern philosophy, which makes the case that evil poses questions about the intelligibility of the world that are even more basic than those that have been engaged under the heading of "the problem of evil" by the varied religious and secular forms of modern theodicy. Evil presents a problem so fundamental to the efforts of human reason to render the world intelligible – including efforts of a reason disciplined to function within the self-imposed limits of a Kantian critique – that it makes the standard modern distinctions among the genres of philosophical inquiry break down:

> Every time we make the judgment *this ought not to have happened*, we are stepping onto a path that leads straight to the problem of evil. Note that it is as little a moral problem as it is a theological one. One can call it the point at which ethics and metaphysics, epistemology and aesthetics meet, collide and throw up their hands. At issue are questions about what the structure of the world must be like for us to think and act within it.[9]

On the deeply ruptured conceptual terrain she sees as the philosophical inheritance that modernity has bequeathed to us from the efforts of its thinkers – including those of Kant – to make sense of evil,

[9] Neiman, *Evil in Modern Thought*, p. 5. See KpV 5: 146–8 ("On the Wise Adaptation of the Human Being's Cognitive Faculties to His Practical Vocation") for one text in which Kant engages the issue of "what the structure of the world must be like for us to think and act in it" in a way that suggests the aptness of Neiman's characterization of the problem of evil as "the point at which ethics and metaphysics, epistemology and aesthetics meet, collide and throw up their hands." Kant argues here that if the moral structure of the world were transparent to the theoretical use of human reason, it would become impossible for us to lead morally worthy lives; we would do what is right in view of the reward we know accrues to it, rather than in view of recognizing that its rightness makes it fit for us to do. This is part of what Neiman calls "one of his greater arguments: if we knew that God existed, freedom and virtue would disappear" (Neiman, *Evil in Modern Thought*, p. 327). See KrV A818–19/B846–7.

Neiman offers two tropes to orient us, first, to what the fractured world of the aftermath of modernity "is like" and, second, to the manner in which we must "think and act within" that world. The first trope – which stands for what the world "is like" – is "homeless." She offers this to frame our human circumstances of a "conceptual helplessness" in the face of evil that seems to have taken intellectual hold in the aftermath of the massive horrors that humans have inflicted on each other since at least the start of the twentieth century – and continue to do so in the twenty-first. The second – which stands for how we must "think and act" in the world – is the insistent "Why?" of a child's questioning. She offers this as a model for the hope in which we are called to persist as we seek our human way through the inhospitable terrain of a disenchanted world. In keeping with the remark in her first chapter – "Immanuel Kant has already appeared in this book, and will accompany it to the end"[10] – Neiman imparts to these tropes a tonality resonant with the regulative demand for intelligibility that Kant understood to be at work in the principle of sufficient reason: "that the *is* and the *ought* should coincide,"[11] that "the real should become the rational."[12]

The two tropes around which Neiman centers her account of evil thereby function as coordinates, rooted in Kant's articulation of the practical use of reason, for locating the source of the fault line running through human experience, as well as the dynamics that shape its contours, within the ambit of the exercise of finite human freedom as it is embodied into the contingency of the spatio-temporal world. This line demarcates the fracturing of human intents, purposes, and meanings as they move athwart the radical contingency that, as Neiman notes, the workings of nature present to us as the context in which we strive to make sense of the world and to satisfy our aims within it:

> our power over the consequences of our actions is really very small[13] … The gap between our purposes and a nature that is indifferent to them leaves the world with an almost unacceptable structure.[14]

[10] Neiman, *Evil in Modern Thought*, p. 61.

[11] Ibid., p. 322.

[12] Ibid., p. 323. See KrV A542–57/B570–85 for an extensive discussion of the regulative use of reason precisely with respect to judgments regarding what "is" and what "ought to be."

[13] Neiman, *Evil in Modern Thought*, p. 74.

[14] Ibid., p. 75. In the section of the *Critique of the Power of Judgment* noted above (n. 3), Kant articulates this gap in terms of a "righteous" unbeliever (explicitly mentioning

In addition to providing bearings on the fault line between "is" and "ought" upon which evil confounds human intent, each trope also captures a distinctly different modulation – one resonant and one dissonant – sounded in Kant's claims and hopes for the reason that is relentless in its pursuit of unity across the fault line demarcating what "is" and what "ought to be."[15] The insistent "Why?" of the questioning child is powered by a dogged expectation that all will, in the end, fit together in measured order. The sense that we are "homeless," a sense that, at its deepest level, the world cares not to welcome us – because, it seems, the world *is such as not to care at all* – draws us into a din where all that there is may turn out to be only unrelieved, terrifying dissonance. Attention in turn to each trope – the child's insistent "Why?" in the rest of this section, "homeless" in the next – and to the modulation each displays will provide markers along which this essay will then track the route that human reason hopes to open by persisting in the one effective mode it has for forging moral sense from and in a world fractured by evil: steadfast resistance. This route is one along which we may start to open a space upon which to learn how, even in the absence of a lasting "home" provided by the world as its "is," to make one another "at home" by welcoming each other in all our human

Spinoza) who experiences the indifference of nature even to persistent human moral efforts (KU 5: 452): "But his effort is limited; and from nature he can, to be sure, expect some contingent assistance here and there, but never a law-like agreement in accordance with constant rules (like his internal maxims are and must be) with the ends to act in behalf of which he still feels himself bound and impelled. Deceit, violence and envy will always surround him, even though he is himself honest, peaceable and benevolent; and the righteous ones besides himself that he will encounter will, in spite of all their worthiness to be happy, nevertheless be subject by nature, which pays no attention to all that, to all the evils of poverty, illness and untimely death, just like all the other animals on earth and will always remain thus until one wide grave engulfs them all together (whether honest or dishonest, it makes no difference here) and flings them, who were capable of having believed themselves to be the final end of creation, back into the abyss of the purposeless chaos of matter from which they were drawn."

[15] Charles Taylor is another interpreter who sees Kant's project fundamentally engaged with a "fault line" between reason and nature: "Just because it is a theory of freedom, Kantian moral philosophy finds it hard to ignore the criticism that the rational agent is not the whole person. This didn't lead Kant to want to alter his definition of autonomy, but he did see that the polar opposition between reason and nature was non-optimal; that the demands of morality and freedom point towards a fulfillment in which nature and reason would once more be in alignment." C. Taylor, *Sources of the Self: The Making of the Modern Identity* (Cambridge, MA: Harvard University Press, 1989), p. 385.

circumstances in a manner befitting the shared fragility and dignity of our finite, embodied human freedom.[16]

Both tropes – which Neiman takes to serve as indispensable coordinates for orienting ourselves not just to the fault line, but also to the hope by which to shape efforts to traverse it – have their origin in her Kantian reading of "the principle of sufficient reason" as a dynamic of practical intelligibility. This principle articulates reason's drive not simply to make sense of the world, but to make sense of the world as *a field that human moral activity has power to shape*. It is a demand for making sense that reason places with at least equal force upon our decision and action as it does upon our thought: "Belief that there may be reason in the world is a condition of the possibility of our being able to go on in it."[17] She characterizes this demand as "transcendental," i.e., as "located neither in normative nor descriptive space"[18]; it is one that lies inseparably at the root of both metaphysics and ethics as demands for making sense of the world that we, as beings endowed with the powers of finite reason, place upon ourselves. Inasmuch as reason, as practical, determines our possibilities for acting in a world so shot through with radical contingency, what we do, not merely what we think, in response to that contingency is crucial to the project of "making sense": "Belief that the world should be rational is the basis for every attempt to make it so."[19]

How the trope of the child's persistent "Why?" issues from the principle of sufficient reason understood as a human dynamic demanding that the world make sense is not too difficult to see:

The urge to greet every answer with a question is one we find in children not because it's childish, but because it's natural. Once you begin the

[16] Kant's discussion of "the cosmopolitan right to hospitality" both in *Perpetual Peace* and *The Metaphysics of Morals* suggests that recognition of our common human identity also involves respect for the very difference and otherness of "the foreigner" that, were we to follow self-protective inclination, we would otherwise make the basis for hostility: EF 8: 357; MS 6: 352–3.

[17] Neiman, *Evil in Modern Thought*, p. 324.

[18] Ibid., p. 323. See KrV A808/B836: "I call the world as it would be if it were in conformity with all moral laws (as it **can** be in accordance with the **freedom** of rational being and **should** be in accordance with the necessary laws of **morality**) a **moral world**. This is conceived thus far merely as an intelligible world, since abstraction is made therein from all conditions (ends) and even from all hindrances to morality in it (weakness or impurity of human nature). Thus far it is therefore a mere, yet practical, idea, which really can and should have its influence on the sensible world, in order to make it agree as far as possible with this idea."

[19] Neiman, *Evil in Modern Thought*, p. 325.

search for knowledge, there is no obvious place to stop. The fact that the desire for omniscience cannot be met does not make it either foolish or pathological. Indeed, it is embodied in the principle of sufficient reason itself.[20]

Less immediately evident, however, is the manner in which the *moral* intelligibility of the world is at stake in such persistent questioning. Neiman elucidates this point by noting that the child's persistent questioning is directed not simply at discovering how the world works but at finding reasons why the world works the way it does:

> The principle of sufficient reason expresses the belief that we can find a reason for everything the world presents. It is not an idea we derive from the world, but one that we bring to it … Kant called it a regulative principle … Children display it more openly than adults because they have been less often disappointed. They will continue to ask questions even after hearing the impatient answer – *Because that's the way the world is.* Most children remain adamant: *But why is the world like that, exactly?* The only answer that will truly satisfy is this one: *Because it's the best one.* We stop asking why when everything is as it should be.[21]

The child's persistent "Why?" is thus a marker of human reason's engagement with the fissure that runs between the world as it is and the world as it should be. As we explore that fissure – especially in the light of disappointment that the world too often turns out to be not as it should be – we begin to find that the fissure also runs within us, for we find ourselves standing on each side of the fracture between the world as it is and the world as it ought to be. The principle of sufficient reason thus also articulates a drive to find ways to bring into alignment the fundamental duality we experience in seeking to make full sense of the world our reason engages: On one hand, reason in its theoretical use, renders the world to us in terms of causal dynamism in which we are ourselves inextricably enmeshed; on the other hand, as moral agents, despite those capacities to grasp the causal dynamism of the world as it is, reason in its practical use renders the world to us in terms of possibilities for shaping the world in accord with what it

[20] Ibid., p. 320. See also P 4: 367: "That the human mind would someday entirely give up metaphysical investigations is just as little to be expected, as that we would someday gladly stop all breathing so as never to take in impure air."

[21] Neiman, *Evil in Modern Thought*, p. 322.

should be, possibilities to which the world as it is all too often manifests itself as recalcitrant.

Neiman thus frames the overall question of intelligibility – and the unity of the reason that seeks to make sense of the world and the place of humanity within it – as a question of our human "capacities to find and create meaning in the world" and pointedly asks whether they are "adequate to a world that seems determined to thwart them?"[22] She takes these capacities to function in terms of the distinction that Kant makes between the theoretical and the practical uses of reason, i.e., the former as the manner in which reason engages the world as it "is," the latter as reason engages the world as it "ought to be."

Yet, even as she follows Kant in taking the theoretical and the practical to be uses of one and the same reason upon one and the same world, the different forms of reason's engagement with the world make manifest to us a distance between "is" and "ought" that stands as a challenge to reason's fundamental task of rendering that world fully intelligible. The world "as it is" presents itself to the theoretical use of reason as the "appearance" of a nature that in its causal dynamism works, at best, indifferently to the ends and purposes that the practical use of reason proposes as befitting the dignity of our finite human freedom. Neiman notes:

> It would be easy to acknowledge that not controlling the natural world is part of being human, were it not for the fact that *things go wrong*. The thought that the rift between freedom and nature is neither error nor punishment but the fault line along which the universe is structured can be a source of perfect terror.[23]

So as mightily as Kant labors in the *Critique of the Power of Judgment*, as well as his occasional essays on history, politics, and culture, to legitimate the application of categories of purpose to the workings of nature, that legitimation is not put forth as the basis for a claim about how the world "is": Whatever purposes, if any, the world of nature may have as it "is" – "in-itself" – remain opaque in principle to the theoretical use

[22] Ibid., p. 318; see also p. 322: "the drive to seek reason in the world – even, or especially, at the points where it seems most absent – is as deep a drive as any we have."

[23] Ibid., pp. 80–1.

of finite human reason.[24] Even more important for Kant's account of evil and for the anthropology of finite freedom that forms its context, is the fact that whatever *moral* purposes we may think are necessary for our making sense of the world are not features *of* the world but rather a demand that our reason *brings to* the world. Bringing to the world as it "is" the demand of practical reason to fashion the world as it "ought to be" is central to what Kant affirms as the primacy of the practical use of reason.[25] The exercise of our finite reason brings those purposes to the world not in the mode of theoretical knowledge but in the mode of a practical hope that, by heeding the dictate of practical reason to do as we ought, we make it possible for the world to have, in a least some small measure, a moral order of which it would otherwise seem devoid.

The Unity of Reason: Finding Home on Fractured Ground

This point about the primacy of the practical use of reason provides a crucial link for elucidating the bearing of the principle of sufficient reason upon the trope "homeless." "Homeless" is a figure that, in the first instance, expresses how the world as it is presents itself as seemingly inhospitable to the hopes to which the principle of sufficient reason gives rise about the sense and meaning that we may exact from that world. It is also a figure, however, of how *we* engage a world that presents us with such a blank and bleak face. This figure thus also indicates a central formative mode for our *use* of the principle of sufficient reason in such a context. It situates our finite, embodied rational agency upon the radically fractured metaphysical and moral terrain upon which reason is nonetheless called to enact, precisely in the face

[24] See also *First Introduction to the Critique of Judgment*, KU 5: 181–7, especially 186–7: "The power of judgment thus also has itself an *a priori* principle for the possibility of nature, though only in a subjective respect by means of which it prescribes a law, not to nature (as autonomy), but to itself (as heautonomy) for reflection on nature, which one could call the **law of the specification of nature** with regard to its empirical laws, which it does not cognize in nature *a priori* but rather assumes in behalf of an order of nature cognizable for our understanding in the division that it makes of its universal laws when it would subordinate a manifold of particular laws to these."

[25] Important affirmations of the primacy of practical reason can be found in KrV "The Canon of Pure Reason," Second Section, KrV A804–19/B832–47; KpV 5: 119–21, 236–8.

of such "dissonance and conflict," a unity to its uses. As is the case for the insistent "Why?" reason's demand for intelligibility as expressed in the trope "homeless" is primarily "practical": It bears on how, in the world that "is," we are to shape what we do to accord with the world as it "ought to be." Reason's demand bears most centrally upon the manner in which our responses to the question "What ought we to do?" appropriately engage the exercise of our practical (moral) reason, i.e., our freedom, in a world in which the course of modernity and its aftermath has made manifest – even more so than was manifest to Kant – that, in the world as it "is," we stand "homeless." That world runs its course indifferently – perhaps even inhospitably – to human efforts to exact from it – under the insistent pressure of asking "Why?" – a meaning that can be ordered to our purposes.

On this terrain, the principle of sufficient reason thus becomes that in virtue of which we, as embodied agents of reason, seek to *enact a unity to reason* that will at least make possible a space for us to dwell with one another on such inhospitable terrain: the fact that the world turns an inhospitable face to us does not require that we be inhospitable to one another. Being "homeless" need not be inevitable.[26] Reason's demand that moral intelligibility be brought to the world is inextricably united with its demand for a metaphysical intelligibility of the world as the place we inhabit as embodied agents of finite reason. We enact the unity of reason in meeting the demand of the practical use of reason that we act to make the world as it "ought to be." This trope thus provides a signpost to an important feature of the anthropology of finite freedom at work in Kant's account of evil: This is an anthropology of the hope that finite reason offers us for putting back together what evil has fractured, a hope that has the sturdiness that

[26] The principle that Kant invokes in *The Metaphysics of Morals* with respect to envisioning our human capacities for making peace possible over against the putative "inevitability" of war is instructive here. Being "homeless" is no more inevitable than war, once we grasp (in hope) the possibilities that lie within our power *for making it not so*: "Now morally practical reason pronounces in us its irresistible veto: *There is to be no war*, neither war between you and me in the state of nature nor war between us as states … So the question is no longer whether perpetual peace is something real or a fiction, and whether we are not deceiving ourselves in our theoretical judgments when we assume that it is real. Instead, we must act as if it is something real, though perhaps it is not; we must work toward establishing perpetual peace and the kind of constitution that seems to us most conducive to it … and even if the complete

comes only from a recognition of the fragility of freedom from which it issues.

The two tropes are thus connected to one another through the practical use of our finite reason. Exploration of this connection will provide a context for subsequently articulating the centrality of the fragility of freedom for the anthropology at work in our enactment of the unity of reason as resistance to evil. The persistent "Why?" – which Neiman understands as a demand of reason that we refuse only at the peril of demeaning our humanity – is one that we now pose in conditions that, more starkly than did Kant, we must confront as "homeless," bereft of secure places on which to anchor a comprehensive, abiding intelligibility that makes sense beyond question of our human place in such a world. The conditions of human life at the outset of the twenty-first century provide little from which we may glean firm assurance that we have yet learned how to make the space on which we dwell a fitting "home" for one another as fellow humans, let alone for other living beings with whom we share the earth. The workings of the world of nature provide little guarantee – and we seem to provide even less to one another in the social worlds we construct to affirm "our" identity against "their" identity – that we have mastered the skills to share, in a modicum of peace, even some little space side by side with fellow human beings who are not "us." It has also started to become more apparent that even modest expectations we may have about our own security and the well-being of the generations to succeed us may fail to be satisfied on a planet on which the effects of our resource depleting human modes of living increasingly crowd and even render uninhabitable the life space of many fellow creatures.

Locating the connection between the trope of "homeless" and the principle of sufficient reason in the practical use of reason thus suggests that "homeless" stands as more than just a trenchant image of the influence that understandings (and misunderstandings) of Kant's treatment of practical reason have historically had on later depictions of the character and circumstances of the exercise of autonomous human moral agency. There may very well be sound reasons for taking Kant's articulation of "autonomy" as central to the character of moral

realization of this objective always remains a pious wish, still we are not deceiving ourselves in adopting the maxim of working incessantly toward it" (MS 6: 354–5).

reasoning and agency to stand at the head of a stream of intellectual history leading (most notably through Hegel) to later claims about "alienation" as a defining feature of the human condition and for which "homeless" could then be taken as one apt descriptor.[27] Yet Neiman's discussion implies that this trope has a connection to Kant's thought about the form that human finite reason takes that is conceptually stronger than what may be provided by even indisputable claims of historical influence. I take her to be at least suggesting – if not advancing the first stages of an argument – that this trope aptly expresses a central dynamic in Kant's understanding of the demand of human finite reason that we render the world intelligible: the trope is apt inasmuch as reason's demand for intelligibility arises in virtue of its engagement with "homeless" *as the given condition from which* our human efforts to make sense of the world begin and as the *condition to which* the demand for moral intelligibility is addressed.[28] The principle of sufficient reason is "reason's attempt to be at home in the world,"[29] an effort that arises when what "is" and what "ought to be" fail to coincide:

> For as Kant implied, but never actually stated, behind the principle of sufficient reason itself is the assumption that the *is* and the *ought* should coincide. The principle of sufficient reason starts its work where they fail to meet. When the world is not as it should be, we begin to ask why.[30]

On Kant's account human reason's demand for making sense is both relentless – there is always another "Why?" to pose – and thoroughgoing – it seeks to put the response to each "Why?" into connection with every other one. Neiman faults Kant, however, for confounding the first with the second: "Kant's greatest error was to mistake the

[27] The Kantian roots of such an "alienation" – and the "liberation" that it consequently demands – lie in the central value given to autonomy and the respect due to it. Cf. Taylor, *Sources of the Self*, pp. 363–7.

[28] O'Neill comments upon Kant's image of building a shelter (KrV A707/B735) to characterize the project of critique: "Like Descartes, Kant uses metaphors of construction to explain his view of philosophical method; but he starts with a more down to earth view of building projects … The result is in some ways disappointing, especially when matched against the rationalist ambition to build 'a tower would reach the heavens' … We may not need a lofty tower that reaches the heavens, but we need at least a modest cottage" (O'Neill, *Constructions of Reason*, pp. 11–12).

[29] Neiman, *Evil in Modern Thought*, p. 323. [30] Ibid., p. 322.

demand for reason with the demand for system."[31] With Kant she thus affirms the unity of reason, but distinguishes that unity from a "will to system." To the extent that the latter became identified as "the heart of rationalism," she considers it to be "the miserable, unspoken legacy of German philosophy."[32] As we shall see in the next and final section, recognition that the locus for "reason's attempt to be at home in the world" is constituted by a dynamic of "fracture" rather than of "system" is crucial to the articulation of an anthropology of finite freedom adequate for moral engagement of the conditions of intelligibility provided by a "metaphysic of permanent rupture." Such an anthropology, I will argue, provides the context in which the practical use of human finite reason can be the locus from which to shape a fragile but nonetheless effective hope that envisions and enacts possibilities for rendering humanly habitable for one another the fractured terrain of modernity and its aftermath.

Freedom: The Sturdy Fragility of Practical Reason

Even as she rejects Kant's association of reason's demand for making sense with a demand for system, Neiman strongly affirms Kant's view that satisfaction of the demand for intelligibility must exhibit a unity to the theoretical and the practical uses of reason, a unity in which practical use has primacy:

> Belief that the world should be rational is the basis of every attempt to make it so … the demand that reason and reality come to meet is the source of whatever progress occurs in actually bringing them together. Without such a demand, we would never feel outrage – nor assume the responsibility for change to which outrage sometimes leads.[33]

Human reason places its demand for making sense upon a world that, even as it presents itself as yielding an intelligible order of causal necessity to the theoretical use of reason, stands resistant in its radical contingency to yielding a stable unity of "is" and "ought" that is at the heart of the demand for moral intelligibility required by and for the practical use of reason. Human finite reason's engagement with the world as a demand for making sense of it all – including making sense

[31] Ibid., p. 326. [32] Ibid. [33] Ibid., pp. 325–6.

morally – can thus result only in partial satisfaction. To the extent that it has yet to result in making full moral sense of the world as a whole (including the inconstancy of our own moral efforts within it) pressing the demand seems an exercise in futility that offers grounds for contesting the primacy Kant assigns to the practical use of reason. If we cannot make moral sense of all of it, why continue efforts to make moral sense of any of it? Let us settle for making sense of the world "as it is" and be done with it. Perhaps the most we can expect is to figure out how, for the most part, the world presenting itself to our senses works; then, in emulation of Hume, we may put aside as idle any question about what purpose, if any, we serve as part of its workings. Selective attention to the principle of sufficient reason might make life less vexing, at least for those for whom the workings of the world have provided more fortunate circumstances.

Against this objection, the suggestion that we untangle human reason's demand for making sense from a demand for system – which seems a way to re-articulate Kant's distinction between metaphysics as a disposition and metaphysics as a science – provides a basis for understanding the principle of sufficient reason in terms of Kant's affirmation of the unity of reason: the principle of sufficient reason is the juncture at which the moral and metaphysical demands for making sense of the world as a whole meet. It is at peril to the integrity of reason embodied in our humanity that we ignore either side of its demand or the dynamics of their juncture with each other. This suggestion, moreover, provides a needed gloss for understanding Kant's affirmation of the primacy of the use of the practical within the unity of human finite reason. Kant's affirmation of the unity of the reason that demands we "make sense" is not also an affirmation that the finite reason that makes such a demand will finally reach the comprehension it seeks of how it "all" makes sense. Articulation of this limitation to reason's demands is central to Kant's enterprise of critique, which he sees precisely as a discipline for effecting human finite reason's self-appropriation of this limitation in each form of its exercise. The practical use of reason has primacy in this regard in that this use of our finite reason most clearly manifests the difference between "making sense" and "making system": What the practical use of our reason enjoins here and now is a making of "moral sense" with regard to specific actions and their maxims – which, for Kant, always require

resistance to a maxim of self-preference, the fundamental form in which evil presents itself to finite reason – not a comprehensive making moral sense "of it all."[34] The latter is an object of hope – which Kant takes as rationally founded – but for the immediate exercise of practical reason such hope is as much an acute awareness of the *absence* of "moral sense" in the totality of the world as it is an expression of confidence that our moral action helps bring the world closer to being as it ought to be.

To the extent that we confound – as many of Kant's successors tended to do – reason's demand for making sense with a demand for system, we are likely to overlook fragility, fracture, and incompleteness as central to the anthropological structure of the moral freedom that is the practical use of our finite reason. We are likely to miss that it is a particularly important consequence of a central point that Neiman sees Kant making insistently:

> Of the many distinctions Kant took wisdom and sanity to depend on drawing, none was deeper than the difference between God and the rest of us. Kant reminds us as often as possible of all that God can do and all we cannot. Nobody in the history of philosophy was more aware of the number of ways we can forget it.[35]

This consequence – simply put – is that while "making sense of it all" lies always beyond our grasp, that does not doom this human project to a futility that renders pointless our specific efforts to make sense of "this" or of "that." We still may make sense of the part, and put various parts together, even though comprehensive grasp of the whole ever exceeds our farthest horizon. In the exercise of the theoretical use of our reason, this sense for the limitation of reason can be a spur to ever widening the field of theoretical inquiry to find

[34] P. Guyer, "The Strategy of Kant's Groundwork," in *Kant on Freedom, Law, and Happiness* (Cambridge University Press, 2000), pp. 207–31, notes that in the *Groundwork* Kant had already identified giving priority to a maxim of self-preference as fundamental to the structure of what he will later term, in *Religion within the Boundaries of Mere Reason*, "radical evil." See G 4: 424: "If we now attend to ourselves in any transgression of a duty, we find that we do not really will that our maxim should become a universal law, since that is impossible for us, but that the opposite of our maxim should instead remain a universal law, only we take the liberty of making an exception to it for ourselves (or just for this once) to the advantage of our inclination."

[35] Neiman, *Evil in Modern Thought*, p. 75.

out how and why the world works as it "is": recognition that efforts to "make sense" of the spatio-temporal workings of the world will never be complete is far more likely to be a source of the exhilaration that prods further inquiry than a cause of the discouragement that leads to its abandonment.

It seems the opposite, however, for the exercise of reason's practical use. In that case, our inability to "make moral sense of it all" in the face of evil has had a variety of consequences, one of which has been what Neiman notes as the virtual abandonment of inquiry about evil as a central intellectual problem by much of twentieth-century philosophy: "If any one feature distinguishes twentieth century philosophy from its predecessors, it is the absence of explicit discussion of the problem of evil."[36] Of at least equal importance is the fact that the breakdown of distinctions that once offered promise for headway in making sense of evil – most notably the one between physical and moral evil that makes possible the location of evil in human intention – provides an impetus for abandoning any hope that we have power to do more than limited and local "damage control" in the face of evil that presents itself as, at once, capricious and inevitable. It seems that it will always be the case that (at least some) "things go wrong," that, at best, justice is (mostly) served imperfectly and far too often not well at all, that it is purely contingent for what "is" and what "ought to be" to converge and coincide. Discouragement about the possibility of contending with the "whole" of evil may lead to reluctance to contend with any particular instance beyond those few that appear most tractable.

Yet, as Neiman astutely notes, the fissure between the world as it is and the world as it ought to be, articulated as Kant's distinction between reason and nature, is not equivalent to the distinction between physical evil and moral evil that has been a staple for many of the arguments over theodicy. One line of argument she pursues in *Evil in Modern Thought* is that this last distinction has lost much of whatever usefulness it may once have had in consequence of the ways in which modernity has apparently accomplished a thoroughgoing naturalization of the human as itself a product of the processes of the world. What once looked to be a promising strategy for properly

[36] Ibid., p. 288.

apportioning responsibility for evil between the human and the divine has lost effectiveness once full realization that "God is dead" finally took hold in the main precincts of Western intellectual culture, and humanity could thus be conceived as itself nothing more than one more part of nature:

> The very naturalism that was the pride of those who sought to disenchant the world undermines the very distinctions they sought to establish. The more human beings become part of the natural world, the more we, like earthquakes, become one more unfortunate fact about it. The more evil itself seems explicable in terms of natural processes, the more nature itself is implicated.[37]

When there no longer is a God whose ways need justification by a theodicy, the "anthropodicy" that takes its place almost inevitably slides into a "cosmodicy." Having first disenchanted the workings of the world of nature into indifference to human purposes, we have proved ourselves no better at clearing space upon which to welcome one another's flourishing:

> Science may have abolished the sense that the world is inhabited by forces with will of their own, and in this way reduced the *unheimlich*. But the price is enormous, for all of nature stands condemned. Human beings themselves become walking indictments of creation.[38]

Bleak as Neiman's assessment may initially seem, it nonetheless helps to articulate a feature of the anthropology of the embodied freedom of finite human reason that serves well as a primary link to Kant's metaphysics of "permanent rupture." The human role in this ruptured landscape is to exercise in steadfast hope the fragile power our finite freedom has for bringing what "ought to be" to bear upon what "is." This fragility of human freedom is embodied in conditions of spatio-temporal finitude. It orients the larger anthropological framework of the critical project that Kant constructs to delimit the unique position human beings occupy in the cosmos as the embodied juncture of nature and freedom.[39] As embodied, our freedom is

[37] Ibid., p. 236. [38] Ibid., pp. 236–7.

[39] For a more extensive treatment of the manner in which Kant understands human finite reason to stand at the juncture of nature and freedom see, P. J. Rossi, S.J., *The*

rendered fragile not simply by the inconstancy of intention that Kant marks out as the "inversion of our maxims," nor only by the inattention and distraction with which we thoughtlessly descend into evil's banality. It is also rendered fragile by a vulnerability of body and spirit to violence and violation.

Yet within that larger framework, *the fragility of human freedom stands coordinate to its dignity*: As we each stand alone, our embodied state provides thin and tenuous protection to our core dignity of spirit; its ultimate bulwark is mutual recognition, the respect we accord each other for the fragile and vulnerable freedom we each embody. Kant's recognition of the inestimable dignity of the power of human freedom to effect good is equally a recognition that such power resides in agents who are themselves profoundly fragile, whose exercise of that power is correspondingly fragile, yet who are capable of empowering each other's freedom in mutual respect for one another's fragility. Exercising finite human freedom in a manner responsive and responsible to both its dignity and its fragility empowers human agents to bring the "ought" of a moral order of mutual respect to bear upon the "is" of the world. It is thus within and by the fragility of human finite reason that the unity of reason is enacted. The enactment of the unity of reason brings forth conditions that open possibilities for freedom and nature to work together effectively for the attainment of "the highest good." The human power for bringing about good in a world of shattered meaning thus thoroughly pertains to, and is rooted in, the fragmentary, fragile exercise of a finite embodied practical reason.

Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind (Albany: State University of New York Press, 2005), pp. 19–65.

2

Kantian Moral Pessimism

Patrick Frierson

> Those valiant men mistook their enemy … They sent forth *wisdom* against *folly* instead of summoning it against *malice*.
>
> Kant, *Religion within the Boundaries of Mere Reason (6: 57)*

> The human being is by nature evil.
>
> Kant, *Religion within the Boundaries of Mere Reason (6: 32)*

Whether people are evil is not a popular topic among contemporary moral theorists. Nonetheless, assumptions about whether humans are generally good or evil play widespread and unnoticed roles in moral theorizing. In this paper, I show some of the ways that moral optimism – the view that humans are generally good – affects contemporary ethical theory. I start with recent work by Gilbert Harman and John Doris, in which empirical psychology plays important roles in ethical reflection. Wherever empirical work is taken to have normative implications, the issue of whether people are fundamentally good contributes to thinking about how empirical studies relate to normative conclusions. I then turn to Barbara Herman's work to show how optimism informs discussions of central issues in contemporary moral philosophy. I end with Kant's "moral pessimism."

Throughout, I primarily contrast moral optimism with moral pessimism. Moral optimists need believe neither that people are omniscient nor that they always do the best thing, but only that the main

failings of most people are not primarily moral, but have to do with ignorance or incompetence or social conditions or (non-culpable) negligence or lack of self-control. Moral pessimists, by contrast, believe that people are not basically good, that (at least) most people (at least) most of the time are *morally* deficient, and that many human misdeeds are due to moral deficiency.[1] Optimism and pessimism are not exclusive options for assessing people's moral status. Both depend upon a robust conception of morality that takes moral obligation seriously. Nietzsche is neither a moral optimist nor a moral pessimist; his optimism or pessimism lies "beyond good and evil." Both views also depend upon applying categories such as "morally good" to persons, rather than merely actions or states of affairs. Finally, I leave out alternatives like moral agnosticism (one cannot know whether people are morally good)[2] and moral ambiguity (people are good in some respects and evil in others). Here I focus on optimism and pessimism in part because my claim that commitments regarding people's moral status play a role in moral theorizing is more forceful when I can show surreptitious commitments to a more extreme position (moral optimism) than to a more moderate one (like moral ambiguity), and in part because both agnosticism and ambiguity typically slide towards optimism or pessimism in particular cases, so discussion of optimism and pessimism is relevant to assessing other views.

1. Situationism and Optimism

Recently, Gibert Harman and John Doris have invoked social psychology against "character-based virtue ethics."[3] They use empirical research that shows human behaviors determined by situation rather than character. For example, in the Milgram experiment,[4]

[1] Kant thinks that *all* people are *radically* evil, but a moral pessimist need not take such a strong position.

[2] Kant's pessimism is so infused with agnosticism that many see Kant as morally agnostic. In response, see P. Frierson, *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge University Press, 2003).

[3] G. Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford University Press, 2000), p. 176.

[4] See J. Doris, *Lack of Character: Personality and Moral Behavior* (Cambridge University Press, 2002), pp. 39–51; and S. Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology*, **67** (1963), 371–8.

an experimenter got subjects to administer what they thought were deadly electric shocks to actors posing as fellow participants. Another experiment invited seminarians to participate in a study of religious vocation.[5] Subjects filled out a questionnaire and were asked to give a verbal presentation in another building. After the questionnaire, subjects were told that they were either late, on time, or early for the presentation. Along the way, the subjects passed an (apparently) extremely distressed person. Whether students stopped to help correlated strongly with their level of hurry, with only 10 percent of the "high hurry" subjects stopping and 63 percent of the low hurry subjects stopping.

Harman/Doris use such empirical studies to critique character-based virtue ethics, claiming that they show that human behavior is better explained by appeal to circumstances (an authority figure present or being in a hurry) than by character: "The experimental record suggests that situational factors are often better predictors of behavior than personal factors … To put it crudely, people typically lack character."[6] Since "virtue ethics presupposes that there are character traits of the relevant sort, that people differ in what character traits they have, and these traits help to explain differences in the way people behave,"[7] virtue ethics seems empirically false.

Unfortunately for Doris and Harman, this argument against virtue ethics depends for its plausibility upon moral optimism, at least to the extent of denying that most people are morally evil. A moral pessimist looking at the data might read not a refutation of character's importance, but a moral indictment of people: "perhaps there was no virtuous person among the subjects of these experiments: if virtue requires practical wisdom, *one would expect virtuous people to be rare.*"[8] This point can be strengthened given Kant's conception of moral character. Kant identifies "good character" with the "good will" (VA

[5] See Doris, *Lack of Character*; and J. M. Darley and C. D. Batson, "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology*, **27** (1973), 100–8.

[6] Doris, *Lack of Character*. Also see Harman, *Explaining Value*, p. 178.

[7] Harman, *Explaining Value*, p. 168. Also see Doris, *Lack of Character*, pp. 5–6 and 15–22. For a defense of ancient virtue ethics, see R. Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," *Ethics*, **114** (2003), 458–91.

[8] Kamtekar, "Situationism," p. 485.

25: 648) and claims: "The person that ought not to trust himself with respect to his resolutions is in a state of hopelessness of all good" (VA 25: 1387–8). The character so important for a good will is precisely the "stability and persistence in principles" (VA 7: 294) that social psychology calls into question. Kant explains: "the most important part of character" is "that a human being has a constant will and acts according to it" (VA 25: 1386). However, while Kant highlights character's importance, he insists on its rarity: "the formal element of will as such, which is determined to act according to firm principles (not shifting hither and yon like a swarm of gnats), has something precious and admirable to it, *which is also something rare*" (VA 7: 292, emphasis added; VA 7: 294; VA 25: 630–1; MS 6: 651–2). For Kant, the Milgram and Princeton experiments quantitatively confirm an empirical claim Kant already affirms. That few act consistently from good principles does not imply that consistent action is an inadequate moral ideal, but that moral virtue is an accomplishment that is, at best, rare.

In response to such interpretations, Harman sometimes explicitly invokes moderate optimism: "can we really attribute a 2 to 1 majority response to a character defect? … Does *everyone* have this character defect?"[9] And Doris suggests, with respect to a more troubling case, that

> virtually all Auschwitz doctors performed selections [deciding who would be killed and who would do forced labor]; did only men of bad character find their way to the camp? … Unfortunately, it does not take a monster to do monstrous things.[10]

Doris tries to make his optimism palatable by explaining that "[t]he problem the empirical work presents is not widespread failure to meet heroic standards – perhaps this would come as no surprise – but widespread failure to meet quite modest standards."[11] Doris's argument does not depend upon the claim that ordinary people are moral heroes, only that they are morally decent. While making it more moderate and palatable, this nonetheless simply highlights

[9] Harman, *Explaining Value*, p. 171. [10] Doris, *Lack of Character*, p. 54.
[11] Ibid., p. 30.

Doris's commitment to moral optimism. Kant can respond that the data, instead of requiring revision to morals, simply require abandoning even *moderate* optimism. Kant *agrees* with Doris/Harman that experiments show widespread lack of stable character traits. Yet, for Kant this lack results from widespread moral failing. Lack of character is not something to build a moral theory around, but a *problem* to combat in order to bring about moral reform.

The difference between Kant's insistence upon character's moral importance and Doris's dismissal of it has profound effects on how each conceives of moral education. For Doris, "Rather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance, we should invest more of our energies attending to the features of our environment that influence behavioral outcomes."[12] Against this, Kant first could argue, on purely normative grounds, that Doris's program for moral education leads people *deeper* into corruption. By avoiding morally difficult situations, people preserve corrupt volitional structures while becoming increasingly morally self-satisfied. Doris rightly asks: "which moral psychology is better suited to effecting the practical aims of ethical reflection?" But Doris fails to sufficiently defend what those ethical aims *are*. If Kant is correct that ethical reflection is oriented toward good *wills* (rather than good actions),[13] Doris's program of moral education is disastrous. Secondly, Kant argues (and some recent research confirms[14]) that attention to a fixed dutiful disposition best inspires people to emulate the virtuous life (KpV 5: 156). Kant's focus on pure moral principles is not merely for philosophical clarity, but also to illuminate the rigorous, sublime moral law in order to inspire "the greatest veneration and lively wish that [one] could become such a [good] person" (KpV 5: 156).

[12] Ibid., p. 146.

[13] Doris claims, "ethical reflection is in the business of helping people behave better" (Doris, *Lack of Character*, p. 166), but offers little argument for this (cf. pp. 15–20). For Kant's extensive argument that the structure of one's will, rather than one's actions, is the "business" of ethical reflection, see G, KpV, and R.

[14] See Doris, *Lack of Character*, p. 50 ("obedience in the Milgram experiment was facilitated by perceptions of diminished responsibility") or p. 37 ("individual tendencies to accept rather than deny responsibility are positively related to a range of prosocial behavior").

Social science cannot arbitrate between Kant's and Doris's interpretations of the data, because these interpretations turn not on data but on their normative *implications*. Kant's conception of ordinary moral virtue requires character, so widespread lack of character reflects widespread lack of virtue. Doris/Harman are unwilling to allow that moral corruption is widespread, so widespread lack of character must reflect character's moral irrelevance. Deciding between Kant and Doris regarding optimism requires doing (pure) moral theory *first*, that is, getting straight on moral ideals. Only then can empirical research help one discern how ideals apply to people and the extent to which people actually live up to them.[15]

2. Optimistic Neo-Kantianism

Unlike Doris/Harman, Barbara Herman operates solidly within Kantian moral philosophy. She agrees with Kant's commitment to principled action and highlights character's importance.[16] Nonetheless, like Doris, Herman often operates under morally optimistic background assumptions. While optimism is inessential for Herman's key arguments, this section examines three areas where optimism shapes her emphases: the role of rules of moral salience in judgment, nonmoral motivation, and integrating morality with human identity.

[15] One final point: it might seem unfair to blame people for lacking character when even Kant admits that character must be acquired over a long period of time. Given the widespread lack of character, it might seem better either not to hold people responsible at all or to develop accounts of localized moral responsibility (Doris, *Lack of Character*, chapter 7). However, even as Kant claims "the human being is evil *by nature*," he argues that one is evil "through one's own fault" (R 6: 32). Kant's reconciliation of these claims appeals to his critical concept of freedom, such that free choice explains one's empirically/observable nature. More particularly, when Kant explains why character is rare, he shows how rarity is due to moral failing – reliance on inclinations – for which individuals are rightly held accountable (VA 7: 294). R. Kamtekar articulates a similar point to explain situational variations with respect to deception: "It may require a strong interest (in the consequences of deceiving or not, or in the activity of deceiving or not) to lead one to extend one's strategies (of deception or non-deception) across situations … [T]he absence of a strong enough interest … may help to explain cross-situational inconsistency (Kamtekar, *Situationism*, pp. 269–70). Situationists emphasize, in particular cases, that "the deeds in question do not require heroic commitment or sacrifice" (Doris, *Lack of Character*, p. 31). But Kant and Kamtekar point out that *developing a character* that acts consistently may require substantial (and thus rarely undertaken) sacrifices.

[16] B. Herman, "Making Room for Character," in S. Engstrom and J. Whiting (eds.), *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty* (Cambridge University

Barbara Herman is best known for her work on rules of moral salience (RMS) in the practice of moral judgment. For Herman, RMS "constitute the structure of moral sensitivity"; they "pick out certain aspects [of situations] … with the point of letting the agent see where moral judgment is necessary."[17] Moral agents cannot simply apply the categorical imperative (CI) procedure to determine whether maxims can be made universal. Agents must first formulate maxims by seeing their situations in ways that highlight morally relevant features.

> An agent who came to the CI procedure with no knowledge of the moral characteristics of actions would be very unlikely to describe his action in a morally appropriate way. Kant's moral agents are not morally naïve. In the examples Kant gives of the employment of the CI procedure, the agents know the features of their proposed actions that raise moral concerns *before* they use the CI to determine their permissibility.[18]

The maxim "I will tell a friend that I will repay her in order to borrow money from her" is universalizable, but only because the borrower fails to include the morally relevant fact that she does not *intend* to repay.

No optimism so far. The importance of RMS could even be interpreted pessimistically, as another avenue for corruption. But Herman's discussion of RMS includes three elements that together suggest substantial moral optimism. First, Herman rightly notes that Kantians should not hold people directly accountable for acting from bad RMS: "there seems to be no way to judge actions apart from the way they are willed, [so] … morally defective RMS may not yield morally defective actions."[19] "It can be permissible for agents with mistaken RMS to act in ways that would be judged impermissible if their RMS were correct."[20] Of course, one *can* be held morally responsible for bad RMS insofar as one is responsible for having the RMS that one has.[21] But, and this is a second element of Herman's view, although reflection can provoke change, RMS are "typically … acquired in childhood as part of socialization."[22] People are not directly responsible

Press, 1996), pp. 36–60. Most of this essay is focused on B. Herman, *The Practice of Moral Judgment* (Harvard University Press, 1993). It was written before the appearance of B. Herman, *Moral Literacy* (Harvard University Press, 2007), so I do not address that work.

[17] Herman, *The Practice of Moral Judgment*, p. 78. [18] Ibid., p. 75.
[19] Ibid., p. 91. [20] Ibid., p. 89. [21] Ibid., n14. [22] Ibid., p. 78.

for acting from bad RMS, and people are typically not responsible for the RMS that they have.

Even these features need not imply moral optimism without a third element of Herman's account: most misdeeds seem ascribable to bad RMS rather than bad wills. Herman claims, for example, "the scope of beneficent actions … *will be* greater for persons who can more readily perceive the distress of others."[23] The assumption seems to be that the primary reason for failures of beneficence (and other virtues) is a failure of RMS. Divergent emphases between Kant and Herman regarding the explanatory role of RMS are further reflected in different positions on moral education. Both Kant and Herman advocate promoting better RMS, for instance by visiting "places where the poor who lack the most basic necessities are to be found" (MM 6: 457). But whereas Herman argues *against* education focused on "rational musculature,"[24] Kant *defines* virtue as "strength of will" (MM 6: 405), and Kant's "society … for the sake of laws of virtue" (R 6: 94) not only seeks to improve moral sensitivity, but also works to foster moral *strength*.

Combining the moral exculpability of bad RMS, the typical lack of responsibility for those RMS, and their explanatory power in assessing behavior, one finds a pervasive but inexplicit moral optimism. Most human misdeed are ascribable not to evil wills, but to mistaken RMS. These mistaken rules are a moral problem and should be changed, but they are not a problem *with* the moral agent.[25] Herman seems to assume that people are basically good, but bad RMS lead to bad deeds.

Kantian pessimists need not disagree with Herman's general account of RMS, but only with one or more subsidiary hypothesis. The first – that one acting on bad RMS can nonetheless have a good will – is linked to Kant's fundamental commitment to the evaluation of *maxims* rather than (directly) of actions. This element of Herman's picture is necessary for any plausibly Kantian account, and it helps constrain and thereby clarify Kant's pessimism. Kant need not claim that *all* misdeeds are due to corrupt wills. Kant, like Herman, can allow

[23] Ibid., p. 81, emphasis added; cf. p. 78.
[24] Ibid., p. 43. [25] Ibid., p. 90.

that some misdeeds result from non-culpable ignorance of situations' morally salient features.

But Kant disagrees with Herman's other auxiliary hypotheses, if not strictly, at least by emphasis. It would be absurd to deny that upbringing exerts influence on moral sensitivity, but where Herman *emphasizes* this influence, Kant highlights how deliberate *self*-corruption of RMS is used to protect moral self-satisfaction. For Kant, people cannot avoid moral self-judgment (MM 6: 438), and recognizing one's moral evil causes self-dissatisfaction (MM 6: 394). But people proficient at self-deception manipulate themselves to simply "fail to notice" areas where moral demands interfere with pursuing non-moral ends. Kant describes, for example, how one's "[self–]deceptive" "natural inclination towards ease … makes [one] content with himself *when he is doing nothing at all* (vegetating aimlessly) because he *at least is not doing anything bad*" (VA 7: 152).[26] As a result of self-deception, one fails to notice as morally salient the fact that one accomplishes nothing. In the moment, the inclination to ease is not necessarily *stronger* than one's sense of duty, but one employs a "ruse" (VA 25: 503) that warps one's RMS in the interests of inclination. Or one privileged in society might direct attention away from structural injustices that would require radical changes in one's life: "I have more important things to do"; "This is just too hard to figure out"; "*I*'ve managed to work within the system, haven't I?" Over time, such redirections warp RMS to allow for pursuit of personal happiness without moral qualm. For Kant, apparently innocent failures of RMS are often blameworthy forms of self-deception.

Regarding Herman's third point, the explanatory power of RMS, Kant suggests that emphasizing morally neutral causes of misdeeds is often a means of congratulatory self-deception: "[w]e like to flatter ourselves with the false claim to a more noble motive" (G 4: 407). Given Herman's claim, with which Kant would certainly agree, that "it can be permissible for agents with mistaken RMS to act in ways that would be judged impermissible if their RMS were correct,"[27] people often reinterpret actions in accordance with RMS that make choices

[26] For further discussion see P. Frierson, "The Moral Importance of Politeness in Kant's Anthropology," *Kantian Review*, 9 (2005), 105–27.

[27] Herman, *The Practice of Moral Judgment*, p. 89.

seem permissible: "the human being knows how to distort even inner declarations before his own conscience" (MVT 8: 270).[28]

Moral optimism thus plays important roles in Herman's account of RMS, and given RMS's centrality in her work, it is unsurprising that optimism shows up elsewhere. One important issue for neo-Kantian ethics involves non-moral motivations in ethical life. In *Groundwork*, Kant (in)famously writes that acts of beneficence performed out of "an inner satisfaction in spreading joy" have "no moral content" and get "genuine moral worth" only when performed "simply from duty" (G 4: 398). For many years, neo-Kantians have sought to dull the force of this statement. Herman takes it up in the context of a critique by Bernard Williams. As she summarizes his twofold critique:[29]

> (1) Kantian morality often demands that we care about the wrong thing – about morality – and not about the object of our action and natural concern; (2) it leads to an estrangement from and devaluation of our emotions, especially in the rejection of emotions as morally valued motives.[30]

One who helps from duty rather than compassionate concern for another's well-being apparently cares about the wrong thing and mistakenly devalues healthy emotions. The problem is particularly acute with personal relationships. In Williams's famous example, one who saves his wife from drowning because "it is morally permissible for him to save his wife" has "one thought too many."[31]

Herman's response to Williams involves first distinguishing between "motives" for actions and "ends" promoted by actions: "the end is that state of affairs the agent intends … to bring about. The motive … is the way he takes the object of his action to be good, and hence … reason-giving."[32] When helping another, one's *end* need not be fulfilling the moral law but can be another's well-being. Direct interest in other's

[28] This point deserves particular emphasis given the many who criticize Kant for misunderstanding the complexities of maxim formation. Kant was aware not only that the same actions could be represented under different maxims, but also of the human tendency to use this fact as a means of justifying evil deeds.

[29] Throughout, I use Herman's summaries of Williams. Cf. B. Williams, *Problems of the Self* (Cambridge University Press, 1973), pp. 207–29 and *Moral Luck* (Cambridge University Press, 1981), pp. 1–19.

[30] Herman, *The Practice of Moral Judgment*, p. 24.

[31] Ibid., p. 41. [32] Ibid., p. 25.

welfare is present even when acting "out of duty," because duty is not a further *end*, but the *motive* making another's welfare an end. Secondly, Herman argues that even as motive, duty often functions only as a "limiting condition." "It is not the function of the motive of duty to bring about moral states of affairs … [I]t expresses the agent's commitment that he will not act (on whatever motive, to whatever end), unless his action is morally permitted."[33] So not only is duty generally not one's end, it need not even be one's primary motive: "As a limiting condition, the motive of duty in fact requires the effective presence of some other motive."[34] When one helps another, one's help legitimately has the other's welfare as its end and can even have compassion as a motive. The motive of duty merely limits compassion to being effective only when its expression is not precluded by moral demands.

Herman goes further, considering cases where "duty … [is] sufficient by itself to bring the agent to do what is morally required."[35] Beneficence may be a paradigm case of this, since people are obligated to promote the welfare of others. In these cases, Williams objects, "the kind of help that can come from the motive of duty is not the kind of help that is needed," so "it may be rational to prefer an emotion-based to a morally motivated action [and so] be rational to place higher value on nonmoral than on moral conduct."[36] Herman's response takes duty to the level of character. What matters is whether one is a morally good *person*, not particular actions' moral worth: "We probably will perform more acts with moral worth the better our will is. The number of morally worthy acts performed, however, is not proportional to the will's goodness."[37] For a person with a thoroughly good will, duty will be "ubiquitous" but not necessarily "pervasive" in the sense of being the primary motive for all actions. In fact, Herman suggests, where beneficent action done from compassion better promotes another's welfare, the agent with a good will acts from compassion. A good will implies that duty is always a limiting condition and functions as primary motive when needed. But "it is not morally required that we always set the motive of duty between our feelings and our response to others."[38]

[33] Ibid., p. 31. [34] Ibid., p. 32. [35] Ibid., p. 34.
[36] Ibid., p. 33. [37] Ibid., p. 35. [38] Ibid., pp. 36–7.

As with RMS, Herman's account here is one with which Kantian pessimists need not disagree. In fact, aspects of Kant's account of human evil fit well with Herman's account. Kant emphasizes that evil consists not in "self-love" – the pursuit of contingent ends – but in "self-conceit" – the *unconditional* pursuit of these ends (KpV 5: 73; R 6: 35). Even non-moral ends are grounded in predispositions "*to the good*" (R 6: 26–8). And Kant insists, like Herman, that morally good agents do not *eliminate* non-moral predispositions but act on them in such a way that they are subordinate to one's moral predisposition over one's whole life (R 6: 36). This conception of moral goodness encourages the idea that acting directly on non-moral desires expresses a morally good will when those desires are part of a character that subordinates non-moral incentives to moral ones.[39]

But while Kant could agree with Herman in these respects, he adds something important. An *idealized* good person could act from non-moral incentives and still express a good will, but when *real* human agents act from non-moral incentives, *we* express one or another form of *evil*. While Kant entertains the possibility of perfectly subordinating non-moral incentives to moral ones, he insists that *all* people *lack* this volitional structure; Herman's hypothetical good will does not exist. Human wills are frail (acting from non-moral motives despite moral commitment), impure (doing what is morally correct only because cooperating inclinations are present), or corrupt (explicitly subordinating the moral law to sensuous incentives) (R 6: 28). Of these, the most relevant here are frail and impure wills. Combating frailty requires cultivating strength of will, something both Williams and Herman discount. Combating impurity raises even more complex issues. Impurity occurs when moral commitment coincides with inclinations. Herman and Williams insist that in such cases inclination can acceptably be one's immediate motive. Strictly speaking, Kant *could* agree. For properly ordered wills that act on inclinations *because* inclinations conform with morality, inclination could be an immediate motive. But people lack perfect wills. Impurity involves subordinating morality to inclination by performing morally good actions only because such actions also satisfy inclination. One's character

[39] Herman, "Making Room for Character."

subordinates morality to inclination, but the expression of that character may look morally conscientious.

Recognizing this danger need not imply that only actions from duty alone are morally good. In fact, *impure* wills may particularly portray their actions as motivated from duty alone. But recognizing the danger of impurity has important implications. First, combating impurity requires emphasizing moral "*strength*" (MM 6: 404) and moral *purity* – "the law being by itself alone the incentive, even without the admixture of aims derived from sensibility" (MM 6: 446). Second, one must be astute in self-assessment. Williams and Herman allow self-satisfaction in performing good deeds from inclination, but given the tendency to impurity, Kantian agents should typically discount comfortable virtue in self-assessment. Insofar as one seeks to grow in virtue, one will certainly visit and comfort sick friends. Since comfort "from the heart" is most soothing, one acts from inclination as one's immediate motive. But one should not base moral self-evaluation on such actions, since one cannot reliably distinguish whether they constitute virtue or impurity. Finally, one will seek opportunities to truly test moral resolve, not by "despising [friends] and doing with repugnance what duty bids,"[40] but by particular attention to occasions when conscience requires resisting inclinations. One morally self-satisfied with generosity to friends might be complacent on these occasions, but one cautious of impurity recognizes them as cases in which life is brought into focus. Failing in difficult duties is not merely excusable weakness; it taints easier good deeds, suggesting that they show impurity rather than virtue.

Herman's account of emotions and Kant's account of evil both raise issues about the relationship between personal integrity and morality. Williams objects that Kantian morality "insists on dominion over even our most basic projects and intimate commitments, demanding a degree of attachment to morality that alienates us from ourselves and what we value."[41] Here, while Kant can again agree with aspects of Herman's response to Williams, his approach is radically different. Part of Herman's response is that integrity is preserved, because

[40] From Schiller's well-known and oft-quoted satire. See A. Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), p. 23.

[41] Herman, *The Practice of Moral Judgment*, p. 24.

non-moral concerns can play deep and even motivating roles in virtuous lives. But Williams deepens his objection by arguing that "in order to live at all, a person must have … 'categorical desires'."

> There is surely something true in the thought that our basic commitments and loves may be such that they make us morally vulnerable … we may find ourselves wanting to do something that impartial morality condemns … But Williams wants to claim something stronger. Suppose our ground projects are what give us a reason to go on with our lives at all. Then if impartial morality can interfere with the pursuit of a person's ground project, there will be cases where an agent *could not* have reason to act as morality requires, for the *only* reasons we will have for acting are those that direct him to the impermissible pursuit of his ground project … So the Kantian idea that a rational agent will always have reason to act as morality requires is false. Since having ground projects is a condition of character … the demands of impartial morality and those of character may conflict in deep ways.[42]

Herman's response to this objection is twofold. First, "[w]hile it is (psychologically) true that attachments to projects can be unconditional, it is not a requirement of the conditions of having a character that they be so."[43] People need to have (non-moral) commitments, but these can function in constituting character even when constrained by morality. Second, because proper attachments can be conditioned by morality, "the moral agent is … one who has a conception of himself as someone who will not pursue his projects in ways that are morally impermissible." Virtuous agents have integrity by having various projects subordinated under one unconditional one: morality itself. "Kantian morality can be (and is meant to be taken as) defining of a sense of self."[44]

Strictly speaking, Herman is correct. Perfect moral agents' lives would have integrity provided by governing commitments to morality. But Williams's critique is also essentially correct. For actual human agents, morality *will* require conflict with ground projects that give us reasons to live. However theoretically possible a perfect life in which identity is defined by virtue, no human has actually chosen such a life. Our *actual* ground projects reflect fundamental subordination of morality *to* non-moral goals, a subordination expressed

[42] Ibid., pp. 37–8. [43] Ibid., p. 39. [44] Ibid., p. 40.

in individual choices and actions as well as in our deepest identity, the ultimate ground of these particular choices. Kant calls this deep-seated and categorical commitment to non-moral ground projects "radical evil."

> [Human] evil is *radical*, since it corrupts *the grounds of all maxims*; as a natural propensity, it is also not to be *extirpated* through human forces, for this could happen only through good maxims – something that cannot take place if *the subjective supreme ground of all maxims is presupposed to be corrupted.*(R 6: 37, emphasis added; cf. 6: 45)

Because the choice to subordinate morality to inclinations occurs at the supreme ground of *all* one's maxims, at what Williams and Herman rightly call one's (deepest) "identity," Kant agrees with Williams that "[nonmoral] ground projects … give us a reason to go on with our lives, [so] … the demands of impartial morality and those of character may conflict in deep ways."[45]

Williams is even *almost* right in positing that "if impartial morality can interfere with the pursuit of a person's ground project, there will be cases where the agent *could not* have reason to act as morality requires."[46] From the standpoint of an agent's fundamental commitments, moral reasons seem insufficient to trump ground projects. But Williams is wrong in that even corrupt agents see moral reasons as reasons (albeit not overriding), and moreover, even for such agents morality still *has* authority (if unacknowledged) that *requires* obedience. While agreeing that humans' categorical commitment to non-moral projects ("evil") is inextirpable (through human forces), Kant nonetheless insists, "In spite of the fall, the command that we ought to become better people still resounds unabated in our souls; consequently, we must also be capable of it" (R 6: 45). Like Williams and unlike Herman, Kant does not think that people can categorically choose morality without sacrificing their most fundamental ground projects, their identities, their characters. But like Herman and unlike Williams, Kant still maintains that people have a reason to categorically choose morality.

The implication of Kant's middle position is that, for Kant (unlike Williams or Herman), moral life is a long, slow, painful suicide of one's

[45] Ibid., pp. 37–8. [46] Ibid.

deepest commitments. Kant describes such a life as "conversion," an "exit from evil and an entry into goodness, 'the putting off of the old man and the putting on of the new'" (R 6: 74). We must become "*other* people and not merely better people (as if we were already good but only negligent about the degree of our goodness)" (SF 7: 54), and this transformation is *painful*:

> The emergence from the corrupted disposition into the good is in itself already sacrifice (as 'the death of the old man' …) and entrance into a long train of life's ills which the new human being undertakes … simply for the sake of the good. (R 6: 74)

The "long train of life's ills" is a sacrifice of a sort of integrity, one constructed around non-moral ground projects that were prioritized over morality.

Kant's pessimism about the relationship between morality and integrity captures valuable insights of both Herman and Williams. With Herman, Kant insists that morality's demands are possible in that a human life without non-moral *categorical* desires could have integrity. But with Williams, Kant acknowledges that moral life requires sacrificing one's deepest commitments. Kant shows why Williams's critique feels powerful and is even correct. Herman's easy ethical integrity is untrue to the real-life ethical struggle towards moral betterment, a struggle that really does involve giving up one's deepest commitments. But Williams's complacent acceptance of categorical projects that trump morals does not do justice to the morality that calls for such struggle. Kant's moral pessimism, in this case, seems to get it just right.

3. Kant's Moral Pessimism

Having already elucidated much of Kant's moral pessimism by contrast with Doris and Herman, here I merely outline Kant's argument for pessimism and his response to four pitfalls that seem implied by the view that "the human being is by nature evil" (R 6: 32). Kant's argument for pessimism begins in *Groundwork*. First, *Groundwork* distinguishes between moral philosophy proper, which is a priori and purely normative, and moral anthropology, which considers humans' empirical nature. Because Kant neither derives nor modifies moral

principles based on empirical facts, he is (unlike Doris) open to moral pessimism. Second, *Groundwork* emphasizes that good wills do not "make an exception … for [them]selves" (G 4: 424). Because morality is universal, particular circumstances do not warrant exceptions to it. Only acting from exceptionless morality is acting "from duty."

In *Religion*, Kant uses this universalist ethic to exclude what I previously called "moral ambiguity," which Kant calls "moral latitudinarianism" (R 6: 22). In its place, Kant defends extreme "moral rigorism," denying any intermediate between good and evil.

> [I]f [one] is good in one part [of life], he has incorporated the moral law into his maxim. And were he … to be evil in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim relating to it would be universal yet particular at the same time: which is contradictory. (R 6: 24–5)

Because morality requires *unconditional* and *universal* compliance, acting in conformity with morality sometimes but not always shows that one never *really* makes the moral law one's ultimate motive, since any law whose application depends upon circumstances cannot be the moral law.

Kant then argues for pessimism based on the fact that certain actions cannot be willed in accordance with duty, because they are directly contrary to right, and others are transparently based on morally impermissible maxims. Given rigorism, those who perform such actions do not make duty supreme in their lives and are therefore evil. Kant then defends pessimism based on "the multitude of woeful examples that the experience of human *deeds* parades before us" (R 6: 32–3). In the present context, Kant could invoke the quantitative research to which Doris appeals to clinch the empirical argument for human evil.

Kant's argument then takes a turn that seems to raise serious problems. Kant not only argues that people are evil; he finds evil in human *nature*. While Kant's account of anthropological knowledge allows for inference from the empirical universality of a trait to the (revisable) ascription of that trait to humans generally, the ascription of evil to human *nature* may seem to undermine the notion that virtue is even *possible*. If virtue is literally *impossible*, this seems a serious blow to Kant's ethics. Arguably, much of the appeal of empirical arguments like

Doris's is due to the suggestion that any moral philosophy dependent upon stable character traits is beyond what humans can reasonably require of themselves. And Kant *is* committed to the principle that one is obligated only what to one can in fact do (KpV 5: 30).

So how does Kant reconcile evil in human nature with his commitment to the possibility of acting morally? First, Kant resolves this apparent tension through his transcendental account of human freedom. In his *Critique of Pure Reason*, Kant argues that *all* empirical claims – including claims about human nature – refer to mere "appearances." In his practical philosophy, Kant shows that human agents are free things-in-themselves that ground their appearances in the world. However one interprets these claims, Kant's point is that human freedom is primary over the most basic empirical claims about human nature.[47] In *Religion*, he reiterates this point with respect to radical evil:

> "He is evil *by nature*" simply means that being evil applies to him considered in his species; not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality would be necessary), but rather that, according to the cognition we have of the human being through experience, he cannot be judged otherwise … Now, since this propensity must itself be considered morally evil … something that a human being can be held accountable for … it must … always come about through one's own fault … ([be] brought upon us by ourselves). (R 6: 32)

That people are evil by nature does *not* mean that it is *impossible* for a human to be morally perfect, only that no people are *in fact* perfect. Kant's first *Critique* shows how this could be: empirically, one can reasonably infer universal evil, but this universality is *ultimately* grounded, not in empirical causes, but in free choices of human agents.

In addition to this transcendental response to the problem of evil in human nature, it is important to distinguish several ways in

[47] Typically, Kant's idealism is interpreted to refer to either two standpoints: H. E. Allison, *Kant's Theory of Freedom* (Cambridge University Press, 1990), C. M. Korsgaard, *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1997), or two worlds: E. Watkins, *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005). For a discussion, see Frierson, *Freedom and Anthropology*, chapter 1, and "Two Standpoints and the Problem of Moral Anthropology" (unpublished manuscript).

which one might use the dictum "ought implies can." Kant primarily argues *from* obligation *to* possibility: one "judges … that he can do something because he is aware that he ought to do it" (KpV 5: 30). One might also use empirical data to specify details of our obligations. Individuals are not obligated to feed every hungry person when this is beyond the scope of one's powers. And humans are not obligated to *feel* love because "I cannot love because I will to" (MM 6: 401; cf. G 4: 399). Sometimes, human nature *expands* moral requirements: we are obligated to help others in part because we need help, and to be polite because politeness alleviates certain moral failings. In all these cases, empirical knowledge of physical-biological possibility helps specify particular duties.

Another use of "ought implies can" would go further, using empirical claims about human capacities to moderate the demands of morality *in general*, such as when Doris argues that morality is not applicable to people in its purity because of human limitations. Kant considers this use of "ought implies can" as an abstract but disastrous possibility (KrV Bxxiii–xix), and he avoids it by articulating an account of freedom that makes perfect virtue possible even if never actualized. Even Doris admits that he has "given no reason for thinking that the realization of virtue is strictly impossible."[48] This concession is all Kant needs. Kant not only provides a detailed working out of the strict possibility of human realization of virtue, he is also willing to accept the moral pessimism that Doris is determined to reject. For Kant, people are evil precisely because we act in ways that conflict with a moral law that we *could* obey.

But even if Kant responds to *philosophical* concerns about universally transgressed obligations, there remains an important *existential* concern. One convinced of his own deep moral corruption, even (perhaps especially) if he recognizes that corruption is his own fault, may collapse into paralyzing moral despair. This despair is especially likely given Kant's moral rigorism: there is nothing that can make it so that one *always* obeys the moral law, since one has *already* failed.

In *Religion* (6: 72), Kant raises and responds to this concern. The first aspect of Kant's response to the existential problem of human

[48] Doris, *Lack of Character*, p. 112.

evil appeals to religious concepts: God, immortality, and (especially) grace. Kant emphasizes that evil cannot be extirpated "through human forces" (R 6: 37) and suggests, "Some supernatural cooperation is also needed to [a person] becoming good or better" (R 6: 44).[49] Although supernatural appeal seems excessive, Kant makes it more palatable by emphasizing the inscrutability of supernatural aid and by insisting that grace does not absolve one of responsibility to actively promote one's own virtue (R 6: 44).[50]

The second aspect of Kant's response to this problem is his affirmation of the enduring presence of what he calls the "predisposition to personality," "the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*" (R 6: 27). Evil involves subordinating that moral predisposition to non-moral ones, but people never eliminate it entirely:

> [T]here is still a germ of goodness left … a germ that cannot be extirpated or corrupted … The restoration of the original predisposition to the good … is not therefore the acquisition of a *lost* incentive for the good … [but] only the recovery of the *purity* of the law, as the supreme ground of all our maxims. (R 6: 45–6; cf. R 6: 49; VA 7: 43, 58–9)

Even when one subordinates moral to non-moral incentives, the force of morality is still felt; anxiety over radical evil even shows the enduring presence of one's predisposition to good. However clear one's subordination of morality, one still has resources to recognize its supremacy and act from respect for it.

Of course, one still might wonder "how it is possible that a[n] … evil human being should make himself into a good human being," but Kant points out that "since the fall from good into evil … is no more comprehensible than the ascent from evil back to good, then the possibility of this last cannot be disputed" (R 6: 45). To be morally good, one requires a basic capacity for respect for the moral law, but all people have this capacity. The existential problem of moral despair comes when one confronts one's own free choice *not* to act out of this

[49] For discussion and further references, see Frierson, *Freedom and Anthropology*, pp. 114–22.
[50] R 6: 191; SF 7: 43–4.

respect. But the solution to this problem lies in the use of one's free choice, and how such choice is determined is incomprehensible. As long as one has the capacity to recognize the moral law as binding, prior evil choices do not warrant despair about prospects for obeying that law.

A third aspect of Kant's response follows from the previous two. Although Kant is a moral pessimist, he is also a philosopher of moral *hope*:

> Assurance of [moral transformation] cannot of course be attained by the human being naturally … [y]et he must be able to *hope* that … he will attain to the road that leads in that direction. (R 6: 51)

Hope is essentially related to pessimism. One cannot "naturally" be assured of moral goodness, but "duty commands that [we] be good, and duty commands nothing that we cannot do" (R 6: 47). This fervent Kantian hope in moral goodness, rooted *not* in self-satisfied cognition of virtue but in recognition that even evil does not absolve one of responsibility, is not *easy*: one "is a good human being only in incessant laboring and becoming" (R 6: 48). By grace, however understood, one can hope for the best. But the best is "an *endless progress* toward complete conformity" with the moral law (KpV 5: 122), a "battle … against the attacks of the evil principle" (R 6: 93) and a constant "struggle" (R 6: 78) to be good against self-wrought evil tendencies. In place of moral despair, Kant offers a realistic but challenging moral hope.

Even if it avoids hopeless despair, however, Kant's pessimism might seem conducive to gloomy misanthropy focused on others' failures. Kant does, in fact, recognize that realistic assessments of evil can cause misanthropy: often "someone becomes a misanthrope due to the sensation of virtue, not because he despises people, but because he does not find them to be how he wants them to be" (VA 25: 553; cf. VA 25: 106, 813, 932). But misanthropy is caused by misplaced *optimism*, a disconnect between expectations and reality. If people are evil, pessimism inoculates against misanthropy rather than causing it. Moreover, what is worthy of respect is *not* perfect virtue, but the *capacity* for virtue. Although good wills are the only things good "without qualification" (G 4: 393), "the human being … exists as an end

in itself" (G 4: 428); even the most wicked are worthy of respect.[51] Given the difficulty of respecting those known as evil, Kant recommends avoiding slander and construing others' actions favorably (MM 6: 466). Without deceiving oneself, one can avoid excessive attention to others' failings. Relatedly, Kant endorses "polite" interaction in which "signs of well-wishing and respect … lead to genuine dispositions of this sort" (VA 7: 152; cf. MM 6: 473–4). Through politeness, we accustom ourselves to giving others respect, and we draw attention to others' likeable qualities. Finally, Kant suggests that the proper (pessimistic) way to think about evil is precisely the opposite of what leads to misanthropy: "Misanthropy comes from a perverted concept of one's own importance and out of a black representation of [other] people" (VA 25: 1364). In place of self-inflating attention to the *others*' evil, Kant directs such attentions towards oneself. At a rational, cognitive level, one recognizes both the radical evil of all and the fact that all are nonetheless worthy of respect, while at the imaginative and affective level, one remains agnostic or even optimistic about others while deeply cautious about optimistic self-deception regarding oneself.[52]

Finally, even if Kant separates pessimism from misanthropy in general, pessimism seems to undermine valuable sorts of social interaction; recognizing human misdeeds that result from bad RMS (Herman) or bad situations (Doris) encourages social struggle towards a better world. People can cultivate better virtue *in each other* through dialogue, inquiry, and social networks conducive to good choices. By emphasizing personal corruption, Kant seems to undermine such arenas for moral improvement.

In fact, however, Kant's pessimism has the opposite implication. Although evil is self-wrought,

> the causes and the circumstances that draw one into this danger and keep him there … do not come … from his own raw nature, so far as he exists in

[51] For further defense of this claim, see P. Frierson, "Review: Richard Dean, *Kant and the Value of Humanity*," *Notre Dame Philosophical Reviews*, 04.17 (2007), online at http://ndpr.nd.edu; Korsgaard, *Creating the Kingdom of Ends*, and Wood, *Kant's Ethical Thought*.

[52] For elaborate discussion of this "humility," see J. Grenberg, *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005).

isolation, but rather from the people to whom he stands in relation or association. (R 6: 93)

The primary means by which one cultivates one's worst tendencies are social.[53] Evil manifests itself not merely in individual wrongdoing, but in the cultivation of vice-conducive social climates. Hence for Kant, struggle against self-wrought evil tendencies involves reform of *society*: "The dominion of the good principle is not otherwise attainable … than through the setting up and the diffusion of a society in accordance with, and for the sake of, laws of virtue" (R 6: 94). Rather than leading to withdrawal into individual responsibility, Kant's pessimism leads to proactive social engagement. Kant's moral community seeks to improve RMS (like Herman) and reduce circumstances that tempt to vice (like Doris), but it also goes further, actively promoting consistent character and moral resolution to act rightly in the light of one's RMS. In that sense, Kantian ethical life is even *more* socially engaged than Doris and Herman's proposals. Pessimism leads, not to disengagement, but to aggressive and focused engagement in social-cultural-political change.

4. Conclusion

Kant's view of the human species is not particularly happy. Bad actions are rooted in fundamental failures of character. People choose to subordinate unconditional demands of morality to shifting inclinations. We usually fail to act consistently, and what consistency we have usually results from pursuing non-moral grounding projects. But Kant is neither *hopelessly* pessimistic nor absolving of people's responsibility to deal with evil. Instead, Kant's pessimism orients us to real moral threats and thereby makes both moral philosophy and moral reform more relevant to actual conditions of human life.

By accepting the ubiquity of evil, Kantian pessimists can use empirical psychology not to revise moral demands but to show where evil must be combated. By diagnosing evil partly as self-deception, Kant

[53] See S. Anderson-Gold, *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (Albany: State University of New York Press, 2001), Frierson, *Freedom and Anthropology*, A. Wood, *Kant's Ethical Thought*, and "Unsociable Sociability: The Anthropological Basis of Kantian Ethics," *Philosophical Topics*, **19** (1991), 325–51.

shows how often blaming evil on situation-responsiveness or defective RMS are subtle strategies for preserving moral self-satisfaction while satisfying non-moral interests. By recognizing evil's depth, Kant does justice to the real *struggle* of a life of moral hope by showing why moral reform requires sacrificing one's deepest self-conception. And by drawing attention to how society and situation facilitate self-corruption, Kant orients social reform.

Kant's pessimism allows and requires a shift in emphasis in moral philosophy. Against Doris, Kant keeps accounts of morality's nature free from attenuation by facts about human behavior: "Any high praise for the ideal of humanity in its moral perfection can lose nothing in practical reality from examples to the contrary" (MM 6: 405–6). But against Herman, Kant does not merely articulate an ideal of a morally good human will; he insists upon a moral anthropology that highlights and works to remedy the pervasive evil that prevents people from realizing this ideal.

3

Kant, the Bible, and the Recovery from Radical Evil

Gordon E. Michalson, Jr.

I

A familiar feature of Kant's religious thought is his moral theory of biblical interpretation. His position enjoys a capsule statement in Kant's reaction to the contemporary biblical critic J. D. Michaelis, who comments in the following way on a prayer for revenge in Psalm 59: "The psalms are *inspired*; if they pray for revenge, then it cannot be wrong: *We should not have a holier morality than the Bible*." In his rather wry response, Kant notes: "I pause here at this last statement and ask whether morality must be interpreted in accordance with the Bible, or the Bible, on the contrary, in accordance with morality."[1]

Embedded in Kant's obviously rhetorical question is his utter lack of interest in the running debates over the Bible's historical and literal accuracy that were consuming the scholarly energies of the day. Referring to such debates, Kant would argue that we "should not quarrel over an issue unnecessarily, and over its historical standing, when, however we understand it, the issue does not contribute anything to our becoming a better human being" (R 6: 43). Properly read and interpreted, the Bible simply reminds us of what we already know, since morality for Kant needs no instruction from the "outside." In particular, we need no assurance regarding its historical accuracy to rely on such moral lessons that the Bible may convey. One commentator

[1] R 6: 110 (emphasis Kant's).

rightly observes that Kant's "remark about the dogma of the divine Trinity – that it 'has *no practical relevance at all*' – is typical of his general lack of interest in specifically doctrinal questions of Christianity" and, we may add, historical questions as well.[2]

Kant's stance thus suggests that the Bible is, at best, an auxiliary or an aid that has as its chief positive role the encouragement and improvement of the moral life. Any other interpretative interest has for Kant the whiff of the driest sort of scholasticism, with no potential for making human beings "better."

The rigorous consistency of Kant's moral theory of biblical interpretation makes all the more striking his rather abrupt appeal to the Bible in an altered way at the very moment in *Religion within the Boundaries of Mere Reason* when he attempts to resolve one of the deepest problems of his entire philosophy, which concerns the recovery from radical evil. When Kant writes of moral conversion in terms of the Pauline "new man" and the Johannine motif of a "rebirth," he is not merely providing a biblical gloss to a previously worked out conceptual account, for the simple reason that he has offered no such account (R 6: 47). Quite to the contrary, the biblical references in fact serve as a *substitute* for further argument – in effect, as pictorial filler for a conceptual lacuna. The reversal here is more than slightly ironic. Whereas the Bible for Kant typically serves as the illustration or reminder of what we already know, in the case of the recovery from radical evil it serves as a needed substitute for what Kant himself cannot state or argue more directly.

One implication here is that we confront in a fresh way the truly profound nature of an evil that is radical. It is so profound that the depiction of overcoming it requires special means. Another implication is that biblical imagery provides Kant with what his concepts cannot, which is a way of rendering moral change over time. The Bible provides Kant a means of depicting the chronological features of the moral life, badly needed in the crucial account of moral conversion yet impossible to frame in purely Kantian terms. Since the biblical

[2] H. Bielefeldt, *Symbolic Representation in Kant's Practical Philosophy* (Cambridge University Press, 2003), p. 3. Kant's remark is from *Conflict of the Faculties* (SF 7: 38–9), emphasis Kant's.

element in question thus functions as a needed *narrative* feature in Kant's effort to depict the transition from depravity to virtue, his use of the Bible is potentially suggestive of the idea of a "shared" narrative that might animate or otherwise inspire the moral community as it endeavors to grapple with the historical realities of a fallen world. Whatever Kant's deepest intentions in the matter may be, there is here the suggestion that the moral community might benefit from viewing itself in terms of a shared narrative, not only as it struggles with the need for moral change, but also as it confronts the challenge of imagining or depicting the needed change itself. In effect, the capacity to *imagine* moral change is the first step in bringing it about.

In order for these interpretative possibilities to come more clearly into view, it will help to sketch out further the contours of Kant's moral theory of biblical interpretation against which his use of the motif of a moral "rebirth" stands in considerable contrast.

II

Kant's reductionistic view of the Bible hardly arises in a cultural vacuum, and the eighteenth century is the fertile soil out of which most ongoing debates about biblical interpretation grew. In particular, Kant's response to Michaelis is indicative of a highly complex set of historical changes sustained by a series of suspicions toward traditional religious belief and its trappings: Newtonian suspicions toward all appeals to supernaturalism; deism's codification of these suspicions in a stripped down, simplified view of religious belief emphasizing sound moral conduct over appeals to supernatural intervention;[3] and the widespread suspicion that Christian orthodoxy was intellectually bankrupt, perpetuated largely by a silent conspiracy of priests and princes in an increasingly desperate effort to shore up their legitimacy.[4] Very generally speaking, these suspicions emerged in turn from

[3] The exact nature of Kant's relationship to deism is controversial. For one view, see A. W. Wood, "Kant's Deism," in P. J. Rossi and M. Wreen (eds.), *Kant's Philosophy of Religion Reconsidered* (Bloomington: Indiana University Press, 1991), pp. 1–21.

[4] A helpful snapshot of this historical setting is provided in the "Introduction" to T. Pinkard, *German Philosophy 1760–1869: The Legacy of Idealism* (Cambridge University Press, 2002), pp. 1–15.

the subversive effects of a rationalist criterion of truth embraced by Kant and others, with its mathematically driven emphasis on universality and necessity as the authentic marks of truth.

When applied to the Bible, such a view of truth generated devastating effects. In particular, the so-called "truths" of the Bible that were tied to revelatory historical events were undercut for two reasons. First of all, such truths were "contingent" – not universal and necessary – since they were connected to such utterly idiosyncratic, unrepeatable historical events as whales swallowing men, Israelites escaping Egyptian armies through parted seas, virgins having babies, and dead men emerging from tombs. Moreover, as these and countless other examples clearly suggest, such biblical truths were very often tied to events involving a disruption of natural law, which is to say, they were miracles. As a result, the "offense of particularity" involved in the issue of historical contingency was made only more offensive by the accompanying affront to Newtonian regularity.

The interesting point here is that the emerging progressive emphasis would be on a truth that is anterior to the occurrence of revelatory historical events – a position that effectively makes altogether moot the question of a religious interest in whether a particular miracle occurred. Kant would surely endorse G. E. Lessing's remark that "the Bible isn't true because the evangelists and apostles taught it; the evangelists and apostles taught it because it's true."[5] In effect, the criterion of biblical truth is something separable from the Bible itself, in terms of which biblical truth is then judged. Since nothing of religious importance thus hangs on whether a particular historical event actually occurred (miraculous or not), there is no authentically *religious* need to devote intellectual energy to the question of the literal accuracy of the Bible.

Kant's specific way of putting this general interpretative strategy into practice is of course governed by his conviction that the necessary and universal touchstone of religious truth is morality. The Bible is "true" only to the extent that it is conducive to moral improvement. Since everyone has natural access to the moral law simply by virtue of being rational, we are not dependent on the Bible for the discovery

[5] Quoted in H.E. Allison, *Lessing and the Enlightenment* (Ann Arbor, MI: University of Michigan Press, 1966), p. 96.

or discernment of religious truth. In other words, the integrity of the moral life does not hang on the historical accuracy of any particular feature of the Bible. Once again, the issue of the literal accuracy of the Bible falls into religious irrelevance. In what is in fact a fairly dramatic moment in the history of biblical interpretation, Kant makes explicit the sheer religious irrelevance of the search for the literal accuracy of the Bible:

> We can explain how we put a historical account to our moral use without thereby deciding whether this is also the meaning of the writer or only our interpretation, if this meaning is true itself, apart from all historical proof, and also the only meaning according to which we can derive something edifying from a text which would otherwise be only a barren addition to our historical cognition. (R 6: 43)

The clear implication seems to be that, in his eagerness to exploit the biblical text for moral purposes, Kant is willing to ride roughshod over the original author's intention and the text's apparent or most evident meaning. With disarming candor, Kant pleads guilty to exactly this charge:

> This [moral] interpretation may often appear to us as forced, in view of the text … and be often forced in fact; yet, if the text can at all bear it, it must be preferred to a literal interpretation that either contains absolutely nothing for morality, or even works counter to its incentives. (R 6: 110)

In Yovel's apt summary of the matter, the study "of the Bible is for Kant antiquarian and devoid of value unless it serves the moral ends and reinforces the will to realize [these ends]. As a purely scientific object the Bible is dead."[6] In biblical matters, Kant's characteristic insistence on the priority of the practical over the theoretical culminates in the disentangling of religious meaning from historical accuracy.

III

Within the context of *Religion within the Boundaries of Mere Reason*, Kant's moral theory of biblical interpretation is a specific instance of

[6] Y. Yovel, "Bible Interpretation as Philosophical Praxis: A Study of Spinoza and Kant," *Journal of the History of Philosophy*, 11 (1973), 189–212, p. 191.

his more generic intention to derive moral lessons from historical or ecclesiastical faith – as he puts it, to "start from some alleged revelation or other" and "hold fragments of this revelation" up to the light of "moral concepts" (R 6: 12). He maintains that "we require an interpretation" of a revealed faith "in a sense that harmonizes with the universal practical rules of a pure religion of reason" (R 6: 110). Kant's most familiar approach to biblical interpretation is a specific application of this more general rule.

Consequently, it is all the more striking when, at a pivotal point in the *Religion*, Kant appeals to the Bible in a manner that cannot simply be subsumed under his usual moral interest. The context is Kant's obviously strained effort to provide an account of our recovery from radical evil. In Kant's own terms, the moral agent could only initiate such a recovery by generating maxims in keeping with such an intention. The problem is that Kant has explicitly defined an evil that is "radical" in terms of the perversion of the underlying moral "disposition" – or "supreme maxim" – that is the sole source of the needed maxim. "This evil is *radical*, since it corrupts the ground of all maxims … [and] constitutes the foul stain of our species" (R 6: 38). Consequently, by lodging radical evil in the disposition, Kant has apparently crippled the moral agent's capacity ever to produce the new maxim that would transform the agent's disposition and constitute the recovery from an evil defined precisely in terms of a polluted disposition – not unlike trying to draw potable water from a poisoned well. In the (understandably cumbersome) formula of one commentator, "radical evil is an evil in which the means of overcoming evil are themselves contaminated by the evil that is to be overcome."[7]

Associated with this difficulty is the limited maneuvering room Kant's own epistemology allows him in depicting the sequencing – the "before-and-after" – associated with any change in the moral life, such as the change associated with moral regeneration. Within Kant's framework, temporality, involving the very idea of before-and-after, implies determinism. Consequently, the conceptual cost of protecting freedom and, thus, the moral life from the causal sequence

[7] A. Hewitt, "The Bad Seed: 'Auschwitz' and the Physiology of Evil," in J. Copjec (ed.), *Radical Evil* (London and New York: Verso, 1996), p. 84.

associated with the phenomenal world includes the impossibility of explicitly framing the moral life in temporal terms. The free act associated with a moral undertaking can never be shown to be "caused," as surely follows for Kant from locating such an act in a before-and-after sequence. Such an act arises instead out of a rational but noumenal gesture suggestive only of the mysteries of moral improvement and not of the predictable regularities of the physical universe. By Kant's ground rules, an act of genuine freedom is, quite literally, "not in time."[8]

This well-known constraint on Kant's efforts to depict the moral life has implications for his overall ethical theory that far exceed the narrow confines of the issue of moral conversion.[9] Nonetheless, this constraint suggests particular pressures on Kant's strained efforts to describe the transition from an evil to a good disposition. My central point here is that, at precisely the intersection of these seemingly intractable difficulties in his account of moral conversion, Kant invokes biblical language. "And so a 'new man' can come about only through a kind of rebirth, as it were a new creation (John 3:5; compare with Genesis 1:2) and a change of heart" (R 6: 47).

Such language, clearly suggesting that "a kind of conversion experience serves as Kant's model for this notion of a revolution in conduct of thought,"[10] would have been altogether familiar to Kant through his lifelong exposure to the idiom and sensibility of pietism. Allen Wood, among others, has made the case that, despite the fact that "[m]ost of Kant's explicit pronouncements about pietism are negative," his mature "conception of true morality and religion amounts to a rationally purified version of pietism."[11] In the current instance,

[8] See Kant's account of the "Second Analogy of Experience" ("Principle of Succession in Time, in Accordance with the Law of Causality"), in KrV A189–211/B232–56.

[9] An excellent summary and analysis of the several ways in which temporality is a problem for Kant's moral theory is provided by P. Stern, "The Problem of History and Temporality in Kantian Ethics," *Review of Metaphysics*, 39 (1986), 505–45. For an illuminating effort to resolve the morality/temporality issue through an emphasis on the idea of "character," see G. F. Munzel, *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment* (University of Chicago Press, 1999).

[10] Munzel, *Kant's Conception of Moral Character*, p. 160.

[11] A. W. Wood, "General Introduction" to Wood and G. di Giovanni (eds.), I. Kant, *Religion and Rational Theology*, Cambridge Edition of the Works of Immanuel Kant (Cambridge University Press, 1996), p. xii.

Kant's remarks about rebirth arise in the context of his labored effort to reconcile the gradualism associated with the phenomenal appearance of a moral life undergoing "reform" with the invisible (because noumenal) "revolution" in the underlying moral disposition that presumably generates the observed reform. He needs the revolutionary motif to protect freedom from the causal clutches accompanying temporality, and he needs the gradualist motif to remain true to our actual experience of moral improvement. The revolutionary motif finds expression in biblical imagery, imagery that manages simultaneously to contribute a narrative element.

The conflation of the revolutionary idea and the biblical reference becomes explicit in Kant's insistence that it requires "a single unalterable decision" for a human being to reverse "the supreme ground of his maxims by which he was an evil human being (and thereby puts on a 'new man')" (R 6: 47). The motif of a "revolution" provides the needed chronological element while simultaneously minimizing it by compressing the implied time frame. The motif thus conveys the incommensurability between moral change and temporality while still offering language that helps us to represent the change. Kant's biblical allusion precedes a complex paragraph that pursues the theme of reconciling gradual (phenomenal) reform with sudden (noumenal) revolution by appealing to a dual, human/divine standpoint. In his effort to formulate just such reconciliation, Kant proposes that God's "intellectual intuition" enables God to "see" as a completed whole what can only appear to finite beings as "an ever continuing striving for the better" (R 6: 48).[12] The confusion here between how moral regeneration *occurs* and how it *appears* is less important than Kant's concentrated (and, thus, revealing) effort to domesticate the problems inherent in providing a chronological sequence for moral change.

In short, Kant's allusion to the biblical idea of being "born again" appears at a crucial pivotal point in his effort to keep his dark account

[12] Elsewhere, I have suggested that this "solution" to the problem of moral conversion is beside the point, since Kant answers his own question of "how" this revolution in the disposition comes about with an account of what it "looks like." See G. E. Michalson, Jr., "Moral Regeneration and Divine Aid in Kant," *Religious Studies*, 25 (1989), 259–70, p. 269. In an interesting discussion of the vexed notion of "forgiveness" in Kant, David Sussman challenges my way of framing the issue. D. Sussman, "Kantian Forgiveness," *Kant-Studien*, 96 (2005), 85–107, pp. 101–2.

of radical evil from being the last word on our moral destiny. While the moral concern associated with Kant's more familiar use of the Bible remains evident in a very general sense, his chief need here is to complete a thought without transgressing certain boundaries, and the Bible can evidently do both. In other words, the striking feature concerns the way the biblical theme of a rebirth effectively substitutes for an extension of Kant's actual argument depicting the recovery from radical evil.[13] When Kant states that moral conversion "can come about only through a kind of rebirth" (R 6: 47), he deploys an image that serves as a needed proxy to capture the temporal transition latent in the motif of a revolution, a proxy made necessary by the missing conceptual splice between time and freedom.

By providing the needed narrative element, the biblical reference thus gives us a picture to imagine, not a concept to argue. Biblical allusion thereby becomes a kind of placeholder – an apparently indispensable placeholder – for the narrative element that Kant's philosophical position requires but cannot provide.

Kant himself explicitly reinforces the indispensability of the narrative element in his separate remarks about the fall from an original state of virtue into radical evil – the same issue viewed this time on the front end. In his effort to account for the propensity to evil in Part One of the *Religion*, Kant confronts the issue of the relationship between a free action and time as he struggles with the question of "when" to impute responsibility for the fall into moral evil (R 6: 40–4). The entanglements unleashed by his own epistemology nearly strangle his prose as Kant relates the moral agent's past and present in the depiction of the accountability for evil.

However evil a human being has been right up to the moment of an impending free action (evil even habitually, as second nature), his duty to become

[13] Gary Branham claims that, in *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge University Press, 1990), I have completely misunderstood Kant's theory of freedom when I refer to Kant's "long-windedness" on both the issue of the "fall" into radical evil and the "recovery" from it (e.g., "Kant can no more explain the 'fall' than could Augustine, his long-windedness on the matter notwithstanding," p. 65). Yet my own point – evidently framed in misleading irony – is the very one Branham, too, is making, since I am attempting to underscore the incommensurability between Kant's theory of freedom and all possible "explanations" (with their causal sequencing) for a free act. Indeed, the whole point of what I am suggesting about Kant's use of the Bible is that biblical references fill the gap left by Kant's

better himself was not just in the past; it still is his duty *now*; he must therefore be capable of it and, should he not do it, he is at the moment of action just as accountable, and stands just as condemned, as if, though endowed with a natural predisposition to good (which is inseparable from freedom), he had just stepped out of the state of innocence into evil. Hence we cannot inquire into the origin in time of this deed but must inquire only into its origin in reason. (R 6: 41, emphasis Kant's)

At the point of this self-confessed limit to his ability to reconcile moral agency with temporality, Kant immediately alludes to "the mode of representation which the Scriptures use to depict the origin of evil, as having a *beginning* in human nature … for the Scriptures portray this beginning in a narrative, where what must be thought as objectively first by nature (without regard to the condition of time) appears as a first in time" (R 6: 41, emphasis Kant's).

In other words, moral agency, including the fall into evil, is framed in terms of an a-temporal rationality, not in terms of time-bound nature. Yet any inquiry into the "origin" of evil virtually begs for an answer expressed in chronological terms, which reason cannot legitimately provide. But the Bible can. Once again, biblical allusion provides the needed narrative framework, helping to depict what Kant calls the "*beginning in time*" of the propensity to evil, not so much illustrating a deeper moral truth as providing a needed account that cannot be rendered in non-narrative terms. As in the parallel account of moral regeneration, Kant links together the impossibility of conceiving the ground of the fall into evil and the help provided by biblical narrative. The propensity to evil "remains inexplicable to us," Kant claims, as "there is no conceivable ground for us … from which moral evil could first have come in us."

The Scriptures express this incomprehensibility in a historical narrative, which adds a closer determination of the depravity of our species, by projecting evil at the beginning of the world, not, however, within the human being, but in a *spirit* of an originally more sublime destiny. (R 6: 43–4, emphasis Kant's)

As in the case of being born again morally, the biblical reference here provides a needed chronological element without committing Kant to a temporal sequencing of the moral life.

inability to "explain" either the fall or the recovery from it. See G. Branham, *Kant's Practical Philosophy: From Critique to Doctrine* (Basingstoke and New York: Palgrave Macmillan, 2003), p. 247 n.4.

In effect, by once again providing the needed narrative feature, biblical allusion "schematizes" the rational idea of moral change. In the *Critique of Pure Reason*, Kant had devised the notion of schematization to account for the way something entirely rational can be "represented" to the senses – in the original context, schematization accounted for the transformation of pure concepts (the "categories") into instances of perception, giving pure concepts "referential" as well as "logical" significance.[14] In the *Religion*, Kant adapts his teaching about schematization through several appeals to what he calls the "schematism of analogy" (R 6: 65n.). The schematism of analogy provides Kant with a means of "representing" what otherwise remains purely rational and non-empirical, enabling us "to make supersensible characteristics comprehensible to us" without actually moving "outside" the "boundaries of mere reason" (R 6: 65n.). Kant's chief example is Jesus, understood as the "personified idea of the good principle" – that is, the historical embodiment of a moral disposition wholly pleasing to God (R 6: 60). As Kant explains, we "use a schema for a concept to render it comprehensible to us (to support it *with an example*)" (R 6: 65n., emphasis added).

The biblical references designed to convey the "before-and-after" of moral conversion mimic the function of the schematism of analogy in a less technical yet still significant sense. That is, they do the hard work of making something "comprehensible to us" through the aid of an "example," yet without transgressing the clear boundaries established by Kant's epistemology. Especially provocative in the light shed by this comparison is the fact that, beginning with the first *Critique*, Kant maintains that the "source" of the schemas – or of those "images" that help us to represent something rational to ourselves – is the "imagination."[15] The puzzling but powerful role played in Kant's philosophy by the "imagination" assumes fresh importance as we reflect on the imaginative train of thought leading Kant to the examples of a "rebirth," a "new creation," and a "change of heart." The imagination – in this case, we might even say, the "moral imagination" – becomes the source of the motifs and representations that not only enable Kant to elude a technical dead end, but also provide the inspiration for the hope of recovery from radical evil.

[14] KrV A137–47/B176–87. [15] KrV A141–2/B181.

IV

My admittedly aggressive reading of the distinctiveness of Kant's use of the Bible in this specific context relies heavily on the idea that biblical narrative substitutes for conceptual argument in a way that neither requires a literalistic approach to the Bible nor points to a deeper meaning. Instead, the biblical reference itself is the element indispensable to Kant's completing his argument, particularly and most importantly the argument concerning moral regeneration. Appeal to the Bible is itself the stand-in for philosophical reasoning rather than the illustration of a truth that can be articulated through philosophical reasoning alone apart from the biblical reference. We have virtually the flip side here of Kant's more familiar position concerning the potential *dispensability* of the Bible when seeking its moral meaning.

The function thus performed by the biblical allusion can perhaps be clarified by means of a provocative parallel suggested by Hans Frei's effort to recover "realistic narrative" as the lost hermeneutical option in eighteenth and early nineteenth century debates about biblical interpretation.[16] In a story that in fact involves Kant in complicated ways, Frei argues that, under the impact of such intellectual forces as deism and historical criticism, biblical critics gravitated toward either of two all-inclusive accounts of how biblical narratives "refer" and, thus, gain their religious meaning. On the one hand, there would be the very traditional view that biblical narratives gained their religious meaning through ostensive reference, which is the same thing as their depending on being literally true. Religious meaning would thus be closely aligned with historical likelihood. In the case of this option, a lot obviously hangs on being able to show that the narratives are historically accurate, or at least on believing that they are.[17]

Alternatively, there would emerge a range of progressive thinkers sensitive to the negative implications of critical thinking for the Bible's literal accuracy, particularly with respect to the element of miracle often associated with biblical narratives. In what amounted to a sustained recovery of long-standing allegorical readings of the Bible, this group of thinkers would devise various means of showing

[16] H. W. Frei, *The Eclipse of Biblical Narrative: A Study in Eighteenth and Nineteenth Century Hermeneutics* (New Haven and London: Yale University Press, 1974).

[17] Ibid., pp. 11–12.

that biblical narratives "illustrate" religious truths or deeper meanings that are true independent of the Bible's accuracy – which, of course, is precisely Kant's position in his familiar moral approach to biblical matters. The illustrative power of biblical narratives implies their dispensability as actual history, since religious truth turns on the deeper, usually moral, meaning rather than on historical accuracy.

These two options thus result either in the view that biblical narratives are literally true and convey their meaning through their factual power, or in the opposing view that the narratives mean something other than what they say and simply require decoding by an interpretative theory that provides access to the deeper meaning.

Where Frei's account and my current interest overlap is in his claim that these two hermeneutical options do not exhaust the possibilities and actually collude to hide from view a lost third option. Indeed, the point of his book is to identify the lost option and recover it. Frei refers to the lost option in terms of a realistic or "history-like" element in the Bible that depends for its meaning on *neither* its literal accuracy *nor* on its capacity to illustrate a deeper truth, but on its expressly *narrative* feature.

> By speaking of the narrative shape of these accounts, I suggest that what they are about and how they make sense are functions of the depiction or narrative rendering of the events constituting them – including their being rendered, at least partially, by the device of chronological sequence.[18]

In effect, "sense making" in much of the Bible is a function of its narrative rendering – including the *device of chronological sequence*. The narrative does not need to be literally true, nor does it need to evoke or point to something beyond itself, to be meaningful. For all their differences, biblical literalists and non-literalists alike mistakenly collapsed "history-likeness" into history itself, thus forfeiting the possibility of discovering this third option. Curiously, claims Frei, "the realistic character of the crucial biblical stories was actually acknowledged and agreed upon by most of the significant eighteenth-century commentators."

> But since the precritical analytical or interpretive procedure for isolating it had irretrievably broken down in the opinion of most commentators, this

[18] Ibid., p. 13.

specifically realistic characteristic, though acknowledged by all hands to be there, finally came to be ignored or – even more fascinating – its presence or distinctiveness came to be denied for lack of a "method" to isolate it. [19]

Influenced heavily by Erich Auerbach's theory of realistic narrative,[20] Frei has chiefly in mind the "simplicity of style, the life-likeness of depiction, the lack of artificiality or heroic elevation in theme" in the first three chapters of Genesis, the account of Abraham's willingness to slay Isaac, and, especially, the Synoptic Gospels. For example, in establishing the "identity" of Jesus, meaning turns on narrative depiction as Jesus' full identity is rendered only through the shape of the story about him – though not in a sense requiring either historical proof of the story's accuracy or an account of a deeper meaning that the narrative about Jesus' identity illustrates. Consequently, the narrative does not have to be literally true to be meaningful (in contrast to the position of the literalists). At the same time, however, *nothing else can substitute for the narrative* for the simple reason that there is no "deeper" meaning in relation to which the narrative is merely evocative (in contrast to the position of the non-literalists). The meaning of the text reflects the "indispensability of the narrative shape," for the narrative is not "a shadow of something else more real or more significant."[21] Indeed, for Frei, "even the miraculous accounts are realistic or history-like (but not therefore historical and in that sense factually true)," which is to say, "even such miraculous accounts are history-like or realistic if the depicted action is indispensable to the rendering of a particular character, divine or human, or a particular story."[22]

A potentially fruitful parallel here with Kant's appeal to the moral agent's "rebirth" hardly requires a definitive assessment of Frei's account, since the key point here is simply the insight that a narrative element – taken neither as literal nor as pointing beyond itself – is indispensable for making sense of something. As we have seen, the narrative element conveying the before-and-after associated with

[19] Ibid., p. 10.
[20] E. Auerbach, *Mimesis: The Representation of Reality in Western Literature* (Princeton University Press, 1968).
[21] Frei, *The Eclipse of Biblical Narrative*, p. 13; p. 14.
[22] Ibid., p. 14.

moral change is precisely what Kant needs as the explanatory power of his moral philosophy reaches its limit point in the account of the recovery from radical evil. In the case of moral conversion, sense-making involves putting into narrative form something we cannot state conceptually. Again, the appeal to the Bible here is not an illustration of something we could say more directly so much as it is the rendering of a transformation in terms involving the needed chronological sequence. Without the element of before-and-after, Kant's account of moral regeneration cannot make complete sense.

V

In a world in which the expression "radical evil" seems less like a dated philosophical abstraction than a description of actual states of affairs, anything that Kant might say on the matter of how to recover from it is of more than passing interest. In recent years, a growing body of Kant scholarship has redirected our attention from the purely individualistic or personal dimensions of moral regeneration to the more broadly *social* dimensions, a seemingly arcane interpretative issue that is in fact of pressing contemporary importance.[23] In Allen Wood's capsule statement of the matter, our "moral vocation is a social one, which must be pursued through membership in a *community*."[24] Personal responsibility assumes its rightful context in the setting of society, the latent premise all along of the third version of Kant's own categorical imperative: "Act in accordance with the maxims of a universally legislative member of a merely possible realm of ends" (G 4: 439).[25] Apart from the interesting questions such a viewpoint raises about the demarcations typically constructed to frame the debate between "liberals" and

[23] Examples include P. J. Rossi, "Autonomy and Community: The Social Character of Kant's Moral Faith," *Modern Schoolman*, 61 (1984); S. Anderson-Gold, "Kant's Ethical Commonwealth: The Highest Good as a Social Goal," *International Philosophical Quarterly*, 26 (1986), 23–32; Anderson-Gold, *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (Albany: State University of New York Press, 2001); and A. W. Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999).

[24] A. W. Wood, "Rational Theology, Moral Faith, and Religion," in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge University Press, 1992), pp. 394–416, p. 407.

[25] I. Kant, *Groundwork of the Metaphysics of Morals*, trans. L. W. Beck (Indianapolis and New York: Bobbs-Merrill, 1959).

"communitarians," this view of the matter highlights the importance of understanding what connects and motivates people of "good will" in the face of such real-world horrors as genocide, torture, suicide bombings, or wars instigated by deception and misrepresentation.

One thing that may connect and motivate them is a *shared narrative* that draws on the moral imagination, showing that there could indeed be "another way" to organize the world and lead our lives. As we have seen, one possible way to view Kant's appeal to the Bible is as a "before-and-after" story that sustains hope for moral recovery by representing it imaginatively. The narrative feature that he so desperately needs and that his moral philosophy is challenged to provide is precisely the feature that yields the regenerated moment. One might say that the capacity to imagine and relate a story about moral change constitutes the condition of the possibility of bringing about that change. The element of "imagination" here is certainly no weakness, for the same reason that the absence of historical accuracy is never a threat to Kant's use of the Bible. Kant's appeal to the motifs of "rebirth" and a "change of heart" do not explain how the transformations occur, but they do underwrite the hope that radical evil is not the final word – the narrative feature is not crucial for its literal accuracy but for its capacity to deliver us from evil. Collectively relating the imagined narrative about moral recovery may thus become a prime resource for the community's response to realities that might otherwise leave it in the profoundly discouraged and immobilized state of a community that can "imagine" no other world than the very one in which it lives.

Curiously, recent events have demonstrated the power of the imagined narrative working quite well for those prone to murder and destruction. In a compelling account of recruitment methods for Islamist suicide bombers and terrorists, George Packer has recently detailed the power of a compelling narrative account that simultaneously contextualizes public events, personal grievances, and a higher cause in a way that spurs individuals to actions of an extraordinary nature. Drawing on the work of Olivier Roy and Marc Sagemann, Packer suggests that the "common ground" for much of the most repugnant terrorist activity on the world scene lies "not in personal pathology, poverty, or religious belief but in social bonds" that create uncommon commitment through "information strategies" that

assume narrative form.[26] As a matter of fact, the entire point of Packer's analysis is to expose the futility of counter-terrorist methods that rely too heavily on military means rather than on deploying alternative information strategies, grounded in social networks, that would depict an alternative vision capable of generating similar levels of motivation. World events amply suggest that narrative is more powerful than military might – or, more to the point, military might is weaker than the appropriate narrative.

In such a setting, the subtle interconnections Kant draws among moral regeneration, narrative form, and the imagination surely warrant further reflection. Since Kant's own cosmopolitan hope for humankind, deeply grounded in Enlightenment ideals, otherwise appears today to be in complete shambles, there is some irony attached to my suggestion. Kant offers us the idea of a shared narrative arising out of the community's moral imagination and depicting the transition from evil to good in a manner that sustains hope for that very possibility. The narrative element is a proxy requiring little or no attention to the accuracy of its details, yet capable of creating the new world that follows a "before-and-after." Kant's is no doubt a profoundly challenging idea – but it is surely not an empty one.

[26] G. Packer, "Knowing the Enemy: The Anthropology of Insurgency," *The New Yorker*, December 18 (2006), 60–9.

4

Kant's Moral Excluded Middle

Claudia Card

1. Prologue

The common sense view is that good and evil, unlike right and wrong (not right), are not contradictories. They are contraries. It is not possible to be both in the same respect at the same time. But it is at least logically possible to be neither. Common sense finds that many people who are not particularly good are not downright evil either. Immanuel Kant rejected that view along with the view that a person can, at the same time, be good in some respects and evil in others. This essay challenges that excluded middle of his, defends common sense, and suggests ways to be between good and evil that preserve much of what is best in Kant's ethics.

The deep and endlessly fascinating issues Kant identifies in Book I of *Religion within the Boundaries of Mere Reason* (hereafter, *Religion*) make that work a natural place to begin thinking about evil even today. As a theory, however, Kant's analysis of "radical evil in human nature" is seriously incomplete. It offers, more specifically, a conception of radical *culpability*. A fuller theory would include a conception of radical *harm*. To be fair, Kant's objective in Book I of the *Religion* is an account of evil in human nature, which makes his focus on culpability appropriate. Yet even his understanding of culpability needs to be deepened with a conception of harm that goes beyond the failure merely to respect humanity. For culpability increases, other things equal, with increase in the harm the perpetrator is

wrongfully willing to inflict. *Destruction* of the humanity of those who still live is a deeper harm than misuse, abuse, or disrespect of it. Hannah Arendt's conception of evil in her work on totalitarianism includes such destruction. Highly influenced by Kant, she appropriated his term "radical evil" and then later abandoned it, partly for fear that "radical" made evil sound attractive or exciting and partly from being convinced at Eichmann's trial in Jerusalem that the psychology of the perpetrators of such horrors was not deep at all but utterly superficial.[1] She had used "radical evil" not to mark deep culpability, however, but to mark the radical harm of being turned into a "living corpse."[2] Some such notion of radical harm might supplement Kant's theory of evil to yield a fuller, more complex, and more balanced view that would still be in important ways Kantian. Destruction of our moral agency strikes at the very capacity that Kant regarded as giving us dignity, a value for which he insisted there is no equivalent. If agency can be destroyed, it can develop and mature, or fail to do so, in ways imputable to agents. This idea suggests the possibility of somewhat Kantian intermediacies between good and evil.

Six controversial theses emerge from my reading of Book I of *Religion*. They offer an overview of his conception of evil and set a background against which I will argue with Kant's moral excluded middle. The six theses might naturally be thought of as responses to the following questions: (1) Is there anything intermediate between a good will and an evil will? (2) Can human beings do evil for its own sake (just because it is evil)? (3) What is the worst principled form that evil can take in human beings? (4) Can we ever know, ultimately, why someone opts for good or for evil? (5) Are we responsible for our own good or evil will? (6) Can evil ever be purged from humanity? These are excellent questions. A good theory of evil can be expected to address them. Here, in brief, are Kant's positions.

First is Kant's excluded middle. Kant denies that there is anything intermediate between a good will and an evil one (R 6: 20–5). A will

[1] See the conclusion of Arendt's response to Gershom Scholem reprinted in P. Baehr (ed.), *The Portable Hannah Arendt* (New York: Penguin, 2000), p. 396.

[2] H. Arendt, *The Origins of Totalitarianism* (New York: Schocken Books, 2004), pp. 565–92.

cannot, he argues, be partly good and partly evil. He does acknowledge grades of evil (R 6: 29–31). But for Kant a will that is not good is thereby evil. This thesis will shortly become the focus of this essay.

Second, Kant denies that we ever embrace immorality as such (R 6: 30, 35–6). We never will "evil *qua* evil" – that, he says, would be *diabolical* (R 6: 35, 37), not human (R 6: 35). This denial points to an asymmetry between Kant's conceptions of good and evil wills. A good will is committed to duty as such. An evil will is not committed to immorality as such. Rather, its commitment to duty is qualified by a more basic commitment to self-interest.

A third, related thesis is that self-love, a principled and unqualified pursuit of self-interest, is the worst *principled* form of evil in human beings (R 6: 36–8). Even an evil will aims at goods, albeit prudential ones. In *Metaphysical Principles of Virtue*, Kant distinguishes some vices as worse than others. He says lying is "the greatest violation of a human being's duty to himself" and that gratitude is "a *sacred* duty."[3] Yet, the principle underlying all vices is the same, a prioritizing of self-interest over duty.

Fourth is Kant's "inscrutability thesis." Our ultimate grounds (reasons) for acting are inscrutable, unknowable, he holds (R 6: 21, 25, 43–44, 51). Why are some people good and others evil, or the same person sometimes good and other times evil? Ultimately, we cannot know.

Fifth is Kant's "imputability thesis" (R 6: 21, 25, 29, 42–3, 45–8). Our good or evil wills are imputable to us; *we* determine what we will, and we can always change ourselves (R 6: 42). We are responsible not just for our evil *deeds* but for our very *propensity to* them. In contrast, our predispositions to good are not imputable. They are given, defining elements of human nature.

Sixth is Kant's "inextirpability thesis." He says the propensity to evil imputable to us is nonetheless inextirpable (R 6: 37). We cannot root it out, although we can overcome it. His thought would have been clearer had he said simply that we cannot get rid of our human vulnerabilities to weakness and impurity, although they need

[3] MS 6: 429, 455.

not defeat us, which he believed. But he seems actually to say that the *propensity to evil*, which he asserts is a choice, is inextirpable. The best sense I can make of this is that we cannot alter the fact that we always have a choice of how to order our incentives. A holy being would not have such a choice, as it would not have the same incentives that we have.

Kant does not number or label these theses as I have. But he asserts and elaborates them in more or less this order. Richard Bernstein has offered an excellent discussion of the tensions between Kant's imputability and inextirpability theses.[4] I have discussed elsewhere Kant's self-love thesis, his denial of diabolical evil, and his inscrutability thesis.[5] That leaves the excluded middle thesis, which interests me in connection with assessing the depth of gravity in wrongs.

2. The Excluded Middle

Kant says our "disposition as regards the moral law is never indifferent (never neither good nor bad)," and also that it is impossible for us to "be morally good in some parts, and at the same time evil in others." He argues thus: "if [a man] is good in one part, he has incorporated the moral law into his maxim," and "were he, therefore, to be evil in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim relating to it would be universal yet particular at the same time: which is contradictory" (R 6: 24–5).

The point is not that one's will must be consistent over time. We can always choose to accept or reject the moral law as our supreme practical principle (R 6: 41). His point seems to be that in prioritizing the moral law, we are committed to applying it to all of our (material) maxims, not just some. If we have reservations, morality is not our fundamental commitment.

[4] R. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press, 2002), pp. 11–45, argues that Kant is "at war with himself" in holding both the imputability and inextirpability theses.

[5] C. Card, *The Atrocity Paradigm: A Theory of Evil* (New York: Oxford University Press, 2002), pp. 77–95.

Kant offers a three-stage account of the will's descent into radical evil and then a two-stage account of the ever-present possibility of overcoming evil. The first stage of overcoming is a revolution in fundamental principle. The next stage consists of the slower and difficult tasks of reform, purifying and strengthening the will to act in accord with its new fundamental principle (R 6: 46–8). Although Kant does not acknowledge anything intermediate between good and evil, he acknowledges degrees, grades (R 6: 29), of evil by way of the three stages. The first he calls frailty. A person with a frail will has adopted the moral law as basic but fails to live up to it on particular occasions. Commitment to the moral law is an act, but, Kant says, it is not performed in time (R 6: 25, 31, 43). Observable acts in time need not reveal our underlying commitments, as the same observable act might conceivably flow from different commitments. Observable acts can also go against one's commitment, as when one acts wrongly from frailty. Kant writes, "'What I would, that I do not!' i.e., I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed" (R 6: 29). I do wrong, but I do not endorse the wrong that I do. I lie but do not rationalize it, as a thoroughly wicked person or even a person with mixed incentives, Kant notes, would tend to do. My will only needs strengthening. The best we can hope to achieve by strengthening is a (merely) good will. A being with a *holy* will would have no frailties. Hence, even the best of the rest of us, vulnerable as we are, have a propensity to evil (R 6: 30).

The second stage is "impurity": mixed incentives. Here, duty is insufficient to move me, and so I rely partly on interest or inclination. "Although the maxim is good with respect to its object (the intended compliance with the law) and perhaps even powerful enough in practice," Kant writes, "it is not purely moral, i.e. it has not … adopted the law *alone* as its *sufficient* incentive but, on the contrary, often (and perhaps always) needs still other incentives besides it in order to determine the power of choice for what duty requires; in other words, actions conforming to duty are not done purely from duty" (R 6: 30).

It is easy to foresee a slide from stage one to stage two, frailty to impurity. Instead of capitulating to temptation or strengthening one's moral commitment, a frail person might begin to rely on self-interest to buttress the motive of duty. Instead of cheating, one might remind oneself not only that it is wrong but "besides, there is the danger of getting caught" and rely partly on prudence. In this second stage, our acts may still (even always) accord with the moral law, whereas in stage one, they violate it. Yet Kant finds stage two, impurity, worse, because it begins a process of corruption in the will, a process of ever greater reliance upon self-interest.

The end of that process is stage three, which Kant calls depravity, corruption, or perversity. The ordering of one's practical principles is now reversed: instead of the moral law taking priority and limiting self-interest, self-interest takes priority and limits morality. We are willing to do our duty only as long as it does not conflict with our interests. Such a willingness would rule out only wrongdoing that is gratuitous in that it does not even appear to promote our interests. Kant's denial of diabolical evil suggests that he would not have admitted such a possibility. In any case, at stage three, evil is finally rooted in the will as a fundamental commitment, truly radical. Kant says a corrupt will "reverses the ethical order as regards the incentives of a free power of choice; and although with this reversal there can still be legally good (*legale*) actions, yet the mind's attitude is thereby *corrupted at its root* [emphasis mine] (so far as the moral disposition is concerned), and hence the human being is designated as evil" (R 6: 30). Because he says here that "the mind's attitude is thereby corrupted at its root," and because he presents this rootedness as the deepest form of evil in the will, I read him as holding that radical evil is reached specifically in this third stage. This reading also makes good sense of the full title of Book I of *Religion*: "Concerning the Indwelling of the Evil Principle along with the Good, Or Of the Radical Evil in Human Nature" (R 6: 18). The evil *principle* appears to be the prioritizing of self-interest *over* duty, which goes beyond stage two.[6]

[6] Some commentators treat the whole three-stage sequence as radical evil, apparently because it is evil in the will. But then, with what would "radical" be contrasted?

Even at stage three, there can be *legally* good (outwardly right) actions, owing to fortuitous coincidences between duty and prudence. And so, Kant says, we tend to deceive ourselves. We rationalize. We endorse our choices. They are no longer made in weakness. In the first two stages, Kant says, our guilt is unintentional (R 6: 38); in the third stage, "it is deliberate guilt (*dolus*), and is characterized by a certain *perfidy* on the part of the human heart (*dolus malus*) in deceiving itself as regards its own good or evil disposition, and, provided that its actions do not result in evil (which they could well do because of their maxims) in not troubling itself on account of its disposition but rather considering itself justified before the law" (R 6: 38). "This is how so many human beings derive their peace of mind" (ibid.).

Whether it makes sense to suppose that one's will can be partly good and partly evil depends on how the will is conceived and how evil is conceived. Kant's conception of the will is complex: the will has a legislative aspect and a decision-making aspect. Apparently, these "parts" of ourselves do come apart in the case of frailty, as our decisions violate our principle. The will is not defined simply by one's maxim and its incentive on a particular occasion. In the opening paragraph of the first section of the *Groundwork*, Kant says of the will that is to make use of our talents that its "distinctive constitution is therefore called *character*," and he contrasts character with temperament. Neither temperament nor character is defined simply by one's choice and its incentive on a particular occasion.[7] Temperament is defined by dispositions and character by commitments. In the more complex account of character in the *Religion*, one's commitments impose an ordering on a given set of predispositions. But what is to rule out an agent's being ambivalent or divided with regard to any particular ordering? An agent who confronted such ambivalence or division within itself might have still another layer, or aspect, of will that takes responsibility for achieving integrity at the level of self-legislation, takes responsibility for *producing* a *coherent* self (will).

Kant's conception of evil presents a difficulty distinct from those of his conception of the will. Like so many moral philosophers, he does not distinguish evils from lesser wrongs. It is not that he takes "evil" simply to *mean* "wrong" (contrary to duty). For Kant, evil (like

[7] G 4: 393.

morally good) includes the incentive as well as the material maxim. But for Kant *every* wrong is part of an evil as long as it has an incentive given in his analysis of our propensity to evil. That analysis recognizes some deeds (such as those done from weakness) as evil to a *lesser degree* (a lesser grade). But it offers no way to distinguish atrocities (torture or mass murder, for example) from trivial culpable wrongs (such as petty theft or trivial lies). Kant's analysis simply picks out some forms of culpability as worse (more thoroughly evil) than others. Acting on an evil principle is worse than acting from frailty.

Kant may have believed that there are no trivial wrongs. But suppose I were a closeted homosexual and that to protect my privacy, I lied about it without causing any foreseeable suffering. Such a lie would be evil on Kant's theory. But one can dispute that judgment, without disputing the lie's wrongness, my culpability, or even Kant's view that should harm result, the harm would be imputable to me, and further, I will argue, without resorting to utilitarian ethics.

In light of these issues, and against Kant's excluded middle thesis, I offer five kinds of intermediacy worth taking seriously, five commonly recognized ways that we often judge individuals to be either "not quite good but not quite evil, either," or "partly good and partly evil." The first intermediacy requires but a simple modification in Kant's terminology. The second requires a substantive change in his ethics (although less drastic than he might have thought). The third and fourth require further complexity in Kant's metaphysics of the will, building on what he already has. And the last challenges most ethical theories, not just Kant's.

In brief, here are the five intermediacies. First is frailty. "Evil" is ordinarily too severe a judgment for frailty. Most frailty is better regarded as intermediate. Possible exceptions are weaknesses easy to overcome when those weaknesses lead foreseeably to intolerable harm to others without comparable harm threatening the agent. Still, a frail will is not evil without qualification. Kant admits that a frail will is divided against itself, acts against its own good principle: "'What I would, that I do not!' i.e. I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed" (R 6: 29).

Good reasons not to regard most frailty as evil are (1) one's basic commitment is uncorrupted and (2) even the best of us has frailties, but the seriousness of "evil" is diluted when applied to the best of us. President William Jefferson Clinton's notorious frailties led to impeachment by the US House of Representatives in 1998.[8] Yet even in lying to cover his tracks he should not be judged *evil*. He fudged the labeling of his philandering (what counts as "having sex"?) but did not, in the end, endorse his conduct. He apologized. Publicly. In contrast, if Susan McDougall's memoir is accurate, US independent counsel, special prosecutor Kenneth Starr, was evil to engineer her eighteen months' confinement in progressively worse prisons that harmed her health, even jeopardized her life, all for her refusal to implicate the Clintons falsely in Whitewater wrongdoing.[9] To my knowledge, Starr has admitted no wrongdoing.

The Clinton/Starr contrast takes us to my second intermediacy, which it may also illustrate, namely, fully culpable wrongdoing that is not foreseeably harmful (or not very). Common sense distinguishes evils from minor wrongs not just by the agent's strength of commitment (as in frailty) but by the depth of harm wrongfully but willingly done. *Other things equal*, more harmful wrongs are more reprehensible, if the harm was reasonably foreseeable. According to Kant's grades of an evil will, culpable wrongs that are similarly motivated are ethically equally grave regardless of differences in foreseeable harm, although the vices they exemplify may not be equally grave and the *lex talionis* dictates punishments of different severities. Ethically, although not legally, petty theft is on a par with murder when the incentive of both is self-interest. Ethically, for Kant, harm is not only not definitive but irrelevant. That is bizarre. I will suggest a way to acknowledge the moral relevance of harm without utterly abandoning a Kantian moral framework (although Kant might not have liked it) and without invoking any form of utilitarianism, not even rule utilitarianism.

A third intermediacy is ambivalence or indecisiveness at the level of principle. Kant assumes that we are never without a fundamental

[8] On 19 December 1998, the House of Representatives approved two of the proposed articles of impeachment. Without confirmation by the US Senate, no further action was taken.

[9] See S. McDougall and P. Harris, *The Woman Who Wouldn't Talk* (New York: Carol & Graf, 2003).

commitment regarding the moral law. Yet, appearances in everyday life and in ordinary moral development fly in the face of this view.

A fourth intermediacy is illustrated by persons who appear to exercise good (even exemplary) moral judgment in certain contexts and, during the same time period, astonishingly poor moral judgment in others. Such people seem to have a good side and an evil side, a characterization Kant rejects. Perhaps we can always interpret a person's acts in accord with some imagined single fundamental commitment, as geocentric astronomers can always add epicycles to the paths of the planets. Yet, a more complex view of the will may do equal if not better justice to appearances, enable us better to raise questions we should want to raise regarding responsibility and agency, and allow us to hold some such agents responsible for their lack of an integrated character.

The fifth intermediacy, which presents a challenge for most ethical theories, is suggested by Primo Levi's "gray zone."[10] Here victims of oppression under duress operate the machinery of oppression. Theirs is no ordinary weakness. Some kinds of duress do not so much reveal frailties in human nature that a good person would strive to overcome as challenge one's very principles. The meaning of "good will" can become unclear in such circumstances. Yet some degree of moral choice clearly remains in that some options are far worse than others. Despite their best efforts, gray-zone survivors often feel morally compromised. That sounds intermediate.

Of these five kinds of intermediacy, frailty is sufficiently clear. So let me elaborate the remaining four, beginning with the distinction between evils and other fully culpable wrongs.

3. Evils vs. Lesser Wrongs

In popular paradigms of evil, such as genocide or torture, *harm* is what is most salient. Evils, unlike lesser wrongs, are thought to do reasonably foreseeable intolerable harm. Harm becomes radical, intolerable, when it jeopardizes access to basics ordinarily needed to make a life (or a death) tolerable or decent. Such basics include uncontaminated

[10] P. Levi, *The Drowned and the Saved*, trans. Raymond Rosenthal (New York: Vintage, 1989), pp. 36–69.

air, water, and food, sleep, the ability to move one's limbs, spheres in which one can exercise effective choice, and freedom from such things as severe and prolonged pain, humiliation, debilitating fear, disabling and disfiguring diseases, extreme and prolonged isolation, and so forth.[11]

In her work on totalitarianism, Hannah Arendt used the term "radical evil" to describe practices that produce a very deep *harm*, whereas Kant used that term to characterize deep *culpability*. In principle, the two can be combined: one can be radically culpable in inflicting radical harm. We might have expected Arendt to assert later that Eichmann exemplified just that. But she did not. What struck Arendt about Eichmann was the shallowness of his motives. She refused to describe as radical the conduct of a man who struck her as totally superficial. She then claimed that Eichmann had no intent to do wrong, that he lost the ability to distinguish right from wrong, that "nothing would have been further from his mind than to determine with Richard III 'to prove a villain,'" and concluded that he was guilty of sheer thoughtlessness.[12]

We can take issue with Arendt's readiness to believe Eichmann's protestations of not being anti-Semitic. But many of her conclusions can be contested without disputing her take on the facts. By his own admission, Eichmann deliberately coordinated trains whose destinations he had seen and understood firsthand from visits to the Eastern front and to death camps. There is no reason to think he ever forgot what he saw. Because, knowing what he knew, he did his work meticulously, "thoughtless" is a description that might mislead anyone who is not attuned to Arendt's special understanding of "thinking." Eichmann fits Kant's third degree of evil. He prioritized his own advancement (self-interest) over the Categorical Imperative (which he quoted at his trial with fair accuracy, admitting he had ceased to follow it). But his priorities are not the most salient thing. More salient are the depth and extent of the harm that he knowingly and willingly furthered, the fate to which he sent trainloads of people. Yet

[11] What I here call "radical harm" I call "intolerable harm" in *The Atrocity Paradigm* (throughout).

[12] H. Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, revised and enlarged edn (New York: Penguin, 1994), p. 287.

even mass death is not what Arendt meant by "radical evil," when she was still willing to use that term.

Radical evil, for Arendt, literally *dehumanizes* victims, producing "living corpses." Radical evil was suffered not by those killed immediately but by those who survived long enough to lose dignity. This thought, although not Kant's, is in the spirit of Kantian values. Wrongful willingness to inflict such harm aggravates the culpability of even Kant's third degree. But Arendt did not pursue that thought. Her view at that time was that such evil dehumanizes perpetrators as well. When perpetrators become living corpses, lacking spontaneity and interchangeable with each other, then responsibility and culpability have disappeared, a state of affairs more chilling than Kant's radical culpability.[13]

In contrast with wrongs that do extreme harm, wrongs that foreseeably do no harm (or little harm) do not warrant the gravity of the judgment "evil." Kant's view makes sense if he reasons that one who would commit petty theft not through frailty but from calculated self-interest must also be prepared to commit murder should it become profitable since, barring fundamental character change, the underlying principle would prioritize self-interest in both cases. But that last premise may be wrong: even without fundamental character change, the underlying principle may be more complex. A self-interested thief might have scruples against murder or torture, drawing the line not from fear of detection but because of the depth of foreseeable harm to victims. Imagine a Robin Hood who would *steal* from the rich but would not *murder or torture* even the rich. The case is different if Robin Hood thinks he is morally right. My "Robin Hood" concedes the wrongness of his thefts but keeps on anyway, partly because he enjoys being the instrument of others' good fortune (self-interest is one incentive) but also partly because he does not think the wrong very *serious*, since it does no harm that victims cannot easily absorb. A scruple qualifies his self-interest: he would not murder or torture to become a still greater benefactor. Maybe he would not steal from the rich who have physical disabilities. This plausible character's scruples are captured neither by Kant's moral law nor by simple self-interest but by a principle of self-interest that makes real concessions to morality.

[13] Arendt, *Origins*, pp. 437–59. She, of course, changed that view of the perpetrators after reporting on the Eichmann trial in Jerusalem.

"Evil" is too strong a condemnation. "Less than scrupulous" (maybe "slippery"), but not evil. Or, not yet.

To cast the point in Kantian language, the *contents of one's material maxims* might distinguish evils from lesser wrongs. How to state a maxim is a well-known difficulty in Kant's ethics. He does little to clarify material maxims beyond noting that (unlike formal maxims) they do not abstract from ends, and he offers widely disparate examples. Maxims are subjective principles that state our intentions in general form, including at least an act and a purpose, if not also circumstances. Kant states his material maxims in value-neutral terms; that, at any rate, seems his intent. In some of his best cases, harm is truly not foreseeable in the individual instance. But in other cases, such as torture and many killings and rapes, radical harm is not only foreseeable but caused deliberately. If foreseeable radical harm were included in a maxim to be subjected to Kant's universality test, the harm would not determine the rightness or wrongness of the maxim but would indicate in a morally relevant way what one is willing to do. Including reasonably foreseeable harm in the statement of one's material maxim would also fit nicely with Kant's theory of imputation, which recognizes that we are responsible for harm caused by our violations of duty.[14] The significance of such harm in one's maxim would be that should the maxim fail the universality test, acting on that maxim would be not merely wrong but evil. And that seems correct. Such an emendation would make Kant's ethics less stoic but would not reduce judgments of evil to a calculus of utility. Harm would not determine duty, only whether a culpable violation of duty was not merely wrong but also evil.

Why distinguish evils from other wrongs? First, evils are the most important wrongs to avoid. Second, for moral blame (not only punishment), it matters not just whether agents are frail, impure, or corrupt but also how deep the harm they willingly inflict is. Evils are more reprehensible than other wrongs. Moral vocabulary should center and encourage this distinction.

[14] Near the end of his general introduction to *The Metaphysics of Morals*, Kant states two principles of imputation. First, we are not responsible for the good or bad results of a morally required action or of omitting a meritorious action. Second, we are responsible for the bad results of a wrongful action and for the good consequences of a meritorious one (MS 6: 227–8).

4. Two Ways to Lack Unity in the Will

My next two forms of intermediacy are (1) indecisiveness or ambivalence and (2) internal conflicts that are not reducible to frailty. Kant assumes a basic coherence or unity of the will. He implicitly rejects the common-sense views that some of us lack basic commitments and that some of us even have plural and incompatible but equally basic commitments. The appearance of ambivalence and of the lack of a single underlying commitment is created by conflicting patterns of action, each with its own apparent priorities, no pattern clearly dominant over the others. These appearances can lead us to press the questions what it means to have a fundamental commitment, how deep our commitments can go, and what evidence we have for our commitments. For simplicity, I ignore here the distinction between evils and other wrongs.

Consider someone who is unpredictably irresponsible. Some days, she feels like not getting up for work (or like getting up and playing hooky) and so calls in sick, not from weakness but because then inclination just seems more important. Other days, she is moved by obligation, despite feeling it would be a great relief to stay home and unplug the phone. She does the right thing then because that is what seems most important *then*. This woman appears ambivalent – not frail, not even committed to self-interest, but basically uncommitted. Kant would have to say that her fundamental commitment changes often. The common-sense view is that she is immature, has not "got her act together," has not yet developed a fundamental commitment (and possibly never will). Yet we also tend to hold that against her.

Kant might object that she does not exhibit a good will and a bad will at the *same* time, which might seem to be all that he denies. In his account of how we can overcome evil, he acknowledges that one can exhibit a good will and a bad will at *different* times. He also stresses our fallibility with regard to our own deepest motives. Appearances can mislead.

Well, what counts as "a time"?[15] Must it be a moment? Can it be a *stretch?* For a basic commitment, a stretch seems right. Kant says our

[15] Thanks to Marilyn Frye for raising this question ("What counts as a 'time'?") long ago in a very different context. See M. Frye, *Willful Virgin: Essays in Feminism 1976–1992* (Freedom, CA: The Crossing Press, 1992), pp. 109–19.

basic disposition is an act, a choice "an intelligible deed, cognizable through reason alone apart from any temporal condition" (R 6: 31). Choices can be made in a moment. But the choice of a commitment needs a stretch of time for its realization. As Aristotle says of friendship, the wish for it arises quickly, but friendship does not.[16] Friendship requires time and trials. Likewise, moral commitment; quickly chosen, it does not materialize overnight. In discussing the overcoming of impurity and frailty, Kant sees this but does not acknowledge that the time required to translate commitment into action opens a space for ambivalence or indecisiveness *in commitment*, not just in practice.

When conflicting *patterns* of behavior appear in the same stretch of time and in the same contexts, they suggest a will that is fundamentally *un*decided, *un*committed. Such a person might be called a moral "flip-flopper." To insist that at any moment one's commitment is fundamentally either good or bad is to dismiss the salient pattern in flip-flopper agency. What we ordinarily call *lack* of commitment is not revealed in a moment. Such ambivalence seems common among children, adolescents, adults with troubled pasts, and those who live under stressful conditions. To describe it as a weak commitment to morality suggests that, in contrast, the commitment to self-interest is strong. But that may not be true.

The moral flip-flopper exhibits *unpredictably* different patterns in the *same* contexts, a fairly common case. A less common but by no means unusual case that also appears to exemplify lack of unity in the will is the person who exhibits *systematically* different and conflicting patterns in *different* contexts. Some people behave predictably well in some contexts but equally predictably poorly in others. That difference may not always be explainable by the agent's relative ignorance of facts in the second context. To illustrate, some people have excellent judgment in positions of public responsibility but not at home in family life or in intimate relationships. This is not an unusual case, although it can become extreme enough to seem bizarre and raise questions regarding the person's status as a responsible agent. I choose an extreme case because it illustrates so clearly the challenges for a Kantian interpretation.

[16] Aristotle, *The Nicomachean Ethics*, trans. D. Ross (New York: Oxford University Press, 1925), p. 197 (VIII: 3).

Consider this example from real life. Sue William Silverman, an incest survivor, writes the following about her father in her memoir (published after his death). He was Chief Counsel to the US Secretary of the Interior from 1933 to 1953 and played key roles in establishing statehood for Alaska and Hawaii, Philippine independence, the creating of the Puerto Rican Commonwealth, home rule for the Virgin Islands, Guam, and Samoa, and civilian rule of Japanese possessions after World War II. From 1954 to 1958 he was president of large banks. He was photographed with President Harry Truman, Adlai Stevenson, and other influential political figures. And he was a child molester. For many years, he assaulted his daughter sexually, severely, locking her door at night in her bedroom, beginning when she was less than five.[17]

Were those who placed this man in positions of public trust *totally* deceived about his character? Or did he have a good side and an evil side? He appears at first to have embodied the contradiction Kant thought impossible. If, however, we regard him responsible for both patterns of behavior, and if they truly do manifest conflicting principles or priorities that he has, his character is not at its *most* basic level defined by these principles (hence, does not exhibit the contradiction Kant rejected). Rather, at its most basic level his character is defined by his failure to take responsibility for himself in a way that people with more coherent or conventional inclinations might never have to. This kind of failure is not captured by a formal maxim prioritizing self-interest. The task facing this man is to create a coherent self. Nor is his failure well captured by frailty. How much strength could it take not to rape one's five-year-old daughter and continue doing so behind a locked door for years? There is a policy here, not a lapse.

This man's "good side" and his "evil side" suggest different potentialities for an integrated character, something this man never achieved. Whether we should hold an individual responsible for that kind of failure is a matter on which not even psychiatrists are agreed.[18] If not,

[17] S. W. Silverman, *Because I Remember Terror, Father, I Remember You* (Athens, GA: University of Georgia Press, 1996). Her father's political career is summarized on p. xv.

[18] Silverman indicates in her memoir that her father was probably also subjected to sexual abuse during his own childhood. What, if anything, did he remember about that? There is probably no simple answer to questions of responsibility that would do for all such divided persons.

he is better described as neither good nor evil, no "moral personality" at all. For myself, I see no conceptual barrier to holding responsible a man who shouldered the morally complex political responsibilities that this man did. Yet holding him responsible for failure to achieve overall integrity presupposes complexities in the will that go beyond, although they could build upon, Kant's legislating and deciding aspects of the will. A man who did achieve his own integration, who confronted conflicting ideals that were warring in his soul, is African American sociologist and philosopher W. E. B. DuBois, who reflects on his experience of "twoness" in the first chapter of *The Souls of Black Folk*.[19] His case might be instructive for developing a view of the will that goes beyond Kant's distinctions, extending further the capacity for rational agency.

If we read the character of Sue William Silverman's father as Kant apparently would have, then we take his treatment of his daughter to reveal the real man under the sham of his public face. We must then wonder what unexposed abuses he perpetrated in his public trust. Not likely he could occupy such positions for so long without confronting moral conflicts. He does not appear to have undergone deep character change but, rather, continued this apparent Jekyll/Hyde pattern over many years. Possibly his "good" public behavior was motivated by the rewards of reputation and salary, and his public duties luckily coincided with what was prudent.

And yet, is it not also *possible* that he made moral decisions conscientiously on the job (asking seriously whether he could universalize the maxims of his actions), despite his failure to negotiate family life honorably?; that he was a moral model in public and a moral monstrosity at home? To look at him moment by moment and ignore the patterns is to dismiss what is most striking, even sinister, in his agency: his failure to heal that deep and persistent split.

If appearances do *not* mislead, moral flip-floppers are neither fundamentally good nor fundamentally evil, because of the absence of an enduring basic disposition. Yet we often hold adults responsible for such immaturity. Failure to develop a stable basic commitment can be due to culpable negligence rather than bad priorities or weak

[19] W. E. B. DuBois, *The Souls of Black Folk* (New York: New American Library, 1969), pp. 43–53.

commitment. In Arendt's language, there may be a moral failure here to think about what one is doing. Culpable negligence is a well-known problem for Kant, as there may be no maxim (intention) to fail the universality test.

In contrast to moral flip-floppers, Jekyll/Hyde characters seem to have two enduring but opposed dispositions underlying their observable deeds, for which they fail at a deep level to take responsibility. Their failure also seems a moral one but more troubling and puzzling than either immaturity or ordinary self-interestedness. For clearly they often do what is right in the face of strong pressures to do otherwise. Other Jekyll/Hyde cases may include slave owners, such as President Thomas Jefferson, in the ante-bellum American South, and the Nazi doctors interviewed by Robert J. Lifton.[20] Yet, if DuBois was able to get it together, can we not expect at least some others to do so as well and hold them responsible if they fail?

Kant admits that our best evidence of the nature of our will consists in patterns of choice that we observe over time in our conduct.[21] Yet some patterns seriously challenge his faith that underlying the appearances at any particular time is a single, coherent fundamental commitment. It is more likely that individual instances of our behavior might mislead us as to the true nature of our will than that large-scale patterns over long periods of time would do so. Kant does not explicitly consider lives in which more than one pattern appears repeatedly over the same period of time. His analysis has the virtue, however, of acknowledging layers in the will: the layer that legislates and the layer that carries out the legislation (or fails to). Perhaps there can be also layers that confront ambiguity or divisions within the self at the level of legislation, or that refuse to confront them, or neglect to do so.

5. Gray Zones

In his reflections on Auschwitz, Primo Levi identified a "gray zone" in which some victims of evil become perpetrators of the very evils they suffer.[22] They do so by accepting positions of power over others and

[20] See R. J. Lifton, *The Nazi Doctors: Medical Killing and the Psychology of Genocide* (New York: Basic Books, 1986).

[21] See, for example, R 6: 69 and 77.

[22] I discuss gray zones at greater length in *The Atrocity Paradigm*, pp. 211–34.

in that capacity inflicting harms on others in exchange for rewards that may be only a postponement of their own suffering or death. Levi cites prisoners in death camps who became *Kapos* (the term in German means a junior non-commissioned officer) and ghetto prisoners who served as ghetto police or on ghetto councils, charged with selecting or rounding up prisoners to be sent to their deaths and with enforcing other rules that predictably resulted in deaths. Some prisoners refused such service. Many (perhaps most) of those who refused did not survive. Others, knowing this, were torn: should I refuse on principle? Should I exploit an opportunity that might enable me to save others, not necessarily myself?

The realities were that accepting such a position almost certainly made one complicit in evils, wrongs that foreseeably did intolerable harm. But did it make the complicit person's *will* evil? Or is it genuinely unclear how one ought to confront such a choice? If unclear, is that because there are right responses about which reasonable people might disagree? Or is it that *sometimes* there is no unambiguously right choice, that the ambiguity is in the nature of the case, not just in its appearance, which would suggest the possibility of a will that is not unambiguously either good or evil? Multiple ambiguities in Levi's gray zones seem to offer substance to the idea of "gray areas" between good and evil, not areas of indifference and also not just ambivalence.

In conclusion, my own suspicion is that most of us most of the time are somewhere between good and evil, if only because our common moral failings usually fall short of producing reasonably foreseeable major harm. But also, the processes of moral maturation, like those of moral deterioration, can involve ambiguities and uncertainties of principle, not just of judgment regarding empirical facts. On our way to becoming good or evil, we may pass through more or different stages than the three levels of the Kantian descent into radical evil or Kant's two-stage recovery program. Yet Kant's analysis remains a fruitful beginning for a more empirically and phenomenologically sensitive analysis.[23]

[23] I am grateful for comments and suggestions on earlier versions of this essay from audiences at the 2005 Chapel Hill Colloquium, the philosophy departments at the universities of Victoria, Michigan State, and Illinois, the 2005 conference on evil at the University of Purdue, and in particular from Sharon Anderson-Gold, Paula Gottlieb, Thomas E. Hill, Jr., and David Sussman.

5

Evil Everywhere

The Ordinariness of Kantian Radical Evil

Robert B. Louden

> If someone like that did all this, then there is really no chance. That's the biggest evil, that it was not someone from far away. It was one of us.
>
> *Ahmed Kulenovic, a Muslim, commenting on his childhood friend, Dusan Tadic, a Serbian. Tadic, currently in prison, is the first person to be convicted of crimes against humanity by an international court since the Nuremberg trials after World War II.*

> The human being is by nature evil.
>
> *Kant, Religion within the Boundaries of Mere Reason (6: 32)*

Since 9/11, American politicians, preachers, journalists, and academics have all invoked the word "evil" with a frequency that has not been seen since the Holocaust. Philosophers too have contributed to this phenomenon in their customary way, issuing a small spate of monographs and anthologies on the topic.[1] Most of these recent books do

[1] See, e.g., R. J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press, 2002), Bernstein, *The Abuse of Evil: The Corruption of Politics and Religion since 9/11* (Cambridge, MA: Polity Press, 2005), C. Card, *The Atrocity Paradigm: A Theory of Evil* (New York: Oxford University Press, 2002), R. W. Grant (ed.), *Naming Evil, Judging Evil*, with a Foreword by A. MacIntyre (University of Chicago Press, 2006), M. P. Lara (ed.), *Rethinking Evil: Contemporary Perspectives* (Berkeley, CA: University of California Press, 2001), S. Neiman, *Evil in Modern Thought: An Alternative History of Philosophy* (Princeton University Press, 2002), A. D. Schrift (ed.), *Modernity and the Problem of Evil* (Bloomington: Indiana University Press, 2005). NB: I do not mean to

not devote serious attention to Kant's account of radical evil,[2] in part because of the authors' shared belief, as one critic puts it, that "when faced with the question of evil," Kant, "the quintessential modern Enlightenment philosopher," is "confused, … eventually confesses defeat," and offers only "confused chatter about the rooting of radical evil in human nature."[3]

My own view is that we still have much to learn from Kant's account of radical evil. While no author will ever have the last word on such a perplexing and pervasive feature of human existence, Kant's discussion of radical evil is still very much relevant today. For as others have noted,[4] his theory of radical evil is the first distinctively *modern* account of evil: he coined the evocative expression "radical evil," a term which has been repeatedly invoked (albeit sometimes in ways that differ strongly from his intended meaning) in contemporary efforts to make sense of the horrors of the twentieth and twenty-first centuries; he is "the modern philosopher who initiates the inquiry into evil without explicit recourse to philosophical theodicy"[5] (albeit in a manner that definitely does not foreclose religious responses to evil), and he is the first writer to place human responsibility for moral evil at the center of his account of evil.

imply that each of these books was directly inspired by 9/11: this is clearly not the case. But the recent surge of philosophy books on evil is noteworthy, whatever the precise causes of the phenomenon might be.

[2] Exceptions include Neiman, who assigns "a central place" to Kant in her narrative (S. Neiman, *Evil in Modern Thought*, p. 61 – see also her earlier book, *The Unity of Reason: Rereading Kant* [New York: Oxford University Press, 1994]); Bernstein, who notes that it was Hannah Arendt's reference to Kant that initially aroused his interest and curiosity in the expression "radical evil" (Bernstein, *Radical Evil*, p. 3) and who opens his interrogation with a chapter devoted to Kant's account of radical evil ("Radical Evil: Kant at War with Himself," pp. 11–45 – an earlier version of this chapter is included in Lara (ed.), *Rethinking Evil*, pp. 55–85); and Card, who also includes a chapter on Kant ("Kant's Theory of Radical Evil," pp. 73–95 in *The Atrocity Paradigm*). Part of my aim in this essay is to respond to several of Bernstein's criticisms of Kant's position on radical evil, criticisms which he often develops in the course of discussing appraisals that other commentators have offered of Kant's account.

[3] W. L. McBride, "Liquidating the 'Nearly Just Society': Radical Evil's Triumphant Return," in Schrift (ed.), *Modernity and the Problem of Evil*, pp. 28–38, pp. 29, 36.

[4] See P. Dews, "Disenchantment and the Persistence of Evil: Habermas, Jonas, Badiou," in Schrift (ed.), *Modernity and the Problem of Evil*, pp. 51–65, p. 52; Neiman, *Evil in Modern Thought*, p. 218.

[5] Bernstein, *Radical Evil*, p. 4.

Kantian radical evil is a huge topic, and my aim in what follows is by no means to cover all of this difficult and bewildering terrain. Rather, I wish to focus on four basic criticisms of his account of radical evil. In each case, I have two goals: first, to defuse the criticisms by explaining Kant's position in a way that is consistent both with his texts as well as with common sense; second, to show – contrary to what critics claim – that Kant's position on these matters, correctly interpreted, is not a weakness but in fact a strength.

Explanatory Impotence and Human Freedom

One common criticism of Kant's doctrine of radical evil is that ultimately it does not explain anything. Once we have worked our way beneath its forceful rhetoric, we are left with something obvious and unenlightening. For all that remains beneath the surface is the simple claim that radical evil refers to our propensity to knowingly do what is morally wrong, to intentionally violate fundamental moral principles. As Kant himself puts it: "The statement 'The human being is *evil*,' cannot mean anything else than: he is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it" (R 6: 32).

We have already seen an oblique version of this criticism in William McBride's complaint (see n. 3, above) that, when faced with the question of evil, Kant "eventually confesses defeat." Gordon E. Michalson, Jr. also hints at it when he notes that despite the "complex conceptual and terminological gridwork" attending Kant's account of radical evil, in the end we are left with "an unhelpful result"[6] – one that does not ultimately explain why we choose evil. But Richard Bernstein has developed the most detailed version of the *explanatory impotence criticism*. In *Radical Evil*, he writes:

> The more we focus on the details of Kant's analysis of radical evil, the more innocuous the concept seems to be … We do not always follow the moral law *because*, as human beings, we have an innate propensity to evil. Our wills are corrupted at their root. But does this "because" really explain anything?

[6] G. E. Michalson, Jr., *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge University Press, 1990), p. 61.

Does it do any conceptual work? I do not think so. When stripped down to its bare essentials, it simply reiterates the fact that human beings who are conscious of the moral law sometimes (freely) deviate from it … In short, radical evil – the alleged propensity to moral evil which is a universal characteristic of human beings – does not have *any* explanatory force (practical or theoretical) at all![7]

A few pages later, Bernstein raises the ante further by claiming that in his analysis of radical evil Kant has unconsciously caught himself in a dialectical illusion – the diagnosis of which is one of the main themes of the *Critique of Pure Reason*, indeed of Kant's work in general. In the first *Critique*, Kant warns readers that in addition to empirical or optical illusions (where our imagination leads us to think we see things that are not there) and logical illusions (where, because we have made a fallacious inference, we make a false judgment about something), there also exists a much more fundamental and dangerous kind of illusion – viz., one that occurs when we try to employ concepts of the understanding beyond the limits of possible experience. This latter dialectical illusion is "*natural* and unavoidable" – it is "attached irremediably to human reason, so that even after we have exposed the mirage it will still not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed" (KrV A298/B354). Similarly, Bernstein claims, the concept of radical evil has caught Kant (and all those who accept his doctrine) in a dialectical illusion, because "it seduces us into thinking that we can *explain* something that we cannot possibly explain." It is an illusion to think that the doctrine of radical evil "enables us to explain or account for why we adopt evil maxims, why we sometimes succumb to his temptation. This alleged explanation turns out to be vacuous."[8]

However, the explanatory impotence criticism misses its target completely. For Kant is quite clear in stating that his doctrine of radical evil is in no way intended to explain *why* human beings choose to adopt evil maxims. The adoption of evil (or good) maxims is always a free choice; one for which each person is responsible. Without this assumption of freedom, it does not make sense to hold us morally

[7] Bernstein, *Radical Evil*, p. 33. [8] Ibid., p. 35.

responsible for any of our actions. As Kant states in the *Religion*: "The human being must make or have made *himself* into whatever he is or should become in a moral sense, good or evil. Both conditions must be an effect of his free choice [*freie Willkür*], for otherwise they could not be imputed to him and, consequently, he could be neither *morally* good nor evil" (R 6: 44; cf. KpV 5: 98). In short, on Kant's view it is fundamental to our own self-conception as moral agents that we are free-acting beings (cf. R 6: 37; VA 7: 119). In any given choice situation, we can hypothesize about the roles that various environmental and genetic factors may have played in leading a person to make the choice that he or she made, but ultimately no causal explanation is fully satisfactory, for the simple reason that the choice was free. In many cases, we simply do not know why people choose to do what they do – this is something, as Kant famously remarks, that "remains inscrutable [*unerforschlich*] to us" (R 6: 43). For each human being, "the depths of his own heart (the subjective first ground of his maxims) are to him inscrutable [*ihm selbst unerforschlich*]" (R 6: 51; cf. MS 6: 392). Our assessment of others' motives and acts is also always fallible. Indeed, at one point in the *Groundwork of the Metaphysics of Morals* Kant insists that we do not know "with complete certainty" whether anyone in the entire history of the human race has ever succeeded in performing an act from the motive of duty (G 4: 407).[9]

Ironically, Bernstein himself fully endorses this particular aspect of Kant's position on radical evil. On the last page of his book, he states:

> *The ultimate ground for the choice between good and evil is inscrutable.* We initially encountered this thesis in Kant's reflections on radical evil, when he claimed that the ultimate subjective of the adoption of moral maxims is inscrutable. I consider this to be one of Kant's most profound and important insights about morality.[10]

However, in making this assertion, Bernstein completely undercuts his earlier explanatory impotence criticism. As he himself

[9] Kant's doctrine of the opacity of intentions is a major theme in Onora O'Neill's work. See, e.g., O. O'Neill, *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge University Press, 1989), pp. 7, 77, 85, 88, 98, 130, 151–2.

[10] Bernstein, *Radical Evil*, p. 235.

notes: "Human beings are responsible for the choices they make, but *ultimately*, we cannot explain why they make the moral choices they do … Not only is this inscrutable; it *must* be inscrutable, because this is what it means to be a free and responsible person."[11] Ultimately, the explanatory impotence criticism will only persuade hard determinists who, because they assume that every event in the universe is caused by antecedent causes, conclude that a complete and accurate causal account of every human action is in principle always available and that moral responsibility is therefore impossible.[12]

Kant's position regarding the ultimate inscrutability of human motives is a strength rather than a weakness in his doctrine of radical evil. Human action often does have an indecipherable character. Particularly in cases where people have committed horrendous acts of moral evil, we are often simply at a loss to explain definitively why they did what they did. Even the most ordinary people are capable of the most horrendous deeds, and it is to Kant's credit that he recognized this disturbing fact of human life.[13]

Finally, Kant's stance with regard to the inscrutability of freedom is by no means novel or extreme. It has a long and distinguished pedigree that tracks back as least as far as Augustine. In *The City of God*, Augustine too warns readers against attempts to offer causal explanations for moral evil: "the truth is that one should not try to find an efficient cause for a wrong choice." Trying to find explanatory causes in cases where people have made a free choice, he adds, in a famous analogy, "is like trying to see darkness or to hear silence."[14]

[11] Ibid., p. 45.

[12] For a contemporary defense of this position, see G. Strawson, "The Impossibility of Moral Responsibility," *Philosophical Studies*, 75 (1994), 5–24.

[13] As others have noted, here there appears to be a link to Hannah Arendt's later thesis concerning the banality of evil. In the Epilogue to *Eichmann in Jerusalem: A Report on the Banality of Evil*, revised and enlarged edn (New York: Penguin Books, 1977), she writes: "The trouble with Eichmann was precisely that so many were like him, and that the many were neither perverted nor sadistic, that they were, and still are, terribly and terrifyingly normal" (p. 276). Cf. S. Anderson-Gold, "Kant's Rejection of Devilishness: The Limits of Human Volition," *Idealistic Studies*, 14 (1984), 35–48, esp. p. 48 n.30; and H. E. Allison, "Reflections on the Banality of (Radical) Evil: A Kantian Analysis," in *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge University Press, 1996), pp. 169–82. Allison's essay is also reprinted in Lara (ed.), *Rethinking Evil*, pp. 86–100.

[14] Augustine, St., *The City of God*, trans. Henry Bettenson (Harmondsworth: Penguin, 1977), XII.7, p. 480. For a recent appreciation, see S. Hauerwas, "Seeing Darkness,

Ultimately, free actions – whether for good or for evil – are fundamentally inexplicable.

Self-Love and the Moral Law

In the previous section I argued in part that Kant's account of radical evil is not primarily a theory about *why* people commit acts of evil, and that he has good reasons for not offering a theory of this sort. Because of his dual commitments to human freedom and to the ultimate inscrutability of our motives, he does not have a lot to say about what specifically drives people to do evil. For some readers, this result is understandably unsatisfying. What they want from a theory of evil is an explanation of why people commit acts of evil. However, on Kant's view this is an illegitimate request that is foreclosed by our awareness that we are free beings whose actions are not causally determined. When it comes to human motivations to do evil, all that we can safely and accurately say is that whenever people commit evil, they have intentionally violated fundamental moral norms – they are "conscious of the moral law" but have willfully deviated from it (cf. R 6: 32).

However, this is not quite the complete Kantian story regarding human motives. Kant does secondarily address issues of motivation in his discussion of radical evil, but critics have not been happy with this part of his analysis either. For instance, at one point in the *Religion* he asserts bluntly that "self-love [*Selbstliebe*]," "when adopted as the principle of all our maxims, is precisely the source of all evil [*gerade die Quelle alles Bösen*]" (R 6: 45; cf. 30–1, 36). This assertion that self-love is the sole source of evil has led many commentators to criticize Kant for his allegedly simplistic and naïve account of human nature.

Perhaps the most famous example of the *self-love criticism* is to be found in Hannah Arendt's *The Origins of Totalitarianism* (1951). In

Hearing Silence," in Grant (ed.), *Naming Evil, Judging Evil*, pp. 35–52. Kant's inscrutability of freedom position has also received significant post-Kantian support. Schelling, for instance, in *Of Human Freedom* (1809), clearly endorses both when he states that "evil ever remains man's own choice … every creature falls through his own guilt. But just how the decision for good or evil comes to pass in the individual, that is still wrapped in total darkness." See F. W. J. Schelling, *Of Human Freedom*, trans. James Gutmann (Chicago: Open Court Publishing Co., 1936), p. 59, and Bernstein, *Radical Evil*, p. 93.

referring to an allegedly new kind of non-Kantian radical evil – one that "breaks down all standards we know," cannot be explained "by comprehensible motives," and occurs within totalitarian regimes "in which all men have become equally superfluous" – she also states that this new type of radical evil can "no longer be understood and explained by the evil motives of self-interest, greed, covetousness, resentment, lust for power, and cowardice."[15] Shortly after her book appeared, Arendt also wrote, in a letter to her former teacher Karl Jaspers (and to whom she had sent one of the first copies of her book), that

> the Western tradition is suffering from the preconception that the most evil things human beings can do arise from the vice of selfishness. Yet we know that the greatest evils or radical evil has nothing to do anymore with such humanly understandable, sinful motives. What radical evil really is I don't know, but it seems to me it somehow has to do with the following phenomenon: making human beings as human beings superfluous.[16]

Bernstein uses this quotation to support his claim that one of Arendt's "most characteristic thought-trains" is the view that "the most evil deeds that human beings perform do not arise from the vice of

[15] H. Arendt, *The Origins of Totalitarianism*, new edition with added Prefaces (San Diego: Harcourt, 1994), p. 459; cf. pp. viii–ix. Arendt briefly refers to Kant's concept of radical evil on this same page, but quickly dismisses it on the ground that Kant mistakenly thought that evil "could be explained by comprehensible motives." Kantian radical evil, as is well known, does not refer to an allegedly new type of evil never before witnessed by humanity, but rather to a universal propensity within the human species, one that is "in all cases somehow entwined with humanity and, as it were, rooted in it" (R 6: 32).

[16] Arendt to Jaspers, March 4, 1951. H. Arendt and K. Jaspers *Correspondence, 1926–1969*, trans. L. Kohler and H. Saner (New York: Harcourt Brace Jovanovich, 1992), p. 166. See Bernstein, *Radical Evil*, p. 207. Arendt includes a second anti-Kantian argument in this same letter to Jaspers, when she goes on to state that making people superfluous as human beings is different from using them "as a means to an end" (p. 166). Here the proper Kantian response, I believe, is to acknowledge that while not all cases of treating people as means to an end are also cases of making people superfluous, all cases of making people superfluous are cases of treating people as means to an end. Treating people as means to an end is a broader category than the category of making people superfluous, but the latter is simply the extreme limit of the former. A professor who requires that students read one of his new essays is treating them as a means to his end; but in doing so he is not making his students superfluous. However, the Nazis' treatment of the Jews during World War II was clearly a case both of treating them as means to an end as well as making them superfluous.

selfishness,"[17] a thought-train that he himself endorses and then uses to make a further criticism of Kant's account of radical evil. Kant is wrong, Bernstein argues, in holding that evil always arises from selfishness. Rather, Bernstein notes, we should

> recognize that there are other incentives that are not easily assimilated to "self-love." It is difficult to see how the incentives that motivate fanatics and terrorists who are willing to sacrifice themselves for some cause or movement can be accounted for by self-love. The horrors of the twentieth century (and not just this century) have opened our eyes to the variety of types of incentives that motivate evil actions.[18]

The self-love criticism of Kant's account of radical evil would seem to have common sense on its side. As I myself remarked in an earlier discussion, "people are evil for many different reasons," and "the possibilities for evil are infinite."[19] Why reduce all of these reasons to self-love? But I think now that Kant's claim that self-love is "the source of all evil," is not so easy to dismiss, once it is placed within the proper context of his moral theory.

First of all, "self-love" in Kant's sense – Arendt and Bernstein to the contrary – is not synonymous with what is normally meant by "selfishness." Kantian self-love is a broader motivational tendency that encompasses a wide variety of desires and inclinations, many of which themselves can be and are used to promote decidedly non-selfish purposes. As Andrews Reath notes:

> Self-love is a concern for well being which modifies an inclination only when it conflicts with one's overall happiness. It is opposed to the moral disposition,

[17] Bernstein, *Radical Evil*, pp. 207–8.

[18] Ibid., p. 42. Cf. Allison, who (in commenting on an argument of John Silber's), writes: "Great evil, it would seem, can involve as much self-sacrifice (at least as it is usually conceived) and intensification of personality as great virtue" (Allison, "Reflections on the Banality of (Radical) Evil," p. 176).

[19] R. B. Louden, "'On the Radical Evil in Human Nature'," in *Kant's Impure Ethics: From Rational Beings to Human Beings* (New York: Oxford University Press, 2000), pp. 132–9, p. 139. One passage from the *Religion* that I used to support my position was Kant's claim that "whenever incentives other than the law itself (e.g. ambition, self-love in general [*Selbstliebe überhaupt*], yes, even a kindly instinct such as sympathy)" (R 6: 30–1) determine our actions, such actions are evil. I took Kant here to be distinguishing self-love from other non-moral incentives such as sympathy and ambition. But I think now that this interpretation is wrong, or rather (since the particular passage quoted still seems to me to support this interpretation), that the normal way of reading this passage does not in fact represent Kant's considered view.

not due to the inclinations involved, but because it recognizes no moral restrictions. The inclinations may be good in that they can ground morally permissible ends, when properly limited. But in recognizing no moral restrictions, self-love makes the moral law a subordinate principle.[20]

The main problem with self-love, according to Kant, is simply that it does not recognize the supremacy of the moral law. Whenever we act from a maxim of self-love, we have freely chosen to subordinate the incentives of morality to those of inclination. In some cases (e.g., with people who are "so sympathetically attuned that without any other motive of vanity [*Eitelkeit*] or self-interest [*Eigennutz*] they find an inner satisfaction in spreading joy around them" [G 4: 398]), the purpose of the act may even be to *help other people*. Kant's naturally kind-hearted person is not selfish, and Bernstein's fanatics and terrorists who are "willing to sacrifice themselves for some cause or movement" probably are not either. ("Probably not" because, again, human motives are ultimately inscrutable. Some fanatics and terrorists do seem to be selfish, but we can certainly imagine others who are not.) However, neither Kant's naturally kind-hearted persons nor Bernstein's hypothetical fanatics and terrorists, despite their non-selfish motivations, are acting from moral maxims; that is, from maxims derivable from the categorical imperative. For instance, they all clearly violate the first formula of the categorical imperative: "*act only in accordance with that maxim through which you can at the same time will that it become a universal law*" (G 4: 421). Their maxims are not universalizable, and because they have subordinated the moral law to their own non-moral inclinations they are all acting from self-love in Kant's sense of the term.

One very common type of evil person in Kant's sense (and for him the term "evil" has extremely wide scope: there are *a lot* of evil people in the world, only some of whom count as evil according to non-Kantian theories of evil) is simply someone who "makes the incentives of self-love and their inclinations [*die Triebfeder der Selbstliebe und ihre Neigungen*] the condition of compliance with the moral law" (R 6: 36). Such a person says, in effect: "I will do what I desire and what

[20] A. Reath, "Kant's Theory of Moral Sensibility," in *Agency and Autonomy in Kant's Moral Theory* (Oxford: Clarendon Press, 2006), pp. 8–32, p. 16.

is morally required, as long as the moral law doesn't conflict with my self-love." The morally good person, on the other hand, says: "I will do what is morally required and what I desire, so long as my desires don't conflict with the moral law."[21]

When the self-love criticism is placed within the larger context of Kant's moral theory, the claim that self-love "is precisely the source of all evil" loses much of its counterintuitive air. What it amounts to is simply the claim that morally good people put the moral law first, while the rest of us don't. This claim, I submit, is true, even if it does not count as one of the most profound statements of moral psychology to come from a philosopher's pen. Admittedly, Kant here "has not dared to descend into the depths,"[22] but this was not his intent. For again, his account of radical evil is primarily a theory about *what* evil is (and *how* we should respond to it) – not a theory about *why* people do evil. However, given the indecipherable character of much human action, perhaps it is best not to speak presumptuously about why people commit evil. Those who think they have succeeded in descending into the depths here are often mistaken. Unfortunately, it is very difficult to ever reach bottom in this particular line of work, for the depths of human evil are unfathomable.

Diabolical Evil and Moral Responsibility

Another area within Kant's analysis of radical evil where the issue of motives comes up secondarily concerns his claim that the concept of "a *diabolical* being [*ein teuflischer*[23] *Wesen*]" is not "applicable to the

[21] I have borrowed this contrast from C. M. Korsgaard, "Morality as Freedom," in *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1996), pp. 159–87, p. 165. See also A. W. Wood, *Kant's Moral Religion* (Ithaca, NY: Cornell University Press, 1970), p. 213.

[22] Friedrich Nietzsche, *Beyond Good and Evil: Prelude to a Philosophy of the Future* (1886), trans. Walter Kaufmann (New York: Vintage Books, 1966), § 23.

[23] T. M. Grene and H. H. Hudson, in their English translation of Kant's *Religion* (La Salle: Open Court, 1934; 2nd edn, New York: Harper & Row, 1960) render "*teuflisch*" as "devilish," and earlier discussions of this topic (e.g., J. R. Silber, "The Ethical Significance of Kant's *Religion*," in *Religion within the Limits of Reason Alone*, trans. Grene and. Hudson, pp. lxxix–cxxxiv; Anderson-Gold, "Kant's Rejection of Devilishness") follow their lead. However, partly because "devilish" is an ambiguous term, one of whose meanings is "mischievous, teasing, or annoying" (a meaning

human being" (R 6: 35). Here too, he has been criticized by many; here too, the central criticism is that his moral psychology is shallow and naïve, and does not adequately reflect the true depths of human depravity.

John Silber has developed the best-known version of the *diabolical evil criticism*. In "The Ethical Significance of Kant's *Religion*" (1960), he writes:

> in dismissing the devilish rejection of the law as an illusion, Kant called attention to the limitations of his conception of freedom rather than to the limits of human freedom itself . . . Kant's insistence to the contrary, man's free power to reject the law in defiance is an ineradicable fact of human experience. [24]

And in his subsequent essay, "Kant at Auschwitz" (1991), he reiterates the criticism in the following remark:

> Kant's ethics is inadequate to the understanding of Auschwitz because Kant denies the possibility of the deliberate rejection of the moral law. Not even a wicked man, Kant holds, can will evil for the sake of evil. His evil, according to Kant, consists merely in his willingness to ignore or subordinate the moral law when it interferes with his nonmoral but natural inclinations. His evil is expressed in abandoning the conditions of free personal fulfillment in favor of fulfillment as a creature of natural desire.[25]

Similarly, Bernstein, after citing Silber's criticisms, concludes that Kant's "analysis of evil and radical evil is disappointing," that one of the primary reasons it is disappointing stems from a failure to consider that there do exist people who "incorporate into their maxims the primary incentives to *defy* the moral law," and that this failure ultimately "is rooted in Kant's limited moral psychology, in the narrow range of types of incentives that he acknowledges."[26] Finally, Claudia Card has

occasionally rendered in cartoons and Valentine's Day cards), I don't think it takes us very close to what Kant is talking about. "Diabolical" is a better translation for Kant's "*teuflisch*." But even here, articulating what exactly *teuflisch*/diabolical in Kant's specific sense means is no easy matter.

[24] Silber, "Ethical Significance of Kant's *Religion*," p. cxxix.

[25] J. R. Silber, "Kant at Auschwitz," in G. Funke and T. M. Seebohm (eds.), *Proceedings of the Sixth International Kant Congress* (Washington, D.C.: Center for Advanced Research in Phenomenology and University Press of America, 1991), pp. 177–211, p. 198. Cf. p. 194.

[26] Bernstein, *Radical Evil*, pp. 36, 41, 42.

also argued recently that Kant was "wrong in his insistence that evil is never 'diabolical' in human beings (that we never do wrong for its own sake)," and that "diabolical evil in human beings is very real."[27]

In order to assess the diabolical evil criticism, we need first to get an accurate sense both of what Kant means by "diabolical evil" and why he rejects its possibility in human beings. I believe that once these preliminary tasks have been accomplished, some versions of the diabolical criticism can be dismissed, simply because they are off-target.

As is well known, Kant divides the human propensity to evil into three different degrees or steps (*Stufen*) (R 6: 29). The lowest or most common degree is *frailty* (*Gebrechlichkeit*, *fragilitas*), and corresponds roughly to what is traditionally meant by weakness of will. Agents at this first level of evil intend to act from moral motives and often succeed in doing so, but sometimes at the last moment they weaken their resolve and act from non-moral motives. The second degree of evil is *impurity* (*Unlauterkeit*, *impuritas*, *improbitas*), and is basically a case of acting from mixed motives. Here agents want to do the morally right thing and to do so from morally right motives, but the presence of moral motives alone is often not sufficient to get them to do the right thing. As a result, they frequently need to turn to non-moral motives in order to do the right thing (e. g., help others not from the motive of duty but out of sympathetic feeling). Because they need non-moral motives in order to get themselves to do the right thing, their motives are "not purely moral [*nicht rein moralisch*]" (R 6: 30). Finally, the third and most severe degree of evil, is *wickedness* or *depravity* (*Bösartigkeit*, *vitiositas*, *pravitas*). Agents at this level deliberately and consistently act on non-moral maxims in virtually all of their behavior – they simply do not want to act from moral maxims at all. As a result, they represent

[27] Card, *The Atrocity Paradigm*, pp. 91, 212. See also M. B. Matuštík, who complains that Kant "rejects without explanation" the possibility that human beings can be diabolical beings ("Violence and Secularization, Evil and Redemption," in Schrift [ed.], *Modernity and the Problem of Evil*, pp. 39–50, p. 41). Card goes on to develop and defend a non-Kantian conception of diabolical evil, one that "focuses on the *harm* one is willing to inflict rather than on the reasons why" (p. 211). She also holds that her "understanding of diabolical evil comes closer than Kant's does to the classic view of Satan as a corrupter, as one who tempts others to abandon morality or demote it to a low position on their scale of values" (p. 212). However, as I argue below, this notion of "Satan as a corrupter" is not what Kant's argument against diabolical evil is about.

"the *perversity* [*Verkehrtheit, perversitas*] of the human heart," and their mental attitude is "corrupted at its root" (R 6: 30).

We can see already – contra some versions of the diabolical evil criticism – that Kant by no means denies that some people do "reject the moral law in defiance"[28] and do "deliberately and consistently reject the moral law."[29] This is precisely what the third *Stufe* of evil is all about. People at this third level do openly, directly, regularly, and intentionally reject the moral law, and this is why they are wicked and corrupt.

However, it remains the case that Kant does assert that the concept "diabolical being" is not "applicable [*anwendbar*] to the human being" (R 6: 35). If, as I argued above, he does not simply mean here that human beings are unable to openly and defiantly rebel against the moral law, then what exactly does he mean?

The answer is not terribly complicated, and is located in a simple contrast Kant draws between a diabolical being or "an absolutely evil will [*ein schlechthin böser Wille*]" and "a purely *animal* being [*ein blos tierischer Wesen*]" (R 6: 35) on the one hand, and a human being on the other. Something important is missing in diabolical and animal beings – viz., they lack a moral personality that enables us to hold them accountable for what they do. In both cases, they do not make free choices and hence cannot be held legally or morally accountable for their behavior. Human beings, on the other hand – "even the worst [*selbst der ärgste*]" (R 6: 36) – do possess this capacity, and so can be held accountable for their actions.[30]

Here too (cf. my earlier discussion of the self-love criticism), it is important to keep in mind that Kant's primary goal is not to offer a detailed account of why people do evil. His discussion takes place on entirely different level. As Henry E. Allison notes: "Kant's denial of a diabolical will is not a dubious piece of empirical moral psychology, but rather an *a priori* claim about the conditions of the possibility of moral accountability."[31]

[28] Silber, "Ethical Significance of Kant's *Religion*," p. cxxix.

[29] Bernstein, *Radical Evil*, p. 40.

[30] For further discussion to which I am indebted, see Wood, *Kant's Moral Religion*, pp. 212–14; and Allison, "Reflections on the Banality of (Radical) Evil," pp. 174–7.

[31] Allison, "Reflections on the Banality of (Radical) Evil," p. 176.

Part of the point of Kant's (admittedly somewhat counterintuitive) diatribe against the possibility of diabolical evil in human beings, I submit, is that he does not want us to romanticize evil. We must not be seduced by *schwärmerische* historians, novelists, movie directors, *et al.* into thinking that some human beings, in virtue of their "strong" and "potent" personalities,[32] are somehow above the rest of us, and are not to be judged by the same laws and principles that apply to ordinary human beings. Karl Jaspers, in a letter to his former student Hannah Arendt, put the point well when he criticized an early version of her notion of an allegedly new type of radical evil that "breaks down all standards we know" and "oversteps and shatters any and all legal systems":

> You say that what the Nazis did cannot be comprehended as "crime" – I'm not altogether comfortable with your view, because a guilt that goes beyond all criminal guilt inevitably takes on a streak of "greatness" – of satanic greatness – which is, for me, as inappropriate for the Nazis as all the talk about the "demonic" element in Hitler and so forth. It seems to me that we have to see these things in their total banality,[33] in their prosaic triviality, because that's what truly characterizes them. Bacteria can cause epidemics that wipe out nations, but they remain merely bacteria. I regard any hint of myth and legend with horror, and everything unspecific is just such a hint … The way you do express it, you've almost taken the path of poetry. And a Shakespeare would never be able to give adequate form to this material – his instinctive aesthetic sense would lead to falsification of it – and that's why he couldn't attempt it.[34]

In short, we must resist the temptation to aestheticize evil. This is one reason why Kant rejects the strategy of attributing diabolical or demonic motives to human beings who commit evil. But his central point is simply to underscore the necessity of legal and moral

[32] See Silber's discussion of Hitler, Napoleon, and Herman Melville's Ahab in "Ethical Significance of Kant's *Religion*," p. cxxix.

[33] Kohler and Saner (eds.), *Hannah Arendt/Karl Jaspers Correspondence*, insert an endnote here that reads: "This passage may have influenced the subtitle of A.'s *Eichmann in Jerusalem: A Report on the Banality of Evil*" (p. 702 n.6).

[34] Arendt, *The Origins of Totalitarianism*, p. 459; Arendt to Jaspers, July 9, 1946, in Kohler and Saner (eds.), *Hannah Arendt/Karl Jaspers Correspondence*, p. 54; Jaspers to Arendt, October 19, 1946, Kohler and Saner (eds.), *Hannah Arendt/Karl Jaspers Correspondence*, p. 62. See Bernstein, *Radical Evil*, pp. 214–15.

responsibility when talking about human evil. He does not want to let perpetrators of evil – particularly the most extreme perpetrators of evil – off the hook. As long as they are aware of what they are doing and freely chose to do it, they should be held responsible for their conduct. Even the most wicked and depraved individuals are still rational beings who understand morality and the law, and because they possess this understanding they must held accountable for their deviations from morality and the law. They have the capacity to recognize the criminality and immorality of their acts. To label someone a diabolical being in the sense described by Kant is to grant him or her the status of a being who is no longer legally and morally accountable. And no sane human being should be granted this status.

Anthropology and the Universality of Evil

The last criticism of Kant's account of radical evil that I wish to examine concerns his curious claim that this evil is both "innate [*angeboren*]" throughout the entire human species (R 6: 32; cf. 25, 29, 38, 43, 50) and yet freely chosen by each individual ("brought upon us by ourselves [*uns von uns selbst zugezogen*]" [R 6: 32]), and his even more curious effort to convince readers of the truth of this paradoxical claim by appealing to experience (its truth is allegedly evident "from what one knows of the human being through experience [*durch Erfahrung*]" [R 6: 32]; it "can be established through experiential demonstrations [*durch Erfahrungsbeweise*]." [R6: 35]) How can something be innate and yet freely chosen, and how can the claim that a propensity present in every human being, past, present, and future – "even the best" (R 6: 32) – be established by appealing to experience? For as every student who has read at least the first paragraph of the Introduction to the *Critique of Pure Reason* knows, experience "gives us no true universality [*keine wahre Allgemeinheit*], and reason, which is so desirous of this kind of cognitions, is more stimulated than satisfied by it" (KrV A1). Empirical universality is not true or strict universality (*strenge Allgemeinheit*) but "only an arbitrary increase in validity from that which holds in most cases to that which holds in all" (KrV B4). Strict universality can never be justified by an appeal to experience; indeed, this is why Kant insists that it is one of

the two "sure signs [*sichere Kennzeichen*] of an *a priori* cognition" (KrV B4).[35]

Critics and sympathetic commentators alike have had a field day on this issue. For instance, Allison remarks sensibly that "the *most*" that an appeal to experience "can show is that evil is widespread, not that there is a universal propensity to it," and concludes that Kant's argument is "quite disappointing";[36] while Michalson, noting "the peculiarity of this appeal to experience, one which cannot possibly support the argumentative weight Kant seems to be placing on it," bemoans "the absence of genuine argumentation for this crucial point."[37] And Bernstein, after citing Allison and reminding readers that Kant himself says that "we can spare ourselves the formal proof [*uns … den förmlichen Beweis ersparen*]" (R 6: 33), simply asserts that Kant has thrown in the towel:

> When Kant reaches this crucial stage in his exposition, when we expect some sort of proof or justification of radical evil as a *universal* characteristic of human beings, *no* such proof is forthcoming … Kant never gives – or even attempts to give – a *proof* of his controversial and bold claim that man is evil by nature.[38]

Indeed, criticism of the paradoxical nature of Kant's argumentation on this particular point is by no means a recent phenomenon, and goes back at least as far as 1794. For instance, Johann August Eberhard (1739–1809), one of Kant's early rationalist critics, in his

[35] A related potential problem: because evil is freely chosen, it is not a necessary feature of the human being – as Kant remarks, we can't infer evil "from the concept of a human being in general" (R 6: 32). But this additional claim seems to violate another key doctrine of the first *Critique*; viz., that universality and necessity "belong together inseparably" (KrV B4; cf. A2). In the case of radical evil, Kant appears to be asserting that we encounter universality without necessity. However, as I argue below, the kind of universality relevant to Kant's discussion of radical evil is an empirical, species-universality (one that applies to human beings, but not necessarily to other species of rational being). It is not the "strict" universality that we find in a priori judgments. And it is only the latter kind of universality that belongs together inseparably with necessity.

[36] H. E. Allison, *Kant's Theory of Freedom* (New York: Cambridge University Press, 1990), p. 154.

[37] Michalson, *Fallen Freedom*, p. 46.

[38] Bernstein, *Radical Evil*, pp. 34–5.

essay "Über das Kantische radicale Böse in der menschlichen Natur" (On Kantian Radical Evil in Human Nature), asks:

> Now how does Herr Kant prove that such a radical evil exists, or that the human being is evil by nature? He does not carry out this proof, as would be expected, from principles of pure reason; rather, it rests merely on experience [*bloß auf die Erfahrung*]: he refers first to savage peoples, then to civilized nations, and tries to show through their well-known manner of acting that they are all afflicted with radical evil.[39]

Previous attempts to extricate Kant from these difficulties, while ingenious, have often seemed as paradoxical as the problem for which they are the alleged solution. Thus Allison sidesteps Kant's appeal to experience entirely, encouraging readers to treat Kant's claim as a synthetic a priori postulate,[40] while Robert Merrihew Adams suggests that we can solve the dilemma of a freely chosen but innate propensity by conceiving it noumenally as something that "originated in a free and voluntary act that was not in time."[41]

My own counter-strategy begins by insisting that we take seriously Kant's frequent appeals to experience and anthropology in his discussion of radical evil. These appeals ought not to be jettisoned, regardless of the philosophical troubles they seem to land him in, for the simple reason that they pervade his entire analysis of evil in the *Religion*, and thus cannot be dismissed as an unnecessary aberration. For instance, as I have noted in a previous discussion, we find either the term "human nature" (*menschliche Natur*) or "human being" (*Mensch*) not only in the title of Part One of *Religion* but in all four section titles of Part One as well.[42] In using this language, Kant makes it very clear to readers that in his discussion of radical evil he

[39] J. A. Eberhard, "Ueber das Kantische radicale Böse in der menschlichen Natur," *Philosophisches Archiv*, 2, 2 (1794), 34–47, at pp. 41–2. (I would like to thank the Universitäts- und Landesbibliothek at Westfälische Wilhelms-Universität in Münster, Germany for providing me with a photocopy of this important essay.) Cf. M. Kosch, *Freedom and Reason in Kant, Schelling, and Kierkegaard* (Oxford: Clarendon Press, 2006), p. 63 n.39.

[40] Allison, *Kant's Theory of Freedom*, p. 155. See also Bernstein's criticisms of Allison's strategy in *Radical Evil*, pp. 240–1 n.32.

[41] R. M. Adams, Introduction to Kant, in A. Wood and G. di Giovanni (eds.) *Immanuel Kant: Religion within the Boundaries of Mere Reason and other Writings* (Cambridge University Press, 1998), p. xiii.

[42] Louden, *Kant's Impure Ethics*, p. 132. See R 6: 18, 26, 28, 32, 39.

is concerned specifically and solely with human beings, and not – as is often the case in his canonical ethical theory writings – with the much larger set of rational beings in general, of which human beings constitute only a subset.

Furthermore, as noted earlier, when Kant discusses radical evil, he insists that he is concerned with a kind of moral evil that we actually encounter in our daily experience: "the existence of this propensity to evil in human nature can be established through experiential demonstrations of the actual resistance in time of human choice [*menschliche Willkür*] against the law" (R 6: 35). Indeed, as Michelle Kosch observes: "The claim that evil is given empirically (and only empirically) is reiterated throughout *Religion*."[43] For instance, at the very beginning of Part One, Kant appeals to the collective experience of humanity to support the thesis that human beings are by nature evil (R 6: 19), and he criticizes certain overly optimistic Enlightenment philosophers for failing to offer empirical support for their counterposition that human beings are by nature good – their view, he notes pointedly, has "certainly not been drawn from experience [*sicherlich nicht aus der Erfahrung geschöpft*]" (R 6: 20). And when he does attempt to convince readers that there exists a corrupt propensity to evil "entwined with humanity itself and, as it were, rooted in it" (R 6: 32) (if indeed any convincing is needed), he says that "we can spare ourselves the formal proof" and invites us to look instead at "the multitude of woeful examples that the experience [*die Erfahrung*] of human *deeds* parades before us" (R 6: 32–3) – viz., the horrendous acts of evil that people continually commit against one another in "the so-called *state of nature*,"[44] in the "civilized state" (R 6: 33) (which turns out to be not so civilized), and last but not least in the international arena, where nations (then as well as now) remain in "a state of constant readiness for war (*ein Stand der beständigen Kriegsverfassung*)" (R 6: 34).

[43] Kosch, *Freedom and Reason*, p. 63. Cf. Louden, *Kant's Impure Ethics*, pp. 132–3.

[44] Bernstein holds that Kant's examples here are merely "evidence of his prejudices, based upon limited and highly selective anthropological sources" (*Radical Evil*, p. 240 n.31). That Kant harbored many prejudices I do not deny. However, I think the main point of his brief litany of examples of evil found within "the so-called *state of nature*" is to signal disagreement with Rousseauian romantics who think that we somehow find innocent and uncorrupted human beings outside of Europe. Kant's point is that "they're people too." And because they are human beings, they too have a propensity to evil.

By 1793, the year *Religion* was published, Kant had been teaching an annual lecture course in anthropology for over twenty years. Described early on as an "empirical study" or observation-based doctrine (*Beobachtungslehre*) (Kant to Marcus Herz, toward the end of 1773, Ak. 10: 146) in which "the grounds of cognition are taken from observation and experience [*Beobachtung und Erfahrung*]" (VA 25: 7), Kantian anthropology, as is well known, is also designated by its creator as a *pragmatic* anthropology. Pragmatic anthropology in turn is distinguished from the *physiological* anthropology that the German physician Ernst Platner (1744–1818) and others were already advocating when Kant first began lecturing on anthropology in 1772. In the Preface to *Anthropology from a Pragmatic Point of View* (1798) – essentially his last set of lecture notes for the course that he taught annually for twenty-four years until his retirement from university lecturing in 1796 – Kant distinguishes pragmatic anthropology from physiological anthropology as follows:

> A doctrine of knowledge of the human being, systematically formulated (anthropology), can exist either in a physiological or in a pragmatic point of view. Physiological knowledge of the human being concerns the investigation of what *nature* makes of the human being; pragmatic, the investigation of what *he* as a free-acting being makes of himself, or can and should make of himself. (VA 7: 119)[45]

However, in order to adhere to its self-imposed constraint of being a *Beobachtungslehre*, pragmatic anthropology's investigation of the human being as a free-acting being must be conducted empirically, not transcendentally. Pragmatic anthropology studies the phenomenal effects of human freedom in the empirical world, not freedom's allegedly noumenal origins.

Also central to Kantian pragmatic anthropology is its emphasis on a particular understanding of human nature, which I call a *cosmopolitan* conception of human nature.[46] For instance, in the Preface to

[45] I have recently prepared a new English translation of this text: Immanuel Kant, *Anthropology from a Pragmatic Point of View*, trans. and ed. R. B. Louden, Cambridge Texts in the History of Philosophy (Cambridge University Press, 2006).

[46] For further discussion, see my "Anthropology from a Kantian Point of View: Toward a Cosmopolitan Conception of Human Nature," *Studies in History and Philosophy of Science A*, 39 (December 2008), 515–22.

Anthropology from a Pragmatic Point of View, Kant states that anthropology is only properly called *pragmatic* "when it contains knowledge of the human being as a *citizen of the world* [*Erkenntnis des Menschen als Weltbürgers*]" (VA 7: 120). Similarly in the Preamble to the *Friedländer* anthropology transcription (1775–6), he states:

> [A]nthropology is not however a local but rather a general anthropology. In it one comes to know not the state of human beings but rather the nature of humanity, for the local properties of human beings always change, but the nature of humanity does not. Anthropology is thus a pragmatic knowledge of what results from our nature, but it is not a physical or geographical knowledge, for that is tied to time and place, and is not constant … Anthropology is not a description of human beings, but of human nature. (VA 25: 471)

Essentially, what Kant strives for in his anthropology is a wide conception of human nature that is not tied to time and place, one that focuses on what human beings share in common with one another. Again though, because Kantian anthropology is a *Beobachtungslehre*, this conception of human nature must be arrived at empirically, through collective reflection on the chief tendencies and characteristics of the species as a whole. It is not an a priori cognition but rather a posteriori: something "known from the experience of all ages and by all peoples [*aus der Erfahrung aller Zeiten und unter allen Völkern*]" (VA 7: 331).

But while empirical, the cosmopolitan conception of human nature also has an important normative status within Kant's anthropology. In effect, it functions as a teleological moral map: a practical guide by means of which human beings are to orient themselves toward both the present and the future. For instance, in the final sentence of his *Anthropology*, Kant summarizes the human species's character as "a species of rational beings that strives among obstacles to rise out of evil," but he then adds that we can only be expected to reach the goal (*Zweck*) "by a progressive organization of citizens of the earth into and toward the species as a system that is cosmopolitically united" (VA 7: 333).

Kant's discussion of radical evil in *Religion*, I submit, fits well with his extended anthropological investigations into human nature. In both his anthropology lectures and in Part One of *Religion*, he is concerned with what our experience tells us about the human species as

a whole – with what human beings past, present, and future share in common with each other. But there is also a fundamental difference between the anthropology and *Religion* discussions. In the various anthropology lectures (the following point also holds for Kant's writings on the philosophy of history), the discussion of human nature focuses primarily on the *future* – on humanity's cosmopolitical vocation and the gradual realization of a global society that administers justice universally. However, in Part One of *Religion*, the primary focus is more on the *past*.[47] The propensity to evil "is detectable as the first manifestation of the exercise of freedom in the human being" (R 6: 38); our ancestors saw it "at the beginning of the world [*im Weltanfang*]" (R 6: 43).

In sum, Kant is serious when he encourages us to examine the woeful examples of radical evil "that the experience of human *deeds* parades before us." Empirical "anthropological inquiry [*anthropologische Nachforschung*]" (R 6: 25) of this sort is as close as we are going to get to a formal proof of the universality of a propensity to evil within the human species, and it is no doubt closer than many of us want to get. Throughout history and in every culture, human beings have continually revealed their propensity to evil in their conduct toward one another. Evil is truly everywhere.

Kant's account of radical evil, like many other aspects of his philosophy, is certainly not without its paradoxes and counterintuitive claims. But I hope I have succeeded in showing that four of the most frequently voiced criticisms of this account are often wide of the mark, and that when we take the time to determine what Kant is trying to say, his own position still makes a lot of sense. To summarize my responses to the four criticisms: (1) Even the most ordinary people are capable of the most horrendous deeds, but in many cases we will never know for sure what drives people to evil. Human action, particularly evil human action, often has an indecipherable character, and because of

[47] A future-orientation is nevertheless occasionally detectable in Kant's later discussion of grace, a concept that he regards as "very risky" and "hard to reconcile with reason" (R 6: 191), but which he is also willing to admit "as something incomprehensible" (R 6: 53). See also Part Three of *Religion*, "The Victory of the Good Principle over the Evil Principle, and the Founding of a Kingdom of God on Earth" (R 6: 93–147). This part of *Religion* is closer to the teleological orientation of the later sections of the anthropology lectures and the philosophy of history essays.

this it is unrealistic to demand that a plausible theory of evil explain why people commit evil. In order to hold rational agents responsible for their conduct, we must assume that they have the capacity of free choice. But this very capacity of free choice itself implies that in many cases the ultimate motives of human conduct will be inscrutable. (2) Morally good people put moral principle first, and the rest of us don't. Beyond this, it is unwise to speculate about whether we have succeeded in plumbing the depths of human depravity. When confronted with cases of evil, all that we can safely say is that the perpetrator has knowingly violated fundamental moral norms. (3) We must resist the temptation to aestheticize evil by attributing motives of satanic evil to all-too-human criminals. To label perpetrators of evil demonic or diabolical is merely to offer them an escape route from responsibility for their deeds. (4) Finally, as human beings whose intentions are opaque, we have no choice but to try and extract invisible moral dispositions from the visible deeds before us (cf. R 6: 71, 77). And when we do look carefully at human deeds in different times and places, we find ample evidence of a universal human propensity to evil.[48]

[48] I would like to thank the following individuals, groups, and institutions for their help and support: Richard J. Bernstein, for his comments on an earlier version of the essay; the Idealism Group and the Institute of Philosophy and History of Ideas, Aarhus University, Denmark, for their invitation to present the essay to them and for their warm hospitality during my visit in February 2008; students in a new course (Problems in Philosophy: Evil) that I taught at the University of Southern Maine in spring 2007, for exploring some of the essay's Kantian themes with me; and the Alexander von Humboldt Foundation, for its financial support in June–July of 2007, during which time a draft of the essay was written in Münster, Germany.

6

An Alternative Proof of the Universal Propensity to Evil

Pablo Muchnik

I want to address a vexed question in this essay: does Kant really need a transcendental deduction to justify his claim "man is evil by nature"? Transcendental deductions, Kant is the first to admit it, are notoriously difficult.[1] In the case of the *Religion within the Boundaries of Mere Reason*, whose transcendental argument (if there is one) must be assembled through careful detective work, the difficulty is clearly compounded. I take up the gauntlet here because much of the current debate on this question is fueled, I suspect, by an insufficient grasp of the systematic character of Kant's doctrine of radical evil. Triggered by Kant's own lack of expository clarity at crucial passages, interpreters have tended to conflate the different notions of an "evil disposition" (*böse Gesinnung*) and a "propensity to evil" (*Hang zum Bösen*).[2] A reader of the acuity of Henry E. Allison, for instance, says:

> [T]he distinctive features of the Kantian conception of *Gesinnung* are that it is acquired, although not in time, and that it consists in the fundamental or controlling maxim, which determines the orientation of one's *Willkür* as a

[1] Kant acknowledges this in KrV Axvi.

[2] Since the notions of "*Gesinnung*" and "propensity to evil" share basic formal features (both are the result of imputable and intelligible acts (*Taten*), logically prior to any deed in time, which nonetheless determine the subjective use of freedom in general and have the status of maxims), Kant is at times careless in distinguishing the different scope these concepts have in his own doctrine. See, e.g., R 6: 31–2, 35, and 36.

moral being. Given this, we can now see that this *Gesinnung* is precisely what Kant means by a moral propensity.[3]

But surely this cannot be Kant's considered view. For he cannot possibly mean that the individual's choice of *Gesinnung* is equivalent to the species's choice of propensity.[4] Otherwise, our personal wrongdoing would be explicated (and exculpated) by sheer membership in humanity.[5] This untoward conclusion, however, can be averted once we realize that the notions in question refer to two different units of moral analysis: the *Gesinnung* indicates the fundamental moral outlook of an *individual agent*; the propensity, the moral character imputable to the *whole human species*. Overlooking the logical independence of these analytic units gives the impression that Kant's talk of a universal propensity to evil is inconsistent with his commitment to freedom. For if we consider "*Gesinnung*" and "propensity" to be synonymous, it seems natural to suppose that the choice at the level of the species carries causal efficacy at the level of the individual, and is hence at odds with our autonomy.[6]

The analytic distinction we propose saves Kant from this blunder. Moreover, it suggests a way to justify, on Kant's behalf, the a priori (necessary and universal) character of the attribution of a propensity

[3] See H. E. Allison, *Kant's Theory of Freedom* (Cambridge University Press, 1990), p. 153. Other interpreters have reached a similar conclusion. Daniel O'Connor, for instance, reads Kant's indictment "man is evil by nature" as saying: "all men have, in fact, chosen an evil disposition." See D. O'Connor, "Good and Evil Disposition," *Kant-Studien*, 3 (1985), 288–302, p. 296. Mark Timmons is even more extreme, for he identifies radical evil with lack of moral worth: "[A]n evil disposition, an evil will, a character that lacks moral worth and one who is possessed of radical evil are one and the same." See M. Timmons, "Evil and Imputation in Kant's Ethics," *Jahrbuch für Recht und Ethik*, 2 (1994), 114–44, p. 134.

[4] Kant makes this clear, for instance, in R 6: 25: "by the 'human being' of whom we say that he is good or evil by nature we are entitled to understand not individuals (for otherwise one human being could be assumed to be good, and another evil, by nature) but the whole species …" He makes a similar point at R 6: 29.

[5] The implication of this reading is that Kant's doctrine of radical evil is but a poorly disguised piece of Christian dogmatism, as Goethe himself concluded in a letter to Herder (June 7, 1793). Cf. E. Fackenheim, "Kant and Radical Evil," *University of Toronto Quarterly*, 23 (1954), 339–53, p. 340. Kant's clearest rebuttal appears in R 6: 40.

[6] Richard Bernstein makes this point. See R. J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press, 2002), p. 33.

to evil to human beings. For if an evil *Gesinnung* represents the failure to realize the *good* (i.e., to give primacy to the categorical imperative in one's volitional orientation), the propensity to evil represents the failure to realize the *highest good*. This latter expresses the fundamental obstacle the species faces in its duty to integrate, as a result of its collective effort, the purposiveness of nature and the purposiveness of freedom. At the basis of both doctrines Kant finds the same natural dialectic, and this fact calls for a philosophical justification centered on their structural affinity and complementarity in Kantian ethics.

To begin this argument one must disentangle the different types of moral failure covered by the single appellative "evil," and then systematically connect these failures with the different units of analysis Kant uses in Part I of the *Religion*. This move yields a surprising result: Kant's proof of the propensity to evil lies where no one expects to find it, namely, in the Preface to the first edition. But the coveted proof will disappoint the purists, for it falls short of the strict demonstrative standards of the first *Critique*. There is no denying it: the "transcendental" argument Kant advances in the *Religion* incorporates elements of his moral psychology arrived at by experience – and it is thus unabashedly impure. Yet, Kant's proof goes a long way to justify the subjective necessity, universality, and a priori character of the propensity to evil – features which would be lost from sight without it. Furthermore, the *hybrid* nature of this argument is in line with the general thrust of the *Religion*, a book whose moral anthropology has also a quasi-transcendental ring, neither reducible to empirical observation nor totally severed from it.

A Brief Genealogy of Evil's Radicalism

To appreciate the role the notions of *Gesinnung* and propensity play in Kantian ethics, it is useful to trace them back to their source in the *Groundwork*.[7] The "radicalism" of evil results from Kant's extending two assumptions from the earlier work:

[7] The question about the continuity or lack thereof in Kant's view has triggered controversy from the very beginning. For a different take on the genesis of Kant's doctrine, see M. Kosch, *Freedom and Reason in Kant, Schelling, and Kierkegaard* (Oxford: Clarendon Press, 2006), p. 46ff.

(a) Kant *radicalizes* his doctrine of transcendental freedom to comprise now the *choice* of the principle of maxim-selection. By "radicalization" I mean the extension of Kant's notion of freedom, which in the *Groundwork* was circumscribed to the choice of maxims of action (first-order maxims), to the choice of the principle of maxim-selection (second-order maxim). This extension implies a reflection of freedom upon itself, for it expresses the agent's decision about how she will use her freedom in general. This act of reflection is imputable; by its means, the agent constitutes her moral character by *choosing a rule for choosing*. To refer to this act, Kant develops the notion of *Gesinnung*, i.e., the agent's "first (*erste*) subjective ground of the adoption of maxims" (R 6: 25). An agent's *Gesinnung* is good or evil according to the principle of maxim-selection she has chosen, i.e. according to the fundamental *deliberative* tendency expressed in her second-order maxim.[8] Thus, this notion roughly matches what, in the *Groundwork*, Kant would have called a *good* or an *evil will (Wille)*. Since the conceptual apparatus is already in place, the deduction of an evil *Gesinnung* is (relatively) unproblematic for Kant in the *Religion*, as his two-step inference at the beginning of Book I indicates.[9] What is truly innovative is the host of problems the *Gesinnung* introduces in terms of the justification of maxims and the moral constitution of agents.[10]

[8] Cf. Allison, *Kant's Theory of Freedom*, p. 136ff.

[9] See R 6: 20: "In order to call a human being evil, it must be possible to infer *a priori* from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and from this, the presence in the subject of a common ground, itself a maxim, of all particularly evil maxims." Given the unknowability of the *Gesinnung* (even my own) (cf. MS 6: 392; G 4: 407), it is essential for Kant to start with wrong actions (*gesetzwidrige*) for the two-step inference to succeed. Patrick Frierson rightfully insists on this point. See P. R. Frierson, *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge University Press, 2003), pp. 105–7.

[10] In a nutshell, while the reasons for action can be justified by maxims under the categorical imperative, the choice of the first form of one's will (to which Kant also attributes the status of a maxim) cannot appeal to higher reasons to justify itself. Since the *Gesinnung* (supreme maxim) works as the ultimate (*erste*) ground of the exercise of freedom (R 6: 21), and provides an end to the series of justifying reasons, it must itself be groundless or give in to an infinite regress. Moreover, the process of moral self-constitution contains a lurking problem. Despite Kant's assertions to the contrary, the choice of *Gesinnung* cannot function as a "first" ground. Since the *Gesinnung* constitutes the moral character of the agent, but is itself an act of freedom, it presupposes an already existing will to make that choice possible. There is, so to speak, a transcendental pre-history of the *Gesinnung* which Kant does not fully acknowledge. I deal with these problems in *An Essay on Kant's Theory of Evil: An Essay*

The extension of transcendental freedom to the choice of a superintending meta-maxim points at *one* of the senses of evil's radicalism. With the adjective "radical" Kant does not intend to express the degree of immorality or the intensity of harm an agent produces with her actions, but the location of its source at the level of the individual's *Gesinnung* – the invisible root of evil, not its visible branches. "Radical," in this sense, refers neither to the quality nor to the effects of actions, but to the *locus* of evil; it is a *spatial* metaphor, not one of intensity or magnitude.[11] This distinguishes Kant's account from any type of consequentialism.

(b) Kant *naturalizes* the principles of his moral psychology and develops the idea of a universal propensity to evil. By "naturalization" I do not mean, as is usual in contemporary discussions, the reduction of ethical phenomena to their ultimate biological determination. Such a use would not be strictly Kantian – at least if we take seriously his mature, incompatibilist moral philosophy (dominant in the *Religion*, where "human nature" is itself *chosen* as an act of transcendental freedom). Instead, by "naturalization" I designate Kant's extension of the psychological conflict between competing incentives, which in the *Groundwork* characterized the individual's subjective use of freedom, to the subjective use of freedom one can attribute to the species as a whole. This extension leads Kant to elaborate a more comprehensive sense of agency, i.e., one that refers to the whole class of finite rational beings, instead of to any one of its particular members.

Needless to say, the species is for Kant an "agent" only *figuratively*, in a regulative sense: it is an idea of reason useful to reflect on global historical phenomena and their relation to man's moral destiny.[12]

on the Dangers of Self-Love and the Aprioricity of History (Lanham, MD: Lexington Books, forthcoming), chapter 3.

[11] Allison and Wood make similar points. Cf. Allison, *Kant's Theory of Freedom*, chapter 8, p. 147; and A.W. Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), p. 284.

[12] I owe this clarification to an objection raised by Sharon Anderson-Gold. She agrees with me that the propensity to evil must be linked to the doctrine of the highest good, but doubts that this allows one to speak of a "choice" at the level of the species. It is always individuals that hinder or promote the highest good. I insist on the language of "choice" to underline the logical independence of "propensity" and "*Gesinnung*." This is a necessary move in a justification of the propensity to evil, but is not the last word in the moral anthropology of the *Religion*. Here the notion of "the good or the evil heart" (R 6: 29) is meant to do the job of connecting, at a

That is, the species is a *subject of imputation*, not a metaphysical entity to which we can ascribe actions in a *constitutive*, literal sense, as we do in the case of individuals.[13] The goal of Kant's *extended* sense of agency is to move our attention from the existential obstacles a single person faces in the battle to attain virtue (the main point of the natural dialectic in the *Groundwork*) to the historical struggle the species must wage in the pursuit of the highest good. This conceptual shift requires a delicate balancing act: although Kant sees the will of the species as analogous to that of the individual in all relevant senses of agency, the human species is not analogous to the Hegelian *Geist*. A candidate for moral judgment, the species does not properly act – it is *we* who attribute, for the purposes of evaluation and imputation, global patterns of action and intentionality to the otherwise fortuitous acts of individuals.[14]

"Natural Dialectic" Naturalized

The clearest textual evidence for the genealogy of the propensity to evil we are suggesting appears in a famous passage at the end of *Groundwork* I, where Kant uses the notion of "propensity" so prominent in the *Religion*. The claim there is that, in spite of their irreconcilable character as determining grounds of the will, the demands of

quasi-transcendental level, the different degrees (*Stufen*) of the propensity to evil with the *Gesinnungen* of particular individuals. This connection provides a kind of *moral schematism*, which brings to bear universal anthropological determinations on the particularity of actions, via the individual's *Gesinnung*. See *Kant's Theory of Evil*, chapter 4, and S. Anderson-Gold, "Kant's Ethical Anthropology and the Critical Foundations of the Philosophy of History," *History of Philosophy Quarterly*, 11, 4 (1994), 405–19, p. 413.

[13] See Wood, *Kant's Ethical Thought*, p. 289. Since Kant is an individualist when it comes to moral responsibility, one should distinguish between *subjects of imputation* (e.g., corporations, nations, groups, the human species, etc.) and *agents* in a strict sense (always individual actors). While the former are candidates for moral evaluation, the latter are the *real* locus of responsibility. My loose sense of "agency" covers both cases.

[14] See I 8: 17, 41. The species as a *global agent* becomes prominent in Kant's writings on history and anthropology. See, for example, the second section of the *Contest of Faculties* devoted to answer the question of whether the species (as a whole) is morally progressing, and Kant's concluding section in *Anthropology from a Pragmatic Point of View*, "The Character of the Species," particularly, A: 7, 331, where Kant speaks of the species as a "regulative principle."

happiness and morality present themselves as being equally pressing for a will like ours. Although to be happy is a necessary goal for all finite rational beings, reason holds up every maxim of the will to the standard of the pure will. No matter how inclination might protest to the contrary, the agent recognizes that only the form of universal law makes her principles acceptable for others. From this irreducible conflict of interest arises

> a propensity [*Hang*] to rationalize [*vernünfteln*] those strict laws of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and where possible, to make them better suited to our wishes and inclinations, that is, to corrupt them at their basis and to destroy their dignity – something that even common practical reason cannot, in the end, call good. (G 4: 405)

Kant identifies this tendency with a "natural dialectic." Although this summary identification has momentous consequences, it has been usually underappreciated in the literature.[15] In the context of the *Groundwork*, the dialectic refers to the fact that, in the attempt to reconcile the contradictory and seemingly equitable demands of a finite will, the agent is led to overstep the limits reason has set for its proper employment. To place the motivation for an action anywhere else than in respect for the law undermines the maxim's claim to have objective validity. Thus, the propensity to question the intransigence of the motive of duty and accommodate inclinations is *corrupting*: the satisfaction of one's desires, whatever the consequences, is made the measure of all things. Nonetheless, Kant calls the dialectic "natural": the pursuit of happiness constitutes an essential goal of empirical practical reason – a need that *imperceptively* leads the agent "to rationalize those strict laws of duty" when they oppose the interests of self-love.

The universal propensity to evil in *Religion* I is an outgrowth of the same moral psychology. Reconstructing the process of deliberation that must have taken place for Adam to fall into evil, i.e., in an attempt

[15] Allison is a notorious exception. He refers to the same passage in the *Groundwork* to substantiate his claim that there is a fundamental continuity between this earlier work and the *Religion* (Allison, *Kant's Theory of Freedom*, p. 151). It is remarkable that Allison does not develop this insight any further. Instead, he proceeds to identify the *Gesinnung* and the *propensity*, mesmerized by the picture that holds many Kantians captive – the picture, that is, that confines Kantian ethics to the moral individualism of the *Groundwork*.

to shed some light (through a glass, darkly) on the rational origin of the propensity afflicting the species, Kant uses a language reminiscent of the passage in the *Groundwork*:

> He [Adam] thereby began to question the stringency of the command that excludes the influence of every other incentive, and thereupon to rationalize (*vernünfteln*) downgrading his obedience to the command to the status of the merely conditional obedience as a means (under the principle of self-love), until, finally, the preponderance of the sensory inducements over the incentive of the law was incorporated into the maxim of action, and thus sin came to be. *Mutato nomine de te fabula narratur.* ["Change but the name, of you the tale is told"] (R 6: 42)[16]

The dialectic, which the single individual enacts in the process of moral deliberation, is now said to be "entwined with humanity itself and, as it were, rooted in it" (R 6: 32). That is, the natural dialectic is *naturalized*: Kant considers it part of the makeup of the human species, revealing the fundamental subjective obstacle it faces in the course of its moral development. In Adam's rationalization, we are supposed to recognize our own. Independently of the way in which each individual might have resolved the dialectic within her own *Gesinnung*, Kant believes that anthropological research provides "no cause for exempting anyone" (R 6: 25) from the propensity to subordinate the demands of duty to the claims of happiness. Since this sophistry represents "the subjective ground of the possibility of the deviation of the maxims from the moral law" (R 6: 29), Kant unambiguously now dubs it a "propensity to evil."

The qualification "to evil" ("*zum Bösen*") is decisive. It entails that the natural dialectic, which in the *Groundwork* was assumed to be a sheer fact of our finitude, must now be represented as a result of an act of freedom. Its naturalness can no longer pass as the unavoidable consequence of psychic forces beyond our control; it must be

[16] The biblical narrative Kant resorts to here is *not* meant as an "explanation" of the (rational) origin of evil. This would entail having intellectual intuition and is beyond our ken. It is offered, rather, as an *Ersatz*, as a way to translate into the imperfect medium of time what we must represent as lying beyond time and which would otherwise remain incomprehensible for us (R 6: 43). The story of Adam's fall, therefore, is an illustration "in accordance with this weakness of ours" (ibid.). See Gordon Michalson's chapter in this volume, "Kant, the Bible, and the Recovery from Radical Evil."

understood as a self-imposed condition, as a *consequence* of a choice. In order to have *moral* import, the propensity must have been brought about by the species upon itself – its naturalness is *made*, not *given*.[17]

This inbuilt intentionality has important interpretative consequences. In contrast with the dialectical nature of speculative reason, a faculty whose *fate* (*Schicksal*) unwittingly leads us to generate transcendental illusions (KrV Avii), the dialectical nature of practical reason must be represented as a *willful overstepping* of limits. To distinguish the outcome of this transgressive process from the non-culpable transcendental illusions of speculative reason, let me introduce the notion of a "practical illusion."[18]

Practical Illusion

As in all transcendental illusions, a practical illusion is a subjectively necessary product of human reason that without due criticism takes it to be objectively necessary (KrV A295/B352). What is distinctive of a practical illusion is that this "taking to be" must be represented as the result of a choice, as the outcome of an active process of self-deception. Though all transcendental illusions reflect the dogmatic employment of reason, a practical illusion cannot be glossed over as a cognitive mistake. It is a full-blown moral failure – *evil* is not *error*. Here are its defining features:

- A practical illusion differs from ordinary transcendental illusions because it is itself the voluntary product of the use of freedom. To the extent that the moral law is objectively capable of motivating us, there must be a *choice* (not "fate") involved in its genesis.

[17] It is thus comparable to the "self-imposed (*selbstverschuldet*) immaturity" Kant believes we have a duty to overcome in our slow march towards the Enlightenment (WA 8: 33).

[18] Bernstein also connects Kant's doctrine of radical evil with a "dialectical illusion," though he has a completely different sense in mind. For Bernstein, "the concept of radical evil is a dialectical illusion because it seduces us into thinking that we can *explain* something that we cannot possibly explain – why we freely adopt the maxims (good or evil) that we actually adopt" (Bernstein, *Radical Evil*, p. 35).

- As all illusions of reason, the source of a practical illusion is *subjective*. Yet, "subjective" does not mean here the need of the understanding to overstep the limits of experience, but the inherent contingency of an act that makes imputation possible.
- Finally, as in all other transcendental illusions, a practical illusion is *necessary*. It is not the arbitrary product of a deficient subject, but expression of a fundamental aspiration of human reason.

Kant's "propensity to evil" has the same formal features – just as a practical illusion, it is also *imputable*, *subjective*, and *necessary*. By linking these notions, we can discern another sense of evil's radicalism. While in connection with the individual's *Gesinnung*, "radical" designates the *locus* of evil (a *spatial* metaphor), in connection with the species's propensity "radical" refers to a peculiar kind of *necessity* – one that is *subjective* in character but not for that reason arbitrary or accidental. In this second sense, the radicalism of evil conveys a *modal* metaphor. Each metaphor, then, is connected with a different unit of moral analysis, the individual and the species, and changes meaning in each case. Naturalizing the natural dialectic gives rise to a transcendental illusion that Kant presumes to be entwined with our practical reason – this sense of "evil" does not refer to its location in an individual's will, but to the modality of its presence in the human species.

The Need of a Formal Proof

Insofar as transcendental illusions are bound with reason's necessary aspirations, they haunt us even after their deceptive character has been unmasked by criticism (KrV A297/B354). The same is true in the case of the universal propensity to evil, which subsists even if particular individuals adopt good *Gesinnungen*, i.e., resolve the practical illusion in critically acceptable ways within their will. The persuasiveness of Kant's doctrine depends, therefore, on finding the right relation between these logically independent units of analysis. There are two extremes Kant needs to avoid: too tight a connection and too lax a disconnect. The first extreme takes the evil character of the species to *entail* the evil *Gesinnung* of the individual, leaving the particular

agent no room to exercise her freedom (R 6: 32); the second, divests the propensity to evil of emotional grip, detaching it from the reality of our individual moral struggle.

Kant navigates between the excesses of analyticity and existential irrelevance by postulating a basic *isomorphism* in the conditions of choice at both levels of analysis. As Allison rightly noticed, the choices of *Gesinnung* and propensity are both the outcomes of transcendental acts (*Taten*), independent of the temporal conditions of the second analogy, and hence "innate" (*angeboren*).[19] But isomorphism is not identity. A close reading of the text shows that Kant makes a subtle distinction between these choices: with the choice of *Gesinnung*, the individual establishes "the ultimate subjective ground of the adoption of the [i.e., her] maxims" (R 6: 25); with the choice of propensity, the species determines the "subjective ground of the exercise of freedom in general (*überhaupt*)" (R 6: 21). This difference is important; without it, Kant's reasoning would be flagrantly circular. For if in order to account for the possibility of an agent's evil *Gesinnung*, Kant assumes the evil nature of the species, and that assumption is justified, in turn, by the fact that some agents willingly act against the law, then the *Gesinnung* of those agents functions both as *explanans* and *explanandum* in his argument. If "evil *Gesinnung*" and "man's evil nature" are *identical*, Kant cannot legitimately use them to justify one another.

The text of the *Religion* is hopelessly ambiguous at this crucial juncture. At times, Kant argues as if, although in possession of it, "we can spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human deeds parades before us" (R 6: 32). At times, he asserts the contrary view: "even though the existence of this propensity to evil in human nature can be established through experiential demonstrations of the actual resistance in time of the human power of choice against the law, these demonstrations still do not teach us the real nature of that propensity or the ground of this resistance" (R 6: 35). Kant is well aware that only an a priori type of argument can underwrite his view; yet, instead of providing one,

[19] "Innate" in the sense that they must be "posited as the ground antecedent to every use of freedom given in experience" (R 6: 22), not that we are born with them.

he proceeds to "develop" the concept of evil without justifying the existence of a universal propensity to it.[20] To turn our perplexity into annoyance, at one point Kant argues as if he had already provided the proof we are looking for, but the proof is nowhere to be seen.[21]

Kant's vacillations create serious exegetical problems. As indicated above, the claim "man is evil by nature" cannot possibly be analytic: if the predicate "evil" belonged to the concept "man," the act of ascribing a *moral* character to humanity would be self-contradictory. An entailment would leave the destiny of humanity in the hands of the law of identity, not of the moral law, where it belongs. If Kant's claim is to have any normative status, it must be synthetic. Furthermore, since the "propensity to evil" is said to hold without exception for every individual agent, even the *best*, the claim has a priori pretensions also. Kant's brazen empiricist gestures should not distract us: his infamous condemnation of the species belongs to the heart of his critical philosophy (which is concerned with the possibility of synthetic a priori judgments), and must be defended accordingly. No matter how much empirical evidence Kant may marshal, it will never relieve him from the task of justifying the validity of his moral indictment on (some kind of) a priori ground. His doctrine stands or falls with this ordeal.

Kant's Dilemma: Superfluous or Trivial?

The ambiguities in Kant's text have fueled a recent controversy between Allen Wood and Henry E. Allison on this issue. Their positions can be considered symptomatic of an unfortunate dilemma Kant poses to the interpreter: either to emphasize the widespread social/empirical

[20] Kant argues that an a priori proof is required by the fact that the human will is free (i.e., capable of initiating a new series of events spontaneously) and that the moral law is a purely intellectual concept (belonging to pure practical reason). The propensity to evil, therefore, "must be cognized a priori from the concept of evil, so far as the latter is possible according to the laws of freedom (of obligation and imputability)" (ibid.).

[21] Kant says: "The appropriate proof of this sentence of condemnation [i.e., "man is evil by nature"] by reason sitting in moral judgment is contained not in this section [§ III], but in the previous one. This section contains only the corroboration of the judgment through experience – though experience can never expose the root of evil in the supreme maxim of a free power of choice in relation to the law, for, as *intelligible* deed, the maxim precedes all experience" (R 6: 39n.).

dimensions of evil at the expense of its noumenal origin (the path Wood follows), or to stress its noumenal origin at the expense of its social/empirical dimension (Allison's alternative).[22] Pushed to their limit, both alternatives lead to undesirable results. The first invites us, as Allison complains, to construe evil "purely naturalistically as either a social or biologically conditioned trait, perhaps an unfortunate byproduct of our evolutionary development."[23] The second leads to an individualistic conception of evil, insensitive to the role the propensity plays in Kant's moral teleology and the collective character of the pursuit of virtue.[24]

Allen Wood eschews the need of a "formal proof" of the propensity to evil by downplaying its a priori, transcendental character. Focusing on the role radical evil plays in Kant's philosophy of history, Wood claims that the proposition "man is evil by nature" is an empirical thesis (though not an inductive generalization).[25] The basic idea is that "[s]ince it purports to be a thesis about human nature, it makes

[22] The scholarly landscape is, of course, more complicated. There are at least two alternative lines of interpretation worth noticing. Cristoph Schulte maintains that the absence of a transcendental deduction is the result of Kant's awareness of the *impossibility* of a formal proof. Cf. C. Schulte, *Radikal Böse. Die Karriere des Bösen von Kant bis Nietzsche* (München: W.F. Verlag, 1991), pp. 78–88. Seiriol Morgan, on the other hand, develops a transcendental deduction by associating the primacy of self-love with Kant's conception of "negative freedom." S. Morgan, "The Missing Formal Proof of Humanity's Radical Evil in Kant's *Religion*," *The Philosophical Review*, 114, 1 (January 2005), 80–6. Schulte makes Kant sound disingenuous (if not deceitful) when asserting the need of a proof; Morgan's otherwise illuminating approach unfortunately lacks textual support in the *Religion*, as the author freely admits (p. 87).

[23] H.E. Allison, "On the Very Idea of a Propensity to Evil," *Journal of Value Inquiry*, 36, 2–3 (2002), 337–48, p. 346. I think Allison's complaint is unfair. As Wood makes clear in this collection ("Objection 1" in "Kant and the Intelligibility of Evil"), his interpretation of evil does not deny transcendental freedom, but deliberately connects the exercise of free agency to its social and historical context to explicate the propensity to evil. This, indeed, is Kant's own strategy in the writings on history. My disagreement with Wood lies primarily in the fact that, *pace* Kant, he overlooks the need for an a priori justification of Kant's infamous claim, while I believe that such a justification can (and should) be given.

[24] For a rejection of this individualistic line of interpretation, see S. Anderson-Gold, *Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant* (Albany: State University of New York Press, 2001, chapter 3), and P.J. Rossi, *The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind* (Albany: State University of New York Press, 2005, chapters 4 and 5).

[25] Wood, *Kant's Ethical Thought*, p. 287.

most sense to look for its foundation in Kantian anthropology."[26] On this interpretation, radical evil pertains to our social condition and "is closely bound up with our tendencies to compare ourselves with others and compete with them for self-worth," which Kant identifies with "unsociable sociability." The appeal of this view is that, at first sight, it does justice to the text of the *Religion*, where Kant's examples of evil (i.e., unprovoked cruelty, falsity in interpersonal relations, resentment on account of mutual dependence, and war among states) stem from the competitive tendencies Wood rightly emphasizes (R 6: 33–4).

On closer examination, however, it is clear that this line of thought cannot offer the *whole* Kantian story. There are two important omissions in it. First, Kant associates the "evil" tendencies our social condition brings to the fore with the "predispositions to the good" (*Anlage zum Guten*), more precisely, with the predispositions to *animality* and *humanity*. They are not, in any straightforward sense, linked to the three *propensities to evil* (i.e., frailty, impurity, and wickedness). Wood remains silent on this crucial point, and does not explain how all sorts of vices are "grafted" onto these (in principle) progressive and beneficent tendencies.[27] *That* they are grafted is beyond dispute, but the explanatory task lies in showing how, at a transcendental level (i.e., prior to social interaction and as a result of freedom), our practical reason can warp these purposive tendencies and turn them to "evil."[28]

Second, and more importantly, Wood's appeal to the empirical testimony of history and anthropology does not square with Kant's undaunted placement of the choice of the propensity to evil at the noumenal level. *Why* that choice was made is, unquestionably, beyond our comprehension (R 6: 22n.) – but *how* it must have taken place

[26] Ibid., p. 286.

[27] Wood's silence echoes Kant's own – my point is that the *Religion*'s text can be successfully prodded and made to speak, a task Wood does not himself undertake. I try to do so in *Kant's Theory of Evil*, chapter 4.

[28] Since, according to Kant, the drive to society is fundamentally purposive and beneficial, in order to wreak havoc in inter-subjective relations the propensity to pay unwarranted attention to the demands of self-love must be already in place when we come in contact with other self-centered competitive agents. See J. Grenberg, *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005), p. 34ff.

can be illuminated (given our knowledge of human conduct) by some kind of philosophical argument, without thereby indulging in "metaphysical garrulousness."[29] The social dimension of evil is, without a doubt, an essential *expression* of that a priori choice, but it offers no surrogate for its justification. The choice of propensity must be understood in its own transcendental terms as an act "antecedent to every act (*Tat*) that falls within the scope of the senses" (R 6: 21), i.e., an act which provides the conditions for the possibility of the sad spectacle of evils that history parades. This account is not to be found in Wood. By wisely refraining from ignoring our ignorance of the noumenal, Wood is led to equate radical evil with "unsociable sociability." This latter may well be an empirical thesis; the former doctrine is not. The propensity to evil, although referring to the whole species, is not an anthropological (observational/empirical) claim in any straightforward sense – and Wood's social approach is ill suited to accommodate this fact.

Henry E. Allison, on the other hand, is uncompromising about the a priori character of Kant's view. Indeed, he offers a transcendental deduction in three steps:[30]

(1) It is impossible to attribute a propensity to good to a finite will like ours. Such a propensity would consist in the spontaneous preference of the incentive of morality over that of happiness. Yet, this is a trait of holiness and is unavailable to the human will.
(2) Given Kant's rigorism, this impossibility entails the necessity of attributing the contrary propensity to our species (i.e., a propensity to evil).
(3) Since the impossibility of a propensity to good is not logical (for the notion is not self-contradictory), the conclusion "man is evil by nature" has synthetic a priori status.

[29] "Metaphysically garrulous" is Wood's turn of phrase in private communication.

[30] H.E. Allison, *Kant's Theory of Freedom*, pp. 152–61, particularly p. 155. With small variations, he reiterates the proof in "Ethics, Evil and Anthropology in Kant: Remarks on Allen Wood's *Kant's Ethical Thought*," *Ethics*, 111, 3 (April 2001), 594–613, pp. 609–10, and in "On the Very Idea of a Propensity to Evil," p. 342. To make the contrast with Wood starker, I favor the original formulation in the book, where Allison is less clear about the role of happiness and the meaning of holiness in his deduction.

The elegance of this argument is very appealing. However, I doubt it proves what it is supposed to. The fact that human beings are not holy is a necessary condition for having a propensity to evil, but not a sufficient one.[31] Besides not being holy, it must be possible to ascribe to the species (to every individual without exception) the inversion the ethical order of priority among the incentives. *That* the human will is not holy only means that it cannot escape the strictures of obligation, not that it *actually* has a propensity to evil. The conclusion vouched by Allison's argument, then, is uninformative about the moral character of the species. The fact that the objective demands of reason are subjectively contingent does not preclude the possibility that, counterfactually, every human agent had chosen the primacy of duty as her ultimate motivating ground. It is the necessity of assuming the absence of such a choice, not its mere possibility, which the deduction must justify. As Wood pithily puts it, the fact that our will is not holy reduces the import of Kant's doctrine of radical evil to "a trivial practical corollary of our finitude."[32]

Furthermore, Allison's interpretation of holiness in step 1 is misleading: the notion of a "propensity to good" cannot properly characterize the motivational structure of the holy will – at least in the traditional view of the *Groundwork*, which Allison seems to endorse in the original version of this proof.[33] Kant defines a propensity as "the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general" (R 6: 29). The holy will, whose reason infallibly determines the subjective form of its maxims (G 4: 412–13), cannot possibly have a propensity. Its holiness resides precisely in not being affected by sensible incentives. "Goodness" is a matter of course – not of choice – in this type of volition. It is true (as Allison indicates in step 2) that Kant is a rigorist. Yet, from the fact that man does not

[31] Mark Timmons makes a similar point in "Evil and Imputation," p. 138.

[32] Wood, *Kant's Ethical Thought*, p. 287.

[33] In the *Metaphysics of Morals's* version of "holiness," of course, things get more complicated, for Kant admits the possibility of *human* holly wills (MS 6: 383). But this is not the version Allison seems to endorse in his *Kant's Theory of Freedom*. Nor will this correction fix the larger philosophical problem: the conceptual counterpart of the propensity is not virtue, a concept that Kant uses at the level of individual morality, but the highest good, which plays itself out as a collective moral project throughout human history.

have a "propensity to good," it does not follow that he must have a "propensity to evil," for the first alternative was not available to begin with. The synthetic a priori character of Allison's conclusion, therefore, is purchased by conceptual imprecision: Kant's moral anthropology does not allow for "*propensities* to good" and "*predispositions* to evil" in human beings.[34] The propensity, although referring to an a priori choice, must be understood in light of Kant's sweeping anthropological observations – and Allison's individualistic approach is ill suited to accommodate this fact.

Conceptual Stratification

These remarks do not mean to suggest a wholesale rejection of Wood's and Allison's interpretations. My goal is to show that they are one-sided: the social dimension of the propensity should not be obtained at the expense of providing an account of its transcendental origin, nor should this origin turn its back to the empirical dimension of evil. We must devise a view that preserves the unnerving complexity of Kant's position.

It might be helpful at this point to resume the distinction with which we started this inquiry. Whereas radicalization locates the source of an agent's wrongdoing in her ultimate principle of maxim-selection (her *Gesinnung*), naturalization describes the moral frame of mind we can attribute to the species (the propensity to evil). Although both notions operate at a transcendental level, i.e., they provide knowledge of the a priori conditions for the possibility of certain (moral) phenomena, they can be represented as being hierarchically organized. The *Gesinnung* (good *or* evil) warrants the attribution of the manifold of observable actions to a single moral character; the propensity to evil, on the other hand, warrants the attribution of a single form

[34] Kant is careful in distinguishing "propensities to evil" from "predispositions to good." They play very different roles in the process of attaining our moral destiny: *predispositions* present the purposive arrangement of the will; *propensities* are counter-purposive and explain the deviation from such a purpose. The former are "original" (i.e., not chosen); the latter are self-imposed. Thus, in Kant's framework, a "propensity to the good" is a category mistake. Daniel O'Connor complains about a "lack of symmetry" between these notions; I take this complaint as a misunderstanding of their role in Kant's moral anthropology. See O'Connor, "Good and Evil Disposition," p. 297.

to all *Gesinnungen*, and is necessary to explain the possibility of an *evil Gesinnung* as such. Both concepts are transcendental, but one – so to speak – ranges over the other. This relation I call "conceptual stratification."

Reconstructing Kant's position in these terms renders a twofold service. It undercuts the charge of circularity, which is based on the assumption that the *Gesinnung* and the naturalized principles of Kant's moral psychology are synonymous. More importantly, it helps clarify the kind of empirical evidence we may expect to receive in support of Kant's doctrine. While observable wrongdoing can furnish *indirect* evidence for an agent's evil *Gesinnung* (the two-step inference *must* start from actions against the law), when it comes to the propensity to evil empirical appeals are almost futile. Since an evil *Gesinnung* (a concept itself with transcendental status) is made possible by a higher-order transcendental concept (the propensity to evil), empirical evidence of wrongdoing is twice removed from what it is supposed to illustrate.

As Kant puts it, it must transpire "from anthropological research that the grounds that justify us in attributing [the propensity to evil] … are of such a nature that there is no cause for exempting anyone from it" (R 6: 25). Yet the importance of this confirmation cannot be overstated: "So far as the agreement of actions with the law goes … there is no difference (or at least there ought to be none) between a human being of good morals (*bene moratus*) and a morally good human being (*moraliter bonus*)" (R 6: 30). Anthropological research can at most corroborate – in the minimalist sense of not falsifying – what Kant needs to prove by other means. Kant's empirical threshold is so low that its contribution seems negligible: actions according to duty (*pflichtmäßig*) and against duty (*pflichtwidrig*) equally serve to make the point.

Kant's Failed Deduction

Before offering my version of Kant's argument, it is instructive to follow Kant's own reasoning in *Religion* I to detect where it falls short:

(1) The "multitude of woeful examples" of man's observable immorality gives ample (indirect) proof of the presence of an evil *Gesinnung* in the perpetrators. (This claim is justified by Kant's two-step system of inferences [R 6: 20].)

(2) If we can justify the presence of evil *Gesinnungen*, then we have grounds to assume the transcendental concept that makes them possible; hence, the propensity to evil is justified as a necessary assumption and "we can spare ourselves [its] formal proof" (R 6: 32). (This conclusion results from a logical operation: an inference from actual to possible is valid.)

(3) Since (a) the propensity to evil is not analytically entailed by the concept "man" (R 6: 32), nor (b) is its assumption contradicted by available empirical evidence, the proposition "man is evil by nature" has synthetic a priori status. ([3 a] is a corollary of Kant's doctrine of transcendental freedom, while [3 b] expresses the minimal empirical requirement to validate a higher-order transcendental concept.)

I find steps (1) and (3) convincing, but step (2) problematic. The logical operation Kant performs here presupposes a *symmetrical* relation between the notions of "*Gesinnung*" and "propensity." It assumes that radicalization and naturalization operate at the same transcendental level. Yet this assumption flattens the difference between the units of moral analysis and the types of agent they involve. Furthermore, it disregards the fact that, for the purpose of imputation, the moral character that individuals and the species give to themselves must be construed as independent acts of freedom. The fact that there is proof of the immorality of particular *Gesinnungen* does not demonstrate that the species is evil – only that *those agents* are. All Kant's inference could support is a claim about the widespread generality of the propensity; of its supposed universality, it proves nothing at all. Step (3) might allow Kant to fill the gap, but at the price of inanity: the good, the wicked, and the morally mediocre would be bundled in an empty universal condemnation.[35]

Kant's failure, however, is illuminating. It delineates what is needed for success, namely, an account of the necessity of attributing the propensity to all human beings, independently of the *Gesinnung* (good or evil) each individual adopts for herself. This is the type of necessity conveyed by the *modal* metaphor of evil's radicalism. Framing things

[35] Card and Bernstein reach this conclusion. See C. Card, *The Atrocity Paradigm: A Theory of Evil* (New York: Oxford University Press, 2002), p. 82, and Bernstein, *Radical Evil*, p. 19.

this way allows us to appreciate what was in front of us from the very beginning: the key for the so-called deduction does not lie where it is usually sought, i.e., in the first book of the *Religion*, saturated as it is with examples of immorality, but in the Preface to the first edition, where the highest good conceals all reference to radical evil.[36] This should not be surprising: due to their common dialectical origin in human practical reason, the propensity to evil and the highest good combine anthropology and aprioricity precisely in the way required to solve Kant's dilemma.

The Challenge of Teleology

In the Preface, Kant suggests that the doctrine of radical evil responds to the same subjective necessity that generates the doctrine of the highest good. Both doctrines stem from the same anthropological limitation: "in the absence of all reference to an end no determination of the will can take place in human beings at all, since no such determination can occur without an effect, and its representation" (R 6: 4).[37] Although the categorical imperative binds us "through the mere form of universal lawfulness of the maxims to be adopted" (R 6: 3), and hence bids us to bracket all representation of ends, it is unavoidable for a finite rationality like ours to envision an end as a consequence of our action. Without this end, we would know *how* we ought to act, but ignore "*whither*" (*wohin*) and hence "obtain no satisfaction" in our moral pursuits (R 6: 5). That is, we would understand that a morally relevant action must be based on the motive of duty, irrespective of what our inclinations might say, but lack the representation of the state of affairs we intend to achieve in acting out of duty. Given our limitations, apathy and despair (frames of mind inimical to moral action) would necessarily follow.

[36] This situation is more common than it seems at first sight. We find it somewhat repeated in the *Groundwork*, where Kant's doctrine of the "good will" also forces his understanding of "evil" into the background. I develop this point in my "On the Alleged Vacuity of Kant's Concept of Evil," *Kant-Studien*, 4 (2006), 430–51, and in *Kant's Theory of Evil*.

[37] Although Kant submits this claim as a point about the purposive character of action, his view can also be interpreted in light of his moral psychology: the need of representing an end gives rise to a subreption in the order of priority of happiness and virtue, as I argue below.

Although elicited by an empirical and contingent feature of the human mind, the question "whither" touches morality at its very core. Kant could (momentarily) postpone dealing with it in the *Groundwork*. He was concerned there with "the search and establishment of the *supreme principle of* morality" (G 4: 392), and thus had to resort to the purity of a priori considerations and leave anthropology aside (G 4: 389). However, the question of whether finite rational agents will in fact have the power to pursue the taxing demands of duty "cannot possibly be a matter of indifference to reason" (R 6: 5). Even if morality, "without promising anything to the inclinations, and so, as it were, with disregard and contempt for those claims" (G 4: 405), commands actions whose consequences escape the immediate consideration of such a command, human beings are nonetheless compelled to represent those very consequences.

Kant's doctrine of the highest good provides a critically acceptable answer to this anthropological matter of fact. It aims at satisfying the human need to conceive of some sort of final end comprising all actions and abstentions – an end which can be justified by reason and whose absence would create an impasse in our moral determination (R 6: 5). This final end can be justified because it overrides the immediate answer empirical practical reason gives to the question of teleology, namely, the promotion of one's happiness. As the idea of a sum-total of satisfaction of our needs and inclinations, happiness is also a totalizing end that underlies every human being's subjective relation to her manifold desires (G 4: 418). Yet, to the extent that it is an expression of sensibility, happiness is incompatible with a priori morality and would expedite decision by forfeiting our autonomy. Therefore, the challenge for Kant consists in finding a teleological organizing principle that, though containing an end, can remain in harmony with the conditions of morality. This principle is the "highest good," the complete/total (*Volendete*) object of pure practical reason. It is "total" because it combines in a single volition the demands of happiness and morality – or, as Kant puts it in his more Promethean moments, of nature and freedom (R 6: 5). And it is "of pure practical reason," because, unlike happiness, the highest good presupposes the existence of virtue as its condition for the possibility.

In contrast with the "good" (*das Gute*), which is the object of volition that results from embracing the categorical imperative and

excluding happiness as determining ground of the will (KpV 5: 63), the highest good (*das höchste Gute*) systematically incorporates happiness as part of (the *object* of) our volition.[38] The *ground* of the action (the unconditional commitment to morality) has as its end a state of affairs that incorporates happiness as one of its components. This combination represents the *complete* (*consummatum*) good for a human being: having made herself worthy of happiness, as the good required from her, the individual is now justified in expecting happiness to follow proportionally to her worth. This expectation is articulated in a moral command: *act so that the highest good becomes possible through your actions*. The new imperative does not simply tell us *how to act*, as did the traditional formulae of the *Groundwork*. It also tells us *what should be realized*.[39] That is, the highest good introduces a material/teleological component, whose acceptability rests on a new principle. Let me call it the *demand for subordination*. It stipulates that, to be objective, the inclusion of the claims of happiness (the subjective end of humanity) depends on their prior *subordination* to the formal constraints of morality.

By means of this condition, the priority of pure practical reason is preserved, not simply in isolated actions, but also at the level of an overall moral strategy that organizes the consequences of our actions and directs them to the final end of all moral endeavors. Kant conceives of such a strategy as something intrinsically expansive: it starts with the final end of an individual and concludes with the final end of the whole creation. That is, it starts with the allotment of an agent's happiness in proportion to her obedience to duty, and grows into a

[38] The distinction between "ground" and "object" has been variously wielded by commentators in order to make sense of the doctrine of the highest good and reply to Beck's double charge of its being ultimately inconsistent and empty. Cf. L.W. Beck, *A Commentary on Kant's Critique of Practical Reason* (University of Chicago Press, 1960), pp. 243ff. Among the most important defenders of Kant's view are A. Wood, *Kant's Moral Religion* (Ithaca, NY: Cornell University Press, 1970), F. Beiser, "Moral Faith and the Highest Good," in P. Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge University Press, 2006), pp. 588–629, particularly p. 616, and Y. Yovel, *Kant and the Philosophy of History* (Princeton University Press, 1980). The details of this very important debate need not occupy us here – our focus is on the "architectonic" features of Kant's system.

[39] I am following Yovel in this analysis. See Yovel, *Kant and the Philosophy of History*, p. 33.

global project of reshaping the world according to the demands of morality.[40]

A Failure in Subordination

The expansive stages in the promotion of the highest good confirm our initial contention: Kant is operating with a dual conception of agency, the single individual and the whole species. This duality coincides with the one we assigned to the notions of radicalization and naturalization. Just as at the level of the individual an evil *Gesinnung* fails to subordinate the ethical incentives in a critically acceptable way, at the level of the species the propensity to evil hinders the realization of the highest good.

> We cannot start out in the ethical training of our connatural moral predisposition to the good with an innocence which is natural to us but must rather begin from the presupposition of a depravity of our power of choice in adopting maxims contrary to the original ethical predisposition; and, since the propensity to this [depravity] is inextirpable, with unremitting counteraction against it. (R 6: 51)

This assumption explains why Kant links the doctrine of radical evil to "moral discipline" (*Asketik*) (ibid.). Awareness of the propensity to rationalize the strict commands of duty is supposed to trigger a permanent counter-action on our part, which will (eventually) lift the obstacles towards our moral destiny. The propensity to evil, then, appears as the necessary starting point in Kant's narrative of moral progress: without initial frustration, there can be neither incentive to, nor satisfaction in, realizing the unconditioned in experience.

The complementarity of Kant's doctrines rests on their structural affinity. Happiness and morality present incommensurable, yet equally compelling, demands in a will like ours – demands that generate a "natural dialectic." The condition of *subordination* presents the criteria for its critically acceptable solution. It offers a totalized object

[40] As a result of Kant's strict dualism, the unification of nature and freedom in the highest good sets in motion the mechanism of postulation, whose most notorious example is the existence of God. It is within this framework that we must inscribe Kant's provocative assertion in the Preface to the *Religion*: "Morality thus inevitably leads to religion" (R 6: 6).

(the highest good) in which these interests are connected according to an objective causal rule; furthermore, it makes the acceptability of such an object depend on our prior compliance to the motive of duty. Kant's solution is tailored to avoid that (i) the subjective desire for happiness be taken as the cause for virtue, and (ii) that the object of volition (the end) antecede the moral law (thereby making the will heteronomous).

The propensity to evil represents a failure on both accounts. The *objective order of causation* between virtue and happiness is substituted by the *subjective order of association*, producing an inversion in the ethical order of priority between the incentives. This perverts, in turn, the motivational structure: compliance with the demands of duty depends on their compatibility with the goals of inclination, which have been determined by self-love independently of the moral law. This double distortion is the demise of Kantian morality. It represents an *insubordination*, a revolt of empirical practical reason against the conditions that guarantee the supremacy of pure practical reason. This revolt expresses the non-critical stance our reason adopts towards the unconditioned (the highest good as a total object) and is the source of its dialectic and accompanying practical illusion. To the extent that the revolt must be represented as a *willful* overstepping of limits, the propensity to evil cannot be exculpated as a kind of psychological/anthropological determinism. It is a moral, not a cognitive failure.

I take this to be Kant's main insight in the Preface of the *Religion*. Since "in the absence of all reference to an end no determination of the will can take place in human beings at all" (R 6: 4), "it is one of the inescapable limitations" of our faculty of practical reason to invert the objective order of connection (between intention and end) and replace it by our subjective order of association. The consequence of the action (the end), "though last in practice (*nexu effectivo*) [is yet first] in representation and intention (*nexu finali*)" (R 6: 7n.). Such an inversion is the condition for the possibility of the highest good: agents find in the final end (the idea of a *possible* world where happiness accompanies morality) an occasion to "prove the purity of their intention" (ibid.), i.e., their fundamental commitment to the motive of duty. They hope that happiness will be allotted as a consequence of their virtue, and this belief sustains their moral commitment

throughout their life, keeping in check the apathy and despair that would otherwise overtake them.

Yet the source of this belief can also be morality's downfall. For the claims of happiness, objectively integrated into the total object of pure practical reason (as the expectable consequence of virtue), can be seen, thanks to the same psychological limitation, as being *first* in the order of intention and representation. Radical evil, the inversion of the order of priority between the incentives, is based on the same psychological limitations that the "highest good" comes to satisfy in a critically acceptable way. There is plenty of room for the will to juggle with, and rationalize, what counts as "*nexu effectivo*" and "*nexu finali*" – if it concerns the volitional content, the highest good results; if it touches and corrupts the motivational structure, what ensues is radical evil. In the game of moral evaluation and imputation, *we* must assume a choice in either case.

Conclusion

Despite appearances to the contrary, Kant's proof of the propensity to evil is not really missing, but misplaced and buried in the Preface to the *Religion*, where no one expects to find it. The Preface is important, according to this reading, because it brings an anthropological matter of fact to bear on the transcendental framework of pure practical reason, providing thus the piece missing in the puzzle Kant left for us in the first book. It consists in the realization that the dialectical nature of human practical reason generates both the doctrine of the highest good and of radical evil. These doctrines represent opposite answers to the challenge teleology introduces into Kant's pure morality.

Since the incentives of happiness and duty present seemingly equitable claims for a finite will like ours, and observable behavior does not give us ground for exempting any agent from having undermined their objective order of connection, Kant concludes that there is a propensity to evil in all human beings. It consists in having accepted the illusion that happiness could trump virtue as a motivational ground, i.e., that the result of our conduct (the *nexu finali*) could dispense with the moral constraints required to achieve it (the *nexu effectivo*). Although this illusion stems from our anthropological limitations (and is in a

sense *natural*), it must be represented as something "brought upon us by ourselves" (R 6: 32). It is *evil*, because the propensity frustrates pure reason's need of realizing the unconditioned in experience. More substantively, it is *evil* because self-centered reasons (organized by the principle of self-love) displace considerations of duty (the source of reasons we can share). Since observation of human conduct gives us no "cause (*Grund*) for exempting any one" (R 6: 25) we can assume that all human beings have adopted this mode of deliberation.

The choice of propensity we are ascribing to the species is isomorphic, but not identical, with the individual's choice of *Gesinnung*. Although both stem from acts of transcendental freedom, they concern different moral objects, the "good" and the "highest good," and fall under different types of commands, the *formal* and the *material* categorical imperatives. The latter involves global cooperation and transcends the individual's intention and control; the former falls within the scope of a single will. One imperative tells the agent *how* to act, the other tells her *whither*. In both cases, the evil they represent is compatible with the legality of actions – to turn itself into atrocity, the inversion of the ethical order of priority needs the beckoning of circumstances. In competitive social conditions, opportunities arise all too often. But this unfortunate fact does not lead us to conflate the wicked and the good: it is at least possible for particular individuals to have good *Gesinnungen*, even in the midst of the most corrupt environment. Indeed, they *ought to* – though we have no certainty about how they resolve these matters in their own will. The reason why individual agents choose the *Gesinnungen* they have is as inscrutable (*unerforschlich*) as the reason why the species can be said to have adopted a propensity to evil.

Combining the insights in the Preface and in *Religion* I, we may reconstruct Kant's proof along these lines:

(1) There is a natural dialectic between the claims of happiness and morality in the human will.
(2) It is part of our psychological limitations to give precedence "in representation and intention" to what in fact comes "last in practice", and hence we tend to substitute the objective order of connection with the subjective order of association.
(3) There is a natural propensity, then, to place the claims of happiness over those of morality.

(4) Since the observable actions do not give us any cause for exempting anyone from this tendency, we can conclude that the propensity to evil is present in all human beings, even the best.

Step (1) is a basic assumption of Kant's morality, constant in both the *Groundwork* (G 4: 405) and the *Religion* (R 6: 42). Step (2) expresses what Kant considers to be an anthropological fact, a general feature of *human* practical reason. Step (3) is a consequence of (2) and (1). Step (4) states the minimal empirical requirements to validate a higher-order transcendental concept (the condition for the possibility of a lower-order transcendental concept, such as an "evil *Gesinnung*"). From these steps, it follows that there must be a universal propensity to invert the order of priorities between the ethical incentives, that is, that all human beings are radically evil.

Kant's conclusion has a *sui generis* synthetic a priori status. The notion of "evil" is not analytically contained in the notion of "man," yet it is universal and necessary (in the peculiar sense specified above). Furthermore, Kant's argument appeals to observable action (empirical evidence) *negatively*, i.e., only insofar as it does not contradict his positive philosophical argument, which bears the burden of the proof. Yet, Kant's vindication is not a priori in any traditional sense: anthropological assumptions about the workings of the human mind and the patterns of observable action throughout history play a major role in it. This sets Kant's proof in the *Religion* apart from a classic *transcendental deduction*. It gives a philosophical justification to what otherwise would appear as no more than observational, but contains empirical elements that make the proof less than transcendental. Kant might have called this line of argument "metaphysical," for it offers a principle "by which we think the a priori condition under which alone objects whose concepts must be given empirically can be further determined a priori" (KU 5: 181). I prefer to call it "quasi-transcendental," a name better suited to capture its peculiarly hybrid nature.

No matter this terminological quibbling, the point of finding something like this proof is that it shields Kant's view from unnecessary objections (e.g., circularity, triviality, etc.). If nothing else, the

argument preserves the consistency of Kant's doctrine of radical evil and its explanatory power to account for the sources of immorality, accepting its ultimate inscrutability and the narrow boundaries within which we are forced to make sense of it.[41]

[41] Sharon Anderson-Gold, Robert Louden, and Oliver Thorndike read and made valuable comments on an earlier version of this paper; Dmitri Nikulin made me first aware of the impure dimension of Kant's proof. I want to express my gratitude to them.

7

Kant and the Intelligibility of Evil

Allen W. Wood

Kant's reasons for inquiring into the radical evil in human nature are very different from those that might now lead us to ask questions about evil. The aim of *Religion within the Boundaries of Mere Reason* was to explain to an audience of Christians (of eighteenth-century Lutherans) how their faith might be reconciled with a rational Enlightenment morality. Radical evil is the book's point of departure because of the religious importance of the Christian doctrine of sin. In Part One of the *Religion*, Kant's aim is to articulate that doctrine in rationalistic terms, so as to show in the other three parts how the Christian doctrines of justification and atonement, as well as the function of the church and revelation, might be articulated within the framework of a moral philosophy based on the autonomy of reason.

Today such aims make Kant far more enemies than friends. Christians, and religious people generally, typically charge him with "watering down" the faith, even with offering their religion a philosophical Trojan horse concealing within it the entire army of modern secular unbelief.[1] On the other side, unsympathetic secular philosophers view the *Religion* as proof that Kantian ethics is at bottom nothing but traditional superstition. Both reactions seem to me utterly wrongheaded, but here I will not address either of them directly.

[1] This metaphor is drawn from G. E. Michalson Jr., *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge University Press, 1990).

Instead, my purpose will be to see how Kant's reflections on evil might speak to concerns that are more likely to interest us.

The "evil" in question here is not the bad things that *happen* to people – the pain, grief and sorrow, injury, starvation, death, even their feelings of violation and humiliation. For Kant, all these would all fall under the heading of "ill" (*Übel*) or human unhappiness. Thus our concern is not with the theological "problem of evil." Instead, "evil" (*Böse*) refers to what human beings *do*. More precisely, it consists in actions that they should not do but choose to do, and the principles that lead to these choices. "Evil" includes acts of violence and cruelty – war, rape, conquest, torture, terrorism, genocide – as well as lesser acts of cruelty, callousness, degradation, and disrespect for humanity. "Evil" also refers to our social practices. It includes the obscene gap between rich and poor, both within each society and between different societies, and the oppression of the powerless, based on these economic evils, on social customs, or the abuse of power built into political systems. Evil certainly includes what the human species inflicts on itself and other living things through its irresponsible relation to the natural environment. In fact, "evil" means anything people do when they violate their duties and fail to live up to the dignity of their rational nature.

When I speak about *our* questions concerning evil, what I mean is such questions as these: What, at bottom, does such conduct consist in? And how, if at all, can we make sense of it? How can people do such things? How should we understand the power and prevalence of evil? And how should this understanding influence our struggle against evil? The answers to these questions are what I mean in the title of this chapter by "the intelligibility of evil."

Can evil be made intelligible? Even as we ask these questions, however, we also have to ask ourselves whether they make any sense. Perhaps our questions are nothing but a rhetorical expression of anger and despair and, taken literally, admit of no answers at all. According to one way of looking at the matter, "evil" is simply a word we use to express attitudes of disapproval, blame, or horror at certain deeds. These deeds, and the choices of those who perform them, are natural or social facts that have their physical, psychological, or social explanations. For those who hold that rationality is only a matter of whether you select the right means to satisfy whatever aims or desires

you happen to have, evil deeds may even be completely rational. Once we understand why these deeds occur, the only thing left to explain is why we take the negative attitudes toward them that we do: why we consider them "evil." There will of course be psychological explanations for our attitudes too. Put the explanations for "evil" deeds together with the explanations of our attitudes toward them and you have made evil intelligible in the only sense evil could be made intelligible.

This "deflationist" view of evil, however, remains totally unresponsive to our questions. It cuts the Gordian knot by making our questions about evil disappear, because in the most straightforward sense it makes *evil itself* disappear from the world. It tells us, in effect, that *evil* does not really exist: all that exists are events (in themselves neither good nor evil) and our subjective attitudes toward them. So our questions do not constitute any genuine inquiry at all. Maybe, as already suggested, they are only rhetorical outbursts, expressing an all-too-human attitude or mood that we must simply shrug at, recognizing it as part of our psychology. Or we might even accept the deflationist view of evil as having a certain sublimity, having something in common with a view of life we find in philosophers such as the Stoics and Spinoza, who take it to be the part of reason to rise above our emotional attitudes toward evil, to overcome them through what Spinoza called *amor intellectualis Dei*. Either way, with such a view our inquiry into evil would reach a dead end or be diverted into another inquiry.

I won't try to refute such views directly on their own terms, but I will proceed on the assumption that they are wrong. For the apparently superhuman sublimity of the Stoic or Spinozist transcendence of our attitudes toward evil and the evident shallowness and untenability of the deflationist rejection of our questions about evil have exactly the same cause. None of these views, namely, is anything that we human beings could ever unite with our reflective experience of human life. They are philosophical views suited only to gods, or perhaps rather to robots – bungled experiments at replicating humanity or spectacularly successful attempts at constructing something simpler: beings whose artificial intelligence has been truncated (or, if you prefer, purified), so that it entirely lacks those rational capacities that make us moral

beings who care about our lives and those of others, and for whom, therefore, evil has to be both real and deeply questionable.

I will take for granted, therefore, that there really is evil, that we are right in asking what it is and why it occurs and wrong to think that it is the part of reason to rise above these questions or dismiss them as meaningless. The main thing it means to accept the reality of evil is to assume, along with Kant, that doing evil is *contrary to reason* – that is, that evil is something we have decisive reasons for not doing. If an action is what I have most reason to do, then there seems no longer to be any rational force in the assertion that I should not have done it. If I have no good reason not to do it, then calling it evil, and saying I should not have done it, seems only to express a certain negative attitude toward the action – a reaction, moreover, that I have no reason to take seriously. So we find ourselves back with the view that there really are no evil actions, only a set of natural occurrences (in themselves neither good nor evil) and our non-rational attitudes toward them.[2]

Yet as soon as we grant the reality of evil, we immediately face some familiar and serious problems about how it could ever be explained or made intelligible. For evil is a species of *rationally motivated unreason*. In this respect it is like self-deception or *akrasia*, which notoriously give rise to paradoxes about how to understand them or perhaps even how to provide them with a coherent description.[3] For Kant, in fact,

[2] Even if we grant the rationalist assumption that there are decisive reasons against doing evil actions, it may still be true that evildoing is often "rational" in some perfectly obvious acceptable sense – an evil action may, for instance, be the best means available to the thing the agent wants most (such as that agent's own happiness). Further, even if evil is always contrary to reason, because there are always decisive reasons for not doing it, we may still grant that evil, or at least a lot of it, should not be considered "irrational." For we are not in the habit of applying that term in cases where the reasons someone is acting against are clearly known to the agent but the agent simply refuses to see them as reasons. We who do recognize the rejected reasons, however, must still judge the agent open to criticism on rational grounds.

[3] Deliberate evil, unlike self-deception and *akrasia*, probably does not count as a case of "irrationality" in the usual sense of the term, where we call thinking or behavior "irrational" only if it runs contrary to reasons or standards of rationality that the agent explicitly accepts, or would accept (if the issue were put to her). We do not usually call "irrational" the deliberate refusal to act according to the best reasons one has, or to recognize them as valid reasons at all. (But I think if we regarded the reasons as obvious enough, we might treat this too as a case of irrationality; so it says something about *us* – probably something pretty unflattering – if we do not see the

self-deception and *akrasia* fall under the heading of evil. About self-deception (or what he calls "the inner lie") Kant acknowledges that there are difficulties understanding how it can be possible, but he has no doubt that it plays a large role in human life and also that it is a violation of a duty to oneself (MS 6: 430–1). *Akrasia* belongs to that "frailty" of the will Kant cites as the first or lowest degree of the radical evil in human nature (R 6: 29). The paradoxes involved in *akrasia* and self-deception are therefore included, at least in part, in the more fundamental problem of the intelligibility of evil.

The basic problem about the intelligibility of evil can be stated in the form of a simple dilemma: There are apparently only two things we might mean by "explaining" evil or "making evil intelligible." One would be an explanation of it as an action that is done for reasons. The other would be a causal explanation of it as arising from antecedent conditions. Either explanation, however, if fully successful, would abolish what is evil about the action or display it as something that is not evil after all.

An evil action can be understood as done for reasons in a limited sense. For instance, it can be the action that is the best means to what I want most or the action that will contribute most to my happiness. But if the action is really evil, then whatever reasons I might have for doing an evil action, there are moral reasons for me not to do it, and these reasons are decisive. In principle, therefore, there could never be a fully satisfactory explanation of an evil action as an action for reasons. An explanation that is not a rational explanation, however – a causal explanation, for instance – would be incapable of making intelligible precisely what is *evil* in the action, because evil is conceived precisely in *rational* terms – as a rationally motivated yet

overridingness of moral reasons as obvious). This point about the ordinary usage of "irrational" has been taken by some (e.g., by Bernard Williams) to call into question whether there are any genuine reasons applicable to an agent that the agent does not acknowledge (stigmatizing these as "external" reasons). It is therefore sometimes taken to be an argument for the substantive thesis that we have no reason to act as morality requires unless we acknowledge such reasons. But if it is true that morality provides us with reasons to meet its requirements, then conduct that refuses to recognize such reasons, though perhaps not "irrational," does nevertheless exhibit a clear failure of rationality, and it is open to rational criticism. For a good discussion of this point, see T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), pp. 25–30.

contra-rational action. Hence a causal explanation, whatever it might accomplish, must bypass precisely what is *evil* in the action. Further, a causal explanation would apparently show how antecedent conditions made the action necessary, hence any other action impossible for the agent. But that would also do away with the agent's responsibility for the action – and with it, the possibility of the action itself *as evil*.

These intractable difficulties are no doubt part the appeal of the deflationist view of evil, as well as the perennial appeal of Socratic paradoxes about *akrasia* and the endless problems philosophers have in conceptualizing self-deception. But if someone were to argue on such grounds that self-deception is impossible, then the inevitable rejoinder – utterly decisive, in my opinion – would be to accuse this person of being self-deceived in that denial. And I regard a similar blunt reply to deflationism about evil as more convincing than any objection that could be brought against the project of making evil intelligible on the basis of the dilemma just presented. We have no choice, then, but to persevere in our assumption that evil is real and then try to understand how far, and in what ways, evil actions might still admit of being made intelligible. The point to appreciate going in is that no particular attempt to make evil intelligible should be dismissed simply because it runs afoul of the formidable difficulties just mentioned. These difficulties simply come with the territory; they are not defects of any specific attempt to understand or explain evil. Those who demand that evil to be made fully intelligible in either rational or causal terms cannot even have a coherent conception of what they are asking for.

Kant exhibits a full awareness of these difficulties. He repeatedly emphasizes the limits in principle of both the attempt to conceive and to explain evil. He says the source of evil must lie in the free choice of the rational being, the choice to adopt an evil maxim. But he also insists that "there cannot be any further cognition of the subjective ground or the cause of this adoption (although we cannot avoid asking about it)" (R 6: 25). "We are just as incapable of assigning a further cause for why evil has corrupted the very highest maxim in us, though this is our own deed, as we are for a fundamental property that belongs to our nature" (R 6: 32). Empirical evidences of the existence of a propensity to evil "do not teach us the real nature of that propensity or the ground of [our power of choice's] resistance [to the

moral law]" (R 6: 35). Our choice of evil in time "cannot be derived from some *preceding* state or other" (R 6: 39). We cannot even "inquire into the origin in time [of an evil deed], but must inquire only into its origin in reason" (R 6: 41). That is, we cannot inquire into the *cause*, but only into the *character*, of a freely adopted maxim of evil choice.[4]

The first thing we need, then, is a coherent conception of what might count as making evil intelligible. There are, in fact, two things that might count as doing this. The first we might call "forming an intelligible concept of evil." It consists in conceptualizing evil choices as following a highly general pattern that, although not fully rational, is nevertheless to a degree rational, at least sufficiently so that it is familiar to us as a way that human beings do in fact commonly choose. The Kantian name for this task is *identifying the fundamental maxim of evil* – or, for short, the *maxim problem*. Second, evil might be made still more intelligible if we could understand this general pattern of less than fully rational choice as fitting into human nature as it shows itself under the conditions in which human life has developed on earth. This would help us to understand the persistence and prevalence of evil as a fact of human life, and also enable us to attach a meaning to evil, which might orient both our understanding of it and our struggle against it. This task is what Kant sets himself when he tries to identify evil *as a human propensity* (*Hang*), and to determine why we have such a propensity (how it fits into our psychology and our human life on earth). So we can call it the *propensity problem*. Let us consider Kant's solution to these two problems in turn.

The maxim problem. In order to understand Kant's approach to the maxim problem, we first need a bit of background. Kant distinguishes three original "predispositions" (*Anlagen*) that belong to human nature: (1) animality, (2) humanity, and (3) personality. None of these, he says, is inherently evil, and all may be regarded as present

[4] This is not to deny that we can inquire into the kinds of situations in which people are likely to make evil choices, perhaps with a view to avoiding those situations and thus avoiding evil and its effects. This is a point sometimes emphasized by ethical "situationists," such as John Doris and Gilbert Harman. They are certainly correct that it is important to know what situations these are, and to do what we can to prevent them. The point, however, is that these situations do not *cause* evil choices (what a situation could cause would not be *evil*, but only some event, to which we might take a negative attitude) but only provide the occasion for human beings to make them, or for evil propensities in people to show themselves.

in us *for good* (R 6: 28). *Animality* is the original source of our natural or instinctive impulses, hence of all our empirical desires or inclinations, including, first, for "mechanical self-love" (self-preservation), second, for "propagation of the species" (the sexual drive), and, third, for "community with other human beings (sociability) (R 6: 26–7). *Humanity* is the rational capacity to set ends and devise means to them and also the capacity for rational *self-love*, or the pursuit of our empirical ends as a whole, under the heading of happiness (R 6: 27). This is the predisposition that first achieves development in society, through the *cultivation* by education of our skills to pursue ends and then through the *civilization* of our nature through association with others, which shapes and modifies our conception of our well-being by comparing our state with that of others. Finally, *personality* is our capacity to respect the moral law, as the fundamental rational principle of the will, and to make that respect a sufficient incentive for obedience to the moral law (R 6: 27–8). It too is a predisposition that is developed in the social condition, by the process that (parallel to those of cultivation and civilization) Kant calls *moralization* – but this is a process, in his view, that human history has barely begun (VA 7: 326–7).

Kant insists that none of the three predispositions is in itself evil. Evil must arise from a propensity we display in their use or exercise. Yet evil cannot be traced either to the first predisposition (animality) or to the third (personality). Our natural instincts involve no principle of choice – and only that can be good or evil (R 6: 34–5). Instincts, and the inclinations based directly on them, are in themselves innocent and are capable of being involved in evil only insofar as we incorporate them as incentives into a freely chosen maxim (R 6: 24). But then it is this choice, and not its instinctive source, that is good or evil. Of course it is evident to Kant that a being (such as God) whose only incentives are rational must have a holy will (a will that cannot go against reason's law) (VpR 28: 1075). Perhaps this is why, in some places, Kant seems to explain our temptation to transgress the moral law by our finitude as beings of need (e.g., KpV 5: 25). In the *Religion*, however, he provides the clearest possible rejection of that explanation, when he locates the human propensity to evil not in our natural inclinations but in our use (or misuse) of reason. The enemy of morality, he says, is "not to be sought in the natural inclinations, which merely lack discipline and openly display themselves

unconcealed to everyone's consciousness, but is rather as it were an invisible enemy, who hides behind reason and hence is all the more dangerous" (R 6: 57).

It is consistent with this, however, to see inclinations themselves as resisting morality when they are the expression not merely of natural instincts but also of our free choices. For then they are already manifestations of the propensity to evil. This is especially clear in the case of what Kant calls the "passions" – inclinations which it is difficult for reason to master (VA 7: 265–7; cf. R 6: 93). Every passion rests on a maxim, hence on a choice, for which the agent is responsible. But then these inclinations are not manifestations merely of our animality (or our finitude); our will is already complicit in them, and they are expressions of a propensity grafted onto our rationality.

Evil also cannot be traced to our predisposition to personality – our original relation to the moral law. Evil is no doubt a failure to actualize or exercise the moral predisposition (a failure to respond to the reasons it gives us to do what we ought), but it cannot be traced to the original constitution of this predisposition, or even to some relation in which we might stand to moral reason as such – as if our rational faculty might contain a basic incentive to disobey the moral law rather than to obey it. This Kant calls a "diabolical will" or an "evil reason," and declares it to be impossible (R 6: 35).

Evil must arise, therefore, from something about the way we use (or misuse) our rational predisposition to *humanity* – with a propensity attaching to the way our reason regards our inclinations and self-love. More specifically, Kant concludes that it must consist in a propensity to invert the correct rational incentives of reason and inclination, giving preference to the latter. As rational and moral beings, we have rational incentives to action both in the moral law and also in our inclinations and self-love. The fundamental maxim of evil – Kant's solution to what I have called the *maxim problem* – is that evil lies not in which incentives we incorporate into our maxim, but in the order of priority among them. Evil is conceivable only in the form of a maxim, or freely chosen subjective principle of the will, which involves the preference of the incentives of inclination or self-love over those of morality (R 6: 36–9).

Is Kant's maxim of evil really *evil* enough? Kant's view here, especially his rejection of the possibility of a "diabolical will," is sometimes

criticized for not allowing for the possibility – as it is put – that people can do "evil for evil's sake." The objectors think that Kant is denying we can choose an action not because it promotes our self-interest or satisfies some contingent desire, but simply *because it is wrong*.[5] But I think they have misunderstood him. Kant's argument is that it would be incoherent to suppose a being could be responsible for obeying the moral law and yet lack any rational incentive to obey the law, possessing originally *only* a rational incentive to *disobey* it. It would also be incoherent to think that a being might originally have two directly contrary rational incentives, which would involve the supposition that the being's rational faculty itself is self-contradictory. These impossibilities are what Kant rejects under the heading of a "diabolical will" – not because it represents something "too evil" for human nature, but because it would be incoherent to condemn as evil the choices of a being that could recognize no decisive reason to choose in favor of morality. Whatever harm to human or other beings might be caused by the actions of such a being, they could not be considered *evil*.[6]

When Kant denies that human beings can "incorporate evil as evil for an incentive into their maxim" (R 6: 37), we easily misunderstand this if we assume a certain moral psychology, and a conception of moral reason, that is very different from his. Kant holds that for a rational being, the moral law simply as such is a rational incentive; no distinct (empirical) inclination (such as sympathy or some desire for

[5] See J. R. Silber, "The Ethical Significance of Kant's *Religion*," in T. M. Grene and H. Hudson (trans.), *Religion within the Limits of Reason Alone* (New York: Harper & Row, 1960), pp. cxxv–cxxvii, and R.J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press, 2002), pp. 36–42.

[6] Elsewhere Kant speaks of the vices of culture, though in their "extreme degree, that surpasses humanity," as "diabolical vices" (R 6: 27; cf. MS 6: 461), which clearly are evil in the proper sense of the term. Here he seems to have in mind the vices of hatred – envy, ingratitude, and malice, as well as the vice of rejoicing in others' misfortunes. But in saying that in their extreme degree they "surpass humanity," Kant means to discourage us from thinking of such extremes of evil as actually found in human beings, just as he does not encourage us to think of the contrary ("angelic") virtues as found in actual human beings (MS 6: 458–61). In both cases, I think, the point is that we would do well not to project onto others either our resentment at vice or our admiration of virtue, but to concentrate instead on what we have in common with other human beings (in the way of both virtue and vice) and recognize both the best and the worst of our fellow human beings as not all that different from ourselves. Kant's reluctance to admit "angelic" and "diabolical" extremes in human conduct is not the same as his denial of the "diabolical will."

social conformity conditioned in us by our upbringing) is needed to give us an incentive to obey it. This lies at the heart of Kant's thesis that "reason is of itself practical" (KpV 5: 24), that the moral law is a law of *autonomy*, self-legislated by our own reason. Kant's denial that "evil as evil" can be an incentive for us is a denial that anything parallel to this could be true in the case of evil – in other words, that we might have an original *rational* incentive to disobey the law – that we could have an "evil reason" (R 6: 35). Unlike the original rational incentive to obey the moral law, he is claiming, our incentives to disobey it must take the form of inclinations providing incentives to disobedience.[7]

Kant does not deny, however, that these inclinations can attract us to conduct that is directly contrary to what morality requires (that they might be *empirical desires* for "evil as evil"). For example, the moral law requires us to make the happiness of others our end and so forbids us to take their unhappiness as an end for its own sake. What Kant calls the "vices of hatred" or "diabolical vices" – envy, ingratitude, and malice – are vices because they involve making the unhappiness of another directly an end (MS 6: 458–61, cf. R 6: 27). This looks like "evil for evil's sake" if anything could be.

Further, we all know that people can act "self-destructively" in the sense that they systematically do the very opposite of something they fundamentally will. For example, there are people who directly will to frustrate their own happiness – by becoming addicted to drink or drugs or getting involved in abusive relationships with others. Likewise,

[7] Kant's view at this point therefore involves (contrary to the mistaken claims of Bernstein, see note 6 above) no restriction on the scope of human freedom – it places no limitation on *what* human beings may choose. It is rather a view about the structure of human incentives – a view about what they must be if any choice human beings make is rightly to be called "good" or "evil" at all. Perhaps there is sometimes the temptation to think that some people are *so evil* as to be literally incapable of good, lacking any incentive to be good. But from a Kantian standpoint, this would make sense only if we mean that they absolutely refuse to respond (in their actions or feelings) to the reasons they have to be good, and not if we mean literally that they have no such reasons – since in the latter case, they could not be moral beings at all. Of course someone might take an "externalist" view about moral reasons and moral motivation and adopt a corresponding theory of moral responsibility that enables us to hold people responsible even if we say they have no reason or motive to do what morality demands. But it would be a misunderstanding to think that compared to such a view Kant is restricting the possibility of moral choices. The issue is rather about how we should conceptualize evil choices in relation to reasons and moral responsibility.

"doing evil for evil's sake" could be considered a case of *moral* self-destructiveness, where someone chooses to disobey the moral law simply because they know that obeying it is what they ought to do. The choice so to behave would be based on an inclination to defy the moral law, and it would be this inclination that the agent has given priority over the rational motive to obey the law. Kant's denial of a "diabolical will" involves no denial of such moral self-destructiveness. We can see that this is so once we realize about self-destructive patterns of motivation that the person always also (and more fundamentally) *wills* the thing their self-destructive behavior acts against, so that self-destructive behavior exemplifies precisely the pattern of choosing the rationally weaker incentive over the stronger one. Those who act self-destructively in regard to their own happiness do it because they *also* (and more fundamentally) will to be happy. And the morally self-destructive person, who does something "because it is wrong," must likewise have a fundamental incentive to do the right thing, even if he stubbornly refuses to respond to it. An act of malice, for example, is malicious precisely because the agent knows that morality tells us to benefit others and not to harm them. Far from denying the possibility of "doing evil for evil's sake," Kant's account of evil yields precisely the correct account of what this is.[8]

It is true that in the *Religion*'s discussion of evil Kant does not bother to distinguish between evil actions we might perform because they benefit us at another's expense and evil actions that we do precisely in order to harm the other (whether we benefit from them or not or even are ourselves harmed by them). Here as elsewhere, he sometimes tends to emphasize the contrast between moral motivation and non-moral motivation, at the expense of various other contrasts between different species of non-moral motivation. His aim, after all, is to capture the most fundamental maxim of evil, which necessarily

[8] For a similar recent discussion, see M. Caswell, "Kant on the Diabolical Will: A Neglected Alternative?" *Kantian Review*, 12, 2 (2007), 147–57. Caswell does a good job of making the point that viewing someone as so fundamentally evil in their motivations that they are *incapable* of good is not only morally incoherent but it is also to imagine their evil as "radically alien" to ours in a way which tends to blind us to what is truly evil, especially in ourselves (p. 156). Cf. "[Such views] are best explained as the projection of irrational hatred and resentment, which are not uncommon." Allen Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), p. 401.

involves bringing all contra-moral motivation under a single heading ("self-love," "inclination," "the incentives of our sensuous nature") (R 6: 36–7). Perhaps this makes it easier for us to think that he is reducing all evil maxims to some one type – that of the impulsive, self-indulgent hedonist, for instance, or the cool, self-interested schemer – and that he is excluding others, such as the self-righteous hypocrite or the malicious person consumed by spitefulness or hatred. But the aim in identifying the underlying maxim of evil is only to conceptualize what is involved in acting against moral reason – to provide the most abstract concept of that volition of which the motivated unreason of evil consists. We miss the point of Kant's account if we don't recognize that it is entirely consistent with acknowledging that non-moral incentives take very different forms, some shrewdly prudential, some vengeful and malicious, some involving disguise or self-deception, as when evil assumes the cloak of arrogant self-righteousness or hypocrisy.

There is a quite different reason, however, why people may think that Kant's treatment of evil does not deal sufficiently with "the diabolical." This is that Kant's solution to the maxim problem is simply a general account of evil choices irrespective of the degree of evil involved in them. It fits minor or trivial violations of duty just as much as it does extreme cases of evil. We may find this disappointing, because one thing we may want from a philosophical concept of evil is some special insight into the extreme cases of evil – of some especially uncanny or monstrous mindset that we think must have led to the Holocaust, for example – or what distinguishes (as some people have put the objection to me) the "really evil" from the "merely bad."

Kant of course recognizes that some cases of evil are worse than others. He distinguishes three "degrees" of evil – frailty, impurity, and depravity (R 6: 29–30), and within each of these he recognizes lesser and greater degrees of evil. The "diabolical vices" of hatred, which take the unhappiness of another as an end for its own sake (R 6: 27; cf. MS 6: 461), are clearly worse in his view than minor transgressions resulting from the indulgence of an inclination, not discreditable in itself, that prevents us from doing something we should have done. But when it comes to the concept of evil itself, his aim is to bring all cases of it – whatever their degree – under a single concept, a single maxim of evil, a maxim that applies in the same way to minor evils as it does to the worst evils. Evil, after all, is not like a Platonic form: we

do not call something evil only to the extent that it participates in "evil itself" – the most extreme kind of evil. That way of thinking makes more sense applied, as Plato applies it, to virtue or excellence than to vice or evil. Evil is highly heterogeneous – "pure evil," an oxymoron.

I think it is both significant and commendable that Kant refuses to cater to our prurient craving for a special account that applies especially to the most extreme cases of evil. While recognizing that there can be both "diabolical" vices and "angelic" virtues, Kant discourages us from looking at people as exemplifications of them (MS 6: 458–61). He fears that occupying our imaginations with extreme cases of evil may be merely a way of indulging some of our nastier human traits – rationalizing our resentment and vindictiveness by supplying it with an object that would seem to justify it. Further, to think that extreme cases of evil represent something morally, psychologically, or even metaphysically special may be merely a way of rationalizing our own transgressions. We want to think that the true monster of evil (Hitler, Stalin, Saddam Hussein, Hannibal Lecter, Dick Cheney) has little in common with our petty failings and vices. The image of such monsters also helps us to divide all human beings into "good people" and "evil people," providing our worldview with the "moral clarity" conspicuously exhibited by some of these monsters themselves. Kant wants us to be mercilessly clear about right and wrong when it comes to our own actions, but he encourages an attitude of charitable moral ambiguity when it comes to judging others (MS 6: 437–42, 463–6; VA 7: 151–3). Thus Kant's treatment of evil is designed to make us aware of the continuity between different cases of evil, what cases of evil have in common (however they may differ in degree), and therefore aware of our kinship with other evildoers rather than our distance from them. The Kantian view is that to "look evil straight in the face" is not to gaze in voluptuous horror at the visage of Hitler, but instead simply to look in the mirror, asking yourself honestly and soberly what you might do to improve what is there.

The propensity problem. So far we have seen only Kant's solution to the *maxim problem*. The maxim of evil is to invert the rational order of incentives, placing self-love or inclination ahead of morality. This makes evil intelligible to a degree, because it is often consistent with both instrumental and prudential rationality. Moreover, it follows a pattern in human choice that is entirely intelligible in the sense that

it is familiar to all of us, both in our own conduct and in the conduct of others. This provides us with an intelligible account of *what evil is*. What remains, however, is an even more difficult problem – *the propensity problem*, the problem of understanding the prevalence of evil in the world and its meaning in human life.

By the word "propensity" (*Hang*) Kant means "the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*) insofar as this possibility is contingent for humanity in general" (R 6: 29). A common example of a propensity is the propensity to consume intoxicants, which is aroused in some people by acquaintance with them (R 6: 29n.). A propensity is an empirical pattern of choice, or a desire to choose according to a determinate maxim.[9] The propensity to evil is a propensity for the contra-rational choice that inverts the rational order of incentives, placing the incentives of self-love or inclination ahead of those of moral reason. This propensity is familiar enough to us – in fact, so familiar "from inside" that many philosophers think it is perfectly in accord with reason to reject the claims of morality in favor of those of self-interest or even capricious desire.[10] Kant's conception of a propensity to evil must therefore be regarded as a success in making evil intelligible. However, in the attempt to understand evil, the deeper *propensity problem* is that of coming to understand what it means that we have the propensity to evil, and why it is so prevalent among us human beings.

[9] To have a propensity to choose in a certain way is not the same as actually choosing in that way. Hence to say that human nature has an inborn propensity to evil is not yet to say that people do make evil choices, and although it makes it intelligible that they do so, it does not entail that they must, or make their evil choices any less attributable to their use of their freedom. In fact, one might ask whether Kant thinks the "ideal of humanity well-pleasing to God" or "humanity in its full moral perfection" (Kant's philosophical conception of the Christ image, "the Holy One of the Gospel") is afflicted with the human radical propensity to evil. Kant never answers this question explicitly, but he does say that an individual instantiating this ideal would be "afflicted with the same needs, and also the same sufferings and [hence] … the same temptations to transgression as we are" (R 6: 64). So if we assume that these temptations arise from the radical propensity to evil, then we must conclude that this moral ideal of humanity is also afflicted with that propensity. The holiness of will exemplified by the moral ideal consists not in immunity from the radical propensity to evil but rather only in not yielding to it.

[10] The most famous philosophical discussion of this issue, of course, is Sidgwick's treatment of the so-called "dualism of practical reason," in *Methods of Ethics* (Indianapolis: Hackett, 1981), pp. 200–6, 507–9, and 404n.

The social origin of evil. Kant's solution to the propensity problem is not highlighted in Part One of the *Religion*, because the aims of his discussion are those I have described and not *our* aims in asking about evil. The propensity problem is also of only marginal interest to Kant in the Part Two, and begins to play a significant part in his aims only in Part Three of the *Religion*. Nevertheless, Kant's solution to the propensity problem is presented in the *Religion* both clearly and emphatically, and it coheres with his anthropology and philosophy of history as presented in other works. This solution is that *the human propensity to evil arises in the social condition, and develops along with the processes of cultivation and civilization that belong to it.* Though not emphasized in Part One, the social origin of evil is clearly indicated in Kant's remarks about the predisposition of humanity – that predisposition, as we have seen, in which Kant locates the radical evil in human nature.

> The predisposition to humanity can be brought under the general title of a self-love that is physical and yet involves comparison (for which reason is required); that is, only in comparison with others does one judge oneself happy or unhappy. Out of this self-love originates the inclination to *gain worth in the opinion of others*, originally, of course, merely *equal worth*: not allowing anyone superiority over oneself, bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others. – Upon this, namely, *jealousy* and *rivalry*, can be grafted the greatest vices of secret or open hostility to all whom we consider alien to us. These vices, however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure; for nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture. (R 6: 27)

The original meaning of our natural desire for happiness is that we should compare our state with that of others and find it superior to theirs. As culture develops, our original defensive anxiety to protect ourselves against the ascendancy of others is transformed into a desire for superiority over them. This inclination does not issue directly from nature but arises from the development and use of reason, in setting ends and pursuing happiness. When this competitive spirit is set alongside the basic requirements of the moral law – not to make

an exception of ourselves to maxims we will to hold as universal laws, to treat all rational beings as ends in themselves rather than subordinating them to our ends, to follow the laws of a realm of ends, in which human ends are in systematic harmony – we see that it is in direct conflict with these moral demands (G 4: 421–36). Once we see that our natural inclinations, when shaped by our social condition as rational beings, involve this competitive spirit, then we can see that the fundamental maxim of evil, which gives their satisfaction priority over obedience to the moral law, is really nothing except a desire for superiority over others and a policy of esteeming ourselves on the basis of our state or condition, which can be compared with that of others with the aim of validating that superiority.

Much of Kant's working out of the theme of unsociable sociability in his historical, ethical, and anthropological writings has to do with the various ways in which the self-esteem of individuals clashes or in which people seek the three principal objects over which they compete – namely, power, wealth, and honor. These, Kant says, are the means by which we hope to dominate others, making use (respectively) of their fear, their self-interest, and their opinion (G 4: 393; VA 7: 271–3). But his reference to our "hostility toward all whom we consider alien to us" is significant, in that it implies a collective dimension to unsociable sociability – encompassing national, ethnic, or religious forms of hostility between people. Kant points out that religions frequently invoke the power of deities on behalf of one nation or faith in its combat with others and use alleged divine favor as a pretext for claiming dominance for their group over another (VpR 28: 1124–5). And of course it is war between nations that Kant regards as the form of evil that, at this stage of human history, poses the greatest obstacle to the further progress of the human species (EF 8: 360–8; I 8: 24–7).

Kant is even more explicit about the social origin of evil at the beginning of Part Three of the *Religion*, where his aim is to show that the struggle against evil cannot succeed so long as each of us fights the moral battle apart from others, but has a chance of success only when people join together in an ethical community, taking the highest good as a shared or collective end and recognizing the moral law as a public (though non-coercive) law.

If [the human being] searches for the causes and the circumstances that draw him into this danger [of subjection to the evil principle] and keep him there, he can easily convince himself that they do not come his way from his own crude nature, so far as he exists in isolation, but rather from the human beings to whom he stands in relation or association. It is not the instigation of nature that arouses what should properly be called the *passions*, which wreak such great devastation in his originally good predisposition. His needs are but limited, and his state of mind in providing for them moderate and tranquil. He is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and will despise him for it. Envy, addiction to power, avarice, and the malignant inclinations associated with these, assail his nature, which on its own is undemanding, *as soon as he is among human beings*. Nor is it necessary to assume that these are sunk into evil and are examples to lead him astray; it suffices that they are there, that they surround him, and that they are human beings, and they will mutually corrupt each other's moral disposition and make one another evil. (R 6: 93–4)

This is as clear a statement as one could ask for that the radical evil in human nature arises and manifests itself only in the social condition. For Kant it is also only in the social condition that our reason is capable of developing, so it is also only in society that people could come to awareness of the moral law and could recognize evil for what it is. So it is human society which constitutes the condition both for evil and for the moral struggle against it. This same vision of the human predicament is present in all Kant's writings on human history, for example, in the Fourth Proposition of *Idea for a Universal History with a Cosmopolitan Aim* (1784):

The means nature employs in order to bring about the development of all their predispositions is their ***antagonism*** *in society, insofar as the latter is in the end the cause of their lawful order.* Here I understand by "antagonism" the *unsociable sociability* of human beings,[11] i.e., their propensity to enter into society, which, however, is combined with a thoroughgoing resistance that constantly threatens to break

[11] This phrase is taken from Montaigne: "*Il n'est rien si dissociable et sociable que l'homme: l'un par son vice, l'autre par sa nature.*" Michel Eyquem de Montaigne, "*De la solitude*," in *Essais*, (ed.), André Tournon (Paris: Imprimerie nationale, 1998), I: 388. "There is nothing more unsociable than Man, and nothing more sociable: unsociable by his vice, sociable by his nature." "Of Solitude," in *The Complete Essays*, (trans.), M. A. Screech (London: Penguin Books, 1991), p. 267.

up this society. The predisposition for this obviously lies in human nature. The human being has an inclination to become *socialized*, since in such a condition he feels himself as more a human being, i.e., feels the development of his natural predispositions. But he also has a great propensity to *individualize* (isolate) himself, because he simultaneously encounters in himself the unsociable property of willing to direct everything so as to get his own way, and hence expects resistance everywhere because he knows of himself that he is inclined on his side toward resistance against others. Now it is this resistance that awakens all the powers of the human being, brings him to overcome his propensity to indolence, and, driven by ambition, tyranny and greed, to obtain for himself a rank among his fellows, whom he cannot *stand*, but also cannot *leave alone*. Thus happen the first true steps from crudity toward culture, which really consists in the social worth of the human being; thus all talents come bit by bit to be developed, taste is formed, and even, through progress in enlightenment, a beginning is made toward the foundation of a way of thinking which can with time transform the rude natural predisposition to make moral distinctions into determinate practical principles and hence transform a *pathologically* compelled agreement to form a society finally into a *moral* whole. (I 8: 20–1)

Another Kantian name, therefore, for the radical evil in human nature is "unsociable sociability" – the sociable need that human beings have as *rational* beings for society with others, which, however, is also the unsociable need to gain superiority over them in honor, power, and wealth. Unsociable sociability makes human society the scene of inequality and conflict – all the more so, at least up to this stage of history, insofar as human beings have become cultivated and civilized, and their rational powers have developed through the prodding offered by this same social competitiveness. In the *Religion*'s account of evil, Kant makes unmistakable reference to unsociable sociability when he says that "nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture" (R 6: 27).

All these claims about human nature should be understood in the context of Kant's theory of natural teleology in the study of living things and its application to the history of the human species. Biology and human history present us with phenomena that are governed by causal laws, but they also involve a kind of intelligibility which escapes explanation through these laws. Our best access to this intelligibility is through the regulative employment of the idea of an organized

being, and of species of such beings, each with its own distinctive set of natural predispositions, for whose complete development nature has arranged. Kant's fullest explanation of this theory is found in the second half of the *Critique of the Power of Judgment* (KU 5: 359–484). In a rational species this involves a *historical* process, in which each generation receives the skills and faculties developed by previous generations and then develops them further. The competitiveness of the social condition, which is made possible by the human propensity to evil, or unsociable sociability, is the natural mechanism through which this natural development takes place. Evil is therefore intelligible (to the extent that it can be made intelligible at all) as a mechanism employed by natural purposiveness in developing our species's predispositions in history.

In *Idea for a Universal History*, however, Kant regards the era of history in which unsociable sociability can serve the historical development of human nature as having reached its limit. There he argues that it can continue to do so only if the human tendency to injustice is held in check by "a civil society universally administering right" – that is, a political state coercively enforcing laws of justice (I 8: 22–3). This, in turn, will increasingly depend on the capacity of the human species to achieve a federation of political states maintaining peace with justice between them (I 8: 24–6, cf. EF 8: 360–8). Part Three of the *Religion* develops this thought further, and in a new direction, by arguing that the moral progress of the human species depends on a different kind of human community, an ethical community, grounded on non-coercive moral laws, in principle embracing the entire human species as "a people of God, and indeed in accordance with the laws of virtue" (R 6: 99). This is the function for Kant of a religious community or church. The social model is one of friendship, or of a family "under a common but invisible moral father … a free universal and enduring union of hearts" (R 6: 102). It is this same vision, though stated in more secular terms, that Kant presents at the very conclusion of his *Anthropology from a Pragmatic Point of View*. There he is attempting to articulate the "character" of the human species as a whole, which he says can be done only historically, in terms of its moral vocation. We do this best in that judgment of our species which condemns it for its evil, which judgment also reveals in us the predisposition to good.

So it presents the human species not as evil, but as a species of rational beings that strives among obstacles to rise out of evil in constant progress toward the good. In this its volition is generally good, but achievement is difficult because one cannot expect to reach the goal by a free agreement of individuals, but only by a progressive organization of citizens of the earth into and toward the species as a system that is cosmopolitically combined. (VA 7: 333)

Kant's vision of human history as proceeding by way of unsociable sociability toward a future moral unity is obviously inspired by Rousseau's vision of the human species in his *Discourse on the Origin of Inequality*.[12] In Rousseau too, the development of our rational faculties occurs only in conjunction with the rise of social competitiveness, conflict, and inequality. The development of reason transforms our innocent, animal self-love or *amour de soi* into a new impulse, which Rousseau calls *amour propre*. This distinction is reproduced in Kant's moral psychology, with his contrast between "self-love" (*Eigenliebe*) and "self-conceit" (*Eigendünkel*) (KpV 5: 73) – making "self-conceit," along with "unsociable sociability," yet another Kantian name for the radical propensity to evil. Kant agrees with Rousseau's pessimistic assessment of civilization and the results of developing our rational faculties if one considers only what human beings have made of themselves so far. The challenge for us, however, is to make of humanity something that it never yet has been, to realize our moral vocation:

Rousseau was not so wrong when he preferred to [our civilized condition] the condition of savages, as long, namely, as one leaves out this last stage to which our species has yet to ascend. (I 8: 26)

In this manner one can also bring into agreement with themselves and with reason the assertions of the famous *J.-J. Rousseau*, which are often misinterpreted and to all appearance conflict with one another. In his writing *on the influence of the sciences* and *on the inequality of human beings*, he shows quite correctly the unavoidable conflict of culture with the nature of humankind as a *physical* species in which each individual was entirely to reach his vocation; but in his *Émile*, his *Social Contract* and other writings, he seeks again to solve the harder problem of how culture must proceed in order properly to develop the predispositions of humanity as a *moral* species to their vocation, so that the latter no longer conflict with humanity as a natural species. From

[12] J. J. Rousseau, *Discourse on the Origin of Inequality*, (trans.), Donald Cress (Indianapolis: Hackett, 1992).

this conflict (since culture, according to true principles *of education* of human being and citizen, has perhaps not yet rightly begun, much less having been completed) arise all true ills that oppress human life, and all vices that dishonor it; nevertheless, the incitements to the latter, which one blames for them, are in themselves good and purposive as natural predispositions, but these predispositions, since they were aimed at the merely natural condition, suffer injury from progressing culture and injure culture in turn, until perfect art again becomes nature, which is the ultimate goal of the moral vocation of the human species. (MA 8: 116–18)

Three objections. The centrality of this conception of human history in Kant's anthropological and ethical writings is plain enough. The evidences of the same vision in the *Religion*'s account of radical evil are, as we have seen, equally clear and explicit, even if Kant's purpose in the *Religion* keeps him from giving them, at least until Part Three, the prominence they might seem to deserve. In several earlier writings, I have tried to emphasize the social and historical context of evil in Kant's account, but have encountered several objections to the thesis that Kant regards the social condition as the context of radical evil.[13] This is perhaps a good place to reply briefly to three of them.

Objection 1: "Intelligible freedom." One objection has been that to see the social condition as the context of radical evil is inconsistent with Kant's doctrine that we are free beings only in the intelligible world. This objection is based, in my view, on some very fundamental errors about Kant's treatment of the problem of freedom and the role of transcendental idealism in resolving it. The function of Kant's idea that we might be free as members of the intelligible world is only to show that there is no contradiction in regarding our actions both as free and as subject to the causal mechanism of nature in the sensible world (see KrV A557–8/B585–6).[14] Nothing Kant says

[13] For instance, see *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), pp. 283–320, and "Religion, Ethical Community and the Struggle against Evil," *Faith and Philosophy*, 17, 4 (2000), 498–511. My eyes were first opened to this theme in Kant's anthropological, historical moral and religious thought by S. Anderson-Gold, "God and Community: An Inquiry into the Religious Implications of the Highest Good," in P. Rossi and M. Wreen (eds.), *Kant's Philosophy of Religion Reconsidered* (Bloomington: Indiana University Press, 1991), pp. 113–31, as well as by discussions of this theme with her at the conference on which this volume was based.

[14] See also my paper "Kant's Compatibilism," in Wood (ed.), *Self and Nature in Kant's Philosophy* (Ithaca, NY: Cornell University Press, 1984), pp. 73–101, especially the

could justify ascribing to him the absurd metaphysical fantasy that as free agents we are locked away in little monastic cells somewhere up there in the noumenal world. Even in the first *Critique*, intelligible freedom is explicitly described as an intelligible *faculty* that belongs to the human being *as an appearance* – hence as a part of nature or the world of sense, not a faculty belonging to a separate, noumenal being (KrV A538–9/B566–7). All Kant's writings on history and anthropology confirm that it is his intention to understand our moral condition in a natural, social, and historical context. It is simply the wrong way to take Kant's references to our membership in "an intelligible order of things" (G 4: 451–3; KpV 5: 42, 49) to interpret it as detaching free agency entirely from nature, society, and history. Kant's assertions that evil develops in us only as social beings and as part of a natural teleology in human history certainly give the lie to these sadly prevalent misconceptions. Those who raise the present objection show only that they have fallen prey to some deplorably common errors.

Objection 2: "Duties to oneself." A second objection has been that to ascribe radical evil to the social condition accounts only for those forms of evil that involve the violation of duties to others, and cannot encompass the violation of duties to ourselves. Clearly not every instance of evil action directly involves social competitiveness, and the

final paragraph on p. 99: "In assessing Kant's compatibilism, it may help to remind ourselves that his theory of timeless agency is put forward only as a means of exploiting the burden of proof in the free will problem, which falls on those who would show that freedom is incompatible with determinism. Kant is not positively committed to his theory of the case as an account of the way our free agency actually works. Indeed, Kant maintains that no such positive account can ever be obtained. Kant does not pretend to know how our free agency is possible, but claims only to show that the impossibility of freedom is forever indemonstrable. If what bothers us about Kant's theory is that it seems too far-fetched and metaphysical, then it may help at least a little to realize that once the theory has served as a device for showing that freedom and determinism cannot be proven incompatible, he is just as content to dissociate himself from it and adopt a largely agnostic position on the question how our freedom is possible." Many discussions of this article have apparently proceeded as if I had never written these words, since they have proceeded on the supposition that I meant to defend Kant's notion of intelligible freedom as a dogmatic metaphysical doctrine. If I had anticipated such profound misunderstandings, I would certainly have given this point more emphasis so as to forestall them. But experience shows that the capacity of philosophers maliciously to misunderstand what other philosophers have written is virtually infinite, so perhaps nothing I could have said would have made any difference.

violation of duties to oneself is an obvious place to look for examples of this. But to see these examples as objections to the idea that our propensity to evil lies in unsociable sociability mistakes both the problem and Kant's solution to it.

The propensity problem is fundamentally that of understanding why we have a propensity to give the rationally weaker incentives of inclination or self-love priority over the rationally stronger incentives of morality. Kant's solution is based on the observation that in the social condition what happens to us is, most fundamentally, that we come to value ourselves in the wrong way, preferring the worth of our condition – which can be compared favorably to that of others – over the worth of our person, which is measured not by comparison with others but only by the moral law (MS 6: 435–6; KpV 5: 76–7; VE 27: 349). This is just another way of saying that we prefer incentives of inclination or self-love (which have to do with our condition) over the rationally stronger incentives of morality (that pertain to the worth of our person). The violation of duties to oneself is grounded on the failure to respect one's own worth as a rational being, and this failure is most fundamentally what manifests itself in unsociable sociability and the self-conceit that goes along with it. In that sense, unsociable sociability grounds the evil propensity even in the case of violations of duties to oneself.

We misunderstand Kant's solution if we think that it requires claiming that every individual instance of evil directly involves social competition. Particular violations of duties to oneself as an animal being – cases of suicide, gluttony, or drunkenness, for instance – may have a social aspect or they may not. (I may get drunk or kill myself because I have been humiliated by my social rivals, but I may also violate the same duties from motives having nothing directly to do with social competition.) The point is rather that all such violations fundamentally exhibit the propensity to value one's state or condition more than one's person, and Kant's solution to the propensity problem is that social competitiveness is the sole and sufficient explanation for that *propensity* – whether or not social competitiveness is directly involved in its manifestation in a given case of evil choice.

Kant does think, however, that social competitiveness is involved in evil to a greater extent than we may realize, even in the case of self-regarding vices. A look at Kant's discussion of our self-regarding vices

as moral beings – lying, avarice, and servility – reveals that these too are deeply social in their context and motivation. They are manifestations of social corruption every bit as much as our other-regarding vices (MS 6: 429–37; VE 27: 399–405, 604–7). Even the vices that violate duties to ourselves arising from our animality involve desires that go beyond our innocent animal nature, and involve the social corruption we incur as social, and especially as *civilized* beings: The human being's "needs [Kant says] are but limited, and his state of mind in providing for them is moderate and tranquil. He is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and despise him for it" (R 6: 93). Kant thinks that the sweetness of our enjoyment in yielding to self-regarding vices often comes from the fact that our neighbors cannot afford the same indulgences. Here he shares the shrewd, socially and historically sophisticated Enlightenment view of human nature expressed not only by Rousseau but also by Adam Smith:

> It is chiefly from regard to the sentiments of mankind that we pursue riches and avoid poverty. For to what purpose is all the toil and bustle of this world? What is the end of avarice and ambition, the pursuit of wealth, of power and pre-eminence? Is it to supply the necessities of nature? The wages of the meanest laborer can supply them … It is the vanity, not the ease, or the pleasure [of the higher ranks of life] that interests us … Compared with the contempt of mankind, all other external evils are easily supported … It is not ease or pleasure, but always honor, though frequently an honor very ill understood, that the ambitious man really pursues.[15]

Objection 3: "Blame society, not the individual." A third objection is that ascribing the radical evil in human nature to our social condition involves making society, or other people, responsible for our evil choices, which is inconsistent with Kant's view that each of us alone is responsible for them.[16] But it is one thing to say that the social condition provides the necessary context for developing our radical propensity to evil and quite another to say that society *forces*

[15] A. Smith, *The Theory of Moral Sentiments* (Amherst, NY: Prometheus Books, 2000), pp. 70–1, 89, 83.

[16] J. Grenberg, *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005), pp. 31–42.

us to choose evil maxims, removing or diminishing our responsibility for these choices. Kant's assertion of the first thing is plain in his texts, but he never says the second. If the objectors think that the one commits him to the other, then they are criticizing Kant, not interpreting him.

No doubt a bad social environment can sometimes serve to excuse or even justify someone's conduct, which would otherwise deserve blame. But this is plainly not what Kant thinks about the social origin of the propensity to evil, and it would be deplorably simple-minded (not to say morally bankrupt) to take it for granted that all social influences on a person must equally exculpate the conduct that results from them. When Kant says that in the social condition human beings "mutually corrupt each other's moral disposition and make one another evil" (R 6: 94), he obviously means that the presence of other human beings plays a necessary role in our choice of evil maxims, but not that their influence provides us with any excuse for our wrong choices. Kant understands "a corrupted disposition" as something for which *the corrupted person* is morally responsible – otherwise it would not count as *moral corruption*. When we speak, on a less fundamental level, of bad education or bad company as "corrupting" a person's character, we mean that they play a role in *making the person corrupt or wicked*, not that they release the person from responsibility for their bad character. We saw above how Kant is falsely accused of removing free ethical choice from its natural and social context. I suspect it is no accident that the same people who make that mistake also raise the present objection – especially those who misread Kant because they are tempted to think it might be *right* to think of a choice as truly free only if it is devoid of all natural or social context.

Sometimes, however, the objection seems to be a different one: that the unsociable sociability of others puts me under overwhelming pressure to treat them badly, perhaps as my only defense against their bad conduct. But this suggestion too will not withstand scrutiny. My proper defense against others is to demand justice and respect from them, which is not to do evil. It is not to treat them with cruelty and deceit, injustice and hatred, which is to do evil. Rivalry with others does not justify my bad conduct but rather constitutes only a temptation to

evil. To offer the fact that I was tempted as an excuse for bad conduct is a pitiful ploy, inviting derision as well as added blame. Obviously the moral law commands me to resist such temptations, and Kant frequently insists that my predisposition to personality gives me the capacity to do so (R 6: 49).[17]

Kant's account does explain evil *teleologically* by showing how it serves the natural purpose of developing our species predispositions (I 8: 22–4). Does this explanation contradict the claim that we are responsible for evil? Kant does not think we can provide anything like a causal explanation for our choice of evil, describing this choice as "inscrutable to us" (R 6: 21). It would be a basic misunderstanding of Kant's conception of inner natural teleology to think that it is a species of efficient cause explanation – a determinism operating on the will through the mechanism of nature. Kant's account is *not* that "nature" causally determines us to make evil choices using "society" as its lever.

Conclusion. Kant's account makes evil intelligible in two ways: first, by identifying the fundamental maxim of evil, and second, by locating our propensity to evil within the context of our social condition and the natural teleology in its history. Reason, morality, and the propensity to evil are all fruits of our social condition, and in that condition our vocation is to employ our reason, and the principle of morality that it reveals to us, in the struggle against all the evil propensities of self-conceit, unsociable sociability, tyranny, greed, and ambition – which, however, served as the historical conditions for the development of

[17] "Suppose someone asserts of his lustful inclination, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he knows he ought to do it, and cognizes freedom within him, which, without the moral law, would have remained unknown to him" (KpV 5: 30). Keep in mind too that having a radical propensity to evil does not entail that one yields to it. (See note 10 above.)

reason and even of morality itself. Kant's attitude toward the likelihood of our success in this struggle is hardly one of confident optimism, but it is one of sober, principled hopefulness.

Nowadays we may not find Kant's theory of natural teleology a persuasive setting in which to place either a conception of human history or an attempt to make evil intelligible as a part of it. It would be more fashionable for someone to speculate that our propensity to seek superiority over others, along with the competitiveness this entails and the development of human capacities that results from it, belong to traits that were selected for early in the evolution of the human species during the period when it was socially organized into troops of hunter-gatherers on the plains of Africa. Evolutionary biology has explained many things and may explain still more, but when it comes to the contingencies of human history, I think unconfirmable Darwin-inspired just-so stories are seldom any improvement over pre-Darwinian speculations they were designed to replace. A broadly Kantian theory of history, based on a cautious employment of natural teleology grounded methodologically on a theory of reflective judgment, may still be as good as any approach to human history we have so far come up with.[18]

Kant's account of evil locates evil itself, as well as our struggle against it, in our forms of social organization and their history. It treats the development of human reason as part of our cultural history, and evil as a vehicle in this development, including the genesis of our ability to recognize evil for what it is and to struggle against it. Evil is grounded in social inequality and competition, while the struggle against it is fundamentally a striving for social changes that strengthen human solidarity and bring human purposes into that systematic agreement to which Kant gave the name "the realm of ends." Kant's attempt to make evil intelligible shows us that social

[18] Kant's theory of history, driven by the idea that the basic tendency in human history is the growth of our species-capacities, has much in common with Marxian historical materialism, though it lacks the theory of class struggle, and its foundation in regulative principles makes it more methodologically cautious. See my article "Kant's Historical Materialism," in J. Kneller and S. Axinn (eds.), *Autonomy and Community: Readings in Contemporary Kantian Social Philosophy* (Albany: State University of New York Press, 1998) and also *Kant's Ethical Thought*, pp. 244–9.

antagonism and social change provide the basic framework in which we ought to think about both the sources of and the remedies for the evil people do.[19]

[19] I am grateful for discussions of this paper at Emory University, the University of Chicago, the University of Notre Dame, the North American Kant Society Midwest Study Group at Purdue University, and at the *Von Kant bis Hegel* conference in Montréal. Particularly helpful were the questions and comments of Frederick Neuhouser, Calvin Normore, Daniel Sutherland, Jonathan Lear, Agnes Callard, Anja Jauernig, Lynn Joy, Rudolf Makkreel, and Alvin Plantinga.

8

Social Dimensions of Kant's Conception of Radical Evil

Jeanine M. Grenberg

1. A Reaction to Wood's Account of Kant's Radical Evil

Introduction

Allen Wood[1] and I[2] disagree about how best to understand the relationship of Kant's notion of radical evil to the social realm: is evil something that comes about only in society, or would I be evil even without social interactions?

For Wood, our propensity to evil develops only within society. In quasi-Rousseauian style, he understands Kant to be saying that, left in isolation, we would be "moderate" and "disposed to contentment" (R 6: 93–4, quoted at Wood, *Kant's Ethical Thought*, pp. 288–9). But once we encounter other people, our anxieties increase and lead us to prefer the satisfaction of our own inclinations over the moral demand to treat other persons as ends-in-themselves. We fall into fears of inequality, and this leads us to a comparative–competitive over-assertion of ourselves. Radical evil is thus equivalent to what Kant elsewhere describes as "unsocial sociability"[3]: we can't help but seek out human

[1] A. Wood, *Kant's Ethical Thought* (New York: Cambridge University Press, 1999).

[2] J. Grenberg, *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005).

[3] According to Wood (who, here, cites Sharon Anderson-Gold's discussion), radical evil "is not after all so far from our unsocial sociability," and "would pertain to us insofar as we are social beings" (Wood, *Kant's Ethical Thought*, p. 288). This is because evil in our nature is "closely bound up with our tendency to compare ourselves with

society (that is the "sociability" part). But when we do, our natural tendency is to prefer the concerns of the self over even the morally based needs of others (that is the "unsocial" part). Wood thus asserts that social interaction is a necessary condition for our development of a propensity toward evil and, further, that this propensity is equivalent and reducible to unsocial sociability. If we were entirely isolated beings, we would not have a propensity for preferring self-concerns to moral demands.

This is not my own, nor I believe Kant's, understanding of radical evil. My own published stance on this question explains the development of radical evil without necessary appeal to the social context.[4] In my book, I take issue with Wood's position by raising concerns about the ultimate imputability of a propensity toward radical evil that came about only through contact with society.[5] It is my goal here to articulate further my reasons for rejecting Wood's account and also to provide further details of Kant's own account which provide a preferable alternative to understanding evil, including its social dimensions. I do not, in what follows, retract my own more psychological account of the propensity to evil.[6] But I do point toward the need to place that psychological account within a transcendental grounding for the propensity to radical evil, and reflect on how to distinguish individual and social expressions of that propensity.

Society as a Necessary Condition for Evil

I begin, then, by furthering my concerns about Wood's position. Wood himself does not believe that a loss of imputability occurs on his account, but I assert that, without further articulation of the precise role of society in the development of radical evil, we cannot help but

others and compete with them for self-worth" (ibid., p. 287). For Wood, then, "our propensity to evil belongs to us only as social and historical beings" (ibid., p. 289).

4 Grenberg, *Kant and the Ethics of Humility*, pp. 39ff.

5 Ibid., pp. 35–6.

6 Finite, dependent beings with a genuine need for things outside of us in order to thrive cannot assure that everything needed to thrive will be present. In this wretched state, we find the ground of the basic anxiety of a dependent being, and the basis of our propensity to choose self-love over the moral law. All we need for the development of a propensity to evil is the fact of being a dependent being, a desire for happiness, fear of the frustration of that would-be perfect happiness, from whatever source, and a resulting anxiety.

draw the conclusion that this undermining of responsibility occurs, *contra lui*.

The worry develops in the following way: according to Wood, previous to social engagement, we are "tranquil" and "undemanding," and it is only when we enter society that we become "anxious" and strive for inequality (R 6: 93, quoted at Wood, *Kant's Ethical Thought*, pp. 288–9). The implication: if the societal context weren't there, we individuals wouldn't develop a propensity to evil. But if that is the case, the presence of society has some sort of causal or determining influence on our choice, one that undermines our autonomous choice: what we wouldn't have chosen initially somehow becomes what we choose, and the only obvious difference between the two choices is the introduction of a social context. If the presence of society has this causal influence on our choice, then, autonomy seems undermined: one could put at least some of the blame for the choice of evil not simply on our free choice but on the coincidental fact that we are in a social setting. Society is therefore "blamed" for our development of radical evil.

One might argue in defense of Wood that something being a necessary condition for my performing *x* does not always undermine my responsibility for doing *x*, and this seems right. In order for me to commit an evil act against another person, it is necessary, for example, that another person exist in proximity to me. But the fact that she exists does not undermine my responsibility for shooting her. There are thus cases in which some state of affairs could be a necessary condition for my choice, and I can still be blamed for choosing that act. But such cases do not involve the relevant condition playing an active causal role in the production of my action. There are, then, at least two sorts of necessary conditions for action: one in which the condition provides necessary material means for my action, and another in which the condition acts as a necessary causal force for my action. "Entrapment" in the legal realm is a good parallel for the second, causally efficacious condition. In entrapment, someone encourages me to do what I otherwise would not have done, if not for the enticement; and it is legally determined that the influence of the other person weakens my own responsibility for action.

So, is Wood's appeal to society more like a non-contentious necessary condition, or is it more like the entrapment example? Is there something about the presence of other persons which has causal

influence upon my resulting choices and actions? For Wood, "radical evil … pertain[s] to us insofar as we are social beings", specifically insofar as we "compare ourselves with others and compete with them for self-worth" (*Kant's Ethical Thought*, p. 287). Human competitiveness thus reveals itself only within a social setting. Something about the social setting turns a tranquil and undemanding person into a competitive and self-obsessed person.[7] But if this is the case, it is hard to read the introduction of society as anything but an egregious and causally efficacious condition, one that undermines my autonomy. It is not just that other people are present; further, their presence has some influence on the process of my choice. If this necessary condition of society is more like entrapment in this way, then Wood's appeal to society as a necessary condition for the development of evil undermines individual responsibility for that choice.

Reinterpreting Kant's Social Story

There is some apparent textual defense for Wood's implicit claim that other people have this causal influence on my choice of radical evil. Kant seems to assert that other people in society simply cause my corruption: people "mutually corrupt each other's moral disposition" (R 6: 94, cited at Wood, *Kant's Ethical Thought*, p. 289). This isn't just a tendency in me to choose radical evil when in the proximity of other people; this is the stronger claim that other persons in society quite literally (somehow) cause corruption in me. In the language of the previous section, other persons would be a very strong causal condition of my corruption.

Although Wood appeals to this passage, I don't find him taking this stronger point of view explicitly when he interprets it. Rather, he asserts that we corrupt ourselves and are responsible for it. But his position is unclear and ambiguous, since one of the grounding texts upon which he relies seems to make this stronger point that in fact other people corrupt us. When Kant asserts (and Wood quotes)

[7] I have already raised more precise concerns about this transition in my book (Grenberg, *Kant and the Ethics of Humility*, p. 34). How is it that the mere presence of other people turns contented and undemanding individuals into these comparative–competitive maniacs? Tranquil and undemanding persons encountering tranquil and undemanding persons just doesn't seem a recipe for the development of anxiety and ultimately over-assertion of the self.

a claim that in fact it is "other people" who corrupt our moral disposition, then it seems as if we finally have an explanation for how it is that society plays a causal influence on my choice: someone else made me like this!

These claims Kant seems to be making bothered me, so I went back to reread this entire passage (R 6: 93–5) upon which Wood relies. It is, at best, an obscure passage. The most curious bit (interestingly not cited by Wood) is when Kant asserts:

> If [the moral agent] searches for the causes and the circumstances that draw him into this danger [i.e., the danger of being "constantly under attack" by "evil" and its "dominion"] and keep him there, he can easily convince himself that they do not come his way from his own raw nature, so far as he exists in isolation, but rather from the human beings to whom he stands in relation or association. (R 6: 93)

He then goes on to assert the familiar Rousseauian story about how someone "moderate" and "tranquil" becomes "anxious" once in society, thus resulting in "envy, addiction to power, avarice," etc., culminating in the claim that human beings "will mutually corrupt each other's moral disposition and make one another evil" (R 6: 94).

There is a lot that is interesting in this passage. First, it is a curious point at which to assert this quasi-Rousseauian position: Kant is deep into a discussion of radical evil at this point, and yet he seems to be asserting that humans in their "raw" state aren't evil, radically or otherwise. Such a position would go directly against what he argued earlier in the *Religion*. There is no good explanation I can see for why, having just completed a discussion of a propensity to evil which is both "radical" and "natural,"[8] that he would turn around and assert instead that in their "raw" or "natural" state, humans are not subject to this propensity.

And, in fact, I don't think he does say this. Or at least: this passage is open to an alternative interpretation. The other interesting, if obscure, part of this passage is that Kant presents the ideas in it from the perspective of a person in this "perilous" state, trying desperately to figure out how he got into this "danger." He suggests that this

[8] R 6: 37, 21, and 25.

person "convince[s] himself" that the causes of this condition didn't come from himself at all, but from "the human beings to whom he stands in relation or association." Kant is here describing the thought process of a moral agent in the process of deceiving himself into believing that he is not the cause of his own perilous evil state. This interpretation is further encouraged by the fact that this paragraph begins with the unequivocal claim that "[t]he human being is … in this perilous state through his *own* fault" (emphasis added). The only way to get from that to the claim that human agents "mutually corrupt each other's moral disposition and make one another evil" is to understand the latter as the conclusion of the self-deceived ramblings and wishful thinking of an already radically evil Rousseauian! This tendency toward self-deception (the tendency, that is, to blame others for what is in fact imputable only to oneself) comes from the already existing propensity toward evil in this agent. As such, societal interaction is not so much the condition for bringing about this tendency to think "oh, it wasn't my fault; it was others who made me evil"; rather, it provides the best example of the already present (and "natural") human tendency to prefer the self to morality.

Evil Propensities and Evil Actions

A further problem with Wood's account is that it effaces distinctions that are crucial to Kant's account of evil. For example, although Wood mentions, early in his discussion of evil, that evil must be understood as a disposition or "motivational propensity" (*Kant's Ethical Thought*, p. 285), one that is "prior to every use of freedom" (R 6: 22, cited at p. 286), he does not explain these ideas, nor stick to describing evil in these terms. Rather, by equating the propensity to evil with "the propensity to ascribe greater self-worth to oneself than others, preferring one's own interest to theirs through the delusion that one is better than they are" (ibid., p. 290), Wood collapses the distinction between this original propensity previous to any act of freedom which places concern for self above concern for morality, and a mere empirical propensity to place concern for self above concerns for other persons. Our propensity to evil is thus reduced to "a wickedness in the way of thinking about others" (VE 27: 691, cited at p. 288). But in reducing evil to a tendency in our interactions with other persons, Wood seems

to have forgotten both that choice of this propensity is "prior to every use of freedom" (including those involving other people), and that evil is a tendency to place concerns for self over "morality" or "the moral law" (R 6: 36), not simply over "others."

I will address the latter point in Section 2. To address the former: our chosen propensity toward radical evil is both "[t]he subjective ground … of the exercise of the human being's freedom in general," and "antecedent to every deed that falls within the scope of the senses" (R 6: 21). Kant's story of this propensity is an account of how all humans share the same nature which prefers concerns of the self over concerns of morality, a character or nature chosen previous to any this-worldly exercise of freedom. The language here is so strong as to suggest that, whatever this propensity turns out to be, it is not reducible to any psychological condition; it is, rather, a condition for the possibility of the exercise of our freedom in time. In making this previous-to-time "choice" of our "nature," we are defining that capacity known as human freedom. Indeed, without this propensity to act against morality, we couldn't really speak of the practice of "morality" as we know it at all. To be moral is to do what is right in the face of temptations to the contrary; and it is our choice of a propensity toward radical evil that explains the fact that we do have such temptations. This choice of radical evil is, simultaneously, the choice of ourselves as a certain sort of moral being.

We thus have further reason to reject Wood's suggestion that evil requires a social context. Our propensity to evil is previous to every exercise of freedom, including any social interactions (however defined). To put the point more bluntly: society does not inspire evil in us; rather, we bring evil to society. However ironic, a propensity toward evil is a condition for the possibility of our exercise of freedom as finite moral beings.

Wood's willingness to ignore this assertion of an evil nature previous to the exercise of freedom leads to a collapse of a further important distinction between a propensity toward evil and actual, explicitly evil acts.[9] But there is a large difference between the two. Kant is not interested simply in the fact that human actors choose individual evil

[9] Wood's use of the word "evil" throughout the course of his discussion has a vague reference, wavering between the "propensity" and "act" descriptions of it. He says,

acts. More deeply than that, he is seeking to articulate a quality of person or character – really, of human nature itself – which underlies any act.

We can textually ground this distinction by returning to the only remaining passage upon which Wood relies, R 6: 27, where Kant discusses the "predisposition to humanity." This predisposition is "physical and yet involves comparison … [T]hat is, only in comparison with others does one judge oneself happy or unhappy" (R 6: 27). It is thus a predisposition to seek our happiness as physical beings through comparison of our state to that of others. This originally "good" predisposition to seek happiness has "grafted" upon it the "diabolical" "vices of culture" such as "envy, ingratitude, joy in others' misfortunes, etc." Wood uses this passage, especially the claim that these vices arise "only in comparison with others," to support his claim that society is a necessary condition for radical evil.[10]

Yet, once we accept that a propensity toward radical evil of humans generally is different than evil acts, we can understand this passage differently. To describe a predisposition, and then explain how vices are grafted onto it, is not an explanation of the development of our propensity toward radical evil. Rather, it is an explanation of the influence of an already existing propensity to evil upon these predispositions (which are, in themselves, predispositions "to the good", not evil).[11] If these predispositions really are predispositions to the good, then we have to appeal to something beyond them in order to explain their development into "vices." This is the best way of understanding Kant's language of "grafting": these predispositions would not, of their own natures or energies, grow into something vicious; it is only when something else (our already existing propensity to evil) is "grafted"

for example, that "evil is a product of society" (Wood, *Kant's Ethical Thought*, p. 286), that "evil has its source in social comparisons and antagonisms" (ibid., p. 288), and that "Kant can be called an individualist about moral responsibility for evil" (ibid., p. 289). To speak so generally of "evil" in these passages leaves one wondering whether Wood is still speaking of the development of a propensity at all, instead of just the development of the choice of particular evil acts.

[10] Ibid., pp. 288–9.

[11] "All these predispositions in the human being are not only (negatively) good (they do not resist the moral law) but they are also predispositions to the good (they demand compliance with it)" (R 6: 28). This has to mean that comparison of ourselves with others is not, contra Wood, always necessarily a bad thing.

upon them that these vicious states result. The vices associated with these predispositions "do not of themselves issue from this predisposition as a root" (R 6: 26). Rather, a naturally good predisposition toward assessing my own happiness in comparison with others turns bad and "gradually" becomes an "unjust desire to acquire superiority for oneself over others" (R 6: 27) when put into contact with our already existing propensity toward evil. Kant is thus not explaining the development of our propensity to evil here; rather, he describes how this propensity can be the ground of further choices via the influence that this already existing propensity has upon a predisposition in the human. Radical evil is a claim about a propensity in one's nature; but here we appreciate how this propensity expresses itself in individual acts through its interaction with our other predispositions.

As such, Kant's appeal to the evil comparative tendencies humans engage in when they find themselves in society is an appeal to only one particular form of expression of the propensity to radical evil, and not to the propensity to radical evil itself. The general maxim, already internalized in one's character, of placing concern for self above concern for morality becomes the basis for a more particular maxim by which I guide the choice of a specific envious and spiteful act.

Wood's effacement of the distinction between a previous-to-time propensity to evil and evil acts and his reduction of the tendency to place concern for self above morality to the social tendency to place concern for self above others are thus problematic interpretative moves. Radical evil is at its heart a propensity, formed previous to any empirical exercise of freedom, to place concerns for self over moral concerns. This is a propensity with which we are born, which helps to define the nature of our freedom in time, and which grounds specific evil acts.

I do not deny that many questions remain here: if existence in society – or indeed any empirical, psychological experience – is not a necessary condition for the formation of a propensity to evil, then how do we explain the formation of this strange, previous-to-time propensity/nature? Will such an explanation be any more successful at describing the development of our propensity to evil as a genuinely free act? Further, how does this transcendentally chosen propensity to evil relate to empirically expressed propensities identifiable in our experience, including social expressions? To answer these questions fully, it would

be necessary to supplement the psychological account I provided in my book about the loss of the dream of perfect happiness with a more fully transcendental one, then articulate their relationship. Such an account has complexities, both metaphysical and moral, into which I will not enter here.[12] Suffice for now to say that this choice is not only previous to our entry into society, and previous to any empirically experienced psychological fears about the loss of happiness; it is previous to our birth, and a condition for the possibility of our first exercise of freedom. Even without this fuller account, we can, however, make more sense of the relationship of the propensity to evil and its empirical expressions, both social and individual. Let us turn to that task.

2. Social Dimensions of Evil

The Social, Kantian-style

To make sense of a specifically social expression of evil, the first question to raise is one of what it means, for a Kantian, to be "in society." Yet, in pursuing an understanding of the social for Kant, it is crucial to remember Kant's philosophical commitments that emphasize the value and the centrality of the individual.

Most central among these is his commitment to human autonomy: we legislate the moral law to ourselves (G 4: 432–3); and, when we act on that law, it is through a capacity to be first in a chain of causes, or to "begin … a state from itself" (KrV A533/B561), instead of being merely a member of an already existing natural causal series. A Kantian must thus reject the idea that any of our choices are forced upon us by social influences. Whatever social influences exist, the individual moral agent can, through her freedom, stand above them, by being first in a chain of causes which starts with her choice and ends in her action.

[12] Essentially, we need to speak both of a transcendental and a psychological anxiety at the ground of finite, dependent beings. Through appeal to the former, we affirm that our choice of the propensity toward evil is previous to any empirical, psychological experience, including our individual experiences of the loss of things we need to thrive. This transcendental anxiety is, more abstractly, the anxiety about being a particular kind of being, that is, a being who seeks both happiness and morality, and who cannot achieve the former entirely on one's own. This "choice" of our nature is brought to every empirical exercise of our freedom, finding expression in both individual and social acts of evil.

A related individualistic commitment is the notion of worth applied to every individual agent in virtue of these moral capacities (G 4: 428). Even Kant's notion of duties involves the admittedly counterintuitive idea that we have duties to ourselves which are "previous" to duties to others, thus giving a certain priority to individual concerns in relation to social concerns (MS 6: 417–18).

Finally, and ironically, even a definition of the social as the realm of "unsocial sociability" assumes a strong commitment to the existence of individual purposes to get off the ground. Our desire for "sociability" becomes "unsocial" precisely because we prefer the needs of the self to the needs of community, and this tension between self and society defines the social space. The human's "inclination to live in society" (toward which humans are inclined because they believe they will be able best to "develop [their] natural capacities" is in tension with the same being's "great tendency to live as an individual, to isolate himself, since he also encounters in himself the unsocial characteristic of wanting to direct everything in accordance with his own ideas" (I 8: 20–1, translation slightly altered). It is only because humans have strong tendencies to emphasize individual wants and needs that the social space is as it is.

All these concerns for the individual thus place limits upon the realm of the social in Kant's writing. This does not mean that Kant is unconcerned with society; but it does encourage caution as we move into our understanding of the social realm.

Let us, then, define this realm of the "social." Wood has assumed that we are in society whenever we are in the presence of another person. But Kant himself draws distinctions among our predispositions to animality that discourage such a simple equation. "The predisposition to animality" is broken into three parts, including predispositions to "self-preservation", to "propagation" and "preservation of offspring", and to "community with other human beings, i.e., the social drive" (R 6: 26). Note that only the third aspect of this predisposition – the predisposition to "community" – is described explicitly as a "social" drive. The previous two drives to "self-preservation" and "propagation/preservation of offspring," though they can (one might argue "must") involve interaction with other people, are not described as specifically social. Only the predisposition to community with other human beings is described as "the social drive."

Given this, it is apparently the case that Kant does not understand "society" to be equivalent simply to "interacting with other people." If that were the case, then the self-preservation and propagation aspects of the predisposition to animality would be social drives as well; but they are not. Whatever the social is, it is something beyond the mere interaction between male and female in the propagation of the species, and beyond the interaction between parent and child in the preservation of the species.

How then to define the realm of the social for Kant? One might be tempted to appeal to the distinction between duties to self and duties to others. But even this is too easy a division. Kant himself, when discussing duties to self and others in the *Metaphysics of Morals*, often comments on how a duty targeted toward one in fact has aspects of the other. Suicide, for example, violates one's duty to oneself, but also violates duties toward others (MS 6: 422). Envy, a failure of beneficence, is a failure in one's duties to others, but also a violation of one's moral self. It is a "sullen passion that tortures *oneself*" even as it seeks to "destroy … *others'* good fortune." As such, it is "contrary to one's duty to oneself as well as to others" (MS 6: 459, emphases added). To say, then, that the "social" is defined by that realm of activity within which we are held to duties to others, and that the "individual" is defined by that realm of activity within which we are held to duties to self, is too simple.

Let us try a another possibility for defining the social: the "social" is defined not simply by being in interaction with other persons, nor through appeal to the duties we have to others as opposed to the self but, rather, through appeal to the fact that we share purposes with other persons. There are certain prima facie problems in defining the social in this way, but I believe they can be overcome. The largest problem is that we do not find in Kant's writings the articulation of a set of duties which are explicitly social in this way and which would affirm the moral value of these shared purposes and of society itself. We have duties to self and duties to others, but no moral duties to a group or a group's shared purposes.

Despite the lack of such discussion, Kant does emphasize the value of a moral community in the Kingdom of Ends formulation of the Categorical Imperative. There, it is not just the value of an individual rational being, qua individual, with which we need to be concerned.

Beyond that, moral weight is claimed for a community of rational beings, qua community, albeit an ideal instead of real community:

> For, all rational beings stand under the law that each of them is to treat himself and all others never merely as means but always at the same time as ends in themselves. But from this there arises a systematic union of rational beings through common objective laws, that is, a kingdom, which can be called a kingdom of ends (admittedly only an ideal) because what these laws have as their purpose is just the relation of these beings to one another as ends and means. (G 4: 433)

The crucial point here is that the ultimate "purpose" of the moral law is a social, and not merely individual, purpose. Despite the fact that moral worth is located within the individual rational being, and that the law is enacted through individual autonomous acts, the ultimate purpose of this law is to achieve a "systematic union of rational beings," that is, rational beings related to each other morally "as ends and means." The ultimate goal of our moral selves is to exist in a rational kingdom, that is, in "a systematic union of various rational beings through common laws" (G 4: 433).

Kant asserts that it is only through the actions of individual rational beings that this ideal will be made real: "a kingdom of ends would actually come into existence through maxims whose rule the categorical imperative prescribes to all rational beings if they were universally followed" (G 4: 438). One does not anticipate this realization of the kingdom of ends in time; yet one still sees one's own moral acts as a piece of what moves us toward this systematic union of rational beings, or community. When we think of the "social" for Kant, we should thus think not simply of the empirical realm in which persons in fact interact, or even the empirical realm within which persons in fact share purposes with other person (as, for example, when male and female share the purpose of propagation). Rather, we should think of this ideal moral society in which persons share the purpose of realizing the Categorical Imperative.

Furthermore, Kant asserts that we do have a duty to pursue this ethical social ideal. Achievement of our ultimate moral purpose depends not only upon individual moral acts, but also upon "the setting up and the diffusion of a society in accordance with, and for the sake of, the laws of virtue – a society which reason makes it a task and a duty of the

entire human race to establish in its full scope" (R 6: 94). We do, then, have a certain "duty" to society, a duty, that is, to the development of a specifically ethical society, and a duty to achieve that society through shared moral pursuits, not simply through individual moral acts.

To the extent, then, that an actual community of rational beings takes this ethical community ideal as regulative, guiding its own systematic unity as a community, to that extent can we speak of there being a moral value attributed to a community of rational beings, not just qua individual, but qua community.[13] There is indeed, for Kant, a moral social value. This moral community is pursued in any act whose maxim is in agreement with the Categorical Imperative. In this formal sense, every moral act is an act with social concern. Yet to realize this "systematic union of rational beings," we can also speak of acts with a more explicit social concern, those which involve shared purposes amongst individuals, those acts intended explicitly to constitute, enhance, and support this systematic union of rational beings. The social for Kant is thus the realm within which we share specifically moral purposes with other persons.

It is in entering into such shared purposes that we pursue the systemic union of rational beings. As Kant puts it, humans have a "tendency to come together in society," or an "inclination to live in society" because we feel more able to develop our "natural capacities" in such a social setting (I 8: 20–1). That is, we feel more able to realize our capacities when we enter into the shared purposes definitive of and necessary in society. Indeed, the "moral" end of these "natural capacities" is the systematic union of all rational beings.

It is important to emphasize that we are not saying society acts as one large autonomous being. Rather, society always consists in individual autonomous beings who share the ethical purpose of creating this systematic union of rational beings. Rather, when we say that persons share a purpose, we are making the more modest claim that individual agents each hold a maxim sufficiently similar that we are able to say of them, as a group, that they are concerned with the same goal,

[13] This "ethical community" can "exist in the midst of a political community and even be made up of all the members of the latter (indeed, without the foundation of a political community it could never be brought into existence by human beings). It has however a special unifying principle of its own (virtue) and hence a form and constitution essentially distinct from those of the other" (R 6: 94).

value, or purpose; and that there is some at least minimal awareness that this maxim or maxims are so shared. These shared purposes are just what individual agents enter into when they enter society with others, the first half of our unsocial sociability. Individual, autonomous beings seek out society, that is, a community of persons with whom to share purposes. That group of persons can be variously circumscribed according to differing morally relevant purposes, the largest group being humanity itself, which shares the regulative, guiding moral purposes which Kant attributes to all of us.

Individually Evil Acts

Since the ultimate goal of the moral law is the social goal of the kingdom of ends, there is a sense in which every immoral act is also a socially evil act. But we can now draw some finer distinctions between individually and socially evil acts.

First, we can affirm a realm of individually evil acts. Wood asserts that evil is a tendency to place concern for self above others,[14] thus equating evil with unsocial sociability. Yet Kant asserts that evil is a tendency to place concerns for self over "morality" or "the moral law" (R 6: 36), not simply over "others." Further, our tendency to over-assert ourselves above morality is not always expressed as a tendency to prefer our concerns over the concerns of others. It can sometimes be an interest in preferring our desires over moral constraints upon treatment of ourselves. In other words, there can be individually evil acts.

This point is supported in Kant's texts. Corruption of the predisposition to humanity is clearly an example of how our tendency to put concerns of self above morality can be expressed as putting self above others. But corruption of our predisposition to self-preservation (R 6: 26) reveals that our tendency to prefer the self over morality is not always a preference of self over others. Evil is grafted onto our predisposition toward self-preservation when one is tempted to commit suicide. Kant's *Groundwork* example of the person who considers suicide to relieve current pain is an example of just such grafting. The maxim of action the would-be suicide considers is: "[F]rom self-love I make it my principle to shorten my life when its longer duration

[14] Wood, *Kant's Ethical Thought*, p. 288.

threatens more troubles than it promises agreeableness" (G 4: 422). This man's failure is a failure to cultivate his original predisposition to self-preservation. The undermining of this cultivation occurs through a failure to have concerns related to the self in their proper order. One concern of the self is to relieve pain and "troubles." Another is to become the best moral person one can. This person's consideration of suicide puts the former concern above the latter. But putting the relieving of pain (at any cost) ahead of his interest in his moral development is an improper inversion of the priority of his concerns. He is putting concerns of his physical self above concerns of his moral self; in the language of evil, he is putting concerns for self above morality.

The rejection of "morality" as primary here does not occur as an assertion of oneself over others. This person is not putting his concerns above the proper, morally grounded concerns of others (though he might be doing that too). Nor is he killing himself in order to undermine the cohesion of society (though his act might indeed have serious repercussions within that society). Rather, at the heart of his evil act is his motivation to put lower concerns for self above higher concerns for self. He thus engages in an evil act directed against the obligation he has to himself, which does not take other persons or society as its target. It is possible, then, to express our propensity toward radical evil, this tendency to prefer the self over morality, in a way that is not simply a preference of the self over others or society.

We have, of course, already noted that society is not defined simply by being in relation to others. Conversely, then, the realm of the individual is not defined simply by appeal to this realm of duties to self and their violation. Furthermore, we affirm that society is pervasive, so we do not assert that this violation of duty to self somehow occurred in isolation from all social influences, or lacked all social consequences. Rather, our definition of the realm of individually evil acts is more circumscribed. There are few, if any, examples of actions that have no influence from or consequences for the success or failure of the purposes we share with others. But, as we know, consequences are not the be-all and end-all of morality for Kant. More central is the question of what motivates one's action. Because of Kant's emphasis on motivation as central to the value of any act, and because of his deep underlying commitment to the value of the individual, we can identify acts which, despite their embeddedness in the social space, have a

distinctively individualistic quality to them. An act motivated by support or violation of the value of that individual person who is oneself is an explicitly individual act within an unavoidably social context. When I am motivated to place lower concerns for self above higher concerns for self, I thus engage in an individually evil act.

As such, not all evil acts are explicitly social failures. It is as possible for me to be evil to myself as to others. Violation of self may even have its social dimension (e.g., that I was raised in a particular fashion, according to certain societal standards, that the society I am in tends to reject what is important to me personally, and so on). None of this undermines the idea, however, that in choosing to commit suicide, I am violating (amongst other duties) a non-social duty to myself. I am taking the value of my own person as the target of my evil act, and I value myself improperly relative to my best self. There are thus expressions of our propensity toward evil that are not social at their core. Even if a drive has social conditions for its development, and social consequences in its realization, I can still identify an individual drive to maintain and further my existence, and I can still recognize the value implicit in the drive as a value of the self or individual. Furthermore, it is morally sufficient that I affirm such predispositions from this entirely individualistic purpose: when I choose to maintain my life through recognition of the value of my life, I am acting morally. Conversely it is sufficient, for an act to be evil, that it be targeted only at undermining the value of my own person, without appeal to the value of others or of society. When the motivation of an act involves valuing lower aspects of the self over higher aspects of the self, it is an individually evil act.

Socially Evil Acts

I would not, however, claim that suicide cannot be a socially evil act; conversely, although some acts of evil are targeted against other persons, not every act so targeted is a socially evil act. We thus need to say more in order to define socially evil acts.

We know that we need to focus on the motivation of the act to determine its main focus or target. But we can envision actions which lack an explicitly social motivation (that is, which lack motivation involving concern for the moral purposes we share with others) which

are not thereby "anti-social" or contrary to such purposes. Some acts support the value of the individuals (as already suggested) or are neutral in relation to our social purposes, neither supporting nor detracting from them. It is morally sufficient for me to be motivated purely individualistically to preserve my moral person, though the fact that I have this strength is traceable to my good upbringing and education, and the fact that I do maintain myself has the positive consequence that society can benefit from my capacities and that the regulative ideal of the kingdom of ends is that much closer to actuality. Despite all these social precursors and consequences, the motivation to maintain my moral being can have an individual, not a social, target: I am obligated to treat my own self as worthy, regardless of any social influences or consequences. As Kant puts it: "[A] human being's duty to himself as a moral being only … consists in what is formal in the consistency of the maxims of his will with the dignity of humanity in *his* person" (MS 6: 420, emphasis added). Such motivation is not so much anti-social as it is non-social; it is concerned with the value of the individual qua individual, and not with the individual insofar as she is a social being.

Some motivations are explicitly evil, guided, that is, by the maxim of preferring the concerns of self to the concerns of morality. But, as we have seen, not all evil actions are motivated by socially evil concerns. We can, nonetheless, speak of an evil act, one which is a violation of duties to self or others, as also being an act whose motive targets society, that is, whose motive is intended to undermine the moral value inherent in shared social purposes. A socially evil act is thus one which (in addition to being a violation of duties to self or others), in its motive, challenges the moral value of this moral community of rational beings.

To appreciate this definition, let us return to our case of the suicide. We have already asserted a sense in which suicide can be individually evil: when the motive of the suicide is to place a lower concern for self above a higher concern for self. In our example of the suicide, the most direct duty he was violating was his duty to himself as a moral being to uphold his dignity. That is why it was "evil." We can, however, understand this same act of suicide as coming from a more complex motive, a motive in which, in addition to (or instead of) targeting the value of himself in his act, the suicide targets some social purpose: in

killing himself, he is intending to injure society as much as himself. Consider, for example, a young person intent on "getting back" at his family, his school, or his town for the real and perceived insults and injustices he has received from them by killing himself in a very public way. Or, more apropos of this time in history, consider the suicide bomber who intends to subvert the smooth functioning of society via his dramatic destruction of himself and of otherwise safe modes of public transportation. In these cases, it is not just that certain social consequences will ensue which makes an act socially evil. Rather, it is that this person has, in his motive for action, taken society as his target. It is the value of the shared purposes of the community of which he has been a part that is here violated (whether or not the expected consequences do in fact occur).

Socially evil acts are thus those motivated by insulting or undermining those social purposes which are also moral purposes (that is, which happen also to be consistent both with duties to self and duties to others, and which therefore further our movement toward the ideal community of a kingdom of ends). A socially evil act can thus be a violation of either (or both) duties to self or to others. But in addition to being a violation of either (or both) of these sets of duties, it would also undermine social purposes. Violations of both categories of duty thus have the potential to be "anti-social"; and both have the potential to lack this social dimension (though it might be less common for violations of duties to others to lack this dimension).

This undermining of social purposes is best understood in Kant's language of "unsocial sociability." We thus find ourselves returning to the territory Wood first introduced, but now with some more careful constraints on our understanding of it. To be unsocially social is to be both concerned for and destructive of social engagement. We have already seen the first half of this equation: we can't help but seek out those morally relevant social purposes which define us as a society. But our natural tendency is also to prefer the concerns of the self over the needs of this social entity. We can't get away from our "great tendency to live as an individual, to isolate [ourselves]" (I 8: 44). Our already existing propensity toward radical evil encourages us to prefer ourselves to the satisfaction of social purposes. We are thus in constant tension as moral, social beings. We want, need, and are morally obligated to engage in shared purposes with others; yet we are beings

with a propensity toward radical evil, and thus tend to prefer our own perceived needs above those of society. When a human engages in social interactions, it is "his fellows, whom he cannot bear yet cannot bear to leave" (I 8: 21, translation altered). Our social space is defined as much by our desires to live as individuals as it is by our desires to share purposes with others.

But even in this social space, we do not define socially evil acts simply as those in which we indulge in a comparative–competitive over-assertion of ourselves over others. This is certainly one way in which an individual might undermine the pursuit of the very purposes which we share as a society, but as we have already seen, this is not the only way, because a mis-valuing of concerns for self can have the same undermining of social purposes. Our unsocial sociability is thus defined not as an assertion of the self over others but as an assertion of the self over society and its morally relevant purposes.

We thus reaffirm our earlier point that unsocial sociability is not simply to be equated with our propensity toward radical evil. Rather, that propensity is a choice previous to any exercise of freedom in which we choose our nature or tendency to place concerns for self above morality. Unsocial sociability, on the other hand, is one way this tendency expresses itself, an explicitly social expression of our propensity toward evil found both on the level of character and action. This is unsocial sociability: even as we find ourselves drawn to life in society, we are simultaneously drawn to undermine its purposes in favor of our own; we prefer the concerns of self to the achievement of shared social purposes. This unsocial sociability is one very precise, yet also very prevalent, expression of our propensity toward evil.

Unsocial sociability is thus our propensity to evil, chosen previous to time, now expressed in time and, more precisely, in our social interactions with others, as a natural tendency both to seek the sharing of purposes with others in society and also, at the very same time, to prefer the indulgence of individual concerns in a way that undermines those social commitments. Kant thinks that, because of our evil natures, the social space is this inherently unstable space.

With this clearer definition of the realm of the social in hand, we can return to the original question of whether and to what extent society is a necessary condition for our choice of evil acts. We rephrase the question thusly: to what extent does the fact that we unavoidably exist

as holding shared purposes with others act as a necessary condition upon our choice of acts which express our propensity to evil (i.e., our propensity to place the self above morality)?

We have already acknowledged that the social context for action is pervasive. It is dangerous, however, and too quick, to assume from the pervasiveness of the social context for any action that being in society is a strong or problematic necessary condition for all our choices, evil or otherwise. One can assert that the social, this realm within which we share purposes with other persons, is the unavoidable universal context within which all our choices are made. Yet this does not lead to the stronger conclusion that every choice made within that context required that context in some way and could not have been chosen as it was without the causal influence of some aspect of that context. One would do better here to work on a case-by-case basis, investigating empirically what influence some aspect of the fact that I share purposes with other persons does or does not have on the choices I make. We can thus admit, trivially, that all such choices occur within a social context, since we accept that few, if any, parts of our lives could occur previous to developing shared purposes with others. This is a trivial way of understanding society as a necessary condition for our choices, because all we are saying here is: "My presence in society is necessary for me to choose to do something evil in society." True, but tautological, and morally uninteresting.

But does the presence of this social context play a more meaningful causal role in the choices that I make? Would I not engage in evil acts at all if not for the social context within which I discover myself? Again, the answer here has to be "no." A Kantian is committed to the autonomy of all human agents, and this is a strong trump card to play: we are always responsible for our actions, and even the strongest of social influences does not undermine responsibility for an act that is truly our own. If we hold to autonomy strongly, and especially to the idea that, in an autonomous act, I am first in the chain of causes leading from my choice to my action, we end up insisting that every social condition has to be a non-egregious condition of our choice, more like the idea that being in proximity to another person is a necessary condition for me choosing to injure that person. Society is the unavoidable context for all our actions, but it is not a strong necessary condition for any of our choices. Society has no causal influence upon

my choice such that the choice made is brought about by the fact that I share purposes with other persons. Whatever the social influences to which an agent has been submitted, these influences can be overcome in the name of morality through our powers as rational beings. The question of whether this evil act wouldn't have been done if not for being in society ends up being a rather un-Kantian question.

So, contra Wood, society is not a necessary condition either for our original propensity to evil or for the expression of that propensity in individual acts. It is, of course, a trivial necessary condition for the social expression of evil. But that is to draw a distinction that Wood does not, and is only one of the ways in which evil can be expressed. That it is probably the most frequent, most egregious, most common expression of evil does not mean that it is the only sort of it.

3. Conclusion

We have, then, succeeded in marking out a different way of understanding the relationship of the social to Kant's propensity toward evil. The propensity toward evil itself is something previous not only to social engagement, but to any empirical exercise of freedom. Human beings bring this nature with them into the social world and, within that world, find themselves able to express that general propensity to place the self above morality both in an individual and social form. Society is not so much a condition for the development of this propensity as it is its most egregious expression. Yet even within a ubiquitously social space, we discover also the capacity for individual acts of evil. Evil, sadly, takes many forms.

9

Kant, Radical Evil, and Crimes against Humanity

Sharon Anderson-Gold

Introduction

No philosopher has been more committed to the idea of the moral progress of humanity than Immanuel Kant. Yet, despite great technological progress, the twentieth century has witnessed numerous instances of organized violence, from wars spanning continents to the internal oppression and mass murder of citizens by the state. Because of its collective nature, moral gravity, and scope, no phenomenon challenges our hopes for moral progress more than the targeted slaughter of innocent civilians that we now refer to as "genocide." Given that Kantian ethics directs us to the moral improvement of the species, it would appear that Kant's moral theory bears a heavy burden in terms of our ability to explain and respond to these extraordinary crimes. What can Kantian ethics tell us about the nature of genocide?

Kantian ethics demand respect for humanity as an end in itself. Genocide, insofar as it rejects the value of the humanity of some groups, would appear to be a principled rejection of that principle and so an attack on the very basis of Kantian ethics. Yet, a principled rejection of morality, Kant claims, is not possible for a human will. Evil for Kant is defined in terms of the universal principle of self-love, which critics maintain is too shallow and tame to express the depths of harm that characterize genocide. Does genocide therefore fall outside the parameters of Kantian ethical theory? This paper argues that, far from falling outside the parameters of Kant's ethical theory, Kant's

conception of radical evil, in the form of vices of culture, is well suited to comprehend acts of collective evil such as genocide. In contrast to much traditional interpretation, which views self-love as an "egoistic" manifestation of the desires of an isolated individual, I maintain that Kant does not limit self-love to the interests of a physical self. Individual identities arise in a social context where self-love shapes itself in accordance with the interests of those with whom we identify. Group identity also occurs in a comparative and often competitive social environment, where groups themselves compete for rights, status, or privileges of various sorts. Although the preconditions of genocide are complex, group-based identity conflicts play a central role in shaping group-based acts of aggression. Given that such group-based conflicts are at the heart of Kant's conception of radical evil, the evil that is both innate and universal to humanity, it is my thesis that we can learn much about the nature of genocide from Kant's conception of radical evil.

The "Inhumanity" of Genocide

"To think about genocide," says Bruce Wilshire, "is to accept an invitation from hell."[1] This statement suggests that there is something about genocide that both defies ordinary human understanding and arises from qualities that are nonhuman. Although the wholesale destruction of peoples during times of war has not been uncommon in human history, our evolving conceptions of human rights have placed the indiscriminate killing of enemy noncombatants outside the moral pale. Indiscriminate killing, we like to think, belongs to our moral past. We might liken genocide to a type of war where one group within an organized society defines another as an "internal enemy." But this analogy would be misleading, because, unlike war, genocide often involves ordinary people not trained as killers in acts of violence against others who have not engaged in acts of aggression. Genocide goes beyond physical aggression in attacking the cultural identities of those deemed to be the enemy, thereby adding to violence a challenge

[1] B. Wilshire, *Get'Em All! Kill'Em!* (Lanham, MD: Lexington Books, 2006), p. ix.

to our very definitions of "humanity." Rather than disrupting moral life from the outside, as is the case with war, genocide violently separates social groups from within. In becoming entwined with everyday life, genocide threatens us with the loss of our moral compass and thus becomes an invitation from hell. In understanding genocide as a collective crime, we seek to understand what Alessandro Ferrara has called "evil as a temporarily shared form of life."[2]

Genocide is a "collective evil." As a crime defined in international law, genocide has a group-based character both with respect to the intended victims and with respect to the organized plans implemented by the perpetrators. A systematic plan of group destruction is generally regarded as a necessary component of genocide, distinguishing this crime from the hostile and discriminatory acts of individuals called "hate crimes." Although collective evil so defined takes a legal form different from individual evildoing, Kant's notion of radical evil encompasses them both.[3] It is a common misunderstanding of Kant's conception of evil that it is limited to selfishness or "egoism," and thus has no room for the more extensive or "extraordinary" passions. It is a strength of Kant's notion of radical evil that it illuminates how genocide can arise from a form of evil common to human beings and can therefore be understood and responded to within our ordinary understandings of morality.

Genocide as an "International" Crime

The crime of genocide is distinguished in international law from mass murder under domestic law, because the intention to destroy a designated group is regarded as a "crime against humanity" – a

[2] A. Ferrara, "The Evil That Men Do: A Meditation on Radical Evil from a Postmetaphysical Point of View," in M. P. Lara (ed.), *Rethinking Evil: Contemporary Perspectives* (Berkeley: University of California Press, 2001), pp. 178–9.

[3] Although individual criminal acts may be motivated by hatred for members of a group, I am primarily herein concerned with actions that have been officially promulgated by group leaders and that individuals self-consciously carry out as members of groups. My fundamental theory holds that, in one way or another, all evil – whether collective or individual – is a manifestation of self-love that has been shaped in a social context, and so sees both collective and individual evildoing as two forms of the same genus.

crime that attacks the basic interests of the international community regarded as a community of groups.[4] This definition brings genocide into international jurisdiction. Scholars vary in their analysis of the meaning of this complex term, but two features stand out as common to crimes that bear on "humanity." First, crimes against individuals that are motivated by their group identity aim at devaluing the individuality of the victims according to the negative value ascribed to the group. Secondly, in the attempt to destroy the cultural group as such, an attempt is also made to deny to humanity specific cultural identities. Thus these crimes doubly affect humanity in both denying the humanity of individuals and in denying to humanity its range of identities. In proclaiming humanity an end in itself and prohibiting the use of humanity in any individual as a mere means to the ends of others, crimes against humanity strike at the very foundations of Kantian ethics.

However, the very horrific nature of such crimes seems to many commentators to be telling against Kant's own theory of evil. Despite his designation of evil as something "radical" taking root not only in the individual's character, but also even in the character of the species, commentators have objected that Kant's definition of evil in terms of self-love is too shallow.[5] Egoism, it is objected, is hardly the worst crime. Self-love appears too tame, too hedonistic, too limited in its aims to explain self-negating acts of destruction or the intense forms of hatred underlying ethnic conflicts. Surely destruction in disregard or even in denial of the humanity of others must be motivated by something particularly inhuman or inhumane. It cannot be the case, it is argued, that what motivates ordinary everyday immoralities could in turn generate extraordinary crimes.[6]

[4] L. May, *Crimes against Humanity: A Normative Account* (Cambridge University Press, 2005), pp. 80–95. Chapter 5, "The International Harm Principle," defends the concept that genocide is a crime of international rather than domestic jurisdiction, because group-based crimes harm humanity as such. May calls this the "international harm principle."

[5] R. J. Bernstein, *Radical Evil: A Philosophical Interrogation* (Cambridge: Polity Press, 2002), p. 42. In chapter 1, "Radical Evil: Kant at War with Himself," Bernstein argues that the variety of types of evil calls for a variety of types of non-moral incentives. In particular, he finds the acts of fanatics and terrorists difficult to categorize as "self-love."

[6] Claudia Card takes Kant to task for focusing his theory of evil exclusively on the culpability of perpetrators and ignoring the suffering of victims. She also claims that self-love as a type of "good" cannot be the basis of crimes that have as their objective

But the historical evidence is overwhelming – ordinary people are capable of committing extraordinary crimes. Inspired by Arendt's report on the "banal" and bureaucratic character of Eichmann in her famous *Eichmann in Jerusalem*, genocide scholars have largely focused on the role of institutional mediation to explain the Holocaust.[7] In his classic study, *Modernity and the Holocaust*, Zygmunt Bauman, for instance, has argued that the bureaucratic structures characteristic of modernity are capable of exploiting ordinary desires, which in other circumstances would not result in criminal activity, to produce large-scale atrocities.[8] He argues that the Holocaust was an instance of institutionalized genocide marked by the isolation of the machinery of murder from the sphere of primal moral drives, thereby rendering such drives irrelevant to the task. Professionalism rather than spontaneous affect, and pride in the technical efficiency of the task, characterized the mentality of perpetrators such as Eichmann. The fact that such motivations are not in themselves extraordinary did not lead Arendt to conclude that those who commit these crimes are not "responsible" for what they do. They are just as responsible for their crimes as they are for their ordinary activities. To fail to hold them accountable would undermine the structure of law and morality, which demand that even "cogs" take responsibility for their participation in murder machines.

Extended Self-Love and the Propensity to Evil

Clearly institutional structures can contribute to expanding the scope of group violence, making it easier for ordinary individuals to participate.[9] But do serious evils require extraordinary motivation for their

gratuitous suffering, since suffering can in no way be considered a good. See C. Card, *The Atrocity Paradigm: A Theory of Evil* (New York: Oxford University Press, 2002). Later in this paper I take up the issue of what type of good could be involved in the infliction of suffering on others.

7 H. Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin Books, 1994), pp. 287–8.

8 Z. Bauman, *Modernity and the Holocaust* (Oxford: Polity Press, 1989).

9 In discussing prosecuting individuals for the crime of genocide, Larry May entertains the question: how can an individual be held responsible for something that only a collective can accomplish? Isn't this like holding the individual responsible for the acts of others? May argues that responsibility can be distributed and that individuals

explanation? In participating in serious evils, do ordinary individuals suddenly become transformed?

Kant rejects the claim that serious evils require a distinctly diabolical motivation. This follows from his claim, necessary for moral accountability, that the human will cannot reject the moral law outright. Self-love is the subjective ground of all evil because it forms the primordial basis of disregard, if not resistance, to the moral law. Although self-love is not intrinsically bad, it can lead us to give priority to a variety of inclinations which, when adopted into a maxim, are capable of determining the will to action. Rational self-love arises from a predisposition to the good (humanity) and is a precondition of self-respect. As such it is a legitimate desire for equality with other members of the moral community. It is only when the individual chooses to make self-love her most fundamental commitment, and thereby subordinates the moral law to that incentive, that we call the resulting character "evil." Kant thus concludes: "It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims" (R 6: 36).

Yet the sheer simplicity of this formula – as noted previously – generates the objection that self-love cannot capture the range of phenomena that we term "evil." As Claudia Card has put the objection, self-love, even if not unconditionally so, appears to be something good.[10] In what sense, she asks, are we expressing self-love or pursuing something conditionally good when we deliberately inflict suffering on others? There appears to be a "gap" between the moral bad of pursuing self-love which, while only a conditional good, is still a good, and the deliberate infliction of undeserved suffering, which is deemed to be an unconditionally prohibited evil. What processes close this "gap?"

Self-love does not have as its only object the physically constituted self. Human identity is a complex process that evolves in association

can be held responsible for their contribution toward a collective evil. See May, *Crimes against Humanity*, pp. 170–3.

[10] Card, *The Atrocity Paradigm*, pp. 76ff. Card takes Kant to task for focusing his theory of evil exclusively on the culpability of perpetrators and ignoring the suffering of victims. She also claims that self-love as a type of "good" cannot be the basis of crimes that have as their objective gratuitous suffering, since suffering can in no way be considered a good.

with others. Our conception of our good typically involves the good of others whom we love and upon whom we depend. Social and material goods are typically accessed and distributed according to one's position within various groups. In Part One of the *Religion*, Kant distinguishes between the merely mechanical or physical self-love, rooted in the predisposition to animality and involving no use of reason, and the rational form of self-love, rooted in the predisposition to humanity and arising from a comparative use of reason (R 6: 26–7). Our conceptions of happiness that provide self-love with its contents have a comparative dimension that can render us insecure and unhappy when the goods of others appear to be greater than our own.

These two predispositions, animality and humanity, along with the predisposition to personality, i.e., the capacity for respect for the moral law, are "original" or necessary constituents of human nature that cannot be eradicated (R 6: 28). The predisposition to humanity, then, is not a contingent feature of human social organization that evolved through historical accident and that might have resulted in an "atomistic" sense of self. Humanity is from the beginning defined as a social, moral, as well as a physical species. Thus, while all three predispositions have a role to play in generating the moral character of the species, the predisposition to humanity provides a locus for the propensity to evil in the form of an expanded self-love, which in its corrupted form is a contingent and yet universal characteristic of the species.[11] It is in exercising the predisposition to humanity that the individual constructs the basic orientation of the "self" and either adopts the moral law and humanity as her highest end or gives priority to that which promotes personal "worth" in the eyes of others.

Although self-love is originally rooted in a desire for equality, it is bound up, Kant tells us, with a "constant anxiety" that others will strive to attain superiority over ourselves. Through our comparisons arises an inclination "to acquire worth in the opinion of others" (R 6: 27) that gradually becomes an unjust desire for superiority over them. Given the constant presence of others whose natures are similarly

[11] Since the predisposition to personality can be neither extirpated nor corrupted, this gives morality a leg-up with respect to the historical development of the character of the species – an optimism which somewhat offsets the pessimism of the analysis of the character of the individual members of the species.

constituted, jealousy and rivalry emerge as "stems" upon which a variety of "vices of culture" proliferate and become malignant. Of these vices, Kant states, "in their extreme degree of malignancy (where they are simply the idea of a maximum of evil that surpasses humanity), e.g., in envy, ingratitude, joy in other's misfortunes, etc., they are called diabolical vices" (R 6: 27). Thus, while Kant rejects the idea of diabolical motivation at the incentive level, he clearly recognizes that humans can intend and take satisfaction in inflicting serious harms on one another. Kant was never naïve concerning the manner in which culture can inform and inflame human passions producing "diabolical vices."

Kant's reference to "constant anxiety" complements his notion of a "propensity," which he defines as a predisposition to "crave" an object of experience that then arouses an inclination in us (R 6: 29). In this context Kant describes seeking social status in terms of precaution and safety, suggesting a deep "vulnerability" that arises from our social condition. Such vulnerability is more than a contingent "psychological" fact that would make Kant's conception of evil subject to empirical tests. It is an existential condition built into Kant's moral anthropology.

In her *Kant and the Ethics of Humility*, Jeanine Grenberg has stressed existential features, such as human dependence and vulnerability located within individual experience, in order to safeguard Kant's conception of evil from a "causal" form of sociological interpretation, and hence from a reduced notion of individual accountability.[12] According to Grenberg, it is anxiety over the loss of our original but naïve hope for perfect happiness that provides the internal prompting to overvalue the claims of self-love. Grenberg argues that without an account of how an individual becomes "primed by her own internal conflict," the birth of this "constant anxiety" in the presence of others and her willingness to engage in the gamesmanship of social competition will be unmotivated.[13] While an existential reading of the propensity to evil is valuable in mediating the relationship between the

[12] J. Grenberg, *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005), pp. 35–9.

[13] Grenberg attributes this overly sociological interpretation to Allen Wood. Wood, however, makes clear in his essay in this collection that his social interpretation of evil is not a form of determinism.

individual and the social dimensions of Kant's theory of evil, I maintain that these two dimensions cannot ultimately be meaningfully separated. They cannot be meaningfully separated, not simply because we are always imbedded within social relationships, but also because we are members of a moral community engaged in a common moral task – the creation and maintenance of an ethical community. Thus, the duties we have to ourselves, both perfect and imperfect, stem not from our animal or physical nature, but from the fundamental fact that by virtue of our humanity (as expressed in the predisposition to humanity) we are members of a moral or ethical community. While actions in violations of our duties to our physical or animal nature may appear to be merely "individual" and self-regarding acts without reference to any social condition, these acts are morally evil not because of the harm they do to the body, but because they debase the human person and thereby our common humanity.[14]

Given our common moral destiny, virtue and vice can never be entirely matters of a private or a personal good.[15] Only under the public laws of an ethical community can any individual find "moral orientation" in an otherwise "ethical state of nature." "Human beings," Kant states, "mutually corrupt one another's moral predisposition" (R 6: 97), but not because they are steeped in evil of whatever variety. This mutual corruption, Kant tells us, will occur even among men of good will: "because of the lack of a principle which unites them, they deviate through their dissensions from the common goal of goodness, as though they were instruments of evil" (R 6: 97).

Why should dissension characterize even good wills? Why do we need commitment to ethical community to secure moral orientation? Individual commitment to the moral law is clearly the first step in the formulation of the good will, but it is insufficient. Because the kingdom of ends formula of the moral law requires the integration of our ends with the ends of others, the task of the good will is not completed by a merely internal form of virtue. We cannot integrate ends without

[14] Grenberg uses duties to self to illustrate the possibility of purely individually based acts of evil requiring no reference to any social condition or socially oriented goal.

[15] While duties to self vs. duties to others is a very important distinction, there is a sense in which our overall perfection or virtue is a type of duty to a wider community, i.e., to the moral community. Kant and Mill are quite divergent on this point, and many modern liberals while assuming Kant to be in their camp are closer to Mill.

some form of public principle or social ethic to guide us.[16] It is only insofar as commentators have abstracted from this common goal, and thus from our complete set of duties, that they find room for an individualistic reading of good and evil. In *On the Common Saying* Kant makes an interesting comment on the moral character of individuals as "members of the series of generations" in the context of an argument concerning our innate duty as to influence posterity for the better. He maintains that while we have a duty to influence posterity for the better, we are "yet not so good in the moral character required of me as I ought to be and hence could be" (TP: 8: 309). I read this qualifying comment as stating that the phenomenal character of any individual as a member of a particular generation will reflect her position in the historical progression toward the moral ends of the species. The phenomenal character of particular generations will not be as good as it ought to be given the moral ends of the species. As members of particular generations, we address this deficiency by striving to promote these ends in our own lives and by passing on the duty to continue to promote these ends to the next generation. If this is our "innate duty," then individual virtue cannot be meaningfully understood as occurring in a separate realm of purely personal or private ends. Because our moral life must achieve unification through our contributions to the highest good in the form of ethical community, duty will require that we take our moral orientation from this social goal.

The conclusion of the analysis of evil in Part One of the *Religion*, which reveals evil to be a universal characteristic of the species, leads naturally and logically to the main topic of Part Three, the construction of the ethical community. In fact the two concepts, radical evil and ethical community, are deeply interconnected. Because the species as such is destined for the moral goal of ethical community, which can be achieved only through the collective action of the generations, the phenomenal or historical character of the species is always imperfect

[16] P.J. Rossi, *The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind* (Albany: State University of New York Press, 2005), pp. 71–85. In this work Rossi argues that this social dimension of the moral community is not fully spelled out in Kant, but is clearly indicated by the manner in which he develops the concept of the ethical community as the antidote to evil. Rossi maintains that there needs to be a public institutional form of the ethical commonwealth.

and each individual must adopt this "highest moral good" as his or her own highest end. Furthermore, Kant maintains that "since this highest moral good will not be brought about solely through the striving of one individual person for his own moral perfection" (R 6: 97), our strivings must not be thought of as a disjointed aggregate but as a unified effort. There is therefore a universal obligation incumbent upon each individual to promote the ethical community.

That Kant provides his only in-depth analysis of moral evil in the context of an analysis of the character of the species is no aberration. Individual moral character is constructed in the context of the formation of a social and cultural identity under an obligation to promote an ethical community. Thus, when Kant maintains that the propensity to evil is universal because it is "rooted" in humanity as a historically constituted species, while it is also "contingent" because it must be enacted by each individual in the construction of his/her moral character, he is not generating a simple "contradiction," but articulating the deep structure of social existence. The nature of our moral end thus informs the nature of our moral failings.

Kant recognizes that in working out the complex dynamics of self-love against the demands of the moral law we organize our commitments in ways that discriminate between those who "belong to us" and those who do not belong to us, i.e., those whose endeavors serve our purposes and those whose endeavors do not. In constructing our social identities we decompose the moral community into communities of extended self-love. Thus, a person who inflicts suffering on a devalued other may be generous and kind towards those she values. Evil is not radically "egoistic" or atomistic. It allows for forms of self-love that include social identifications.

Shared Responsibility for Collective Evils

The propensity to evil, then, is the human tendency grafted onto the predisposition to humanity to subordinate the moral law to the demands of a generalized self-love. This includes social and cultural identities that fall short of the requirements of the ethical community. Since a propensity is a property of the will, not of the pre-given anthropological predispositions, it is attributable to the character of the human person. Evil is a choice for something, not a negation of

the incentive provided by the predisposition to personality to internalize the moral law.[17] Pure negation, were it conceivable, would have no character. Human character is always formed in a "context" that is intrinsically social and cultural. Evil is thus an expression of the internalization of social and cultural norms embodying morally corrupt objectives, such as social stratification, subordination and group conflicts.

While individuals may be differently situated with respect to the enactment of specific collective harms and thereby hold different degrees of guilt, individuals nonetheless share responsibility for the identities that they mutually construct. Members of social groups are responsible for the attitudes that they hold and which provide support for the actions of other group members.[18] Ervin Staub, in his classic study of the origins of genocide, *The Roots of Evil*, argues that the role of bystanders is crucial in the formation of the self-concept of perpetrators that allows them to move from less harmful to more extreme forms of violence. He states: "Genocide does not result directly. There is usually a progression of actions. Earlier, less harmful acts cause changes in individual perpetrators, bystanders, and the whole group that make more harmful acts possible."[19]

While this socio-cultural foundation of vice has received some reluctant acknowledgement in recent Kant interpretation,[20] its application to the forms of evil requiring "self-sacrifice" remains contested on the grounds that such premeditated destruction does not aim at

[17] Were negation of the moral law possible, we would lose practical reason altogether. Self-love, were it conceivable without its relation to practical reason, would be without any orientation. The coherence of happiness is contingent and parasitic upon the use of practical reason to achieve a projected unity.

[18] For an interesting discussion of shared responsibility in Hannah Arendt and Larry May, see C. Striblen "Guilt, Shame and Shared Responsibility," *Journal of Social Philosophy*, **38**, 3 (2007), pp. 23–39. I find this approach to the link between identity formation, group membership, and shared responsibility congenial to my thesis on extended self-love as a basis for radical evil.

[19] E. Staub, *The Roots of Evil* (Cambridge University Press, 1989), p. 5.

[20] While originally skeptical of viewing evil as self-love, Robert Louden now expresses agreement. Jeanine Grenberg accepts the self-love formulation, but resists its location in the socio-cultural dimensions of human existence. Claudia Card challenges the capacity of this concept to explain extreme evil. What "good" – it is often asked – is attained by the suicide bomber? Allen Wood has been a long-time exception to this limited view of the nature of self-love in Kant.

any apparent good for the agent. But the processes of cultural identification, even when they require sacrifice of physical gratifications, also answer to deeply held individual and personal needs for social goods, such as security, status, or social recognition. While acknowledging that individuals acting in their capacities as members of groups tend to exhibit some distinctive behavioral characteristics, most social psychologists do not accept an absolute differentiation between the psychology of individuals and that of groups. The fact that individuals are always members of groups does not eradicate their character as individuals. It is because group membership meets basic needs for belonging and security that these needs can be drawn upon and exploited by group leaders to promote group power. Vices of culture can operate between cultural groups as well as between individuals within a cultural group, generating a competitive politics of "identity."[21] This rivalry can produce a devaluing of identities and can serve as a justification for collective actions of subordination, aggression, and even destruction. Members of particular groups locked into such struggles can each come to see themselves as threatened and therefore victimized by the other.[22]

That some persons such as Eichmann have participated in collective projects of evil for purposes of personal aggrandizement without the appearance of great passion suggested to Arendt that evil is "banal" or without great "depth."[23] As noted earlier, the Holocaust has served as the paradigmatic example of collective evil undertaken in a context of institutional and technological mediation intended to distance the perpetrators from the victims, making possible acts of "impersonal" destruction. Zygmunt Bauman argues that every aspect of the Holocaust was "normal" in that it was not "an irrational outflow of the

[21] Paul Gilbert claims that terrorism is best explained as a strategy arising from a "politics of identity." See P. Gilbert, *New Wars, New Terrors* (Edinburgh University Press, 2003), particularly chapter 1, pp. 1–23.

[22] Where groups are competing for territory, resources, or even "identity," acts that aim at destructive outcomes are nonetheless linked to a positive outcome for the winning or surviving group.

[23] And this example, which nicely fits Kant's conception of radical evil but which Arendt calls "banal," is to be distinguished from her earlier view of radical evil, which she had described as going beyond all ordinary motivations and having as its peculiar objective to make humans "superfluous." See H. Arendt, *The Origins of Totalitarianism* (San Diego: Harcourt, 1994).

not-yet-fully-eradicated residues of pre-modern barbarity."[24] Rather, he maintains that the Holocaust was a consequence of modernity's emancipation of the political state with its monopoly of violence and its social engineering ambitions. On Bauman's view, all that need be added to this deadly brew to produce atrocities is the inhibition of any moral sympathy for the proposed victims of state violence through a process of distancing and dehumanization. No further special motives of hatred or desire to inflict suffering on the part of ordinary participants are required.[25] Nazi propaganda often called for the sacrifice of personal sentiments in the pursuit of the higher good of the group without the avocation of cruelty. On this "higher good" interpretation, the extent and degree of the infliction of suffering, because it requires so many otherwise normal participants, depends for its effectiveness upon a mechanism for the suppression of normative empathetic responses. These "mediations" call into question any necessary relation between cruelty as the primary motivation of cruel acts and the actual infliction of cruelty.[26]

Such distancing of motives and consequences characteristic of the institutional model has not, however, been characteristic of the genocides of the late twentieth century in Bosnia and Rwanda, where low-tech tools, personal relationships, and physical proximity marked much of the killing. Bauman's view that physical proximity engenders moral empathy does not appear to hold in these circumstances. Here

[24] Bauman, *Modernity and the Holocaust*, p. 17. Bauman's close identification of modernity with the Holocaust is controversial. Vetlesen rejects this identification on the grounds that Nazi ideology was no product of, but rather an attack on, modernity. See A. Vetlesen, *Evil and Human Agency* (Cambridge University Press, 2005).

[25] Vetlesen argues that Bauman's analysis of the Holocaust is too simple, because as a matter of fact extreme anti-Semitism and overt cruelty did characterize some aspects of the Holocaust, particularly as expressed by some party leaders, and that it cannot be fully generalized, because contemporary genocides have not exhibited the complex institutional machinery of the Nazis. Nonetheless, the potential of the institutionalized if idealized model to produce such atrocities is not denied. See Vetlesen, *Evil and Human Agency*, chapter 1 ("The Ordinariness of Modern Evildoers: A Critique of Zygmunt Bauman's *Modernity and the Holocaust*"), pp. 14–51.

[26] This in no way calls into question the objective cruelty of the acts, nor the responsibility of the perpetrators. This analysis simply serves to elucidate the conditions under which widespread or systematic violence becomes possible given that a principled commitment to cruelty, if it exists, is a rare motivation, and one that on my analysis may sufficiently overlap with irrationality as to call into question accountability.

a dedicated passion appears to override both ordinary empathy and egoism with the result that the perpetrators appear to be anything but "banal" and mechanized cogs. Nonetheless, the majority of those involved in these brutal actions were ordinary citizens who had lived in close physical and social proximity with their victim/neighbors. It would appear that neither the mechanized–institutional nor the more personal–proximate forms of genocide capture exclusively the character of group-based evildoing.

What these apparently different types of group action do have in common is the officially organized, propagated and widely accepted belief that a rival cultural group, planning to commit grave injustices, had become a "threat" to their valued group's survival. Arne Vetlesen in *Evil and Human Agency* has documented the similarities in the manner in which both Hitler and Milosevic constructed and manipulated a collective trauma (past defeat) to mark and manipulate a collective self-identity. He cites the studies of the Copenhagen School on "securitization" to explain how an abstract object such as identity can be made into an existential issue and be translated into political action through the objectification of the identity of the other as a "threat" to one's security.[27]

In light of the perception of a threat to their security, individuals were called upon to sacrifice their "normal" feelings and past relationships in order to protect the valued group and secure its future. The fact that individual members of the devalued group may be powerless in the particular situation in which they are encountered is irrelevant to the significance of the ideological claim that entails that it is their continued existence that is itself a threat.[28] Such encounters with "powerless" individuals throw into relief the social/cultural characteristic of the "threat."

[27] Vetlesen, *Evil and Human Agency*, p. 167. This dynamic is clearly at play also in the current identification of terrorism with a hatred of freedom and thus as a threat to our cultural and political identity.

[28] Wilshire, *Get'Em All! Kill'Em All*, particularly chapter 5, pp. 83–122. Wilshire maintains that the world as experienced is an unconscious cultural construct that can be threatened by interactions with culturally different others under certain conditions. What interests Wilshire is not so much the empirical conditions, but the "genocidal" response implicit in such constructions. He sees conflict, war, and genocide as being on a continuum.

Ideological claims cannot of course operate autonomously. They must first be accepted as valid, which often entails the adoption of an uncritical attitude toward authority in violation of duties we have to our own enlightenment. Ideological beliefs mediate between individual and group identity, and by generating a form of existential threat for those who view themselves as vulnerable, such beliefs can motivate an overvaluation of the group identity.[29] In this context, an object normally valued as good, one's cultural identity, is presented as threatened and the individual is invited to make this value her ultimate end. In this process, members of the rival group are no longer valued in their capacities as neighbors or fellow citizens, but are valued solely in their role as "rivals." The identities of individuals are thus reduced to their group status, and both their individual humanity and the humanity of their group is denied and made available as an instrument for the furtherance of the goals of the valued group. In such a context, the ordinary prohibition on murder becomes subordinated to the end of the affirmation of the group's identity.

While this immersion in a group identity is not "egoism" in its normal definition, such identifications are constitutive of our normal understanding of the "self" and therefore cannot be dissociated from our conceptions of self-love. Nor can our conceptions of our self be dissociated from our conceptions of moral character and personal responsibility. We bear the responsibility for what we value. What has gone morally awry here is the claim for the superior value of one group over that of another and the willingness of individuals in acting upon such a valuation to deny the humanity of members of devalued groups. This dynamic is not unfamiliar to us. We are aware of our own willingness to use status and other forms of social privilege to further advance personal goals at the expense of others. In our ordinary everyday competitions, positive law provides a set of rules within which we can pursue these objectives, while also providing counter-incentives to our temptations to disregard the rights of others. In our ordinary

[29] Existential threats can take many forms and need not always attach themselves to group identities. We can find a parallelism in many forms of individual evildoing such as sadism, where the infliction of extreme suffering on the victim may be a ritualized attempt to deny one's own vulnerability to suffering. In both cases, the moral need to accept one's dependence upon and vulnerability to others is denied, and the overinflated demands for security and control asserted.

pursuance of self-love we are regulated by a set of rules of the road, which, although not always in accordance with morality and justice, nonetheless allow us within these legal limits to be "bad in a way common to all" (R 6: 33).

Cosmopolitan Right and Ethical Community

Historically, therefore, the first step in a political campaign to devalue a rival group is often an attack on the citizenship status and legal rights of that group – often in the form of a claim that the state is inherently the property of a particular cultural group and that groups that do not share that "identity" must be denied any influence on the ideals of the state and/or removed from the territory of that state. The formal equality and independence of the subject/citizen central to the Kantian concept of civil society is violated by any attempt to introduce substantive ethnic content into the idea of legal standing. Birth, Kant states, "is not a deed of the one who is born, he cannot incur by it any inequality of rightful condition" (TP: 8: 293). The attempt to eradicate or render unenforceable legal rights on grounds of ethnicity is, I would argue, an attack on the body and being of the state (even if instigated by political leaders), and hence a situation which ought to raise serious concern in any society of states pledged to protect human rights amongst its members.[30] Denaturalization, the ultimate consequence of an attack on the legal rights of an identity group, is (as Arendt has argued) a denial of the right to have rights, and should be viewed as contrary to any account of human rights.[31]

Although Kant does not have an explicit theory of human rights, the notion of cosmopolitan right developed in *The Metaphysics of Morals* and *Toward Perpetual Peace* is designed to protect individuals from hostile actions, regardless of their national or ethnic identity. Cosmopolitan right is explicitly derived from a universal human entitlement "to present oneself for society" and "to make use of the right

[30] International law forbids arbitrary denaturalization, the denial of the right to have rights.

[31] Although Arendt does not attempt to ground rights in human rights, she does view denaturalization as the equivalent of the denial of the right to have rights. See S. Benhabib, *The Rights of Others: Aliens, Residents and Citizens* (Cambridge University Press, 2004), pp. 56–64.

to the earth's surface, which belongs to the human race in common, for possible commerce" (EF: 8: 358). Given that individuals are owed hospitality as a right of their humanity, violence aimed at the eradication of an identity group, even if sanctioned by national law, is a violation of international right and arguably punishable under international law as a crime against humanity.

Cosmopolitan right is in the service of a dynamic of cultural exchange. Since Kant also maintains that no one originally has more right than another to occupy any particular portion of the earth, the right of any people to possession of its territory must be mutually contracted with other peoples and cannot be derived from original possession or mythic "blood" relations. If national right is ultimately grounded upon cosmopolitan principles, the constitution of a rightful state cannot be based upon ethnic foundations.[32]

Kant's philosophy of history, and the ethical and political theory that it embodies, envision a society of states that are open to intercultural exchange and whose interactions are ordered to the development of a cosmopolitan society. In "What is Enlightenment?" Kant claims that reason has an unrestricted audience and that we may address others as members of a cosmopolitan society. Whether the audience is cross-national or cross-generational (as in the example of a sectarian commitment to a set of unalterable doctrines), Kant grounds the duty of enlightenment not simply in personal perfection, but in the "sacred right of humanity" to progress and improvement (WA: 8: 39). The attempt to exclude particular groups from this interaction is contrary to the "moral destiny" of the species and robs future generations of their rightful multicultural inheritance. This makes such crimes the concern of all humanity.

The dynamics that challenge this moral goal of a cosmopolitan society are the same dynamics that thwart moral development generally; the valuation of the interests of some over the equal valuation of the interests of all required by morality. Because this dynamic is inherent to our condition as dependent and vulnerable social creatures, Kant foresaw that moral development entails a commitment to an ethical form of social life, which he called an "ethical community."

[32] Sharon Anderson-Gold, "The Cosmopolitan Foundations of the Kantian State," unpublished paper.

He states: "An association of human beings under the laws of virtue … can be called an *ethical* and, so far as these laws are public, an *ethico-civil* (in contrast to a *juridico-civil*) society, or an *ethical community*" (R 6: 94).

The ethical community is fully universal and so cannot be constituted by or for any ethnic group. The universality of the ethical community entails that its constitution is derived from a law-giving will that is an unconditioned source of the moral law, a holy will. While empirical humanity cannot thus claim to be the foundation of this community (this would provide grounds for restrictive communities), a morally developing humanity must take its guidance from this "ideal." This does not mean that cultural identities should be overcome or suppressed in the name of an abstract universal; it means, rather, that cultural identities should be viewed as particular expressions of ethical community. Respect for the universality of ethical community is compatible with preservation of cultural heritage.[33] Kant presents the ethical community as not only a regulative ideal for human moral development, but as the ideal necessary for the overcoming of evil. Kant states: "Inasmuch as we can see, therefore, the dominion of the good principle is not otherwise attainable, so far as human beings can work toward it, than through the setting up and the diffusion of a society in accordance with, and for the sake of, the laws of virtues" (R 6: 94). It is because the propensity to evil inheres in competitive forms of social relationships that it must be combated by an ethical ideal of society.

Conclusion

While we may have different names for specific "vices" or manifestations of evil, some of which (as crimes against humanity) we may properly designate as more serious than others and treat with greater severity under the law, we do not require a moral theory with multiple moral principles to capture these distinctions and ground our

[33] Ethical community as a moral ideal regulating intercultural association is, I would argue, an essential aspect of multiculturalism. My own form of cosmopolitan theory is built upon multiculturalism. See S. Anderson-Gold, *Cosmopolitanism and Human Rights* (Cardiff: University of Wales Press, 2001).

moral judgments. Our duty to promote ethical community is also a duty to reform social relationships and to guard against the institutionalization of biases and prejudices. Those who perpetrate either biased crimes or crimes against humanity assume the approval, or at least the passive support of, others who share the social identity of the perpetrators and who are unwilling to risk the devaluations accorded to the social identity of the intended victims. In providing support for and failing to resist social and institutional forms of injustice, individuals participate in and are responsible for various forms of collective evil. Kant does not view the forms of evil associated with subordination, discrimination, or even genocide as stemming from different fundamental incentives, nor need he do so to capture the range of phenomena we call "evil." Once we recognize the imprint of the self in all of our social and cultural formations, we have no need of complex formulas.

10

Unforgivable Sins?

Revolution and Reconciliation in Kant

David Sussman

Moral reconstruction is perhaps the most urgent if neglected problem of contemporary moral and political philosophy. This problem arises in the context of communities that, after a period of significant systemic injustice, are trying to reconstitute themselves justly while properly addressing the wrongs suffered by their members (such as Argentina after the restoration of democracy, South Africa after the dismantling of apartheid, and Rwanda's attempt to confront the massacres of the 1990s). One of the most vexing questions of such reconstruction is how a reformed polity should deal with individuals who, acting under state authority, committed grave injustices against its members. Should these people be tried and punished, or are they properly immune from prosecution in at least some respects? Is there a political role for something like pardon or amnesty, conditional perhaps on the kind of public accounting central to South Africa's Truth and Reconciliation Commission? Do the victims of the old regime have anything like a political or moral obligation to forgive their former tormentors, if such forgiveness is needed to create a stable, just social order?

Kant's practical philosophy may not seem a particularly promising place to pursue these questions, since they lie at the intersection of his two least popular doctrines: his strongly retributivistic understanding of punishment and his absolute condemnation of political revolution. Moreover, the primary arguments for granting some kind of pardon or amnesty to the agents of the old regime seem to be fundamentally

consequentialist. Often these malefactors are sufficiently numerous, or still possessed of sufficient power, influence, and expertise, that any attempt to seriously punish or exclude them from political life would lead to the collapse of the new polity, perhaps even to civil war. Yet in the Doctrine of Right, Kant insists that

> The law of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaimonism in order to discover something that releases the criminal from his 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 the entire people to perish". For if justice goes, there is no longer any value in human beings living on the earth. (MS 6: 332)

While Kant does recognize an executive power to grant clemency, he considers this prerogative to be profoundly circumscribed. A ruler may grant clemency only for crimes against him and his office, so that "with regard to crimes of *subjects* against one another it is absolutely not for him to exercise [the right of clemency]; for here failure to punish (*impunitas criminis*) is the greatest wrong against his subjects" (MS 6: 337). Kant famously claims that even if a society was on the verge of dissolution, it must first make sure to execute every murderer in its jails, so that "blood guilt does not cling to the people for not having insisted upon this punishment" (MS 6: 333).

These remarks seem to rule out any legal immunity or even leniency for agents who have murdered or tortured in the service of the old regime. Yet Kant appears to reach the opposite conclusion when he considers whether an overthrown monarch should be executed or otherwise punished. Kant asserts that "[the monarch], as the source of law, can do no wrong" (MS 6: 321n.). The judicial execution of a deposed ruler is supposedly the worst transgression imaginable. The enormity of this crime is so great that Kant turns to religious imagery to capture the horror he thinks we all should feel at this "complete overturning of all concepts of right" (MS 6: 321). Regicide is "a crime that remains forever and can never be expiated (*crimen immortale, inexpiable*)" such that "it seems to be like what theologians call the sin that cannot be forgiven in this world or the next" (MS 6: 321n.). Here Kant only explicitly considers the judicial execution of a former

monarch, but his worries do not have anything to do with the fact that it is specifically the death penalty that is involved (the penalty that Kant thinks is normally the only morally appropriate punishment for murder).[1] What horrifies Kant is not just the execution of the monarch, but any punishment under the auspices of law, since "a dethroned monarch . . . cannot be held to account, still less be punished, for what he previously carried out" (MS 6: 323; see also 6: 317).

Unfortunately, Kant never explicitly addresses the propriety of punishing subordinates who acted under the authority of the monarch. Yet it is hard to imagine that Kant could hold the agents of the monarch criminally liable, who may have been fulfilling what Kant considers a real duty they owe the executive authority, while at the same time taking the monarch himself to be absolutely unpunishable. The picture that emerges is that for Kant, after revolutionary change we must all start off with a "clean slate" morally and politically. A newly reformed state cannot take any special notice of the wrongs done by its authorities in the past, no matter how severe, so long as these acts really were sanctioned by the executive power as it had been previously constituted.

Kant cannot readily make sense of moral reconstruction in terms of either of the basic kinds of moral/political relation that he recognizes. The problem of reconstruction does not arise within the normal life of a well-constituted political order (what Kant calls a "civil condition"), but neither does this problem involve relations between individuals in the "state of nature" that Kant takes to hold in the absence of any commonly recognized political authorities. This stark dichotomy between the state of nature and the civil condition accounts for the Doctrine of Right's tendency to oscillate fruitlessly between a demand for full punishment for those who acted under state authority and an insistence on their complete immunity from prosecution. However, Kant has unexpected resources for dealing with moral reconstruction in his philosophy of religion, where he considers the "revolution of

[1] Kant recognizes only very limited exceptions to the claim that murder must be punished with death, none of which are relevant to case of the deposed monarch. See my "Shame and Punishment in Kant's *Doctrine of Right*," *The Philosophical Quarterly*, 58 (April 2008), 299–317.

character" that each person must undergo in order to become "a person well-pleasing to God." In what follows, I argue that Kant's understanding of the moral relationship that a person has to herself before and after such a personal conversion experience provides a promising model for the relations that a community bears to itself across a revolution in its basic moral and political character.

1. Before turning to Kant's philosophy of religion, we must first understand why moral reconstruction should pose such a serious problem for Kant's official political philosophy. Kant understands both punishment and political authority as matters of "right" (*Recht*). His practical philosophy in general is grounded in the perspective of the deliberating agent, and in the basic commitments that such an agent must implicitly recognize if she is to coherently see herself as an autonomous, rational being. Through commitment to the moral law rational agents constitute themselves as a special kind of free power, possessed of a spontaneity with respect to their physical and psychological states that makes true "imputation" of their actions to them possible. In his political philosophy, Kant extends this line of thought, considering what commitments a group of agents must jointly recognize if they are to be free, not just with respect to their affects, but with respect to each other. The "internal" freedom defined by morality needs to be complemented with some scheme of "external" freedom that can be realized only by a just political order. Just as each agent must have a personal commitment to morality that frees her from determination by her affects, a community must recognize some set of common laws and institutions that serves to free everyone from dependence on the "arbitrary" choices of any other people in particular. In the individual case, the moral law frees us by assigning us duties. In the political case, we are liberated through an assignment of rights, which mark out areas of choice where we may properly exercise personal discretion without interference from others.

Kant considers it part of the essence of a right that it may be defended by force. Without such authorization, our rights would protect us from subordination to other people only insofar as we could convince those others to recognize the merits of our claims. In that case, our freedom of action would still ultimately depend on the discretion and

good will of others in recognizing and respecting our rights. For Kant, what is morally important about the threat of force is that such threats provide a general way of giving people decisive reasons to respect our rights without having to convince or persuade them of the substantive merits of our claims. Although Kant considers the aspiration to a common deliberative point of view to be the cornerstone of morality, he still thinks that we need enforceable rights so as to define areas of life where we do not have to deliberate with or justify ourselves to one another, or seek out anything like permission or agreement with another person. Despite his retributivist sentiments, Kant does not consider legal punishment to be grounded in the moral demand that everyone get what he deserves. Kant does recognize such a demand, but believes it can be properly acted on only by God. Instead, judicial punishment can be based only in what we need in order to have any significant degree of moral independence from one another, an independence that defines the area of permissible conflict, competition, and simple disagreement in our social lives. For Kant, the fact that we must stand in enforceable relations of right follows from the moral law as a "postulate of practical reason" that must be accepted whenever there is a plurality of individuals who are capable of affecting one another through their actions.[2] Such enforcement is not merely a permissible option that is morally open to us. Instead, each individual has a positive duty "of rightful honor" to defend her rights, as part of her broader obligation to respect humanity in her own person (MS 6: 236).

Kant argues that without a common authority, any attempt to satisfy this duty of rightful honor would itself have to be a violation of the rights of someone else. The problem here is not just of the Lockean "inconveniences" that arise whenever a person has to serve as a judge in his own case, leaving others with little confidence that he will accurately assess and respect what right demands in a particular situation. For Kant, the real problem is simply that in a state of nature, the effective reality of anyone's right must ultimately rest upon individual

[2] For an extensive examination of Kant's understanding of what a "postulate" is in this context, see P. Guyer, "Kant's Deductions of the Principles of Right," in M. Timmons (ed.), *Kant's Metaphysics of Morals: Interpretative Essays* (Oxford University Press, 2002), pp. 23–64.

exercises of judgment and uses of power, even if such judgment is exercised flawlessly. A world in which everyone immediately and correctly comes to agreement about what our conflicting rights require would still be a world in which each person's freedom depends on the good will and the good sense of other individuals, however trustworthy they may be.

A world in which everyone was reliably judicious and fair-minded would indeed be harmonious in a way that allows for extensive and profitable social cooperation. Yet, however fortunate this situation may be from the perspective of self-love, Kant still takes it to be absolutely unacceptable from the perspective of right. Rights presuppose "a will that is omnilateral, that is united not contingently but a priori and therefore necessarily, and because of this is the only will that is lawgiving" (MS 6: 263). Our rights are supposed to define some area of our lives in which we do not have to justify ourselves to others in order to act. In the anarchist utopia of reliable and thoroughgoing consensus, we would still always have to be justifying ourselves to everyone else, and they to us. That this might be easy to do would not change the basic fact that in every dimension of our lives we would be depending on the good will and acceptance of others. For Kant, such dependence means that there would really be no real rights, and hence no meaningful kind of external freedom, and so no morally acceptable ways of living with other people, regardless of how harmoniously we lived together.

To be in a state of nature is to be confronted with an irresolvable moral dilemma. When some matter of right is in dispute, we are morally obligated to defend our rights by force, so as not to acquiesce to any dependence on the discretion and good will of others. Yet insofar we can effectively defend our rights this way, we end up making the rights of our opponents dependent in turn on our own discretion and good will, thereby committing a wrong against them. A general scheme of rights becomes impossible in a state of nature, because the conditions necessary to make one person's right real would have to vitiate the putative rights of another. We would be confronted with a deontic contradiction; we would be obligated to defend our rights, but every possible way of doing so would be morally wrong. Since the moral law is the ultimate ground of both the obligation to and the prohibition of defending our rights, this result would show that the law, by making

inconsistent demands, could not really be a principle of reason at all. Were we ever to really stand in a state of nature with regard to others, morality itself would be overthrown.

The problem here has something of the structure of an antinomy, where apparently contradictory conclusions are derived from equally well-grounded rational principles. The key to the solution, as elsewhere, is to see that the conclusions in question produce a contradiction only given some further assumption, and that it is this assumption that we should reject. In the case of our rights, the hidden assumption is that our duty to defend our rights falls on us first and foremost as individuals, making the defense of them necessarily a "zero-sum" game. However, we can conceive of ourselves as confronting one another morally not just as individuals, but as citizens, as fellow members of a collective body that is the primary bearer of both the duty and the entitlement to enforce a scheme of rights.

Here Kant draws on Rousseau's conception of the general will, as the united will of all citizens which each citizen takes to be the deepest aspect of her own will, with authority over whatever is merely personal in her. Individuals who conceive of themselves as bound to one another through such a joint authority can avoid the contradiction in practical reason that would come about in a real state of nature. Such a general will can specify and defend our rights in a way that does not require that anyone's claim be subordinated to the will of anyone else in particular. If I am to have rights, I must recognize a duty to defend them. However, if I identity with the general will, then when that will acts to defend my right, such action can satisfy my duty of defense, because I do not see my individual will as something that is wholly distinct from or prior to this broader collective kind of agency. Moreover, such enforcement no longer need be an infringement of anyone else's right either. So long as my opponent also identifies his will with the general will, the general will's acts of enforcement do not serve to subordinate his will to any that is truly distinct from his own.

The initial problem posed by Kant's conception of right is that it seemed to require agents to be both distinct and yet identical in order to be properly applied. Questions of right can only arise in the first place where there is some distinction between agents who might affect one another. Individuals do have duties to themselves, but Kant realizes that we could not have anything like enforceable rights

against ourselves (MS 6: 270). However, reflection on the state of nature showed that if we confront one another as wholly separate and distinct agents, each individual's rights could be realized only at the expense of another's. Yet through a common identification with the general will, we gain the ability to address one another as agents who are neither completely identical nor wholly distinct from one another. While the party who advances a claim of right must be in some way distinct from the party who contests it, the power that enforces that claim must serve as an essential aspect of each disputant equally.

Kant does not understand the general will as something that is constructed out of the prior wills and interests of the members of a community by any form of bargaining, voting, or agreement. Instead, the general will has an "intelligible" existence that is prior to how it might be empirically manifested in any real set of procedures or institutions. The general will is an ideal conception of our basic moral and political relations to one another, a normative structure that every person must take to be "always already" in place in any situation where considerations of rights and external freedom can arise. As such, the original social contract that supposedly constitutes the general will could not be any real agreement between people that could be concluded at some particular point in time. Instead, the social contract is only a way of representing how we must understand our basic identities if our relations to one another are to be morally coherent. Kant argues that these identities, as represented by a social contract, must involve basic rights and prerogatives that are determined in a way that respects individuals as free and equal citizens, citizens that can be bound to others in general only in the same ways that those others are bound to them. No scheme of rights, laws, or institutions could be just if it is not possible for a will that is truly general with respect to a community to accept it, consistent with these basic formal constraints of equality and reciprocity.

2. So far, Kant's political philosophy would seem to have profound revolutionary potential. Yet Kant notoriously argues that we must respect any government we find ourselves under. Although we cannot obey immoral laws or commands, we must never challenge the authority of those who issue them: "subjects may indeed oppose this injustice by complaints (*gravamina*) but not by resistance" (MS 6: 319).

These Hobbesian conclusions are certainly surprising in a champion of freedom and autonomy, but they do in fact follow from the peculiar relations that hold between the general will of a polity and its actual political institutions. For individuals to stand in any relations of right, they must collectively identify with a general will as the ultimate sovereign power. As the enforcer of right, this will need to have some sort of determinate manifestation in the world, something that counts as the true exercise of its sovereignty. Ideally, a general will might prefer certain political forms over others, and Kant thinks that republican institutions are always the best fit for a will that is truly general. However, what is most important to any general will is that its authority be expressed through some real institutions, through a government which is actually up and running and which has the power to enforce some determinate scheme of rights, however flawed. Any general will would thus recognize any extant government, no matter how substantially unjust, to be its empirical manifestation, if only because that government is the only real option on the table. Admittedly, a general will choosing *ab initio* would have to opt for a republican government. Yet a general will also always prefers real institutions over ones that do not yet exist, regardless of their relative merits when all are considered as mere possibilities.

Kant argues that no general will would withdraw its identification with a real government in the hopes of creating one that better realizes the equality and freedom of its citizens. The problem is that any revolutionary change would have to deprive the general will of determinate empirical expression, casting the people into a state of nature with respect to each other:

> [A] people cannot offer any resistance to the legislative head of a state which would be consistent with right, since a rightful condition is possible only by submission to its legislative will … The reason a people has a duty to put up with even what is held to be an unbearable abuse of supreme authority is that its resistance to the highest legislation can never be regarded as other than contrary to law, and indeed as abolishing the entire legal constitution. (MS 6: 320)

The problem here is the not just the risk that a stable government will not arise to replace the one that has been overthrown. For Kant, it is enough that there be one moment, however fleeting, in which

we would find ourselves in a real state of nature with respect to one other. For Kant, "this transformation would have to take place by the people acting as a mob … [overthrowing] all civil rightful relations and therefore all right" (MS 6: 340). Once the people become such a mob, they can never become anything else. In the moment that a state of nature becomes real, the demands of morality become self-contradictory and permanently forfeit any claim they had to rationality or authority over us.

A revolution could only be legitimate if it did not involve individuals withdrawing their allegiance from one set of institutions, falling into a state of nature with respect to one another, and then setting out to reconstitute themselves as a new general will given expression by a new set of institutions. There would have to be a way for people, acting united under a general will, to collectively withdraw their support from some set of institutions and collectively bestow it upon another without becoming a mob in the interim: a process Kant likens to "palingenesis" or the direct transfer of a complete soul from one body to another (MS 6: 340). Yet Kant holds that a people can act together in this way only through some such set of real institutions, taken to be the embodiment of the general will. By ceasing to recognize the extant government's actions as the true empirical expression of the general will, a people gives up its only determinate way of acting as a people, and hence any determinate way of recognizing some new government in particular as its proper manifestation.[3] What this means is that for Kant, any revolution would already have to have succeeded before it could even be legitimately begun.

For Kant, these problems afflict the perspective only of those deciding whether to start or support a revolution. Although all revolution is morally wrong, Kant thinks we would still owe our allegiance to any revolutionary government that does indeed succeed in coming to power. This conclusion follows from Kant's insistence that it is the real existence and power of government that makes a general will recognize it, independent of any considerations of the actual process

[3] See C. M. Korsgaard, "Taking the Law into Our Own Hands: Kant on the Right to Revolution," in A. Reath, B. Herman, and C. M. Korsgaard (eds.), *Reclaiming the History of Ethics* (Cambridge University Press, 1997), pp. 297–328. See also S. W. Holtman, "Revolution, Contradiction, and Kantian Citizenship," in Timmons (ed.), *Kant's Metaphysics of Morals*, pp. 209–32.

by which it came into existence (MS 6: 323). A successful revolution may thus be justified retrospectively should it in fact succeed, although it must always be condemned from a forward-looking perspective, where such success, even if highly probable, has yet to be made real.[4]

3. Although Kant allows for an essentially retrospective justification of revolution, he still considers the punishment of a deposed monarch or other executive authority to be "a crime that remains forever and can never be expiated … like what theologians call the sin that cannot be forgiven in either this world or the next" (MS 6: 321n.). Here Kant alludes to the "eternal sin" of blaspheming or denying the Holy Spirit, a sin that supposedly makes salvation and reconciliation with God impossible. Kant portrays regicide as the political analogue of suicide, unpardonable because, if successful, the wrongdoer no longer exists to seek or receive forgiveness: "the execution of a monarch seems to be a crime from which the people cannot be absolved, for it is as if the state commits suicide" (MS 6: 321n.).

Kant's objection here is not simply to the killing (or, presumably, lesser punishment) of a previous authority by a revolutionary government in order to forestall attempts at a restoration. Although such killing is a crime, it is not unforgivable in the way that an attempt to punish an overthrown monarch under authority of law would be. In the case of regicide that does not pretend to be an exercise of lawful authority, "[the monarch's] murder is regarded as only an exception to the rule that the people makes its maxim" (MS 6: 322n.). In contrast, a judicial execution

> must be regarded as a complete overturning of the principles of the relation between a sovereign and his people (in which the people, which owes its existence only to the sovereign's legislation, makes itself his master), so that violence is elevated above the most sacred rights brazenly and in accordance with principle. (MS 6: 321n.)

[4] Kant then does countenance, at least in this one special case, the sort of "moral luck" that Bernard Williams and Thomas Nagel think he categorically rejects. See B. Williams, "Moral Luck," in B. Williams, *Moral Luck* (Cambridge University Press, 1981), pp. 20–39 and T. Nagel, "Moral Luck," in T. Nagel, *Mortal Questions* (Cambridge University Press, 1979), pp. 24–38. Nagel himself gives the example of political revolution as one where our exposure to moral luck is most profound.

Kant does not explain just what the difference is supposed to be between overturning a basic principle and merely making an exception to it. His thought may be that when the former monarch is killed to prevent counter-revolution, his killers, despite whatever power and authority they actually have, do not pretend to be exercising this authority. Instead, the killers act only as private individuals (a group that might be so large is to include all members of the body politic), and their crime remains that of ordinary murder, which is neither unforgivable nor inexpiable in principle.

Such murderous revolutionaries may be properly punished, although Kant suggests that they might still be able to invoke the "right" of necessity in their defense. Earlier in the Doctrine of Right, Kant argued that a particular institution is entitled to punish people only insofar as this legal regime serves to defend their rights. A legitimate regime of punishment is one that realizes some determinate scheme of rights by giving citizens reasons to respect those rights that do not depend upon their acceptance of the substantial correctness of any particular judgment about rights (or about that scheme of rights in general). In cases where coercion fails to give individuals such independent incentives to respect rights, punishment ceases to be justified.

Kant considers the case of shipwreck survivors clinging to bits of wreckage in order to stay afloat. Kant thinks that morally speaking, it would be a wrong tantamount to murder to push another survivor off his plank in order to save oneself. Yet Kant also holds that the plank-taker could not be properly punished for doing so, because no such punishment could give him any real incentive to respect his victim's rights that he did not already have. The state can at worst threaten wrongdoers with death, which is what the plank-taker would suffer anyway if he does not commit the crime. In this case, punishment can give the plank-taker no reasons to respect his victim's rights beyond the substantive merits of those rights-claims themselves. Kant does not think that cases of "necessity" give anyone a real authorization to violate anyone's rights. Instead, they present only circumstances in which the normal consequences of violation, punishment, can no longer be justly applied.[5]

[5] Kant's treatment of the right of necessity may suggest, against what I have argued, that he does admit the possibility of real states of nature coming into being, at least

However, those who would execute a former monarch under color of law cannot invoke these considerations of necessity. Necessity applies only to individuals as they might act within a particular civic order: it cannot apply to the actions of those supposedly exercising the sovereign authority of that order. Kant holds that we must always see ourselves as being bound together in a general will that recognizes whatever political powers that exist to be its empirical manifestation, for only on that assumption are morally coherent conditions of right possible. Suppose then that a revolution occurs, which, although wrongly begun, through its success acquires a retrospective justification. The new government must now be recognized as the empirical manifestation of the general will. If this new government tries and punishes those who acted under authority of the old regime, then we must conclude that either the old regime was not really the representative of the general will, or that the current regime is not. Moral agents cannot accept either conclusion. If the current regime does not express the general will, then, in the absence of any effective alternatives, a real state of nature now exists between us, and morality and practical reason itself are shown to be self-contradictory. On the other hand, if we accept that it is the old regime that had not been properly authorized by the general will, then we must conclude that, however things stand between us currently, in the past we occupied a real state of nature with respect to each other.

If we conceive of the state of nature being merely an unacceptably difficult or dangerous condition, then while we have strong reasons

between people who find themselves in such dire circumstances. However, the shipwreck case is not an example of people who truly have no enforceable rights against one another. Although Kant normally understands enforcement in terms of punishment, it also includes the use of force to prevent the successful performance of a wrongful act. While the state could not properly punish the plank-taker, it could properly use force to stop him (say, if a police officer was also among the shipwrecked). We might also think that, in the absence of a police officer, any general will would deputize the victim to defend himself by force, so that his resistance would no longer count as the sort of wrong that Kant attributes to unilateral defenses of right in the state of nature. If so, then the shipwreck situation does not generate the sort of contradiction in the demands of morality that would characterize true state of nature cases. For further discussion of the idea of such implicit deputization, see my "On the Supposed Duty of Truthfulness: Kant on Lying in Self-Defense," in C. Martin (ed.), *The Philosophy of Deception* (Oxford University Press, 2009), pp. 225–43.

against allowing such a condition to come to be, we need not worry about states of nature that no longer exist. Yet for Kant, the basic problem with a state of nature is not that it would make our lives "nasty … brutish and short." The real problem is that, in such a condition, pure practical reason would contradict itself, and once we recognize that has happened, whether in the past or in the present, we can no longer see such reason as truly lawgiving at any time. To avoid such an "absurdum practicum," we must not countenance any interpretation of our political lives, whether of the past or of the present, that would present them as being without an effective and legitimate public authority. Yet in punishing a past regime, a new government would make such an interpretation inescapable, forcing us into a situation where me must see either our present or past relations as being a real state of nature, where both alternatives reveal morality to be nothing more than a "chimera of the brain." This consequence may explain why Kant treats the judicial execution as a special, unforgivable sin that would involve "the complete overturning of all concepts of right," becoming thereby "a chasm that irretrievably swallows everything" (MS 6: 321n.). By contradicting itself this way, the state, and the general will that animates it, would indeed be committing suicide through "a principle that would have to make it impossible to generate a state that has been overthrown" (MS 6: 321). All this might happen while the outward functions of a polity continue undisturbed. For Kant, a state need not die only through physical destruction or civil war. Here we seem to have a political analogue of what, in the individual case, he cryptically calls "moral death." In the Doctrine of Virtue, Kant claims that were a person to lose his receptivity to "moral feeling,"

> he would be morally dead, and if (so to speak medically) the moral vital force could no longer excite this feeling, then humanity would dissolve (by chemical laws, as it were) into mere animality and be mixed irretrievably with the mass of other natural beings. (MS 6: 400)

It is not entirely clear what moral death means in the case of an individual; Kant says nothing more about it. Yet this sort of living death may be more readily understood in the political context. A people that engaged not only in revolution but in the punishment of

their former authorities could not coherently see morality's demands as being consistent, and hence would have to repudiate the basic principle that provides the grounds for recognizing themselves as a people constituted by a general will in the first place. Even though we might still structure our lives in terms of a certain set of political institutions, there would be no real authority, rights, legitimacy, external freedom, moral duty, or moral responsibility. Practical reason would have to be limited to mere prudence, and the norms of social life could be nothing more than an elaborate *modus vivendi* that emerges from the doings of what remain as merely self-interested, atomistic individuals.

In this case, our humanity would have dissolved into animality in the sense that, without the forms of freedom and responsibility defined by the moral law, human agency would not be fundamentally different in kind from the other kinds of causal power manifest in the natural world. Hobbes would then be right not just about our political and moral life, but about the basic metaphysics of human nature as well. The body politic might still continue to function quite well materially, but the spirit that unifies and animates it morally would be lost. Like the "unforgivable sin" of the theologians we would have repudiated the uncompromising demands of our conscience that Kant takes to be the true content of the idea of the Holy Ghost (R 6: 140n.).

4. In the Doctrine of Right, Kant never considers how a post-revolutionary society might deal with not just a deposed monarch, but with the subordinates who acted under his authority. Yet Kant's reasoning would seem to apply to all officers acting under color of state authority. At least, it is hard to see how such subordinates might be punished or otherwise held criminally liable for faithfully carrying out the directives of a commander, where that commander's authority cannot itself be called into question. If so, then the question of punishment, expiation, pardon, amnesty, or forgiveness cannot even arise, because for Kant no past exercise of state power could be considered wrongful or even prima facie punishable. All such state action would have to be accepted as a proper exercise of the general will, and hence as something that cannot conflict with any rights of citizens. After a successful revolution, all citizens would return to a state of political and legal innocence. There could be no role for any sort of public accounting, with an eye to either punishment or pardon,

because the unified perspective of the people could not recognize any such acts as being wrong in the first place.

This position may be tolerable when we consider relative minor wrongs that a previous regime may have committed, such as the expropriation of property or unfair assignment of offices. However, these claims become incredible when we consider crimes such as false imprisonment, forced migration, enslavement, rape, torture, murder, and genocide. How could any just polity ask any of its citizens to accept on equal terms those who had abused and humiliated them and their loved ones in any of these ways? If this is all that Kant can say about moral reconstruction, it would appear to be a *reductio* of his political philosophy, and perhaps his entire practical thought as well.

The problem here stems not just from common sense, but from sources internal to Kant's thought as well. For Kant, we avoid a morally inconsistent state of nature only insofar as each of us understands her identity not first as an individual, but as fundamentally a member of a collective whose members are bound to one another by certain norms of common deliberation and justification. If I am to adopt this stance toward my former oppressors, then I must acknowledge a will that was able to torture or humiliate me as, to some extent, my real will too. The resulting problem is not just that such acknowledgement would be extremely distasteful, or that it would be a burden too great to ask of any citizen, especially one who has already suffered greatly. Regardless of such concerns, Kantian morality would seem to positively forbid any such identification, even if one could find some way to stomach it. To recognize a kind of common moral identity with your torturer is in part to accept his perspective on you into your own perspective on yourself, and so to become complicit in his previous denial of your own humanity. Although Kant does not himself consider this problem, it seems that in recognizing such former tormentors as fellow members of a general will, we would be ratifying our own degradation, and so be committing what has to be a grave wrong against humanity in our own persons. Thus we have a new dilemma. We must not acknowledge our oppressor as a member of a general will with us, either then or now; but neither may we see ourselves as standing in a true state of nature with respect to him either, on pain of rendering morality incoherent. It seems that morality proscribes both

rejecting and reconciling with those who, acting on state authority, have gravely wronged us. But what other options could there be?

5. In the Doctrine of Right, Kant considers no responses to the wrongs of an overthrown regime other than full punishment or complete immunity. While he properly condemns the former alternative, he does not see that the latter is subject to equally deep moral objections as well. Fortunately, the Doctrine of Right does not exhaust Kant's resources for addressing the problem of moral reconstruction. This political dilemma also finds a close analogue in Kant's philosophy of religion, where he discusses the purely moral revolution that each individual supposedly must undergo during her life.

In the *Religion within the Boundaries of Mere Reason*, Kant takes the will of every human being to be grounded in a fundamental "disposition" that determines the degree of relative importance that the agent attaches to both morality and self-love. Kant understands morality and self-love to be both inherently rational interests, and no autonomous but embodied rational agent could be completely indifferent to either one. In my basic moral disposition, I may subordinate concerns of happiness to those of morality, and recognize the moral law as the final authority over our actions, or I might subordinate morality to self-love and refuse to suffer any real sacrifice of my happiness for purely moral reasons. Our true moral character is determined by the priority we assign to these basic concerns in our fundamental maxim of choice, which finds empirical expression in the particular maxims we act upon throughout the course of our lives. For reasons that are somewhat obscure, Kant thinks that every human being must start out life under the "dominion" of an evil disposition, which she must forever be striving to overcome (R 6: 51). Yet, since we are morally obligated to become morally good people, we must nevertheless be capable of effecting a true "revolution of character" that would make it the case that our lives, considered in their entirety, count as expressions of a fundamental commitment to morality before self-love.

There is much that is puzzling about Kant's understanding of this supposedly timeless and noumenal "change of heart" that is empirically expressed by a steady if gradual improvement in our behavior. In the second book of the *Religion*, Kant draws attention to a paradox that sounds very much like the political dilemma faced by a community attempting to reconstruct itself. Kant considers whether, once

having undergone the appropriate revolution in character, a person is properly subject to punishment for his past wrongs (or his prior wrongful disposition):

> Now since the question here is not whether, also before the human being's conversion, the punishment imposed upon him accorded with divine justice (as there is no doubt about this), the punishment is not to be thought (in this inquiry) as fully exacted before the human being's improvement. Also after his conversion, however, since he now leads a new life and has become a "new man," the punishment cannot be considered appropriate to this new quality (of thus being a human being well-pleasing to God). Yet satisfaction must be rendered to Supreme Justice, in whose sight no one deserving of punishment can go unpunished. (R 6: 73)

The problem here, as in the political case, is in understanding the proper moral relation that holds between the "old man" that precedes the conversion and the "new man" who supposedly issues from it. On the one hand, the new man cannot properly be held responsible for the acts or attitudes of the old man, for this would fail to acknowledge the moral significance of the revolution of character that has in fact occurred. Yet, neither can the old man be simply "written off" morally as another person who no longer exists, so that there is no longer any need for him to suffer punishment for his actions. In both individual conversion and political revolution, we are confronted with a situation where an agent (an individual or a polity) must both own and disown some portion of its past, so that it manages to be both blameless and yet a proper object of punishment for this past.

The *Religion* offers a way of resolving this paradox. Kant argues that the "new man," although he is morally distinct from the old, must nevertheless accept or take on the sins of the old as if they were fully his own, and in so doing render the "satisfaction" that justice demands:

> Physically ([i.e.] considered in his empirical character as a sensible being) he still is the same human being liable to punishment, and he must be judged as such before a moral tribunal of justice and hence by himself as well. Yet, in his new disposition (as an intelligible being), in the sight of a divine judge for whom the disposition takes the place of the deed, he is morally another being. And this disposition which he has incorporated in all its purity, like

unto the purity of the Son of God … bears as vicarious substitute the debt of sin for him. (R 6: 74)

Kant seems to think that, in the course of a real moral revolution, the reformed person must accept some sort of punishment as if it were properly due him, although strictly speaking that punishment is due another for whom the new man only stands surety. Kant sees this situation as a form of vicarious atonement that manages to satisfy the conflicting demands of justice so that a person "well-pleasing to God" might then emerge out of life that begins, like all human lives, in a morally corrupted state.

Strictly speaking, the required punishment is not something that can be accepted after the conversion has been completed (given that it is then wholly inappropriate). Instead, Kant seems to think that the acceptance of the punishment is itself the culmination of the conversion itself, and as such ratifies the new man as someone morally different from the old through the new man's willingness to accept the suffering only properly due the old:

But, since neither before nor after conversion is the punishment in accordance with divine wisdom but is nevertheless necessary, the punishment must be thought as adequately executed in the situation of conversion itself … The emergence from the corrupted disposition into the good is in itself already sacrifice (at "the death of the old man," "the crucifying of the flesh") and entrance into a long train of life's ills which the new human being undertakes in the disposition of the Son of God, that is, simply for the sake of the good, yet are still fitting punishment for someone else, namely the old human being (who, morally, is another human being). (R 6: 74)

Here Kant captures something of the essential paradox of repentance and atonement. To sincerely repent, one must accept punishment (or blame) that, insofar as it is accepted, ceases to be entirely appropriate to inflict. But if it were morally wrong to inflict such punishment, then the penitent should not be able in good conscience to submit to it, and hence he could not sincerely repent. In this case, the penitent would have to remain culpable, making it appropriate to punish him, which he could now in good conscience accept, etc. To escape this situation, Kant allows that a kind of punishment may be called for even in cases where the subject is no longer truly

blameworthy, insofar as both the punishment and its acceptance is itself the final part of what it is to become a new, ultimately blameless person.

This picture of repentance and expiation is hardly clear in the case of an individual, and it admits of no very obvious translation into a political context. Nevertheless, there are some promising parallels. Like the penitent individual, a reformed polity has undergone a constitutional "change of heart" such that it needs to acknowledge some previous wrongs as truly its own while at the same time disowning them entirely. Our deposed regime's acts both are and are not ours in something like the sense in which my former bad acts both are and are not mine in the moment of repentance. For Kant, this peculiar relation is possible only if we are willing to accept responsibility for our acts in a way that goes beyond what could be demanded of us on grounds of strict justice.

What then would it mean for a body politic as a whole to accept some sort of punishment that it does not fully deserve as a sort of vicarious atonement for its own past? Since the body politic is not something distinct from the individuals who compose it, such suffering would have to be borne by and distributed among the citizens of the new order in some sort of morally appropriate way. On the one hand, former state actors who did wrongs, although properly immune from punishment because they acted with legal authority, would nevertheless have a moral obligation not to assert such immunity, thereby accepting a punishment that cannot be properly demanded of them. On the other hand, those wronged by the old regime would for their part have to bear the burden of accepting these malefactors back as full citizens, restored to something like normal civic personality. To do so is to accept a kind of suffering that goes beyond what others can demand of them as well. Yet if the body politic as a whole is going to atone for its wrongs against itself, then some tasks that go beyond justice must fall to those who fundamentally identify with that body, tormentor and victim alike. Just as past oppressors have a right that they nevertheless should not assert, former victims must accept as a political duty something that no one else can claim from them. Here there is a kind of "teleological suspension" of the normal logic of justice, needed to make moral relations between members of community, and a moral relation to the events of its past, possible at the same time.

Kant's practical philosophy may not be able to give any very specific answers to how a polity that has undergone profound moral change should deal with those who, under color of legal authority, committed what are now recognized to be great injustices against its own people. However, Kant's thought can allow that some sort of public accounting and perhaps punishment is appropriate, even if this must fall short of what such wrongdoing, considered apart from the broader political context, would deserve. Despite Kant's official pronouncements about both the necessity of punishment and the immunity of legal authorities, he has the resources to recognize the moral appropriateness of such institutions as South Africa's Truth and Reconciliation Commission, which approximates but does not fully conform to the structure of a proper criminal court.[6]

It is important to see that Kant's account bears only on the dimensions of our lives that involve relations of right and associated questions of "external freedom" and political power. This understanding of moral reconstruction does not address those aspects of our relationships in which questions of right, law, and punishment are not involved. While victims may be under something like a political obligation to accept their former tormentors as fellow citizens, these victims are not thereby obliged to accept them as potential friends or intimates in any way, or to accord them any more concern that what is owed to people in general as a strict matter of right. Although Kant holds that we stand under an imperfect duty to care about the welfare of human beings generally, such concern for any particular person cannot be demanded from anyone as a matter of justice. Victims may then be perfectly entitled to remain completely indifferent to the well-being of their former tormentors, or to even rejoice in their misfortunes. The general will may ask many things of us, but even for Kant, full forgiveness must remain a purely personal prerogative.

[6] For an important discussion of how the TRC might be understood from a broadly Kantian perspective, see B. Herman, "Contingency in Obligation," in B. Herman, *Moral Literacy* (Cambridge, MA: Harvard University Press, 2007), pp. 300–31.

Select bibliography

Adams, R. M., "Introduction," in A. Wood and G. di Giovanni (eds.), *Immanuel Kant: Religion within the Boundaries of Mere Reason and Other Writings* (Cambridge University Press, 1998).

Allison, H. E., *Lessing and the Enlightenment* (Ann Arbor, MI: University of Michigan Press, 1966).

Kant's Theory of Freedom (New York: Cambridge University Press, 1990).

"Reflections on the Banality of (Radical) Evil: A Kantian Analysis," in *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge University Press, 1996).

"Ethics, Evil and Anthropology in Kant: Remarks on Allen Wood's *Kant's Ethical Thought*," *Ethics* **111**, 3 (April 2001).

"On the Very Idea of a Propensity to Evil," *Journal of Value Inquiry*, **36**, 2–3 (2002).

Anderson-Gold, S., "Kant's Rejection of Devilishness: The Limits of Human Volition," *Idealistic Studies*, **14** (1984).

"Kant's Ethical Commonwealth: The Highest Good as a Social Goal," *International Philosophical Quarterly*, **26** (1986).

"God and Community: An Inquiry into the Religious Implications of the Highest Good," in P. Rossi and M. Wreen (eds.), *Kant's Philosophy of Religion Reconsidered* (Bloomington: Indiana University Press, 1991).

"Kant's Ethical Anthropology and the Critical Foundations of the Philosophy of History," *History of Philosophy Quarterly*, **11**, 4 (1994).

Cosmopolitanism and Human Rights (Cardiff: University of Wales Press, 2001).

Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant (Albany: State University of New York Press, 2001).

"The Cosmopolitan Foundations of the Kantian State" (unpublished manuscript).

Arendt, H., *Eichmann in Jerusalem: A Report on the Banality of Evil*, revised and enlarged edition (New York: Penguin, 1977).

The Origins of Totalitarianism (San Diego: Harcourt, 1994).

The Portable Hannah Arendt, ed. P. Baehr (New York: Penguin, 2000).

Arendt, H. and Jaspers, K., *Correspondence 1926–1969*, trans. L. Kohler and H. Saner (New York: Harcourt Brace Jovanovich, 1992).

Aristotle, *The Nicomachean Ethics*, trans. D. Ross (New York: Oxford University Press, 1925).

Auerbach, E., *Mimesis: The Representation of Reality in Western Literature* (Princeton University Press, 1968).

Augustine, St., *The City of God*, trans. Henry Bettenson (Harmondsworth: Penguin, 1977).

Bauman, Z., *Modernity and the Holocaust* (Oxford: Polity Press, 1989).

Beck, L. W., *A Commentary on Kant's Critique of Practical Reason* (University of Chicago Press, 1960).

Beiser, F., "Moral Faith and the Highest Good," in P. Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge University Press, 2006).

Benhabib, S., *The Rights of Others: Aliens, Residents and Citizens* (Cambridge University Press, 2004).

Bernstein, R. J., *Radical Evil: A Philosophical Interrogation* (Cambridge, MA: Polity Press, 2002).

The Abuse of Evil: The Corruption of Politics and Religion since 9/11 (Cambridge, MA: Polity Press, 2005).

Bielefeldt, H., *Symbolic Representation in Kant's Practical Philosophy* (Cambridge University Press, 2003).

Branham, G., *Kant's Practical Philosophy: From Critique to Doctrine* (Basingstoke and New York: Palgrave Macmillan, 2003).

Card, C., *The Atrocity Paradigm: A Theory of Evil* (New York: Oxford University Press, 2002).

Caswell, M., "Kant on the Diabolical Will: A Neglected Alternative?" *Kantian Review*, **12**, 2 (2007).

Dews, P., "Disenchantment and the Persistence of Evil," in A. Schrift (ed.), *Modernity and the Problem of Evil* (Bloomington: Indiana University Press, 2005).

Doris, J., *Lack of Character: Personality and Moral Behavior* (Cambridge University Press, 2002).

DuBois, W. E. B., *The Souls of Black Folk* (New York: New American Library, 1969).

Eberhard, J. A., "Über das Kantische radicale Böse in der menschlichen Natur," *Philosophisches Archiv*, **2**, 2 (1794).

Fackenheim, E., "Kant and Radical Evil," *University of Toronto Quarterly*, **23** (1954).

Ferrara, A., "The Evil That Men Do: A Meditation on Radical Evil from a Postmetaphysical Point of View," in M. P. Lara (ed.), *Rethinking Evil: Contemporary Perspectives* (Berkeley: University of California Press, 2001).

Frei, H.W., *The Eclipse of Biblical Narrative: A Study in Eighteenth and Nineteenth Century Hermeneutics* (New Haven and London: Yale University Press, 1974).

Frierson, P., *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge University Press, 2003).

"The Moral Importance of Politeness in Kant's Anthropology," *Kantian Review*, **9** (2005).

"Review: Richard Dean, *Kant and the Value of Humanity*," *Notre Dame Philosophical Reviews*, **04.17** (2007), online at http://ndpr.nd.edu.

"Two Standpoints and the Problem of Moral Anthropology" (unpublished manuscript).

Frye, M., *Willful Virgin: Essays in Feminism 1976–1992* (Freedom, CA: The Crossing Press, 1992).

Gilbert, P., *New Wars, New Terrors* (Edinburgh University Press, 2003).

Grant, R.W. (ed.), *Naming Evil, Judging Evil*, Foreword by A. MacIntyre (University of Chicago Press, 2006).

Grenberg, J., *Kant and the Ethics of Humility: A Story of Dependence, Corruption, and Virtue* (Cambridge University Press, 2005).

Guyer, P., "Kant's Deductions of the Principles of Right," in M. Timmons (ed.), *Kant's Metaphysics of Morals: Interpretative Essays* (Oxford University Press, 2002).

"The Strategy of Kant's Groundwork," in *Kant on Freedom, Law, and Happiness* (Cambridge University Press, 2000).

Harman, G., *Explaining Value and Other Essays in Moral Philosophy* (Oxford University Press, 2000).

Hauerwas, S., "Seeing Darkness, Hearing Silence," in R. Grant (ed.), *Naming Evil, Judging Evil* (University of Chicago Press, 2006).

Herman, B., *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993).

"Making Room for Character," in S. Engstrom and J. Whiting (eds.), *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty* (Cambridge University Press, 1996).

"Contingency in Obligation," in *Moral Literacy* (Cambridge, MA: Harvard University Press, 2007).

Hewitt, A., "The Bad Seed: 'Auschwitz' and the Physiology of Evil," in J. Copjec (ed.), *Radical Evil* (London and New York: Verso, 1996).

Holtman, S.W., "Revolution, Contradictions, and Kantian Citizenship," in M. Timmons (ed.), *Kant's Metaphysics of Morals* (Oxford University Press, 2002).

Kamtekar, R., "Situationism and Virtue Ethics on the Content of Our Character," *Ethics*, **114** (2003).

Kant, I., *Anthropology from a Pragmatic Point of View*, trans. and ed. R.B. Louden, Cambridge Texts in the History of Philosophy (Cambridge University Press, 2006).

Cambridge Edition of the Works of Immanuel Kant (New York: Cambridge University Press, 1992–).

Critique of Pure Reason, trans. N.K. Smith (New York: St. Martin's Press, 1965).

Foundations of the Metaphysics of Morals, trans. L.W. Beck (Indianapolis and New York: Bobbs-Merrill, 1959).

Immanuel Kants gesammelte Schriften, Ausgabe der königlich preussischen Akademie der Wissenschaften (Berlin: W. de Gruyter, 1902–).

Kant: Political Writings, ed. Hans Reiss (Cambridge University Press, 1970).

Religion within the Limits of Reason Alone, trans. T.M. Grene and H.H. Hudson (LaSalle: Open Court, 1934; 2nd edn, New York: Harper & Row, 1960).

Korsgaard, C.M., *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1996).

"Taking the Law into our Own Hands: Kant on the Right to Revolution," in A. Reath, B. Herman, and C.M. Korsgaard (eds.), *Reclaiming the History of Ethics* (Cambridge University Press, 1997).

Kosch, M., *Freedom and Reason in Kant, Schelling, and Kierkegaard* (Oxford: Clarendon Press, 2006).

Lara, M.P., (ed.), *Rethinking Evil: Contemporary Perspectives* (Berkeley, CA: University of California Press, 2001).

Levi, P., *The Drowned and the Saved* (New York: Vintage, 1989).

Lifton, R.J., *The Nazi Doctors: Medical Killing and the Psychology of Genocide* (New York: Basic Books, 1986).

Louden, R.B., *Kant's Impure Ethics: From Rational Beings to Human Beings* (New York: Oxford University Press, 2000).

"Anthropology from a Kantian Point of View: Toward a Cosmopolitan Conception of Human Nature," *Studies in History and Philosophy of Science A*, **39** (December 2008).

May, L. *Crimes against Humanity: A Normative Account* (Cambridge University Press, 2005).

McBride, W.L., "Liquidating the 'Nearly Just Society': Radical Evil's Triumphant Return," in A. Schrift (ed.), *Modernity and the Problem of Evil* (Bloomington: Indiana University Press, 2005).

McDougall, S. and Harris, P., *The Woman Who Wouldn't Talk* (New York: Carol & Graf, 2003).

Michalson, G.E., Jr., *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge University Press, 1990).

"Moral Regeneration and Divine Aid in Kant," *Religious Studies*, **25** (1989).

Milgram, S., "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology*, **67** (1963).

Montaigne, M.E., *Essais*, (ed.), A. Tournon (Paris: Imprimerie nationale, 1998).

The Complete Essays, (trans.), M.A. Screech (London: Penguin Books, 1991).

Morgan, S., "The Missing Formal Proof of Humanity's Radical Evil in Kant's *Religion*," *The Philosophical Review*, **114**, 1 (January 2005).

Muchnik, P., "On the Alleged Vacuity of Kant's Concept of Evil," *Kant-Studien*, **4** (2006).

Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Aprioricity of History. (Lanham, MD: Lexington Books, 2009).

Munzel, F., *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment* (University of Chicago Press, 1999).

Nagel, T., *Mortal Questions* (Cambridge University Press, 1979).

Neiman, S., *The Unity of Reason: Rereading Kant* (New York: Oxford University Press, 1994).

Evil in Modern Thought: An Alternative History of Philosophy (Princeton University Press, 2002).

Nietzsche, F., *Beyond Good and Evil: Prelude to a Philosophy of the Future* (New York: Vintage Books, 1966).

O'Connor, D., "Good and Evil Disposition," *Kant-Studien*, **3** (1985).

O'Neill, O. "Reason and Autonomy in *Grundlegung III*," in *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge University Press, 1989).

Packer, G., "Knowing the Enemy: The Anthropology of Insurgency," *The New Yorker* (December 18, 2006).

Pinkard, T., *German Philosophy 1760–1869: The Legacy of Idealism* (Cambridge University Press, 2002).

Prauss, G., *Kant über Freiheit als Autonomie* (Frankfurt am Main: Vittorio Klostermann, 1983).

Reath, A., *Agency and Autonomy in Kant's Moral Theory* (Oxford: Clarendon Press, 2006).

Rossi, P.J., "Autonomy and Community: The Social Character of Kant's Moral Faith," *Modern Schoolman*, **61** (1984).

The Social Authority of Reason: Kant's Critique, Radical Evil, and the Destiny of Humankind (Albany: State University of New York Press, 2005).

Rousseau, J.J., *Discourse on the Origin of Inequality*, (trans.), D. Cress (Indianapolis: Hackett, 1992).

Scanlon, T.M., *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

Schelling, F.W.J., *Of Human Freedom* (Chicago: Open Court Publishing Co., 1936).

Schrift, A.D., (ed.), *Modernity and the Problem of Evil* (Bloomington: Indiana University Press, 2005).

Schulte, C., *Radikal Böse. Die Karriere des Bösen von Kant bis Nietzsche* (Munich: W.F. Verlag, 1991).

Silber, J.R., "The Ethical Significance of Kant's *Religion*," in T.M. Grene and H.H. Hudson (eds.), *Religion within the Limits of Reason Alone* (New York: Harper & Row, 1960).

"Kant at Auschwitz," in G. Funke and T.M. Seebohm (eds.), *Proceedings of the Sixth International Kant Congress* (Washington, D.C.: Center for Advanced Research in Phenomenology and University Press of America, 1991).

Silverman, S.W., *Because I Remember Terror, Father, I Remember You* (Athens, GA: University of Georgia Press, 1996).

Smith, A., *The Theory of Moral Sentiments* (Amherst, NY: Prometheus Books, 2000).

Staub, E., *The Roots of Evil* (Cambridge University Press, 1989).

Stern, P., "The Problem of History and Temporality in Kantian Ethics," *Review of Metaphysics*, **39** (1986).

Strawson, G., "The Impossibility of Moral Responsibility," *Philosophical Studies*, **75** (1994).

Striblen, C., "Guilt, Shame and Shared Responsibility," *Journal of Social Philosophy*, **38**, 3 (Fall 2007).

Sussman, D., "Kantian Forgiveness," *Kant-Studien*, **96** (2005).

"Shame and Punishment in Kant's *Doctrine of Right*," *The Philosophical Quarterly*, **58** (April 2008).

"On the Supposed Duty of Truthfulness: Kant on Lying in Self-Defense," in C. Martin (ed.), *The Philosophy of Deception* (Oxford University Press, 2009).

Taylor, C., *Sources of the Self: The Making of the Modern Identity* (Cambridge, MA: Harvard University Press, 1989).

Timmons, M., "Evil and Imputation in Kant's Ethics," *Jahrbuch fűr Recht und Ethik*, **2** (1994).

Vetlesen, A., *Evil and Human Agency* (Cambridge University Press, 2005).

Watkins, E., *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005).

Williams, B., *Moral Luck* (Cambridge University Press, 1981).

Problems of the Self (Cambridge University Press, 1973).

Wilshire, B., *Get'Em All! Kill'Em!* (Lanham, MD: Lexington Books, 2006).

Wood, A.W., *Kant's Moral Religion* (Ithaca, NY: Cornell University Press, 1970).

"Kant's Compatibilism," in Wood (ed.), *Self and Nature in Kant's Philosophy* (Ithaca, NY: Cornell University Press, 1984).

"Kant's Deism," in P.J. Rossi and M. Wreen (eds.), *Kant's Philosophy of Religion Reconsidered* (Bloomington: Indiana University Press, 1991).

"Unsociable Sociability: The Anthropological Basis of Kantian Ethics," *Philosophical Topics*, **19** (1991).

"Rational Theology, Moral Faith, and Religion," in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge University Press, 1992).

"Kant's Historical Materialism," in J. Kneller and S. Axinn (eds.), *Autonomy and Community: Readings in Contemporary Kantian Social Philosophy* (Albany: State University of New York Press, 1998).

Kant's Ethical Thought (New York: Cambridge University Press, 1999).

"Religion, Ethical Community and the Struggle against Evil," *Faith and Philosophy*, **17**, 4 (2000).

Yovel, Y., "Bible Interpretation as Philosophical Praxis: A Study of Spinoza and Kant," *Journal of the History of Philosophy*, **11** (1973).

Kant and the Philosophy of History (Princeton University Press, 1980).

Index