

Kant's Anatomy of Evil, edited by Sharon Anderson-Gold and Pablo Muchnik. Cambridge: Cambridge University Press, 2010. Pp 251, hardcover. ISBN 978-0-521-51432-3.

Kant's Anatomy of Evil grapples with Kant's important but difficult thoughts on the question of evil found in his *Religion within the Boundaries of Mere Reason* (1996). This edited collection contains ten new and significant essays by scholars who have all previously published research on this topic. The overarching theme that unites this collection is the critical interrogation of Kant's radical evil thesis. That thesis states that since all humans have an evil disposition (*Gesinnung*) we may characterise the human species as having a universal propensity (*Hang*) to evil. A disposition is a character-defining highest order maxim. An evil disposition sometimes prioritises self-love over morality, whereas a good disposition always prioritises morality over self-love. The focus of this review will be to briefly examine each contribution in relation to this overarching theme.

Philip J. Rossi, S.J. examines the normative gap between *is* and *ought* that is exposed by the presence of evil in the world and tries to connect this with Kant's claims about the unity of reason. Rossi argues that "the unity of reason is not given beforehand but rather *enacted by the exercise of our finite freedom in resistance to evil*" (15). But this argumentative approach threatens to conflate two distinct issues. As Rossi argues, we can close (or get close to closing) the normative gap between *is* and *ought* by making the world as it ought to be, morally speaking, through action and by overcoming our radical propensity to evil. This will solve the problem of what Rossi calls being "homeless" in a morally inhospitable world (24). But this still leaves the unity of reason problem unsolved, since we would still face the problem of making sense of how the positive freedom that we must suppose that we have, from a practical standpoint, is possible within a world understood theoretically as a causally determined whole. And no amount of enacting resistance to evil can solve *that* unity of reason problem.

Patrick Frierson investigates the "moral pessimism", the view "that (at least) most people (at least) most of the time are *morally* deficient" (34), implied by Kant's radical evil thesis. Frierson looks at empirical evidence which shows that people tend not to display cross-situational consistency in the way that they act. Frierson argues that such evidence does not demonstrate "character's moral irrelevance" (38), as morally optimistic situationists such as John Doris (2002) seek to conclude. Instead it helps to empirically confirm Kant's morally pessimistic thesis that since most people have morally corrupt dispositions they are not likely to act consistently on valid principles across diverse situations. Further, Frierson emphasises the core role for moral communities that Kant envisages as a corrective to such widespread and radical moral failure. The job of such concrete moral communities is to improve the operative rules of moral salience in concrete communities, reduce situations which create a temptation to

vice, and encourage the formation of a consistent and principled moral character among its members.

Gordon E. Michalson, Jr. examines Kant's account of the moral revolution a person undergoes when he replaces his radically evil disposition with a good disposition. Unfortunately Kant's account of this revolution has struck many a reader as confusing, if not utterly confused. Michalson notes these confusions and emphasises Kant's use of biblical references in his characterisation of this moral revolution as the 'rebirth' of a 'new man'. Michalson argues that "the biblical reference itself is the element indispensable to Kant's completing his argument ... Appeal to the Bible is itself the stand-in for philosophical reasoning" (68). How an appeal to the Bible can 'stand-in' for philosophical reasoning is not at all clear to this reviewer. But in any case, Michalson's arguments can only work in the absence of a convincing philosophical account of a moral revolution. Thankfully the rudiments of such an account are not difficult to give. A person undergoes a moral revolution when she goes from seeing the worth of (at least some) persons as comparative to seeing the worth of all persons as absolute, and always consistently follows through in all her actions with this revolution in her views about the value of persons. As such, when she has an incentive which proposes to her a maxim which devalues the worth of any person she fails to see that incentive as providing her with a reason for action, although pre-revolution she sometimes did. No doubt such a revolution in one's disposition is no easy feat for us frail and impure humans to achieve. But it does seem possible to at least give a coherent philosophical account of a moral revolution without requiring biblical references to fill in any gaping philosophical holes.

Claudia Card offers an excellent summary of the six main theses that she reads Kant as developing in Book I of *Religion*. Having previously dealt with some of these theses in her insightful book *The Atrocity Paradigm* (2002), Card focuses here on Kant's excluded middle, the denial that "there is anything intermediate between a good will and an evil one" (75). Card accepts that 'right' and 'wrong' are contradictories (74), so that if an action is not right then it is wrong and vice versa. As such, Card has only potentially identified a problem for Kant's account if we read Kant as meaning by 'evil' something other than 'wrong'. Card thinks that we should read Kant in this way (80), whereas I think that this is a mistake. It is a mistake because in his account of radical evil Kant is examining the preconditions of *wrongdoing per se*, namely the adoption of an evil disposition. Thus Kant, unlike Card, is not trying to offer a theory of evil in the modern sense of the term, that is, a theory of which subclass of wrong acts go beyond the moral pale of mere wrongdoing (see Formosa (2008)).

There seems to be no reason, then, why we cannot supplement Kant's account of wrongdoing, which is his focus, with a theory of evil in the modern sense of the term. But one might doubt whether this is a viable project since, as Card convincingly argues, any plausible

theory of evil must be able to take into account the different degrees of *harm* inflicted, or intended to be inflicted, by wrongdoings (83). But isn't the Kantian unable to take harm into moral account? To resolve this significant problem Card suggests that the Kantian can take harm into moral account by focusing on the fact that the absolute worth of persons derives, for Kant, from their possession of rational capacities for self-government (75). While all wrongdoings express disrespect for the worth of persons, not all forms of disrespect are the same. Some wrongs culpably infringe upon another person's exercise of their rational capacities, some wrongs culpably fail to cultivate, encourage, or protect another person's rational capacities, whereas other wrongs culpably and deeply harm a person's very ability to exercise their rational capacities. Actions in the last group, such as murder and torture, which involve "radical harm" (84) to a person's capacity for self-government, might be thought for that reason to be evil, and therefore morally different to other lesser wrongs, such as lying, which are not radically harmful in this way.

Even so, isn't Card right to claim that there are just two types of persons or wills in the Kantian moral universe, namely those with a good disposition (probably none of us) and those with an evil disposition (probably all of us)? Isn't there an excluded middle here? To illustrate this point Card gives the example of a Robin Hood with scruples (85-86). Card's Robin Hood knows that his acts of stealing are wrong, but has scruples. He would never murder or torture to further his ends. Isn't such a person, even if he is not good, morally better than a Robin Hood without scruples, one who ruthlessly tortures and murders, even children, to further his ends? However, Kant can arguably accommodate a moral difference between the two Robin Hoods at the level of mid-level principles. (It is important to appreciate that Kant's idea of a disposition, or highest-order maxim, introduces a hierarchal understanding of maxims into his account of action). Robin Hood with scruples has a mid-level maxim that endorses robbing but not torturing and murdering, whereas Robin Hood without scruples has a mid-level maxim that endorses robbing, murdering and torturing. Both mid-level maxims imply an identical highest-order evil disposition to not put morality first. But one mid-level maxim is morally worse than the other on the grounds that one endorses merely infringing upon the exercise of rational agency by depriving a person of some of their means through theft, whereas the other also endorses utterly destroying or undermining that capacity through murder or torture. Of course this is merely a sketch of a solution and the very important issues that Card raises deserve further attention.

Robert B. Loudon notes a worry that arises with Kant's claim that radical evil involves the prioritising of self-love over morality, since this seems to imply that all evil is simply a matter of selfishness. But what about self-sacrificing acts of evil, such as those committed by the fanatic or terrorist? Such acts don't seem to be motivated by selfishness. Loudon defuses such worries by arguing that Kant's claim about self-love is not one about the actual *motivation*

of evildoers. Kant clearly thinks that evildoers are motivated in many different ways (as Wood also notes in this volume (154)), not merely by selfishness, but also by malice, envy, rage and a desire for revenge, as well as ideological beliefs about the lack of worth of other persons or groups (see Formosa (2009)). As such, we should read Kant's claim about self-love as a *structural* claim about the necessity of presupposing an evil disposition, a disposition not to "put the moral law first" (103), as the basis of all evil, whether self-sacrificing or selfish. However, at times Louden seems to imply that an evil disposition is one that *always* puts self-love above morality (103). But, as Card's example of a Robin Hood with scruples shows, a person may *sometimes* put morality above self-love without *always* doing so. An evil disposition is therefore best understood as a disposition to *not always* make morality the condition of self-love, rather than as a disposition to *always* make self-love the condition of morality.

Pablo Muchnik examines Kant's proof of his radical evil thesis. Kant's argument for this thesis has two parts (see Formosa (2007)). First, from the presence of a single wrongful maxim Kant argues that we can infer in that person the presence of an evil disposition, since a person with a good disposition (one that always puts morality first) would never adopt such a maxim. Second, given the sheer ubiquity of evil acts and the evil dispositions which underlie them, we must conclude that an evil disposition is universal among human beings, and this justifies us in claiming that there is a species wide propensity to evil which is nonetheless freely self-imposed. The second part of this argument is often seen as problematic since a transcendental deduction or proof, and not the empirical generalisation that Kant seems to offer, is required to defend a claim of universality. To further complicate matters Kant says that a transcendental deduction of his thesis is required, and then says it isn't, and then says that he has already provided such a proof (125-27). Muchnik argues that Kant's sometimes promised proof can be found in the preface to the first edition of *Religion*. However, while this is an interesting suggestion, in the end Muchnik still relies on the claim that "observation of human conduct gives us no 'cause (*Grund*) for exempting any one'" (141) from the charge of having an evil disposition. But such an empirical generalisation can at best support the view that an evil propensity is widespread, and not that it is universal. However, this still leaves us with strong support for the weaker thesis that an evil disposition is very widespread (though perhaps not universal) among human beings.

In his contribution Allen W. Wood revisits his influential claims about the social origins of evil in Kant's theory (159-165). On Wood's view the origin of our radical propensity to evil is our unsocial sociability, our need as rational beings for society with others which, however, is also unsociable as it takes the form of a need to gain superiority over others in terms of the possession of honour, power and wealth (162). In the following chapter Jeanine M. Grenberg challenges Wood's views about the *purely* social origins of evil. On Grenberg's alternative account the origin of humanity's radical propensity to evil is, put simply, the "transcendental

anxiety” brought on by the condition of being the kind of being “who seek[s] both happiness and morality”, but who cannot achieve this happiness entirely on his or her own (182). This existential condition leads to a universal tendency to place our own happiness above morality.

If Wood is right then under ideal social conditions, that is, where there are domestic and cosmopolitan institutions which successfully implement the rule of right and strong ethical communities which successfully implement the rule of virtue, humans would not (at least by and large) adopt evil dispositions. If Grenberg is right then even under such ideal social conditions humans would still, due to their transcendental anxiety, universally adopt evil dispositions. Adjudicating between these two competing views, which both have strong textual support, might simply be an empirical matter that we cannot as yet resolve since we have no experience of ideal social conditions. In any case, given that we can at least conceptually pull Kant’s radical evil and unsocial sociability theses apart, it is a mistake to use the terms as strictly synonymous, as Wood wants to (164). We should keep Kant’s claim about the universal presence of evil dispositions among humans separate from distinct claims about the origin of that propensity in either or both of our social or existential conditions.

Sharon Anderson-Gold examines issues surrounding genocide and crimes against humanity. Genocide is, among other things, a crime perpetrated by one group (or groups) against another group (or groups). Can Kant’s seemingly individualistic account of evil in terms of a corrupted disposition which prioritises self-love over morality shed any light on such large-scale and complex evils such as genocide? Anderson-Gold convincingly argues that it can. She does so by extending the idea of individual self-love, as the prioritising of one’s own condition above the value of the humanity of others, to the level of exclusionary group identities, which involve the members of one group valuing themselves as persons more highly than the members of another group. When this happens the members of one group both devalue the other group as a form of *human* identity and devalue the *humanity* of the members of that group (198). Viewing other groups in this way seems to be a necessary precondition of genocide, and Anderson-Gold’s extension of Kant’s views on radical evil to be able to account for this is an important achievement.

In the final chapter David Sussman looks at the issue of justice after evil by examining Kant’s views on post-revolution punishment. It is Kant’s view that post-revolution it is wrong to punish previous rulers and their functionaries for their perceived wrongdoings while in office since (quoting Kant) “as the source of law, [the ruler] can do no wrong” (216). This would seem to imply, for example, that the Nazis could do no wrong while in office and that after the war it would be wrong to punish them. In order to see how the Kantian might resist this troubling conclusion, Sussman draws an analogy between the case of an individual moral revolution and a political revolution. In his account of the former Kant argues that it is legitimate to punish the

morally reborn new man for the crimes of the old man he used to be, and that the new man should take on his punishment as a form of “vicarious atonement” (233). Applying by analogy the same reasoning to the political case, Sussman argues that “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” (234). This is not, however, a promising argumentative path. Firstly, it implies, for example, that Nazis would be doing us a moral favour in allowing us to punish them. Secondly, it implies that we should punish people whom we think we have no right to punish simply in order to give them the opportunity to engage in vicarious atonement. Thirdly, the analogy Sussman relies on doesn’t hold since what makes punishment acceptable in the moral case is that the new and the old man are numerically the same person, whereas no similar identity holds in the political case of the new and the old government. Of course, this still leaves us with tough questions, which Sussman helpfully raises, about what Kantians should say about justice after political evil.

Overall this is an engaging and important collection of essays on a topic that deserves a wide audience. It is a book that should be read not merely by Kant scholars, but by anyone interested in moral philosophy. Taking into account the sheer ubiquity of human evil is essential to properly theorising moral space, and as this collection of fine essays expertly demonstrates, there is still much we can learn about this by continuing to read Kant.

Paul Formosa

Department of Philosophy, Macquarie University, Australia

References

- Card, Claudia. *The Atrocity Paradigm: A Theory of Evil*. Oxford: Oxford University Press, 2002.
- Doris, John M. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press, 2002.
- Formosa, Paul. "A Conception of Evil." *Journal of Value Inquiry* 42, no. 2 (2008): 217-239.
- . "Kant on the Limits of Human Evil." *Journal of Philosophical Research* 34 (2009): 189-214.
- . "Kant on the Radical Evil of Human Nature." *The Philosophical Forum* 38, no. 3 (2007): 221-245.
- Kant, Immanuel. "Religion within the Boundaries of Mere Reason." In *Religion and Rational Theology*, edited by Allen Wood and George Di Giovanni, 39-216. Cambridge: Cambridge University Press, 1996.