Kantian Review

Date of delivery:

Journal and vol/article ref: krv 1900010

Number of pages (not including this page): 6

This proof is sent to you on behalf of Cambridge University Press. Please check the proofs carefully. Make any corrections necessary on a hard copy and answer queries on each page of the proofs.

Please return the marked proof within 4 days of receipt to:

By email to Richard Aquila at raquila@utk.edu - Please clearly indicate any corrections required by page and line reference or annotate directly onto the PDF with electronic mark up.

Authors are strongly advised to read these proofs thoroughly because any errors missed may appear in the final published paper. This will be your ONLY chance to correct your proof. Once published, either online or in print, no further changes can be made.

To avoid delay from overseas, please send the proof by airmail or courier.

If you have no corrections to make, please email raquila@utk.edu to save having to return your paper proof. If corrections are light, you can also send them by email, quoting both page and line number.

- The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted, unless discussed with the editor of the journal. Only one set of corrections are permitted.
- Please answer carefully any author queries.
- Corrections which do NOT follow journal style will not be accepted.
- A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the typesetter is required.

Copyright: if you have not already done so, please download a copyright form from: http://journals.cambridge.org/images/fileUpload/documents/KRV_ctf.pdf and return to mcox@cambridge.org

- If you have problems with the file please contact krvproduction@cambridge.org

Please note that this pdf is for proof checking purposes only. It should not be distributed to third parties and may not represent the final published version.

Important: you must return any forms included with your proof. We cannot publish your article if you have not returned your signed copyright form.

NOTE - for further information about Journals Production please consult our FAQs at http://journals.cambridge.org/production_faqs
Laura Papish’s new book comes in the wake of a series of studies of Kant’s conception of evil. Two features distinguish her approach: its emphasis on the connection between evil and self-deception (chapters 1–5), and its attentiveness to the role of self-cognition in moral reform (chapters 6–8). Lucidly written and conversant with recent debates in social and moral psychology, Papish’s book expands the range of topics that typically worry Kantians. Its most important contribution is perhaps to have shown that self-deception and self-cognition are countervailing concepts, which together shed light on the neglected, epistemic dimension of Kant’s practical philosophy. My review will adopt the three-part structure of the book indicated in its title.

1. Part I – Evil

Instead of shunning the implications of Kant’s dualistic view of human motivation (the claim that there are only two incentives structuring our conduct, self-love and respect for the moral law), Papish begins her account by developing an interpretation of the self that self-loves which preserves Kant’s commitment to hedonism but avoids its reductionistic aspects. This balancing act is meant to oppose Andrew Reath’s interpretation, according to which pleasure is the source of non-moral action, but is gradually superseded by non-hedonistic considerations. Reath’s picture, Papish believes, overlooks an essential feature of Kant’s moral psychology, namely, the fact that pleasure ‘inform[s] us of our well-functioning, whereas pain or displeasure speaks to the presence of some sort of hindrance to life’ (p. 20). Hedonism, so construed, is not a transient motivator, but the expression of an underlying egoism, of an inextricable ‘drive or tendency to affirm the ego’ (p. 20). The feelings of pleasure and displeasure, therefore, are themselves part of ‘a largely interpretative and manufactured experience’ (p. 23), which an agent uses to determine how her life is going ‘according to the terms she has laid out’ (p. 24). This view has two important interpretable consequences: it loosens the connection between self-affirmation and pleasure-maximization (pleasure is important indirectly, as a sign of flourishing, not as an end in itself), and it...
broadens the type of motivation that can fall under self-love (choosing the path of least resistance can sometimes be the most effective way to protect the self).

With this broader conception of self-love in place, Papish turns in chapter 2 to Kant’s account of evil. To accommodate cases like Eichmann, suicide-bombers or the Milgram experiment, Kantians have scrambled to expand the range of immoral motivation. Papish suggests an alternative: rather than expanding the content of self-love, we should expand its possible formal arrangements instead. Interpreters (myself included) have so far characterized an evil will in terms of an inversion in the order of priority between the ethical incentives. But straightforward subordination, Papish argues, is not the only possible pattern. To the extent that the two incentives are inextricably present in a will like ours, we cannot possibly ‘avoid at least trying to negotiate the competing pulls of respect for the moral law and self-love’ (p. 45). This gives rise to a new volitional arrangement: as we try ‘to incorporate the moral incentive alongside the incentive of self-love’ (p. 47), our will becomes overdetermined.

This kind of arrangement seems to characterize, for example, the subjects in the Milgram experiment. It would be a mistake to conclude (as situationists do) that participants lacked moral character, or that their self-love paid no heed to the demands of morality (their actions were accompanied by regret, pain and protests). More accurate is to say that these subjects were ‘engaged in a frantic effort at overdetermination, caring both about the right thing and a pleasing ease in social interactions (i.e. with the authority figure overseeing the administration of shocks)’ (p. 52). They experienced an ‘agonized state of conflict’ (pp. 45–6) because they were trying to serve two masters. Unlike ‘straight prioritizers’ who at least get what they want, ‘someone with an overdetermined will remains deeply unsatisfied’, for they ‘are not sure to harmonize competing goods ends up in sacrificing both (p. 52). This is why overdetermined agents are so resistant to moral self-criticism: they appear, in their own eyes, as selfless and morally blameless, doing their best in the worst of circumstances.

Such self-conception, however, should give us pause with respect to Papish’s interpretative strategy for two main reasons. First, to reconcile overdetermination with Kantian rigorism (the claim that there are no moral intermediaries), Papish inadvertently conflates an agent’s empirical and noumenal characters, assuming that ‘the agonized state of conflict’ of which we are conscious replicates, and must also be present, at the level of the Gesinnung (the supreme maxim that governs our freedom). Second, this flattening of the volitional structure not only contradicts the Kantian view, but is also notoriously problematic in this case. The illusion of blamelessness nurtured by the overdetermined agent is a pretext, part of a strategy of moral self-protection.
Overdetermination for Kant is not a genuine alternative to subordination – it is, rather, one of the ways subordinators use to preserve the primacy of their self-love intact.

2. Part II – Self-Deception

This objection places us at the heart of Papish’s book, namely, the problem of self-deception (chapter 3). Although Kant himself is the first to do it, Papish thinks it best not to apply the model of lying to self-deception (Metaphysics of Morals (MM), 6: 429–31), since ‘lies to others do not invite the paradox or “contradiction” that deception of oneself involves’ (MM, 6: 430) (p. 70). A person who deceives herself regarding the good qualities of a beloved, for instance, is not only aware of her motives (something the liar does not have to be), but also ‘partly observant of norms of belief formation. She does not fully lack the evidence for the positive she sees in her beloved, nor ... [does] she simply manufacture evidence for her view. Instead, she is selective in how she reads or interprets the evidence available to her’ (p. 71). While external lies flout norms of evidence and contradict what we believe to be true, self-deception misinterprets available evidence to support what we would like to hear. Even if what we believe is not ‘under any form of straightforward control’ (p. 73), we nonetheless exert mediate influence on our belief-formation: we can mitigate ‘the sense of being compelled to accept a certain proposition by exploring, and to some degree constructing, a cognitive basis for assenting to some alternative proposition’ (p. 73).

Although this alternative cognition would not survive the test of public scrutiny, the process of rationalization is not totally groundless. By shifting our attention to the corroborative evidence, it introduces ‘a desirable cognition or hoped for justification into the reasoning process’ (p. 74). Despite the fact that this type of reasoning is a blatant misuse of our rational capacities, it is not thoughtless: ‘[w]hat characterizes rationalization ... is not the search, per se, for new and different grounds of cognition but the improper search for such grounds’ (p. 78). Self-deception, therefore, entails ‘a sophisticated and rationally mindful form of irrationality’ (p. 79), a turning of reason against itself without which evil would not be possible.

Papish’s account of self-deception is sophisticated and insightful. Chapters 4–5 address the question of evil’s radicality. The metaphor of ‘rootedness’, Papish argues, has three different meanings. Kant believes self-deception is ‘rooted’ because (i) it is a necessary condition for immorality, (ii) it has a firm and intractable quality and (iii) it is universal. The two first senses become clear once we understand that self-deception silences the pangs of conscience, and thus removes the agent’s hindrances to self-esteem. This removal increases our feeling of vitality and flourishing, and ‘allows evil to become liked, or subjectively valued, in a manner that is distinct from its
raw instrumental power to bring about our personal, morally impermissible ends’ (p. 105).

To understand the universality of evil Papish suggests we consider the widespread phenomenon of dissemblance or dissimulation. This phenomenon creates conditions of ‘moral hazard’, since it leads us to put forward an appearance that ‘is not fully accurate or representative of one’s motives, intentions, or inner life’ (p. 143). Whether the agent is virtuous or not, social interactions require some degree of opacity and concealment, the production of which inadvertently contributes to the moral corruption of the species. As social creatures, human beings are both ‘fallen and falling’, continuously colluding in the loss of their innocence. This ineluctable complicity, Papish believes, explains why evil for Kant is both universal and freely chosen.

3. Part III – Moral Reform

The last piece of Papish’s triptych is ‘moral reform’ (chapters 6–8). Although self-cognition is necessary to transcend the epistemic failures of evil and transition to the subsequent cognitive accomplishments required for moral progress’, Papish argues that ‘not all obstacles to self-cognition derive from self-deceptive indulgence’ (p. 154). ‘Struggles for self-knowledge are endemic to human life . . . regardless of the specific quality of one’s will’ (p. 156).

Following J. G. Hamann, Kant describes such struggles as ‘hell’ (MM, 6: 441; Conflict of the Faculties, 7: 55), partly because the moral law humiliates our self-love (an experience that is necessarily painful), and partly because the task of self-cognition lacks clear standards of correctness.

Although Kant raises important doubts about our capacity to know ourselves, Papish’s Kant is not a full-blown sceptic about self-cognition. This is so because the point of self-knowledge is not to achieve certainty about who we are in ourselves – the point is self-interpretation. Self-cognition, therefore, is different from other forms of cognition: getting it right about ourselves is not to discover ‘an independent and in principle available fact of the matter’ (p. 166), but to change who we are. Self-knowledge, so understood, is as much an epistemic as a practical activity. Its goal is not contemplation but self-transformation: ‘[w]e seek such self-knowledge not as part of a theoretical inquiry but for the sake of goodness’ (p. 173).

Even if the task of self-interpretation is open-ended, it is not thereby free-wheeling, for it aims at unifying an extended set of experiences under the concepts of good and evil. Although this act of unification is not fully determined, it is aided by moral concepts (the ‘thick vocabulary’ of vices and virtues) which put a limit to our confabulatory tendencies. In this respect, Papish suggests, self-interpretation is akin to reflective judgements (p. 168), and this gives a ‘provisional and experimental character’ (p. 170) to the activity of fathoming the depths of our hearts.
Moral reform, framed in this way, is inextricably linked to self-knowledge. In *Religion within the Limits of Reason Alone*, Kant famously envisions moral transformation as involving two stages: ‘a first stage of moral conversion in which respect for the law alone is incorporated into one’s maxim and a second stage of moral progress in which an agent attends to her behavior and actions for evidence of her new disposition’ (p. 177). The connection between these two stages, however, remains mysterious and has long been a matter of dispute among Kant scholars. Much clarity can be gained, Papish maintains in chapter 7, by interpreting the relation between the phenomenal and noumenal aspects of choice in terms of the notion of ‘commitment’, as well as by emphasizing the cognitive (as opposed to the volitional) dimension of the second, gradualistic stage.

The idea of ‘commitment’ is useful to understand the experience of conversion and its aftermath because a commitment (e.g. marriage) not only involves the adoption of a new principle, but also the ‘reordering of one’s identity and practices’ (p. 191). Temporality is important here, for a person entering a commitment is also seeking opportunities to convey in action what the commitment means for her life (p. 192). This explains why Kant thinks that the person after conversion will find her deeds ‘every time (not generally, but at each instant) defective’ because commitments need to be differently expressed over an indefinite duration in order to come closer to perfection (p. 193). It is impossible for creatures like us, Papish contends, to understand fully what this commitment involves and how it will evolve: ‘It is only in the mode of sense – through living out a commitment, working through tough situations, and having limited means for securing everything we care about’ (p. 194) that we come to terms with the meaning and implications of our commitments. When we fail to live up to them this is not because our past selves exert a causal grip on who we now are, but rather because we have ‘an incomplete or weaker-than-needed understanding of what we have committed ourselves to’ (p. 196) and of how that commitment affects the other goods we care about.

The last chapter of Papish’s book (chapter 8) examines the role of the ethical community in overcoming evil. The primary function of this community, Papish claims, is to address a lingering obstacle to moral reform, namely, the need to overcome the presence of ‘discordant moral judgments’ (p. 205). Such a discord characterizes the ‘ethical state of nature’, which resembles its political counterpart because the individual is here judge in her own case. The source of conflict in both cases, Papish argues, is the absence of a non-arbitrary arbiter. To understand the problem of the state of nature, therefore, we can discard the ‘presumption of badness’ (p. 214) – it is not our natural malevolence, but the problems associated with the proliferation of competing
centres of value judgement, which force people to seek an authority capable of settling disputes and thus enter political and ethical communities.

Here again I must part ways with Papish’s reading. As I see it, abandoning the ‘presumption of badness’ is the counterpart of abandoning self-suspicion in the case of overdetermination: just as blamelessness is the favourite illusion individuals develop to exculpate themselves from the evil they do, discounting malevolence is the way the species protects itself from the radical evil it harbours. After all, does not the primacy of self-love entail the refusal to acknowledge an objective way to judge our desires and inclinations? Is not the propensity to evil a dismissal of the moral law as the limiting condition of our subjective conceptions of the good? Even if by discarding a ‘presumption of badness’ we need not embrace a ‘presumption of goodness’, for a book built upon the notion of self-deception these are questions that cannot be eschewed. For it is at these very junctures that we are most prone to ‘throw dust in our own eyes’ (Religion, 6: 38).

I do not mean with this line of objection to cast doubt on the value of Kant on Evil, Self-Deception, and Moral Reform. On the contrary, I think that Papish’s is an important and well-crafted book, sure to shape the way we interpret how immorality operates within the Kantian framework. My reservations are only a reminder that, just as with the mythical hydra, killing the monster of self-deception is more difficult than cutting off a few of its heads.

Pablo Muchnik
Emerson College
e-mail: pablo_muchnik@emerson.edu