Kantian Review (2023), **28**, 555–574
doi:10.1017/S1369415423000304

KANTIAN REVIEW
CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The Change of Heart, Moral Character and Moral Reform

Conrad Damstra

Brown University
Email: conrad_damstra@brown.edu

Abstract

I examine Kant's claim in part one of *Religion within the Boundaries of Mere Reason* that moral reform requires both a 'change of heart' and gradual reformation of one's sense (*R*, 6: 47). I argue that Kant's conception of moral reform is neither fundamentally obscure nor is it as vulnerable to serious objections as several commentators have suggested. I defend Kant by explaining how he can maintain both that we can choose our moral disposition via an intelligible choice and that we become good through a continuous struggle. I then provide an interpretation of how moral reform occurs in the phenomenal realm.

Keywords: maxims; character; evil; *Gesinnung*; virtue; moral reform; religion; change of heart

Though Kant's most famous claim in *Religion within the Boundaries of Mere Reason* is that humanity is radically evil, his aim is not to condemn humanity but to explain how agents can become good through a process of moral reform. Kant argues that becoming a good human being requires a 'change of heart' (*R*, 6: 47).[1] He maintains this because he thinks that each agent possesses a fundamental maxim that determines their moral character; moral reform requires changing one's fundamental maxim.

Many commentators who have discussed Kant's account of the change of heart have argued that it is deeply obscure or otherwise problematic, with Henry Allison noting that it is 'perhaps the most perplexing feature of Kant's whole discussion of the moral life' (1990: 170). The trouble arises because Kant claims that our character is adopted by an intelligible choice that occurs outside of space and time. Kant's critics argue that this is implausible. Furthermore, such a view is seemingly in tension with Kant's belief that one becomes good only through 'incessant laboring and becoming' and his observation that a change of character 'is to be regarded … as a gradual reformation of the propensity to evil' (*R*, 6: 48). Gordon Michalson writes that because Kant locates the real self outside of space and time 'Kant can string no metaphysical "thread" through the successive moments of the agent's life … he cannot show how a "previous" act of moral condition would be relevant to a "present" act' (1990: 85).[2] Daniel O'Connor agrees that Kant's account of moral reform is threatened by his

transcendental idealism. He claims that Kant cannot make sense of the efficacy of moral education or the cultivation of moral emotions over time that would be part of any plausible gradualist theory of moral reform. These thinkers argue that Kant introduces a conception of character in the *Religion* to explain how a person can have a continuous identity throughout his or her moral life. But because this character is adopted by a choice that itself is outside of space and time, Kant's account is a failure – it cannot serve the explanatory role that he intends for it.

I will argue that Kant's theory of moral reform can coherently encompass both an intelligible choice of character and gradual reform at the empirical level. At the noumenal level, moral reform involves the adoption of a good *Gesinnung* or fundamental maxim which serves as the ground of dutiful actions. However, Kant supplements this account with a rich discussion of how we are to understand moral reform at the phenomenal level. I will argue that for Kant phenomenal moral reform has two stages: it requires both a transformation of attitude in which one makes a commitment to good principles and continuous cultivation of virtue. In describing Kant's views, I will aim to draw out what he thinks about the efficacy and moral significance of these empirical practices. I will focus in particular on how they can yield a stable and coherent moral character. In presenting my argument, I also hope to situate the *Religion* in relation to the other practical and anthropological writings of the critical period.[3]

The structure of this article is as follows. In section 1, I discuss Kant's conception of the *Gesinnung* and the arguments that are meant to support the fact that we must have either a good or evil *Gesinnung*. I also discuss Kant's claim that transformation of one's *Gesinnung* is possible. Section 2 examines Kant's most detailed discussion of moral reform, which is found in *Religion,* 6: 47–8. I present several problems that Kant's account faces and argue that the 'revolutionary' and 'gradualist' aspects of Kant's theory can be reconciled.[4] Sections 3 and 4 examine how moral reform occurs in the phenomenal realm. Investigating Kant's views about phenomenal moral reform will show the positive claims that Kant can make about the stability of one's disposition. The article concludes with section 5.

I

Kant begins the *Religion* by describing a dispute between those who accept the view, characteristic of historical religions, that 'the world lieth in evil' and the recent pedagogues and philosophers who have begun to argue, on the other hand, that 'the world steadfastly . . . forges ahead in the very opposite direction, namely from bad to better' (*R*, 6: 19–20). His strategy for intervening in this debate is to investigate how one can be judged as either good or evil in the first place. He argues that our moral status is determined by our free choice of maxims. Specifically, Kant endorses a view called rigorism which claims that agents must be either good or evil, but not a mixture of both, based on the maxims that they adopt.

Kant argues for this view by appealing to what has become known as the incorporation thesis. He claims that the answer to the question of whether human beings are good or evil 'is based on the morally important observation that freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action *except so far as the human being has incorporated it into his maxim*' (*R*, 6: 24).[5] Briefly put, the incorporation thesis states that incentives cannot

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

determine an agent to act except insofar as she incorporates these incentives into a maxim. By 'incentives', Kant has in mind both the desires that come from the lower motivational faculties and the moral law which has its source in the faculty of reason (Schapiro 2011: 149). He employs the incorporation thesis in the following argument:

1. The moral law can be a sufficient incentive for action.
2. If the moral law is a sufficient incentive for a subject S's action, then their action is good.
3. Self-love can also be a sufficient incentive for action.
4. The distinction between the incentives of self-love and morality is exhaustive.
5. For any particular action A, only one of the incentives of self-love or the moral law can be sufficient for S to A.
6. If S's sufficient incentive for A is something other than the moral law, then it must be self-love. (4, 5)
7. When self-love suffices for S to A in morally significant circumstances, the action A is evil.[6]
8. A person's actions must be either good or evil. (2, 6, 7)

This argument moves rather quickly – it takes up just about half of a page of the Cambridge edition of the *Religion*[7] – largely because it relies on several assumptions that Kant has defended in his earlier writings, and I will not attempt to give a further defence of it here.[8] For our purposes, the main difficulty with this argument is that it does not seem to establish rigorism of one's character. Proposition (8) is a conclusion that a person's actions must be either morally good or evil. However, one may object that a person can do both good and evil deeds and resultingly have a character that is either both good and evil in parts or neither good nor evil but something in between. As presented, this argument does not rule out this possibility, but Kant intends for rigorism to apply to both actions *and agents*. How can he secure this conclusion? The argument for rigorism of character is also quite brief:

> Nor can a human being be morally good in some parts, and at the same time evil in others. For if he 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)

Kant argues by assuming for *reductio* that there is a person who is both good and evil. This person is not necessarily motivated by the moral law, because for at least some actions he takes self-love and not the moral law to be a sufficient incentive to act. However, Kant holds that the moral law obligates us with unconditional necessity (*G*, 4: 416). Accepting the authority of morality means that one must always act out of duty in the relevant circumstances. The person who is both good and evil would simultaneously have both the maxim characteristic of a good will – stated roughly, 'Always perform one's moral duty when this is required' – and the evil maxim 'Self-love can be a sufficient ground to act, even if this contravenes one's duty.' This is

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

contradictory; one cannot rationally hold both these maxims at the same time and so one cannot be good or evil in parts.[9]

Kant does not think that a person could abjure self-love entirely, nor could one fail to be receptive to the moral law.[10] Because we are constantly confronted by the two incentives of morality and self-love in our practical engagement with the world, we must choose how to prioritize them. Furthermore, this rule or principle of action has a propositional structure and hence can be expressed as a maxim that describes which incentive a person 'makes the condition of the other' (*R*, 6: 36). Because this maxim has a higher level of generality in relation to other maxims, some commentators have called it a fundamental maxim.[11] Kant writes that a good fundamental maxim should be understood as making the moral law '*the supreme condition* of the satisfaction of [self-love]' (*R*, 6: 36). Thus, the good fundamental maxim can be modelled as:

> **Fundamental Maxim$_G$**: Act according to self-love only if this does not conflict with any moral duties.

Kant thinks that we have imperfect duties as well as perfect duties. We have a wide obligation to fulfil our imperfect duties, which means that there are various possible ways that we can discharge these obligations. An outsized commitment to self-love that would impede a person from fulfilling their imperfect duties to the extent required would belie an evil character even though the obligation to fulfil imperfect duties is indeterminate.

While the good fundamental maxim makes acting on self-love conditional on such actions being morally permissible, the person with an evil fundamental maxim denies this requirement, holding:

> **Fundamental Maxim$_E$**: Self-love is a sufficient ground to act even when this may conflict with one's moral duties.

Intentionally acting against what morality commands, even one time, would entail that one has an evil character at that time.

Kant notes that these fundamental maxims describe a general attitude that an agent has towards the rational authority of the moral law and calls this an agent's *Gesinnung* or disposition. Because Kant thinks that a person is either good or evil based on their *Gesinnung*, it itself must be freely adopted, as Kant holds that a person can only be morally judged by that which can be imputed to them (*R*, 6: 20). He defines the *Gesinnung* as 'the first subjective ground of the adoption of the maxims' which 'applies to freedom universally' (*R*, 6: 25). The *Gesinnung* is a ground (*Grund*) of particular maxims.[12] Kant writes, for example, regarding the evil fundamental maxim: 'In order, then, to call a human being evil, it must be possible to infer *a priori* . . . an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particular morally evil maxims' (*R*, 6: 20). As Paul Guyer notes, the fundamental maxim serves as the 'basis of which' an agent adopts more particular maxims (2016b: 181). In this sense, the good or evil *Gesinnung* serves as a necessary condition for good or evil actions respectively and is logically prior to these actions.[13]

If Kant conceives of a *Gesinnung* as a ground that applies to our free actions universally, then it seems that a person with a good *Gesinnung* could not act evilly and

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

a person with an evil *Gesinnung* could not act from duty. Most commentators want to attribute a more moderate view to Kant where a person can have a good *Gesinnung* but nevertheless act evilly or have an evil *Gesinnung* but act from duty.[14] One passage suggests such a moderate interpretation:

> So far as the agreement of actions with the law goes, however, there is no difference ... between a human being of good morals ... and a morally good human being except that the former do not always have, perhaps never have, the law as their sole and supreme incentive, whereas those of the latter *always* do. (*R*, 6:30)

This passage appears to say that a person who is not a 'morally good human being' may have the moral law as her sole incentive in action sometimes but not always, while the morally good person always does. However, Kant continues directly: 'We can say of the first that he complies with the law according to the *letter* (i.e. as regards the action commanded by the law); but of the second, that he observes it according to the *spirit* (the spirit of the moral law consists in the law being of itself a sufficient incentive)' (*R*, 6: 30). Kant is saying here that an agent with an evil fundamental maxim would act, at best, in conformity with the moral law whereas only those with a good fundamental maxim can have the moral law as a sufficient incentive, which is what dutiful action requires.[15]

Kant also argues that the *Gesinnung* can be chosen. His commitment to transcendental idealism means that there can be no temporal origin of our moral character, but rather our character must be grounded in a free act that occurs outside of space and time, which he at one point calls an 'intelligible deed' (*R*, 6: 31). Kant accepts this because he holds that no matter how wicked a person has been or what the current natural circumstances are, it must be possible for an evil person to adopt a good *Gesinnung*. This commitment raises a difficulty for Kant because the choice of transforming one's moral disposition, conceptually, cannot have its basis in a person's current *Gesinnung*. In this sense, the choice to transform one's *Gesinnung* must be groundless. The possibility that one can transform their *Gesinnung* is in tension with viewing the *Gesinnung* as a universal ground of one's actions – or so it would seem. Kant is aware of this challenge and responds directly to it:

> How it is possible that a naturally evil human being should make himself into a good human being surpasses every concept of ours. For how can an evil tree bear good fruit? But, since by our previous admission a tree which has (in its predisposition) originally good but did bring forth bad fruits, and since the fall from good into evil ... is no more comprehensible than the ascent from evil back to the good, then the possibility of this last cannot be disputed. (*R*, 6: 44–5)

Kant appeals to the principle 'ought implies can' to explain this, noting that we have a duty to adopt a good fundamental maxim so this must be possible: 'the command that we *ought* to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it' (*R*, 6: 45).[16] Moral reform must be really possible even for an agent who has an evil character because otherwise someone who

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

is evil would have no duty to become good, and this contravenes one of Kant's fundamental philosophical commitments.

Kant concedes that we cannot conceive of how the *Gesinnung* can be transformed. But he notes that what is conceivable does not necessarily track what is really possible. Freedom is, famously for Kant, an 'inscrutable faculty' (*CPrR*, 5: 47), and we can not cognize the grounds by which the moral law can serve as a determining ground of our will (*CPrR*, 5: 72). Nor can we conceive of how an agent who has been endowed with a predisposition to the good has chosen an evil character, any more than we can understand how someone who is evil would choose the good. But we know that both of these choices are really possible, the former because experience has testified to this fact and the latter because it is demanded by pure practical reason. Kant's theory of freedom entails that the choice of either an evil or good disposition is always really possible for human beings.[17]

2

The freedom to choose one's *Gesinnung* is of crucial importance because Kant holds that, because of the universal propensity to evil, all human beings – even those who seem most morally upright – must be thought to initially possess an evil fundamental maxim. Hence no person is exempted from the task of replacing an evil fundamental maxim with a good one.[18] Although the intelligible choice of one's disposition is inscrutable, Kant does not shy away from discussing the process of becoming a good human being in detail. The main passage where he discusses this is quite complex, and because it will serve as the foundation for much of the rest of my interpretation, I will present it in full:

> But if a human being is corrupt in the very ground of his maxims, how can he possibly bring about this revolution by his own forces and become a good human being on his own? Yet duty commands that he be good, and duty commands nothing but what we can do. The only way to reconcile this is by saying that a revolution is necessary in the mode of thought but a gradual reformation in the mode of sense (which places obstacles in the way of the former), and [that both] must therefore be possible also to the human being. That is: If by a single and unalterable decision a human being reverses the supreme ground of his maxims by which he was an evil human being (and thereby puts on a 'new man'), he is to this extent, by principle and attitude of mind, a subject receptive to the good; but he is a good human being only in incessant labouring and becoming i.e. he can hope – in view of the purity of the principle which he has adopted as the supreme maxim of his power of choice, and in view of the stability of this principle – to find himself upon the good (though narrow) path of constant *progress* from bad to better. For him who penetrates to the intelligible ground of the heart (the ground of all the maxims of the power of choice), for him to whom this endless progress is a unity, i.e. for God, this is the same as actually being a good human being (pleasing to him); and to this extent the change can be considered a revolution. For the judgement of human beings, however, who can assess themselves and the strength of their maxims only by the upper hand they gain over the senses in

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

> time, the change is 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. (*R*, 6: 47–8)

At the start of this passage, Kant reiterates his claim that the transformation of one's character is really possible. He then claims that both a 'revolution ... in the mode of thought' and 'a gradual reformation in the mode of sense' are required to become a good human being (*R*, 6: 47). The former makes 'one receptive to the good' but does not yet suffice for possessing a good character; it must be supplemented by 'incessant laboring and becoming' (*R*, 6: 48). While human beings can only hope to continually progress 'from bad to better' they can hope that God, who can determine the nature of one's fundamental maxim, can judge this progress to be 'a unity' (*R*, 6: 48). I will use the term 'change of heart' to describe an agent's adopting a good fundamental maxim. I will call an agent's attempt to possess a good character through incessant labouring 'moral labour'.

Initially, this account may seem puzzling. In the quoted passage, Kant contends that 'reversing the supreme ground of one's maxims' is necessary but not sufficient for being good. One becomes a good person through 'incessant laboring and becoming' (*R*, 6: 48). However, as we have seen earlier, Kant has argued that only 'when a human being has incorporated into his maxim the incentive implanted to him for the moral law, is he called a good human being' (*R*, 6: 45n.; see also *R*, 6: 44). So, Kant appears to give two answers to the question of how to become a good human being. It looks like Kant wants to have things both ways, claiming that the transformation of one's character occurs both through adopting a good fundamental maxim and gradually through a continuous process of moral labour. I will call this the 'inconsistency problem'.

Commentators have tended to emphasize one of these two aspects of Kant's thought regarding moral transformation. Daniel O'Connor claims that a 'change of disposition is by sudden revolution (conversion) not by gradual reformation' (1985: 300). Mavis Biss contrasts the 'revolutionary approach' of the *Religion* with the 'gradualist view' of the Doctrine of Virtue (2015: 3). On the other hand, some commentators have argued that Kant is not committed to the intelligible choice of a disposition. Henry Allison argues that Kant does not think that the choice of a *Gesinnung* is 'like a choice of a disposition or character in a full-blown psychological sense' (Allison 1990: 142). On Allison's view, the *Gesinnung* should be understood as more like a *Denkungsart*, or a set of principles. Julia Peters on the other hand denies that we can choose our *Gesinnung* at any one moment. On Peters' view, the *Gesinnung* is not a disposition that precedes and causally determines us to act either out of duty or self-interest depending on whether it is good or evil. The good *Gesinnung*, for Peters, is not 'fully present at the time at which a particular moral choice is made' (2018: 507). Peters claims, rather, that a good *Gesinnung* must be constructed through a series of dutiful moral deeds over time (2018: 516). Both Peters and Allison reject the idea that one's character and subsequent empirical history is fixed by a choice of disposition.[19]

Kant can hold that moral reform requires both a 'change of heart' and gradual reform by appealing to his transcendental idealism. His rigorism suggests that we must understand the change of *Gesinnung* as a revolution from evil to good; there are only two *Gesinnungen* and, though Kant thinks there are several grades of evil (*R*, 6:

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

29–30), the *Gesinnungen* do not admit of degrees. But Kant never does say that the change of *Gesinnung* occurs instantaneously; this itself is a temporal term that is inapplicable at the noumenal level. The coherence of Kant's programme would indeed be threatened if he were describing two distinct undertakings that both independently suffice for securing a good disposition: one 'noumenal' choice that occurs instantaneously and a second, phenomenal struggle that occurs over time. But Kant need not endorse a 'revolutionary' view that reform happens instantaneously by contrast with a 'gradualist' account. Rather, because the choice of disposition is not spatio-temporal, he is not committed to an understanding of this that would conflict with a theory of phenomenal moral reform. On the contrary, the change of disposition needs to be *compatible* with how we are to understand the efforts to become good that we undertake as empirical agents.

In part two of the *Religion* Kant further clarifies that becoming a good human being requires us not only to achieve but to preserve a good disposition. He notes that although we must always judge our continual moral conduct in this life to be 'defective', because even an uninterrupted series of good deeds is not enough to show that one accepts the unconditional obligation of morality, God can judge the '*disposition* from which [our conduct] derives ... to be a perfected whole' (*R*, 6: 67). Being a good human being, which Kant sometimes describes as being 'well pleasing to God', requires our 'endless progress' in the phenomenal world to be 'a unity' (*R*, 6: 48). That is, to be a good human being an agent's conduct throughout her life (following a moral conversion) must be grounded in a good *Gesinnung*. As I have argued in the last section, Kant's theory of freedom implies that a person with a good disposition can nevertheless relapse into evil, and such a person would not be pleasing to God unless he or she restores and maintains her good disposition. In my following analysis, I will take care to distinguish between having a good *Gesinnung* or disposition and being a good human being (or, equivalently, being 'pleasing to God'), as the latter term describes a person whose post-conversion conduct is grounded in a good *Gesinnung*.

Commentators have questioned whether Kant can say anything plausible about how agents can possess a stable moral character. As previously noted, Michalson argues that because Kant thinks that one's character is outside of space and time, he is unable 'to give sense to any notion of continuity, over time, in the life of the moral agent ... he cannot show how a "previous" act or moral condition would be relevant to a "present" act' (1990: 85). And O'Connor also raises an objection that the commitment to the noumenal will in his critical philosophy leaves Kant unable to account for how temporal factors such as moral education or one's efforts to control their emotions can lead to a stable disposition, because 'nothing outside or *inside*' one's will can determine that it act in one way or another (1985: 294). O'Connor concludes that Kant's account of the change of heart involves an implausible conflation between Sartre's conception of a fundamental project, which implies the possibility of radical change regardless of one's current status, and Aristotle's account of a *hexis*, which is a continuous and stable moral disposition cultivated through moral education and the proper moral practices (1985: 293–6).[20] Because Kant insists on the freedom to choose one's disposition no matter one's previous history, he seems to have no way to explain how an agent's past moral efforts would lead her to preserve her good disposition when she possesses it. I will call this the 'stability problem'.

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

Kant does have resources to respond to this problem. Although he denies that we should look for a temporal origin for a human being's character, he does argue that there are specific empirical practices that can give coherence to an agent's moral life. Furthermore, as we will see, these practices are morally significant – they can play a role in an agent either adopting or maintaining a good moral disposition. To see how Kant can think this, we will have to look in detail at his theory of phenomenal moral reform.

3

For Kant, phenomenal moral reform has two stages: in the first stage an agent undergoes a revolution of thought, which I will also refer to as a 'transformation of attitude'. This is accomplished by making a commitment to the moral law. In the second stage of moral reform an agent continuously labours to live up to her moral commitment; this involves the cultivation of moral emotions.[21] This section will discuss the first stage of Kant's account and will focus in particular on the moral significance of the transformation of attitude.

Much of Kant's discussion of the moral transformation at the end of part one of the *Religion* describes various empirical practices. Kant claims that the moral revolution which leads a person to become good is part of 'moral discipline' (*R*, 6: 51). In the central passage 6: 47–8, he claims that this 'revolution' that must accompany the 'gradual reformation in the mode of sense' is a revolution 'in the mode of thought (*Denkungsart*)' (*R*, 6: 47). So, although Kant contrasts the gradual reformation of sense with the revolution of thought, these both refer to actions that we undertake in the empirical world and are part of a process of phenomenal reform. Kant re-emphasizes this when he notes that 'a human being's moral education must begin, not with an improvement of mores, but with the transformation of his attitude of mind and the establishment of a character' (*R*, 6: 48). A transformation of one's 'attitude of mind' suggests not an intelligible deed but an agent's attempt in the empirical world to radically restructure the way that she relates to her moral duties. Because Kant's discussion is confined to the phenomenal realm here, this should be understood as the establishment of *empirical* character.

Kant discusses the formation of empirical character in most detail in the *Anthropology* where he describes character as the 'property of the will by which the subject binds himself to definite practical principles that he has prescribed to himself irrevocably by his own reason' (7: 292). Because one can bind oneself to both good or evil principles, one can have either a good or evil empirical character. But, of course, what is relevant for Kant's discussion in the *Religion* is the establishment of a good character, whereby a subject forms a commitment to the moral law in its purity. Kant reiterates that this establishment of empirical character can be considered a type of transformation: 'one may also assume that the grounding of character is like a kind of rebirth, a certain solemnity of making a vow to oneself; which makes the resolution and the moment when this transformation took place unforgettable to him' (*Anth*, 7: 294). Education and teaching cannot 'bring about this firmness and persistence in principles *gradually*, but only, as it were, by an explosion which happens one time' (*Anth*, 7: 294). The reason, presumably, why this cannot happen gradually is because it requires an agent to endorse a general and unitary commitment to morality, and this

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

is something that must occur at one time. As Kant argues in the *Religion*, this decision must be 'single and unalterable' (*R*, 6: 48); the agent who makes a genuine commitment to morality must intend to live up to this commitment throughout his entire life as this is what one's moral obligation demands.[22] Such a decision recontextualizes the practical conduct that follows it. The agent who attempts to establish a good character could live with an unwavering adherence to morality and as such would become a good human being. The person who fails to do this is evil, just like he was before this transformation, but this failure is contextualized as a (culpable) inability to follow through with the commitment that he made in utmost seriousness.

We should understand this decision, which Kant sometimes describes using religious terminology as the decision to become a 'new man' (*R*, 6: 48), as expressing a commitment to morality. This commitment, as something done qua phenomenal agent, is not identical with the intelligible choice of a good disposition. In this sense, there is an ambiguity in Kant's discussion of what it means to endorse a good maxim in the *Religion*. Considered at the noumenal level, it describes the intelligible act to make the moral law one's incentive in action. But considered empirically endorsing a good maxim is *aspirational*; this is why Kant claims in *Religion,* 6: 47–8 that changing one's maxims makes one 'receptive to the good' but does not imply that a person is good as such (*R*, 6: 48).

Nevertheless, Kant clearly thinks that such a decision is morally significant – it is a necessary part of becoming a good human being. The reason why he thinks this, however, is initially not entirely clear. At one point, Kant claims that absent making a commitment to the moral law an agent can be at best '*legally* good' (*R*, 6: 47), which he understands in the *Religion* as acting merely in conformity with morality. The person who seeks to become good without a transformation of attitude would not be genuinely acting from duty. But this answer is question-begging; it is not clear why continuous dutiful action requires an agent to make such a moral commitment. Kant comes closer to a satisfying response when he notes that, without transformation of attitude, an agent will 'fight vices individually, while leaving their universal root undisturbed' (*R*, 6: 48). He makes a similar point in the *Anthropology*, noting that 'wanting to become a better human being in a fragmentary way is a futile endeavor, since one impression dies out while one works on another; the grounding of character, however, is absolute unity of the inner principle of conduct as such' (7: 294–5). Kant's suggestion is that by making a moral commitment an agent accepts a principle that serves to unify his or her conduct.

This point can become more plausible still if we understand the transformation of attitude as involving more than an abstract commitment to morality; the revolution of thought grounds specific moral duties which an agent undertakes in an effort to combat certain vices. As Kant makes clear, one of the main sources of evil arises due to agents repressing or otherwise not acknowledging the extent of their moral vocation. Certain commentators, most notably Laura Papish, have argued that self-deception and evil are closely connected for Kant (2018: 87–115). This is plausible because Kant denies that human beings are diabolical, which he takes to mean that they do not incorporate 'evil qua *evil*' into their maxim (*R*, 6: 37).[23] In the *Groundwork,* Kant argues that we 'like to flatter ourselves by falsely attributing to ourselves a nobler motive [than the one that actually guides our actions]' (4: 407). In the *Religion*,

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

he discusses a related form of self-deception which he calls 'deliberate guilt' (*R*, 6: 38). This is guilt in which an evil heart deceives 'itself as regards its own good or evil disposition' by 'not troubling itself on account of its disposition' but instead taking itself to be 'justified before the law' (*R*, 6: 38). Deliberate guilt may characterize someone who knows that she has done an evil deed but quiets the demands of her conscience by considering herself to be good despite these failings. A genuine commitment to possessing a good character requires an agent to reject the incomplete and exculpatory picture of her conduct that she has constructed for herself and take responsibility for her past evil actions.[24] Accordingly, this moral commitment requires an agent to combat the specific vice of downplaying or ignoring one's moral transgressions. The transformation of attitude, then, grounds particular duties of moral self-examination. It is unsurprising that in the *Anthropology* Kant claims that character ultimately requires possessing a maxim of *truthfulness*: 'the only proof within a human being's consciousness that he has character is that he has made truthfulness his supreme maxim' (7: 295). The call to establish one's character requires an agent to examine his motives in order to discern that he is genuinely living up to his moral commitment. Insofar as deliberate guilt is an expression of a vicious state of mind, establishing a good character in an attempt to resist this is itself a necessary part of moral reform.

However, here this proposal faces an objection. Kant claims that we cannot know our fundamental maxim. As such, we can never be certain that we are indeed acting out of duty. This scepticism is best understood as scepticism about the ultimate grounds of our actions, and it threatens to undermine the very coherence of a duty of self-examination.[25] Fully addressing this issue would require a more extensive discussion than what I can give here, so I will confine myself to a few brief points. To start, we should not think that this scepticism by itself undermines the possibility and practical significance of moral self-examination. While Kant thinks that we can never know that we are acting purely from duty, as no matter how dutiful our actions may seem it is always possible that we are being covertly guided by self-love, self-examination can reveal that we are evil. This is because Kant presupposes that we have some basic knowledge of our intentions and our reasons for action. A person can know that they have, for example, made a cutting remark to offend a colleague – knowledge of this sort often does not require much if any sophisticated reflection at all. When reflection reveals violations of the moral law, moreover, it is very rare for wrongdoers to be able to plausibly claim that they genuinely had a good maxim and have unfortunately made an exculpatory error. For Kant, then, there is an asymmetry between good and evil conduct: whether we are genuinely acting from duty is opaque, but evil actions can be transparent to us.[26] Self-examination can reveal that an agent must redouble his efforts to become good.

Here is, briefly, what self-examination involves for Kant.[27] The general commitment to morality engendered by the revolution of thought does not specify what one's particular duties must be; self-examination involves determining what one's commitment to morality requires. Moral self-examination conceived in this way is not primarily about a person peering into the ultimate motives that guide her. But it can help the agent who undertakes it understand the nature and extent of her moral obligations. So, the self-examination that is required by a genuine commitment to morality is not undermined by the scepticism of grounds of action, which for Kant

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

is an inextirpable feature of moral life. Kant's discussion of conscience supports this proposal. Kant claims that 'an *erring* conscience is an absurdity' (*MM*, 6: 401).[28] His point here is not that all of our motives for action are transparent to us but rather: 'While I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not, I cannot be mistaken in my subjective judgment as to whether I have submitted it to my practical reason' (*MM*, 6: 401). While we can be mistaken about whether we are acting for duty or promoting a good end, we cannot be mistaken about whether we are assessing whether our actions are indeed morally permissible.[29] Kant is not surreptitiously overstepping his own epistemic restrictions in his discussion of self-examination.

4

Kant's discussion of a transformation of attitude reveals one part of his response to the stability problem. It suggests that, insofar as we have a coherent moral identity, it must be constructed by forming a commitment to morality, which is a transformative and significant episode in an agent's life, and continuously striving to live up to this commitment. As we will see now, Kant also argues that the agent who engages in this continuous moral struggle aims to instantiate certain ethical ideals, and that such moral labour has an influence on the moral psychology of the agent that undertakes it. Specifically, in the *Metaphysics of Morals* he claims that agents cultivate virtue through continuous moral struggle.

It takes Kant some time to arrive at this position, however, because his views about the nature and efficacy of moral labour shift over time. In the *Critique of Practical Reason,* Kant writes that we must hope to attain 'the *complete conformity* of dispositions with the moral law [as] the supreme condition of the highest good' (*CPrR*, 5: 122). He defines 'complete conformity' to morality as 'holiness' and argues that such holiness may be found through 'endless progress toward' this ideal (*CPrR*, 5: 122). Kant then argues that because we require endless progress to instantiate the ideal of holiness, we must postulate the immortality of the soul. This proposal however has several serious problems. As Guyer points out, it is unclear why Kant presents holiness an ethical ideal (Guyer 2016a: 167). If Kant denies, as he does, that human beings can be holy then it is hard to see how gradual progress can lead a person to become holy even if such progress were extended indefinitely into another life.

It is tempting to think that Kant has rejected this account by the time that he wrote the *Religion.* In the *Religion* Kant makes no explicit reference to the immortality of the soul; his view that God can take one's progress to be a unity can be interpreted as replacing the doctrine of immortality, because it entails that we can be pleasing to God if our continuous moral conduct in this life is grounded in a good fundamental maxim. However, there are signs that Kant has not completely abandoned his views from the second *Critique.* In part one of the *Religion* he writes: 'The original good is *holiness of maxims* in the compliance to one's duty ... whereby a human being, who incorporates this purity into his maxims, though on this account still not holy as such (for between maxim and deed there still is a wide gap), is nonetheless upon the road of endless progress to holiness' (*R,* 6: 46–7). This passage draws a subtle distinction between holiness of maxims and being holy in general.[30] The person who has holy maxims is motivated to act solely out of duty by pure practical reason. Holiness of

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

one's maxims does not make one holy as such, and indeed this is not possible for human beings. However, at the end of this quote Kant returns to the claim that adopting such a maxim would lead one to be on the road to holiness. Thus, in the *Religion*, Kant maintains that holiness is an ethical ideal and the relationship here between endless progress and holiness remains extraordinarily opaque, just as it was in the second *Critique*.

In the Doctrine of Virtue of the *Metaphysics of Morals* Kant once again revises his views about the ethical ideal towards which our conduct is directed.[31] He replaces the ideal of holiness with the ideal of the autocratic will that has the strength to execute its moral duty. Anne Margaret Baxley notes that autocracy is 'the ideal form of moral self-governance for merely finite rational beings' (2010: 49). The autocratic agent possesses virtue, which Kant understands as 'the moral strength of a human being's will in fulfilling his duty ... insofar as this constitutes itself an authority *executing* the law' (*MM*, 6: 405). Unlike holiness, which characterizes an agent who does not experience any impulses that would impel her to act contrary to the moral law, the virtuous person still represents moral laws as commands that she must dutifully follow. This shift implies that we reconceive the continual moral struggle as a struggle to develop the strength required to put one's moral intentions into practice.

Kant describes the person who lacks virtue as follows: 'weakness in the use of one's understanding coupled with the strength of one's emotions is only a *lack of virtue* ... which can indeed coexist with even the best will' (*MM*, 6: 408). While Kant does not employ the language of the *Gesinnung* here, his observation can bear on his discussion of the transformation of disposition in the *Religion.* His thought is that even a person who possesses a good will or, in the parlance of the *Religion*, a good *Gesinnung*, is not virtuous. This is because virtue must be developed over time.

Here is why Kant thinks this. Kant holds that virtue involves the cultivation of moral emotions. The primary moral emotions that Kant discusses in relation to virtue are moral feeling and conscience. In describing moral feeling, Kant notes:

> A human being has a duty to carry the cultivation of his *will* up to the purest virtuous disposition, in which the law becomes also the incentive to his actions that conform with duty and he obeys the law from duty ... since it is a feeling of the effect that the lawgiving will within the human exercised on his capacity to act in accordance with his will, it is called *moral feeling* ... it is a moral perfection, by which one makes one's object every particular end that is also a duty. (*MM*, 6: 387)

For Kant, we do not have a duty to possess moral feeling, since this is a capacity that humans are endowed with in general; we are susceptible 'to feel pleasure or displeasure merely from being aware that our actions are consistent or contrary to the law of duty' (*MM*, 6: 399). Kant's theory of the moral emotions is teleological in the sense that he takes these emotions to develop from our particular dutiful actions. His discussion of beneficence illustrates this:

> Beneficence is a duty. If someone practices it often and succeeds in realizing his beneficent intention, he eventually comes to actually love the person he has helped. So the saying 'you ought to *love* your neighbor as yourself' does not

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

> necessarily mean that you ought immediately (first) to love him and (afterwards) by means of this love do good to him. It means rather *do good* to your fellow man, and your beneficence will produce love of man in you (as an aptitude of the inclination to beneficence in general). (*MM*, 6: 402)

It is not the case that we ought to start with the cultivation of emotions and hope that these emotions would lead to dutiful action. Rather, we should employ our faculty of the understanding to determine how to fulfil particular duties. Dutiful action, Kant claims, would lead to the development of the moral feeling of love. Similarly, we have an indirect duty to not 'avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out, and not to shun sickrooms or debtor's prisons and so forth in order to avoid sharing painful feelings' (*MM*, 6: 457). Though we may experience painful feelings by seeking out those who are less fortunate, this experience is a vivid reminder of our imperfect duties of charity.

There is much more that can be said about Kant's account of the moral emotions.[32] But this brief discussion shows how it informs Kant's theory of moral reform and reveals how he can answer the stability problem. As we have seen, the transformation of attitude requires an agent to combat the vices of self-deception such as deliberate guilt. I have argued that duties of self-examination can be discharged by determining what this moral commitment requires and performing particular good deeds based on this commitment. In Kant's account of virtue, he argues that an agent cultivates moral emotions through particular dutiful actions. By performing good actions an agent will strengthen her faculties of feeling and her responsiveness towards her moral duty, which will allow her to more easily demonstrate virtue when her will is tested in the future.

5

Kant's account of moral reform seems to involve an unstable combination of two core commitments. The first is that one is always free to adopt a good or evil disposition by an act of the noumenal will that occurs outside of space and time and cannot be determined by natural or temporal factors, and the second is that one becomes a good human being through an incessant temporal struggle. I have argued that a person's *Gesinnung* is the result of a free intelligible choice, and that one is always free to adopt either a good or evil disposition. But I have also argued that this thesis is not in tension with Kant's claim that moral reform requires incessant labouring and becoming. If we understand the *Gesinnung* as a disposition to prioritize the moral law over self-love, there is no reason to think that this cannot be secured over time – though not *in* time – through continuous moral struggle. Following a conversion, one is judged by God to be a good human being only if her subsequent action is grounded in a good *Gesinnung*.

In the *Groundwork*, Kant notes that even the good will that achieves no good ends would 'like a jewel ... shine by itself, as something that has full worth in itself' (4: 394). In his later practical writings he argues that stability of character comes from setting and *achieving* morally good ends; Kant thinks that we must construct and preserve our moral identity through the proper moral practices that structure our ethical life. The agent who undergoes a transformation of attitude and strives to put her

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

moral principles into practice will over time acknowledge the 'perceived ... efficacy of these [good] principles' on her conduct, which would give 'cause to infer ... a fundamental improvement in [her] disposition' though – in accordance with Kant's epistemic restrictions – such an inference remains merely a 'conjecture' (*R*, 6: 68).

The account of moral reform that Kant develops does not imply that he abandons the view that we can always choose our moral disposition at any point in our natural lives. Kant's theory of freedom implies that there is always a precarity to human beings' efforts to become good. Because he thinks that cognition is limited to the sensible conditions of space and time, one cannot ever know that he or she genuinely possesses a good disposition, and the human capacity for self-deception suggests that moral failures may become unnoticed or entrenched for agents who do not continuously affirm their moral commitment. In this way, Kant's transcendental idealism informs, rather than threatens, a rich conception of the moral life.

Acknowledgements. I would like to thank Paul Guyer, Michelle Kosch, Charles Larmore, Reed Winegar, Justin Shaddock, Anna Guo, Taylan Susam and the Brown University Dissertation Workshop for their feedback at various stages of developing this article. I also want to thank two anonymous referees at *Kantian Review* whose comments helped me greatly improve it.

Notes

1 Parenthetical references to Kant's writings give the volume and page numbers of the *Akademie Ausgabe* unless otherwise noted. I use the translations found in *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1992–). Abbreviations used are as follows: *CPrR* = *Critique of Practical Reason* (in Kant 1996a); *G* = *Groundwork of the Metaphysics of Morals* (in Kant 1996a); *MM* = *Metaphysics of Morals* (in Kant 1996a); *R* = *Religion within the Boundaries of Mere Reason* (in Kant 1996b); *Anth* = *Anthropology from a Pragmatic Point of View* (Kant 2007). Italics used in quotations represent Kant's own emphasis unless otherwise noted.

2 A similar line of thought can be found in Broad (1952), though Broad is not arguing against Kant's theory of freedom specifically but against libertarianism about free will in general. Thanks to Charles Larmore for this reference.

3 Because my article is focused on interpreting the phenomenal and intelligible aspects of moral reform, there are aspects of Kant's account that I cannot discuss in detail. Kant employs religious concepts or imagery extensively when discussing the topic of moral transformation. In part three of the *Religion*, he claims that the struggle to become good requires participation in an ethical community (6: 93–5), which must take the form of a church (6: 101). In the *Religion* part two, Kant notes that the Christian symbol of the Son of God can also help us adopt a good disposition. He also suggests that divine assistance, or grace, may be required to undergo the change of heart (see 6: 44 and 51–2). I believe that a complete account of Kant's discussion of moral reform would have to explain why he appeals to his moral religion in order to illustrate how the change of heart can be accomplished, but this topic requires much more extensive treatment than I can give in this article. However, see nn. 16 and 22 where I discuss grace and the Son of God respectively in more detail. Recent commentaries by Pasternack (2014) and Wood (2020) contain helpful discussions of all these topics. For a concise and interesting treatment of religion and moral reform, see Vanden Auweele (2015). While I also cannot discuss the ethical community and its relation to moral reform in detail, there is now an extensive literature on this topic, including e.g. Rossi (2005), Wood (1999: 283–321) and (2011), Moran (2012), Guyer (2016b: 275–302), Papish (2018: 203–31) and Pasternack (2021).

4 These terms are from Biss (2015: 3).

5 For an influential discussion of the incorporation thesis, see Allison (1990: 29–53). See also Schapiro (2011) for a more recent attempt to defend the incorporation thesis.

6 The qualification 'in morally significant circumstances' is necessary here because by itself acting out of self-love is not evil. The person who learns an instrument or enjoys a nice meal is acting out of self-love,

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

as Kant understands it. But there is nothing immoral about doing these things provided that one is not avoiding certain moral obligations in doing so.

7 It is found at p. 73 in Kant 1996b and at 6: 24 in the Academy edition.

8 Premises such as (1) and (2) can be contested but are undoubtably central features of Kant's moral philosophy, so the assumptions that bear most of the weight in his argument are (4) and (5). Guyer points out that Kant relies on (4) in the argument for the universal law formulation of the categorical imperative (Guyer 2016b: 132; *G*, 4: 400–2). The most extensive defence of (4) is perhaps found at 5: 22 of the *Critique of Practical Reason*, where Kant claims that all material practical principles, that is, practical principles grounded in the representation that the reality of a particular object is desirable 'belong without exception to the principle of self-love' (*CPrR*, 5: 22). Thanks to Paul Guyer for a discussion of these points.

9 Peters (2018: 501–2, 514) provides a similar analysis of Kant's argument.

10 Such a person would be 'morally dead' and Kant denies that this is possible for human beings (*MM*, 6: 400).

11 For instance, Morgan (2005) and Guyer (2016b: 181–2).

12 As Julia Peters observes, *Grund* can be interpreted to mean either a cause or a justificatory reason. She argues that both of these meanings apply to the *Gesinnung* (Peters 2018: 498–9).

13 The *Gesinnung* can be understood as a disposition insofar as it is the freely chosen ground of particular good or evil actions but is itself not identical with such actions (*R*, 6: 20). Kant most clearly explains this point when he describes the evil fundamental maxim as a ground of particular evil actions; see *R*, 6: 32. See also *R*, 6: 45n. where he describes a good *Gesinnung* as a ground of morally good acts.

14 Allison suggests that we may have both a supreme maxim, which itself is expressed by the propensity to evil, and other higher-order maxims (2020: 502). If I understand this view correctly, then we can have both a good higher-order maxim (the good *Gesinnung*) and an evil supreme maxim. As I have interpreted it, Kant's claim that we cannot be good or evil in parts militates against Allison's interpretation, because Kant develops this point to foreclose the view that Allison wants to endorse. Furthermore, if our supreme maxim is expressed by the propensity to evil, then it is not clear how moral reform would even be possible. Peters suggests that a person with an evil maxim may act from duty, but they will not do so with unconditional necessity (2018: 507–8). Thus, a person with an evil *Gesinnung* is not guided by the 'spirit of the law' (p. 508). Peters cites *Religion*, 6: 30, to defend her view as well. On my reading of 6: 30 the person with the evil *Gesinnung* is acting merely in conformity with duty and not from it.

15 It is possible that Kant's reference to Walpole's thesis, 'Every man has his price, for which he sells himself', serves as his definitive statement of this matter (*R*, 6: 38). Kant thinks the person who has an evil fundamental maxim would choose their own self-interest over what morality requires in any particular situation of choice if the benefits of doing so were great enough. In this sense, acting even in conformity with duty depends on the silence of sufficiently strong countervailing incentives. The person with the evil fundamental maxim has no principled way to follow the moral law when incentives of self-love become sufficiently strong and hence their following even the letter of the law is a matter of luck.

16 Kant follows the quoted passage with the qualification that without divine assistance our moral efforts may be insufficient: 'even if what we can do [to become morally better] is of itself insufficient and, by virtue of it, we only make ourselves receptive to a higher assistance inscrutable to us' (*R*, 6: 45). He elsewhere suggests that divine assistance, or grace, can assist us in the transformation of our disposition (*R*, 6: 44 and 52). Commentators disagree about how to understand Kant's views here. Michalson claims that Kant's position on divine assistance and moral transformation involves a 'set of wobbles' (1990: 9) that indicates an 'unstable conflation of a Reformation emphasis on the fall and an Enlightenment accent on freedom' (Michalson 1989: 265). He argues that, because of the radical evil in human nature, Kant must appeal to divine assistance to explain how transformation of a disposition is possible, which is in tension with Kant's critical commitment to thinking that a person's moral status must be based on actions that can be freely attributed to her. Similar interpretations can be found in Quinn (1984) and Hare (1996). Against this, Chignell (2014) argues that because the transformation of one's disposition occurs in the noumenal realm, we can hope for grace because we are not certain that grace is incompatible with the requirement that we must become good through our own powers. See Wood for a related argument (2020: 140–63). Pasternack gives a convincing argument that Kant denies the Augustinian thesis that we are unable to adopt a good disposition through our own free powers and argues that divine aid should '[support] the use of our own powers … God's aid for Kant is thus likened to that of a protector, teacher or patron' (Pasternack 2020: 115). Mariña (1997) and (2017) also claims that agents can makes use of grace

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

in their moral transformation and provides a taxonomy of the types of grace that, she argues, are operative in Kant's moral religion. While it is beyond the scope of this article to argue for a particular interpretation of Kant's views here, I disagree with Michalson and other commentators who argue that Kant must appeal to divine assistance to explain how transformation of one's disposition is possible. Michalson's interpretation presumes that because the transformation of one's disposition from evil to good is inconceivable it must be impossible without divine assistance, and as I argue above we need not attribute this view to Kant. But while Kant is not required to appeal to grace to secure the real possibility of the change of heart, he may hold that agents can freely make use of grace to assist in their moral transformation, as Pasternack and Mariña argue.

17 For a similar view, see Guyer (2020: 327). Guyer notes that the person who has reformed her character has made two choices: the initial choice of an evil fundamental maxim and the choice of a good fundamental maxim. He argues that Kant has no reason to think that we are limited to making only these two choices.

18 The precise way to understand the propensity to evil remains highly disputed, and I will not attempt to defend a particular interpretation here. My contention is that the universality of the propensity to evil justifies, for Kant, the practical claim that no human should be exempted from the task of improving their character through a revolution in one's fundamental maxim. Kant holds that 'we may presuppose evil as subjectively necessary in every human being, even the best' (*R*, 6: 32). He clarifies that this propensity to evil 'must itself be considered morally evil' and 'consist[s] in maxims of the power of choice contrary to the law' (*R*, 6: 32). This suggests that all human beings are afflicted by a propensity to evil, and, on the basis of this propensity, they can be thought initially to have an evil fundamental maxim. However, at different points in the text, Kant suggests that the propensity to evil is not the same as the evil fundamental maxim. Chiefly, Kant claims that the propensity to evil is inextirpable (*R*, 6: 37). While the propensity to evil is inextirpable, an evil fundamental maxim can be replaced by a good one; hence, they cannot be identical.

19 In the first section, I departed from Peters by claiming that the *Gesinnung* is a fundamental maxim that is logically prior to one's particular dutiful or vicious actions. On Peters' interpretation, the *Gesinnung* is constructed out of a series of moral decisions – one's life conduct – and it is not possible for even God to know a person's *Gesinnung* before her life conduct is completed, as there is no metaphysical fact of the matter: 'the person simply *is* neither good nor evil' (Peters 2018: 511). This is in tension with Kant's rigorist thesis that at any time a person is either good or evil. Furthermore, though Peters views the *Gesinnung* as a ground of one's moral conduct, it is not clear how it can serve this function if it is not present before a person's life conduct is completed. The *Gesinnung* seems to come on the scene too late to serve the explanatory role that Kant intends for it. Nevertheless, I agree with Peters that for Kant our moral duty requires unconditional adherence to the moral law. Much of what I say about the nature of phenomenal moral reform is meant to explain how Kant thinks that this can come about, so this aspect of my interpretation is consistent with much of what she says about transformation of character. The plausibility of Allison's interpretation, on the other hand, rests on his own 'epistemic' conception of transcendental idealism, which I will not argue against here. However, in section 3 I provide an alternative proposal for how Kant's conception of the *Denkungsart* figures into his broader account of moral reform.

20 Gressis raises a similar challenge: '[I]f our underlying intentions constitute our character, it is surprising that we should be able to change them whenever we want to, or that they may be, not only long-term intentions, but also short-term ones as well. Yet this is a tension that results when one wants to claim both that maxims express our character and that, because they are self-imposed rules, we may drop them whenever we want' (Gressis 2010: 223).

21 Biss (2015) and Papish (2018: 177–201) also claim that Kant's theory of moral reform comes in two stages. In Papish's account, the first step involves forming a commitment to morality, while the second step describes continuous labour over time in order to live up to this commitment, where this labour is specifically understood as 'cognitive' not 'volitional' (2018: 192). I agree with many of Papish's points, however there are some differences regarding how we exposit the nature of the moral commitment involved in moral reform. Papish views the commitment as something that 'straddles phenomenal and noumenal perspectives on choice' (Papish 2018: 178, n. 3) and argues that it is analogous to a marriage commitment. The reason that this example is apt, Papish thinks, is because the commitment to marriage happens both at once and must be preserved by continual work. On my proposal a commitment to

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

morality is something that we do qua phenomenal agent. The person who makes this commitment is not necessarily good, but such a commitment is still morally significant because it involves accepting that morality obligates them with unconditional necessity. For humans who are afflicted by radical evil, it involves them admitting that they have done evil and taking on certain duties of self-examination in the hope of avoiding evil in the future. My further exposition of moral reform can be understood as supporting Papish's proposal that moral reform is largely cognitive in nature, although I do not claim that continuous moral labour is cognitive *instead of* volitional – indeed, I think Kant's discussion of virtue resists such a dualism.

22 In part two of the *Religion*, Kant uses religious imagery to illustrate the categorical demands of morality. He claims that the 'ideal of moral perfection' can be understood as a '*prototype*' that 'has *come down* to us from heaven' (*R*, 6: 61). He refers to this prototype as the 'Son of God' (*R*, 6: 61). However, Kant immediately follows this claim by arguing that the Son of God is 'the idea of a human being willing not only to execute in person all human duties' (*R*, 6: 61), and by noting that the 'prototype [of the ideal of moral perfection] resides only in reason' (*R*, 6: 63). Unsurprisingly, Kant's discussion has invited inflationary and deflationary interpretations. Firestone and Jacobs argue that the prototype 'is a type of divine humanity, which constitutes the *telos* of our created species', and that 'Kant cognizes the prototype as coming down to our species via a transcendental incarnation in order to make his own disposition available to our species for adoption' (2008: 165). Vanden Auweele argues that the Son of God is meant to play a role in moral education by serving as a sensible example of our moral perfection which 'can augment the conviction that any human agent can reach such a state of perfection' (2015: 380). For additional commentary, see Pasternack (2012: 37–40; 2014: 133–41).

23 However, Papish herself does not want to put great weight on this particular claim in defending her interpretation (2018: 106).

24 On this point, I am influenced by McMullin (2013) who emphasizes the moral importance of attributing radical evil to oneself.

25 Thanks to an anonymous reviewer for raising this point.

26 Frierson makes a similar point (2003: 107).

27 In this discussion I am indebted to Ware's article on the duty of self-knowledge in Kant, especially his discussion of conscience (2009: 690–7).

28 See also the *Religion* (6: 185–7), which anticipates Kant's discussion of conscience in the *Metaphysics of Morals*.

29 This claim nevertheless may require some qualification. In the *Anthropology,* Kant mentions that the passions 'take root and can even coexist with rationalizing' (7: 265). Someone afflicted with certain passions, it seems, may not just be mistaken about what morality requires but also may be mistaken about whether they are rationally assessing whether their conduct is permissible or not. For a discussion of Kant on the passions, see Wehofsits (2020).

30 For an alternative discussion of this passage, see Guyer (2016a: 170).

31 There are, however, still vestiges of Kant's earlier view that holiness is an ethical ideal, as in one passage he claims that we are commanded to and must strive towards holiness, though we will never reach it (*MM*, 6: 446).

32 Recently there has been renewed scholarship on Kant's account of feelings and moral emotions. For example, see Guyer (2010). See also recent edited volumes by Cohen (2014) and Sorensen and Williamson (2018). For a discussion of Kant's views on the emotions in general, and not just the moral emotions, see Cohen (2020).

References

Allison, Henry E. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

—— (2020) *Kant's Conception of Freedom: A Developmental and Critical Analysis*. Cambridge: Cambridge University Press.

Baxley, Anne M. (2010) *Kant's Theory of Virtue: The Value of Autocracy*. Cambridge: Cambridge University Press.

Biss, Mavis (2015) 'Kantian Moral Striving'. *Kantian Review*, 20(1), 1–23.

https://doi.org/10.1017/S1369415423000304 Published online by Cambridge University Press

Broad, C. D. (1952) 'Determinism, Indeterminism and Libertarianism'. In *Ethics and the History of Philosophy* (London: Routledge), 195–217.
Chignell, Andrew (2014) 'Rational Hope, Possibility, and Divine Action'. In Gordon Michalson (ed.), *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide* (Cambridge: Cambridge University Press), 98–117.
Cohen, Alix (ed.) (2014) *Kant on Emotion and Value*. London: Palgrave Macmillan.
—— (2020) 'A Kantian Account of Emotions as Feelings'. *Mind*, 129(514), 429–60.
Firestone, Chris L. and Jacobs, Nathan (2008) *In Defense of Kant's Religion*. Bloomington, IN: Indiana University Press.
Frierson, Patrick (2003) *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge: Cambridge University Press.
Gressis, Robert (2010) 'Recent Work on Kantian Maxims I: Established Approaches'. *Philosophy Compass*, 5(3), 216–27.
Guyer, Paul (2010) 'Moral Feelings in the *Metaphysics of Morals*'. In Lara Denis (ed.), *Kant's Metaphysics of Morals: A Critical Guide* (Cambridge: Cambridge University Press), 130–51.
—— (2016a) 'Kant, Mendelssohn, and Immortality'. In Thomas Höwing (ed.), *The Highest Good in Kant's Philosophy* (Berlin/Boston: De Gruyter), 157–80.
—— (2016b) *The Virtues of Freedom*. Oxford: Oxford University Press.
—— (2020) *Reason and Experience in Mendelssohn and Kant*. Oxford: Oxford University Press.
Hare, John E. (1996) *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance*. Oxford: Oxford University Press.
Kant, Immanuel (1996a) *Practical Philosophy*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.
—— (1996b) *Religion and Rational Theology*. Ed. Allen Wood and George DiGiovanni. Cambridge: Cambridge University Press.
—— (2007) *Anthropology History and Education*. Ed. Günter Zöller and Robert Louden. Cambridge: Cambridge University Press.
Mariña, Jacqueline (1997) 'Kant on Grace: A Reply to his Critics'. *Religious Studies*, 33(4), 379–400.
—— (2017) 'Kant's Robust Theory of Grace'. *Con-Textos Kantianos. International Journal of Philosophy*, 1(6), 302–20.
McMullin, Irene (2013) 'Kant on Radical Evil and the Origin of Moral Responsibility'. *Kantian Review*, 18(1), 49–72.
Michalson, Gordon E. (1989) 'Moral Regeneration and Divine Aid in Kant'. *Religious Studies*, 25(3), 259–70.
—— (1990) *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge: Cambridge University Press.
Moran, Kate A. (2012) *Community and Progress in Kant's Moral Philosophy*. Washington, DC: Catholic University of America Press.
Morgan, Seiriol (2005) 'The Missing Formal Proof of Humanity's Radical Evil in Kant's *Religion*'. *Philosophical Review*, 114(1), 63–114.
O'Connor, Daniel (1985) 'Good and Evil Disposition'. *Kant-Studien*, 76(1–4), 288–302.
Papish, Laura (2018) *Kant on Evil, Self-Deception, and Moral Reform*. Oxford: Oxford University Press.
Pasternack, Lawrence (2012) 'Kant on the Debt of Sin'. *Faith and Philosophy*, 29, 30–52.
—— (2014) *Kant's Religion within the Boundaries of Mere Reason: An Interpretation and Defense*. London: Routledge.
—— (2020) 'On the Alleged Augustinianism in Kant's Religion'. *Kantian Review*, 25(1), 103–24.
—— (2021) 'The Ethical Community in Kant's Pure Rational System of Religion: Comments on Rossi's *The Ethical Commonwealth in History*'. *Philosophia*, 49(5), 1901–16.
Peters, Julia (2018) 'Kant's *Gesinnung*'. *Journal of the History of Philosophy*, 56(3), 497–518.
Quinn, Philip L. (1984) 'Original Sin, Radical Evil, and Moral Identity'. *Faith and Philosophy*, 1(2), 188–202.
Rossi, Phillip J. (2005) *The Social Authority of Reason: Kant's Critique, Radical Evil and the Destiny of Humankind*. Albany, NY: State University of New York Press.
Schapiro, Tamar (2011) 'Foregrounding Desire: A Defense of Kant's Incorporation Thesis'. *The Journal of Ethics*, 15(3), 147–67.
Sorensen, Kelly and Williamson, Diane (eds) (2018) *Kant and the Faculty of Feeling*. Cambridge: Cambridge University Press.

Vanden Auweele, Dennis (2015) 'Kant on Religious Moral Education'. *Kantian Review*, 20(3), 373–94.
Ware, Own (2009) 'The Duty of Self-Knowledge'. *Philosophy and Phenomenological Research*, 79(3), 671–98.
Wehofsits, Anna (2020) 'Passions: Kant's Psychology of Self-Deception'. *Inquiry* (ahead-of-print), 1–25. DOI: 10.1080/0020174X.2020.1801498.
Wood, Allen (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
—— (2011) 'Religion, Ethical Community, and the Struggle against Evil'. In Charlton Payne and Lucas Thorpe (eds), *Kant and the Concept of Community* (Rochester, NY: University of Rochester Press), 121–37.
—— (2020) *Kant and Religion*. Cambridge: Cambridge University Press.

Cite this article: Damstra, C. (2023). The Change of Heart, Moral Character and Moral Reform. *Kantian Review* **28**, 555–574. https://doi.org/10.1017/S1369415423000304