Book Review


Kantians often assume a connection between self-deception and human evil. Henry Allison says that ‘self-deception … is an essential ingredient in the propensity to evil’ (Allison 1990, p. 159). Sharon Anderson-Gold links ‘self-deception’ with ‘the loss of the power to combat the propensity to evil’ (Anderson-Gold 2001, p. 41). Allen Wood ascribes to Kant the view that ‘human beings in the social condition have tendencies to competitiveness, self-conceit, and self-deception, which combine in the radical propensity to evil in human nature’ (Wood 1999, p. 25). And I have said that the propensity to evil involves the ‘cultivation – largely through self-deception – of frailty, impurity, and depravity’ (Frierson 2003, p. 113). Papish rightly claims that ‘It is assumed’ in the literature on Kant ‘that evil is tightly … linked to self-deception’, and equally rightly objects that ‘the literature has up until now lacked any sustained inquiry into the connection between evil and self-deception’ (p. 4). Her book fills this important lacuna with rigorous argumentation and extensive references to other relevant secondary literature. As the most sustained discussion to date of the relationship between self-deception and evil, it is a must for scholars working on Kant’s account of radical evil. Moreover, with its distinctive take on hedonism and self-deception, it marks out a position worth responding to for those interested in Kant’s moral theory more broadly.

Papish starts in chapter one with an account of self-love that aims to provide ‘a more sophisticated version of Kant’s non-moral psychology … by attending closely to the role and demands of the self that seeks pleasures’ (pp. 5, 15, emphasis original). Arguing against ‘Andrews Reath’s influential attempt to rehabilitate Kant’ (p. 5), Papish aims to ‘lend nuance to hedonistic self-love without abandoning [Kant’s] commitment to the substantive claim that concern for one’s own pleasure is what motivates [all] nonmoral actions’ (p. 15). The nuance arises from four features of Papish’s Kantian hedonism: (1) hedonistic motivation can include welcoming ‘pain … to the extent that doing so facilitates her overall goal of increasing access to the feeling of vitality’; (2) because ‘the satisfaction of hedonistic desires is … largely interpretive and manufactured …, the maximizing principle at work in hedonism is one that pertains to an agent’s striving to determine that her life is going
well according to the terms she has laid out; (3) one can take pleasure directly in ‘these interpretive efforts’; and (4) because ‘central to a hedonistic orientation is an agent’s desire to feel unimpeded’, one can pursue hedonistic satisfaction by ‘conforming passively’ to one’s situation, by ‘pursu[ing] the path of least resistance’ (pp. 23-24, 30-31). These nuances – particularly (2) and (4) – strike me as taking Papish’s account quite close to the expansive accounts offered by Reath and me, but the end result of chapter two is an approach to Kantian hedonism that allows for a wide range of possible ways of being ‘hedonistically’ motivated, from the pursuit of goods one takes to be part of living life well to outright ‘passivity’ (p. 31).

Chapter three turns to Kant’s account of human evil. Kant makes clear in *Religion within the Boundaries of Mere Reason* (hereafter, *Religion*) that evil consists in the subordination of moral incentives to non-moral ones. (See 6:36, cited on p. 40.) Papish draws on her account of non-moral motivation to elucidate this sort of evil, and she develops her approach in opposition to two dominant trends in the contemporary literature. First, she argues against what she calls ‘expansionist’ conceptions of non-moral incentives (for example, those offered by Robert Louden, Wood, and me), which allow that any subordination of moral incentives to other incentives constitutes evil (p. 42). Instead, Papish insists that such subordination must always be for the sake of pleasure (or self-love). Second, against those (such as Allison, Anderson-Gold, Pablo Muchnik, and me) who see subordination in terms of prioritizing non-moral incentives, Papish defends the notion that human evil involves the effort to find a ‘compromise’ between moral demands and self-love.

Chapters three to five focus on self-deception. According to the ‘highly reconstructive’ (p. 69) account of self-deception in chapter three, because human beings are susceptible to evidence, we cannot refuse to believe what we have reasons to believe. Self-deception thus ‘distracts from, rather than denies’ unpleasant truths by ‘diverting our attention’ (p. 75). Moreover, because there are morally benign sorts of shifting attention such as the ‘permissible moral illusion’ by which we consider others to be more morally upright than we know they really are (7:152; p. 77), Papish argues that the concept of self-deception is intrinsically normative; whether a particular shift of attention counts as self-deception depends upon how laudatory (or not) are one’s motives. (My own predilection would be to describe as self-deception any wilful redirection of attention away from truths one knows to be salient, and then to allow for morally benign forms of self-deception, but this is largely a semantic distinction.) After laying out her account of self-deception, Papish considers in chapters four and five three ways that self-deception might relate to human evil:

(1) Self-deception is a necessary condition for evil;
(2) Self-deception facilitates the entrenchment of evil;
Self-deception explains why evil is universal in the human species.

In chapter four, Papish affirms claims (1) and (2). In chapter five, she rejects (3) but affirms a closely related claim, that humans’ tendency to dissemble explains why evil is universal.

Given my view that ‘self-deception is neither a necessary nor a sufficient condition of impurity’ (Freedom and Anthropology, p. 190), it may come as no surprise that I was unpersuaded by her affirmation of the opposite claim, but to be honest, I could not make out just what Papish’s argument for (1) is. She quite rightly raises ‘a few reasons to be cautious about endorsing this view’ (p. 91), including ‘a litany of [apparently] compelling counter-examples’ (p. 92). She then absolves herself of needing to make the prima facie case for (1) as an interpretation of Kant, saying instead that ‘since the majority of commentators see self-deception as a necessary condition of evil for Kant, I will presume that this is the case and consider what arguments can, and can’t, be used to generate support for this claim’ (p. 93). Much of the chapter nicely dismantles arguments one might think support the necessity of self-deception, such as ‘Kant’s stipulation against diabolical evil’ or ‘the claim that the moral law forces itself onto human consciousness as a fact of reason’ (p. 94). Papish’s discussion of these arguments is detailed and nuanced, and she rightly notes for many such arguments that they provide ‘important observation[s] … that can help explain the frequency with which self-deception accompanies evil’ but are ‘insufficient to establish a necessary connection’ (p. 97). The basic problem with all such arguments is that they ‘imply that evil always assumes a particularly intellectualized form … wherein we cloak our practical failures with some sort of … excuse’ (p. 98). But there are clearly cases in ordinary life – Papish mentions cases involving meat-eaters and smokers – and in Kant’s philosophy – most prominently his example of someone tempted to testify falsely against an innocent person to save his own life (5:30; pp. 155–6) – where one can knowingly and deliberately do what is foolish or morally wrong for the sake of some other (non-moral) good.

Nonetheless, Papish thinks that self-deception is necessary for evil. In part, she seems to reaffirm arguments she has shown to be insufficient, such as ‘the ubiquity of distraction alongside evil’ and the (Socratic) notion that ‘knowledge of the good reliably produces … conviction to be good’, but she sees her own use of those otherwise bad arguments as justified from ‘a more extended perspective’ (p. 102). As far as I can tell, this reasoning rests on the claim that someone cannot be evil without diverting attention from ‘how corrupt her decision-making is’ over the ‘long term’: ‘To see the essential role of self-deceptive rationalization with respect to evil, we must shift away from the possibility of an episodic cognition of oneself as evil and toward the question of whether one can steadily balance the virtue of self-knowledge with the vice of evil. This seems to me implausible’ (p. 102). I confess to finding this
conclusion unsupported; just as a person can knowingly do something evil for the sake of happiness in a particular moment, she can be quite self-aware about the fact that, over the long term, she deliberately chooses to promote her own hedonic best interest, even at the occasional cost of moral purity. That’s an unpleasant admission, at least for many, but it doesn’t strike me as an impossibility.

With respect to the role of self-deception in entrenching evil, Papish highlights several roles that evil can play, from the ‘obvious point that… self-deception can help an agent quiet the painful claims of conscience’ to the more contentious claim (bordering on an affirmation of diabolical evil) that ‘the more an agent self-deceives, the more she transforms her evil disposition … into something that is … pleasing in its own right’ (p. 105, emphasis in original). I still find Jeanine Grenberg’s account of how self-deception entrenches evil through ‘enhancing an agent’s comfort level with her own evil’ (p. 102; see Grenberg 2005) more plausible than what Papish adds to that account, but her suggestions about how self-image and ideology ‘expand the territory’ of evil pleasure is worth taking seriously (p. 111).

Finally, with respect to the role of self-deception in explaining the universality of evil, Papish delves into the ongoing debate between those (like Allen Wood and Sharon Anderson-Gold) who argue that Kant’s argument for universality depends essentially upon empirical arguments tied to humans’ unsocial sociability and those (like Allison and Seiriol Morgan) who argue for some sort of a priori argument for evil. Papish explicitly sides with the a priori group (pp. 130-131); her a priori argument depends upon an innate human tendency that is closely akin to self-deception. As she explains, self-deception as such does not show why evil is universal, but human beings’ innate tendency to dissemble does explain this universality.

In my own view, Papish overplays the importance of dissemblance and understates the extent to which her argument brings her into the company of those who advocate more empirical arguments for radical evil. Like those who argue for the universality of evil on the basis of humans’ unsocial sociability, Papish takes a specific feature of human life – our propensity to dissimulation – and argues that given that ‘dissimulation involves a failed use of agency,… a human being… is necessarily evil’ (p. 150). I have three concerns here.

First, the argument is not a priori. Humans’ propensity to dissimulation is an empirical fact about us, just as empirical as our tendency to unsocial sociability. Both are ‘essential’ to human beings, and can even ‘belong to the concept of the species’ (7:332; p. 144) as empirical generalizations incorporated into the empirical concept of homo sapiens. There is nothing essential to finite practical reasoning as such that requires dissimulation; we posit that human beings as such dissemble because, in our experience, we find the practice of dissimulation to be universally present. Since I am personally partial to empirical proofs, this objection is not particularly serious, but it
suggests that Papish is more akin to Wood and Anderson-Gold than to Allison and Morgan.

Second, dissimulation as such cannot be a sufficient basis for supporting the universality of human evil because dissimulation is not necessarily a ‘failed use of agency’. The quality that Kant ascribes to the species is a tendency to ‘explore the thoughts of others but withhold his own’ (7:332; p. 144), a tendency that, in context, is taken to be essential for us to ‘live in peace together’ (7:332) and also for fulfilling moral obligations to participate in polite society (see 6:473) and exercise reserve in personal relationships (see 6:471-2). Dissemblance is often a moral fault, but not always, so a universal propensity to dissemble does not, on its own, imply universal evil.

Finally, Papish argues that while ‘some dissimulation is invulnerable to personal control … regardless of which practical principles any given person adopts or decisions any given person makes’, we should nonetheless see it as ‘a product of her agency’ (pp. 145-146). Without some further explanation of what ‘action’ and ‘agency’ are, I’m not sure how to make sense of this claim in a Kantian context. If I conceal my thoughts without any intention or decision to conceal my thoughts, in what sense is that concealment an ‘action’ for which I can be held ‘responsible’ (p. 146)?

None of these objections should be taken to undermine the excellent work that Papish has done in showing how dissemblance fosters and promotes evil in human beings. My own view is that, for Kant, there are many facts about human beings that have the empirical universality necessary to ascribe them to us ‘as a species’. These include the fact that we (generally) have five external senses, that our imagination is governed by principles of association, that we are susceptible to unsocial sociability and have a propensity to dissimulation, and that we have a propensity to subordinate moral incentives to non-moral ones. For all these claims, there is sufficient empirical evidence to ascribe them to the empirical species-concept of homo sapiens. As empirical claims, none are equivalent to the claim that humans are radically evil; for that, one needs an a priori proof to connect empirically given tendencies with transcendental grounds. And there is an a priori proof – given fairly explicitly in Kant’s Religion – that the tendency to subordinate moral to non-moral incentives constitutes radical evil. Given that argument, the last of the aforementioned empirical generalizations supports the claim that human beings are, as a species, radically evil. But this leaves plenty of room for finding connections amongst our various natural characteristics, including the connections Wood and Anderson-Gold have drawn between unsocial sociability and evil, and – now – the connections Papish has drawn between dissimulation and evil.

In chapters six to eight, Papish turns from diagnosis to cure. If self-deception is a root cause of human evil, then self-cognition is an important part of moral reform. In chapter six, she explains various obstacles to self-cognition, such as Kant’s claims about the limits of introspection (for example, 25:1437; 26:288).
p. 162) or the fact that we can only experience the self as it appears to itself, not as it is (p. 163). She rightly claims that these issues do not wholly preclude self-cognition, and she offers her account of self-cognition as ‘self-interpretation’ (pp. 164 ff.), something that is a ‘hell’ not because it is impossible but because ‘it requires vulnerability and daringness’ as we ‘try out new understandings of ourselves, … cultivate relationships with others, and … find ourselves transformed through our efforts at self-cognition’ (p. 172).

In chapter seven, Papish lays out and defends her ‘two stage-model of moral reform’, according to which moral life starts with a ‘moral conversion in which respect for the moral law alone is incorporated into one’s maxim’ and proceeds with ‘a second stage of moral progress in which an agent attends to her behaviour and actions for evidence of her new disposition’ (p. 178, emphasis original). In this chapter, Papish argues against views (such as mine) according to which moral progress is largely volitional, a matter of strengthening moral resolve and increasingly motivating oneself to subordinate moral to non-moral incentives. On her two-stage account, this volitional work is essentially accomplished in the moment of moral conversion, a conversion that ‘secures’ the ‘purity of the principle of choice’ (p. 191). What happens in subsequent life is primarily ‘cognitive’ (p. 178 and passim) or ‘epistemic’ (p. 193). Drawing on marriage as an analogy, she suggests that the work of building a marriage is not that of ‘overcoming a lack of executive strength’ or of ‘ensuring that [the marriage partners’] commitment fares well against … temptation’ but a lifelong ‘opportunity to convey in action what their commitment means’ through finding ‘a more nuanced way to give voice to and conceptualize [their] commitment’ (p. 192). Moral life, similarly, is established on firm footing with an initial commitment to the moral law (stage one), but must be carried out through ‘moral progress’ as the ‘cognitive work’ of specifying how the moral law plays out in the complex realities of life (p. 194).

Finally, in chapter eight, Papish presents her account of moral community. As in chapter seven, she resists efforts (made by, for instance, Wood and me) to conceptualize community primarily in terms of fostering a stronger or more consistent commitment to the moral law, or working against natural social tendencies that foster vice. Instead, she sees the role of ethical community as primarily epistemic, a matter of ‘publicly shared resources that help individuals clarify their own moral beliefs, articulate them to those who are not inclined to agree, and appreciate beliefs about good conduct held by others’ (p. 225). Such a community thereby fosters a ‘robust mutual understanding’ that makes possible the cognitive tasks of moral progress given the social requirements of life together (p. 231). In this chapter, Papish also offers an extended comparison of ethical community and international law (pp. 205-215) and some specific arguments for the claim that churches and religions based on sacred Scriptures are superior to other forms of ethical life for cultivating the right sorts of moral-cognitive development (pp. 223-227).
In sum, Papish’s book provides important new accounts of non-moral motivation, self-deception, evil, and moral progress. In the rest of this review, I’ll raise a few concerns and clarify some of my positions to which Papish objects. I have relatively little to say about self-deception. While my conception of Kantian self-deception is somewhat more expansive than Papish’s, I quite like her notion that self-deception generally consists in distraction or redirection of attention, her emphasis on the normative dimension of self-deception, and her insistence that self-deception plays a large (even if I would say dispensable) role in entrenching evil in human nature. I disagree with her about the necessity of self-deception for evil and about the role of dissimulation in explaining evil’s universality, but I have already briefly touched on those points above. My focus in the rest of this review will be more broadly on the nature of evil and moral reform in Kant, including how these themes relate to Kant’s transcendental idealism, a point Papish briefly mentions at a few key junctures, but does not discuss in detail.

I’ll start with Papish’s two-stage model of moral reform, since this gives me a chance to better explain how I read Kant’s transcendental idealism in relation to his account of human evil. In *Religion*, Kant clearly talks both about a ‘revolution in the will’ and about moral progress over time. Papish, following commentators like Mavis Biss (2015), reads these as referring to two different events that occur at different times in a person’s life, such that *first* one incorporates the moral law into one’s maxim and *then* one makes moral progress. She notes that ‘neither Kant himself, nor many of his commentators, has been sufficiently explicit about how the moral failures that characterize an evil will, and thus precede a reversal in one’s supreme maxim, are different from the moral shortcomings toward which a person’s post-conversion practical efforts are directed’ (p. 179). For Papish, the difference between these shortcomings is both crucial and somewhat perplexing because the first stage involves a ‘conversion’ and ‘commitment’ to the moral law that ‘secures both “purity of the principle of choice”’ and ‘the “stability” of this principle’ (p. 191; quoting Kant 6:48). Given that the first stage seemingly accomplishes all that needs to occur in order to overcome evil volitionally, it’s not clear what is left to be done with a moral process of ‘striving for the better’, so Papish interprets this process as essentially cognitive.

On my view, there are not two temporally distinct stages of moral reform. Instead, Kant’s reference to a revolution in will refers – as he himself indicates – to a ‘virtus noumenon’ (6:47), an intelligible character that is the noumenal ground of one’s empirical character. Because Kant, in his argument for rigourism, rejects any middle ground between good and evil, he cannot allow that any human being’s fundamental (noumenal) moral condition is intermediate between good and evil or in transition from good to evil. But because he holds out moral hope for each human being, despite the radical evil of each, he needs some way to conceptualize what moral status an agent can have that is compatible with both considering oneself good and
recognizing evil at the root of one’s maxims. The language that he develops for this moral status is ‘revolution’. Admittedly, the language of ‘revolution’, ‘rebirth’ and ‘change of heart’ (6:47) sounds temporal, and it suggests a prior condition of evil followed in time by a later condition of moral goodness. Throughout his discussion of this revolution, however, Kant emphasizes that this single and unalterable decision takes place ‘in the mode of thought’, in the ‘intelligible ground of the heart’ that is directly accessible only ‘for God’ (6:48). What God sees ‘as a unity’ is given to us only as ‘endless progress’ or ‘incessant laboring and becoming’ (6:48). Insofar as this life of progress is thought of as a unity – whether noumenal or intelligible or in the view of God – ‘to this extent the change can be considered a revolution’ (6:48). This moral status is consistent with the observable fact that human beings are evil, since a will-in-revolution includes evil as that in itself against which it revolts, but it is also consistent with its being, overall, a good will, since a will-in-revolution includes evil precisely as something being overcome. But Kant makes clear that the very same change that can be considered noumenally as a revolution is, within human experience, ‘to be regarded only as an ever-continuing striving for the better, hence as a gradual reformation of the propensity to evil, of the perverted attitude of mind’ (6:48). Rather than a two-stage theory of moral reform, we have a two-aspect theory of moral goodness. Intelligibly, one can be morally good insofar as one’s fundamental disposition is one of revolution against self-inflicted evil. Empirically, this disposition shows up as a life of constant struggle against evil tendencies.

Incidentally, this emphasis on intelligible dispositions as grounds of empirical character also clarifies how I think we must understand the ‘initial’ choice to be morally evil; radical evil in human nature is the noumenal ground of particular evil deeds that appear in our empirical character. In arguing against the notion (shared by Jacqueline Mariña and me) that the struggle against evil involves working against what Mariña has called ‘volitional residues’, Papish misinterprets my claim that ‘the limitations of one’s ability to effect changes in oneself are self-imposed limitations’ (Freedom and Anthropology, p. 129; cited on p. 184) to imply that ‘past deeds are given a kind of overriding power with respect to my present conduct’ (p. 185). The point of my argument is that I am – not was – imposing on myself limitations accrued by past misdeeds. Those misdeeds continue to hinder my ability to act purely from moral motives because evil continues to be a part of my will. Even that way of putting it, however, is too temporal, since the ‘self’ that subordinates moral to non-moral incentives, and the self that revolts against that subordination, are really aspects of intelligible character that ground temporally situated appearances. They are not decisions made (and then reversed) in time.

Kant’s text, I think, allows for either a two-stage or a two-aspect reading of the relationship between revolution and progress. The two readings are not even strictly inconsistent, in that a life of moral progress that expresses a
noumenal ‘revolution’ might well include particular moments of decisive resolution (as I think Kant suggests in Anthropology 7:294-5). But there are some significant advantages of the two-aspect reading over the two-stage model. For one thing, this reading allows us to take at face value Kant’s claims that the life of moral progress involves ‘gradual reformation of the propensity to evil, the perverted cast of mind’ (6:48), whereas Papish must read this ‘moral’ progress as merely cognitive, growth in understanding how one’s already good will applies in this or that new situation. Relatedly, there is the problem of backsliding. On Papish’s reading, if one really undergoes a moral conversion, then the volitional work is done, once and for all. But if a person subsequently backslides, clearly doing something evil with full consciousness that they are doing something evil, then Papish seemingly has to say that such a person never really converted in the first place. By contrast, on my account, backsliding and recovery can be a normal part of the ongoing struggle against evil, a struggle that expresses one’s revolution, albeit always only incompletely. I needn’t constantly update the timestamp on my ‘real’ conversion, because my moral conversion has no timestamp. The two-aspect reading also makes better sense – but this is too long a story to tell here – of Kant’s account of atonement in Religion (6:72-8).

Admittedly, my two-aspect (as opposed to two-stage) model saddles Kant with transcendental idealism, a distinction between how things are in themselves (revolution) and how they appear (progress). As I’ve described it here, it involves a metaphysically thick reading of that idealism. At least as a reading of Kant, however, the tight interconnection with transcendental idealism is a mark in its favour; my two-aspect reading thereby incorporates straightforwardly features of Kant’s transcendental idealism that Papish tries to incorporate more indirectly. Moreover, while I cannot develop it here, the distinction between one’s fundamental moral status and the appearance of that moral status can be articulated in terms of first- and third-person perspectives in a way that does not require a particularly thick metaphysical reading of transcendental idealism.

A second area of disagreement between Papish and me relates to my ‘expansive’ notion of radical evil and particularly my response to situationist appropriations of contemporary research in social psychology. Papish takes seriously the question of ‘whether Kant’s moral psychology and theory of evil are complex enough to do justice to … the empirical evidence that Frierson and [John] Doris [in, for example, Lack of Character] rightly demand that any philosopher must attend to’ (p. 46). I should first clarify that one of my central claims about Kant in relation to contemporary situationist psychological research is precisely the priority of moral theory over empirical psychology, so – unlike Doris – I do not demand that any philosopher must attend to the empirical evidence; it’s perfectly appropriate to develop standards of moral excellence independent of empirical evidence and then look to
empirical evidence in order to see whether or to what extent human beings measure up to those standards.

Nonetheless, I agree with Papish that Kant’s theory of evil is complex enough to do justice to what contemporary psychologists have discovered about human tendencies. Oddly, Papish presents my view as the flat and one-dimensional view that we should explain these data simply by positing that ‘evil consists in anything less than fully consistent adherence to the moral law’ (p. 44). Papish presents this view despite citing my forthcoming paper in which I carefully distinguish the ways that impurity, depravity, and various degrees of frailty can explain a range of different results in social psychology that show different kinds and degrees of situation-dependence in human behaviour. Given his rigourism, Kant does take ‘anything less than fully consistent adherence to the moral law’ to be evil, but this claim does not exhaust what he – or I – say about evil. The failure to conform to the moral law has a particular structure, the subordination of moral incentives to non-moral incentives, where these non-moral incentives are always linked with pleasure at their ground, even if not as their end. As I show in the article Papish cites, this structure plays out in quite different ways, which Kant describes as three degrees of evil.

In analysing the situation-dependence of moral (and immoral) action, this complex Kantian account of evil provides a rubric for further research and analysis. Papish is quite right that in the case of Milgram experiments, most subjects who fully succumbed to the pressure of an authoritative experimenter experienced moral conflict. (It’s also the case that some did not succumb to this pressure, for reasons that can be ascribed to features of their moral commitment, a point to which neither Papish nor Doris pays much attention.) For Papish, the moral conflict of those who succumb to situational pressure is evidence that such subjects ‘have something like sincerely held moral allegiances’ (p. 46). For some reason, Papish seems to think that I would deny this claim, but I never deny that human beings have moral allegiance in the sense that they have an awareness of and appreciation for the demands of morality. All evil, on Kant’s account, is a subordination of moral incentives that one values to non-moral incentives that one values more. Papish and I both acknowledge that subordination takes place in the Milgram case.

In fact, to an even greater extent than Papish, I think that the details of Kant’s theory of evil can help make sense of (and distinguish amongst) various cases of situational influence. For the Milgram subjects who succumbed, I think that impurity – which often, though I suspect not here, corresponds to Papish’s emphasis on ‘compromise’ – rather than frailty or depravity was typically the characteristic form of moral evil. As I put it in the aforementioned paper, in this particular case, fear of embarrassment or desire for approval from accepted authority figures, motives that generally support good action, fail to support it. An impure will that generally acts well and
is well regarded loses important supports to upright behaviour and, having never cultivated a character that makes duty sufficient for action, falls short. Even in Milgram cases, frailty or depravity could be at play, and one could even investigate the extent to which this or that form of evil was salient for this or that person. For other experiments, and particularly those where sub-personal variables predict subjects’ behaviours, there is little or no evidence of moral anguish on the parts of subjects affected by situational variables. Even in these cases, we can see subjects’ failures to adhere to moral norms as symptomatic of various kinds of subordination of moral incentives to non-moral ones. For example, an agent may—through frailty—fail to cultivate the strength of character needed to consistently adhere to moral principles to which she has explicitly committed herself; in the moment of choice, she may then subordinate those principles to fleeting inclinations.

I have focused in this review on two of many arguments Papish presents in her book, and even my discussion of those does not deal with all the ways Papish anticipates and partly responds to points presented here. The book offers a wealth of arguments, engagements with secondary literature, responses, and rejoinders to those responses, and I cannot even begin to do justice to all the points she makes in the book. I found myself frequently unconvinced, and my own tendency was to side with the readings of passages that she presents ‘at first glance’ rather than what seem to me the less pluralist and narrower readings she arrives at after processing various ‘perplexing’ features of those views (for example, p. 216). But even where I think that Papish gets wrong the nature of non-moral motivation, or evil, or moral reform, she often brings out important features of motivation or evil or reform. They may not exhaust Kant’s view on the topic, but her positive proposals are quite illuminating about various possibilities for these important features of human life.

References
Frierson, Patrick Forthcoming, ‘Character in Kant’s Moral Psychology: Responding to the Situationist Challenge’, Archiv für Geschichte der Philosophie

*Whitman College*

frierspr@whitman.edu
doi:10.1093/mind/fzy078