




ARTICLE

# Kant on Remorse, Conversion, and the Descent into the Hell of Self-Cognition

Benjamin Vilhauer 

City College and Graduate Center, City University of New York, New York, NY, USA  
Email: [bvilhauer@ccny.cuny.edu](mailto:bvilhauer@ccny.cuny.edu)

## Abstract

Kant’s conception of remorse has received little discussion in the literature. I argue that he thinks we ought to experience remorse for both retributivist and forward-looking reasons. This account casts helpful light on his ideas of conversion and the descent into the hell of self-cognition. But while he prescribes a heartbreakingly painful experience of remorse, he acknowledges that excess remorse can threaten rational agency through distraction and suicide, and this raises questions about whether actual human beings ought to cultivate their consciences in such a way as to experience remorse in the way he conceives it.

**Keywords:** Kant; remorse; conscience; retribution; punishment; conversion; self-cognition; suicide; guilt; hell

## 1. Introduction<sup>1</sup>

We can respond emotionally to the belief that we have acted immorally with a variety of painful feelings, some of which include disgust, embarrassment, shame, remorse, and guilt. Remorse differs from the feelings listed before it on the list just given, in that we can understand (even if we do not endorse) the earlier-listed feelings as responses to anything that can prompt mockery from others, even if it is clearly not a moral failing (e.g. vomiting at the podium during a talk as a result of food poisoning). The difference between remorse and guilt is more complex – terminologically, ‘guilt’ is used to refer both to a painful moral feeling and also to a state of culpability which can be determined by a court or God, which we think ought to have a painful feeling as a feature, but has other features too (for example, courts can ‘find’ people guilty even if they do not feel guilty). Here I will focus primarily upon remorse.

Kant comments on all the feelings in the list given above in various passages. Disgust (*Ekel*) can ‘make moral aversion sensible’ (*MM*, 6: 406),<sup>2</sup> and some specific vices (such as lechery) typically prompt disgust (*Anth-Mr*, 25: 1349), but disgust does not appear to play a significant role in Kant’s account of moral feeling.<sup>3</sup> He emphasizes that there are things which prompt disgust in some people but joy in others (*OFBS*, 2: 207), such as strong smells (*Anth-Mr*, 25: 1246) and French cuisine (*Anth-Mr*, 25: 1402), with the implication that while these things seem disgusting to him, there need be no moral defect in those who feel otherwise.<sup>4</sup> Kant thinks of embarrassment (*Verlegenheit*) as resulting from attention from others which prompts attention to

oneself and leads to self-consciousness (*Anth*, 7: 121, 132), and while this can happen when we act unethically, it can also reflect laudable motivations. He gives an example of embarrassment from being caught in a lie (*Anth-F*, 25: 633), but also says that someone seriously (*ernstlich*) in love may be embarrassed in the presence of the beloved, while a deceiving seducer will not be embarrassed in the presence of his target.<sup>5</sup> Shame (*Scham*) is perhaps more like a moral feeling than the previous ones, but while it can express injury to one's sense of moral dignity (*CPrR*, 5: 88), it can also express injury to one's sense of *honour* in a way that is not bound up with moral dignity but has merely to do with the opinions of others, whether they be right or wrong (*OFBS*, 2: 218).<sup>6</sup> Kant notes that shame can be prompted by forgetfulness (*Anth-Mr*, 25: 1241) and talking to oneself (*Rel*, 6: 195n).

Kant's comments on remorse and guilt indicate that they are always bound up with judgements that actions are immoral, but here analysis of his terminology is more complex. The German word Kant uses which translates as 'guilt' is *Schuld*, which can also mean debt or obligation, and he often uses it in the context of legal or divine judgement. He rarely uses *Schuld* in ways that indicate that he is thinking of feeling, though he does write of feeling guilty in at least one place (*schuldig zu fühlen*: *Rel*, 6: 38). The English word 'remorse' has a helpful simplicity in that it refers just to painful moral feeling, and while it can be uncontroversially matched up to Kant's German terminology, it takes a bit of work. Kant uses a number of terms which the Cambridge edition translates as 'remorse', including *Zerknirschung*, *Kummer*, *Verweis*, and *Reue*. All of these words can be translated in other ways. Kant uses *Zerknirschung* (also 'contrition') and *Kummer* (also 'sorrow') rarely. He uses *Verweis* in several places, but it is more literally 'rebuke' or 'reprimand'. He uses *Reue* at least 40 times, and the related verb *bereuen* at least 10, so this term plays a more significant role, but it has a meaning which is broader than 'remorse'. However, Kant draws distinctions which mark out a kind of *Reue* which I take to be identical to remorse, which he calls *moralische* [moral] *Reue* (*Eth-C*, 27: 353) and *wahre* [true] *Reue* (*Eth-V*, 27: 464). *Reue* has 'rue' as a close cognate in English, and both can mean painful regret for our past actions either because they were immoral, or because they were imprudent and brought negative consequences upon us, and only the former meaning matches 'remorse'. Kant discusses this distinction in multiple places. In the Collins Ethics lecture notes, he identifies *moralische Reue* as *Reue* for behaviour 'in regard to morality' (*in Ansehung der Moralität*) and distinguishes it from *Reue* because one has acted 'imprudently' (*unklug*) (*Eth-C*, 27: 353). Similarly, in his famous 1792 letter to Maria von Herbert (*Corr*, 11: 333, discussed below), he distinguishes *Reue* over 'imprudence' (*Unklugheit*) from *Reue* 'grounded in a purely moral judgement' (*auf bloßer sittlicher Beurtheilung ... Verhaltens gründet*) about one's behaviour.

The Cambridge edition often translates *Reue* as 'repentance', which suggests a theological context, and this context is often part of Kant's discussions, but the account of remorse presented here is relevant beyond this context. I take Kant to hold that remorse is essential for sincere repentance even in a theological context and that it can be isolated from any feelings specifically about God which may also be involved in repentance, for example, fear of damnation. Kant distinguishes the 'inner sorrow' of *wahre* [true] *Reue* (*Eth-C*, 27: 464) from the sorrow of *Buße*, which the Cambridge edition also translates as 'repentance' but can also be rendered as 'penance'. Kant remarks that *Buße* is 'not a good term; it derives from penances and chastisements'

(*Büßungen, Kasteyungen*) which we inflict on ourselves in the hope that we will thereby avoid being punished by God (Eth-C, 27: 464). At MM, 6: 485, he draws a parallel distinction between *moralische zu bereuen* and *büßen*:

[M]onkish ascetics, which from superstitious fear or hypocritical loathing of oneself goes to work with self-torture [*Selbstpeinigung*] and mortification of the flesh, is not directed to virtue but rather to fantastically purging oneself of sin by imposing punishments [*Strafe*] on oneself. Instead of morally repenting [*moralisch zu bereuen*] sins (with a view to improving), it wants to do penance [*büßen*][.]

I think these distinctions allow us to equate remorse with Kant's *wahre, moralische Reue* (though I think the Cambridge translation is helpful in rendering the broader range of expressions noted above as remorse as well). I will therefore typically replace 'repentance' and its derivatives in the Cambridge translation with 'remorse'<sup>1</sup> and its derivatives when it is clear that Kant is referring to *wahre, moralische Reue*.

In this article, I will address the following questions. What are the reasons Kant gives us for feeling remorse? How does Kant think we should cultivate the experience of remorse on the basis of those reasons? And last, but not least: is the experience of remorse Kant thinks we should cultivate an appropriate one to recommend for flawed agents like ourselves?

## 2. Kant's hybrid justification of remorse

There are three ways of justifying the pain of remorse which are helpful in understanding Kant's account. One is *retributive*: we should have painful feelings in response to past wrongful actions because we *deserve* them. Retributivist justifications are often called 'backward-looking' because they refer only to past wrongs. Pure retributivism would be the view that only retributive considerations play a role in justifying remorse. A second is *forward-looking*: we should feel remorse because it improves us and motivates us to act more morally in the future. A purely forward-looking justification would hold that only forward-looking considerations play a role in justifying remorse. A third is *hybrid*, drawing on multiple independent justifying reasons. I argue that if we exercise interpretative care in assembling the sometimes-contradictory remarks in Kant's texts, we have a good case for attributing a hybrid justification to Kant, which has a retributive basis and forward-looking constraints.

Kant's account of reasons for remorse has received little focused discussion in the literature. The most prominent discussion is indirect, in the context of Kant's account of conscience, in a debate between Thomas Hill and Allen Wood about whether conscience *punishes*. A resolution of this debate cannot on its own answer the question about Kant's reasons for remorse, since Kant's reasons for punishment are themselves matters of controversy. However, a review of the debate will situate the present inquiry in the literature and provide helpful context.

On Thomas Hill's interpretation, a 'bad conscience "hurts," and . . . Kant treats this pain as analogous to the suffering imposed on lawbreakers by the system of criminal justice' (2002: 352). He points out that Kant sees conscience as an inner court in which

we ‘think of ourselves as playing several roles: that of accuser, defender, and finally a judge’, who

issues verdicts of acquittal or condemnation. Like a trial judge, who is not legislating or merely informing others about the law, conscience ‘imputes’, ‘reproaches’, and passes ‘sentence’. If it judges us to be guilty, we are made to suffer, and at times the result can be torment. (Hill 2002: 301)

Hill reads Kant as holding that conscience ‘enforces the law, passes sentence’, and ‘makes us suffer for our misdeeds’ (2002: 353).

By contrast, Allen Wood holds that conscience does *not* punish (2008: 187). He claims that ‘Kant denies that it is even *possible* to punish oneself, citing *MM*, 6: 335, which is a discussion of social contract theory and the justification of bodily punishment according to principles of right. There Kant states that ‘it is impossible to will to be punished’, but just as clearly states that there is indeed a sense in which ‘I draw up a penal law against myself as a criminal’: ‘pure reason in me (*homo noumenon*) . . . subjects me, as someone capable of crime and **so as another person**’ to punishment (*MM*, 6: 335, boldface added). Kant’s point here is *not* that I do not punish myself, but that my self-punishment involves a multiplicity of perspectives on myself. Kant also addresses this concern about self-punishment at *MM*, 6: 485 (quoted above), where he faults ‘monkish ascetics’ for aiming at penance (*büßen*) rather than the experience of remorse (*moralisch zu bereuen*), in part because it involves ‘punishment [*Strafe*] chosen and executed by oneself<sup>t</sup>, while punishment ‘must always be imposed by another’. But as Hill emphasizes in his remarks quoted just above, Kant holds that agents must take on multiple perspectives in the experience of conscience, just as we saw they must do in drawing up penal law against themselves.<sup>7</sup> Kant discusses this in detail at *MM*, 6: 438:

[C]onscience is peculiar in that, although its business is a business of a human being with himself, one constrained by his reason sees himself constrained to carry it on as at the bidding of another person . . . [T]o think of a human being who is *accused* by his conscience as *one and the same person* as the judge is an absurd way of representing a court, since then the prosecutor would always lose . . . [A] human being’s conscience will, accordingly, have to think of *someone other* than himself . . . as the judge of his actions . . . This other may be an actual person or a merely ideal person that reason creates for itself.

If punishment must be inflicted by another person, and conscience contains a judge who we must represent as a different person even though that person is an aspect of ourselves, then if that judge inflicts painful emotional punishment on us, we must represent it as inflicted by another person even though it is inflicted by an aspect of ourselves. Kant’s criticism of monkish ascetics is not that it involves self-punishment *as such*, but rather that it involves self-punishment which is not mediated by the inner judge of conscience – that it is about doing penance by imposing unpleasant experiences ‘such as a fast’ on oneself *instead of* experiencing *wahre, moralische Reue*, and is thus a self-deluding substitute for the kind of self-punishment inflicted by one’s inner judge.

The most definitive passage supporting the view that conscience punishes is one that neither Hill nor Wood cite in their texts quoted here, at *MPT*, 8: 260:

[T]he virtuous man lends to the depraved the characteristic of his own constitution, namely, a conscientiousness [*Gewissenhaftigkeit*] in all its severity which, the more virtuous a human being is, all the more harshly punishes [*bestraft*] him because of the slightest indiscretion frowned upon by the moral law in him.

It is reasonable to assume that the virtuous have consciences which function properly, so if their consciences punish, then it is the proper function of conscience to punish.

Wood acknowledges that, on Kant's view, 'conscience can "judge us punishable"', but he claims that Kant thinks conscience does not inflict punishment upon us *itself*, and that 'our happiness or misery is left for the ruler of the world to decide', citing *MM*, 6: 439n, 440, and 460 in support. But the purpose of Kant's discussion of God at *MM*, 6: 439n and 440 is to explain that we must think of the 'internal judge' who 'pronounces the sentence of happiness or misery' (*MM*, 6: 439n) as *God*, while emphasizing that this idea is given to us 'subjectively, by practical reason' (*MM*, 6: 439) and is merely another perspective on 'morally lawgiving reason . . . present in our inmost being' (*MM*, 6: 440). This makes it clear that anything we must think of *God as doing to us* in the context of conscience must also be understood as something we *do to ourselves*.<sup>8</sup> *MM*, 6: 460–1 is a dramatic passage in which Kant talks about the limits of vengeance:

Every deed that violates a human being's right deserves [*verdient*] punishment, the function of which is to avenge [*gerächt*] a crime on the one who committed it (not merely to make good the harm that was done). But punishment is not an act that the injured party can undertake on his private authority but rather an act of a court distinct from him, which gives effect to the law of a *supreme authority* over all those subject to it . . . no one is authorized to inflict punishment and to avenge . . . except him who is also the supreme moral lawgiver; and he alone (namely God) can say 'Vengeance [*Rache*] is mine; I will repay'. It is, therefore, a duty of virtue . . . to refrain from repaying another's enmity . . . out of mere revenge[.]

The kind of punishment at issue here is external punishment we inflict on bodily distinct others, which Kant says cannot be done out of vengeance, because only God can punish out of vengeance. Implicit but obviously present here is the assumption that we *must not think of ourselves as like God* when we punish bodily distinct others. If we follow Wood in applying this passage to conscience, it would seem to imply that, since we *must think of the inner judge as God*, and we *can* think of God as punishing out of vengeance, and this inner judge is an aspect of *ourselves*, we can think of ourselves as punishing ourselves out of vengeance. We may be hesitant to make too much of Kant's remarks about vengeance here, since Kant's God does not always sound vengeful (*L-Th*, 28: 1086). But this passage also refers to desert, so it implies that God punishes *retributively* even if we bracket the reference to vengeance, and thus, if we follow

Wood in applying this passage to conscience, that we punish ourselves retributively too.

However, since Kant does not refer to conscience at *MM*, 6: 460, more direct evidence that conscience is retributive would be helpful. Kant offers more direct evidence in his discussion of the negative duty to promote others' 'moral well-being':

[T]he pain one feels from the pangs of conscience [Gewissensbissen] has a moral source . . . To see to it that another does not deservedly [verdienterweise] suffer this inner reproach [innere Vorwurf] is not my duty but his affair; but it is my duty to refrain from doing anything that, considering the nature of a human being, could tempt him to do something for which his conscience could afterwards torture<sup>t</sup> [peinigen] him[.] (*MM*, 6: 394)

This remark seems to imply quite straightforwardly that to experience pain from the pangs of conscience – to suffer remorse – is to *deservedly* suffer inner reproach.

However, Kant also makes many remarks implying that *forward-looking* considerations play a role in justifying remorse. In the 'monkish ascetics' passage discussed earlier, he states that *moralisch zu bereuen* requires a 'view to improving' (*MM*, 6: 485). In the *Religion*, he criticizes 'remorseful self-inflicted torments [reuige Selbstpeinigungen] that do not . . . originate in any genuine disposition toward improvement' (*Rel*, 6: 77). In the *Anthropology* he warns against regarding our 'record of guilt as . . . simply wiped out (through remorse<sup>t</sup> [Reue]), so that [we are] spared the effort toward improvement' (*Anth*, 7: 236).

In fact, in some passages, Kant suggests such a complete reliance on forward-looking considerations in justifying remorse that he seems to advocate a *purely* forward-looking approach, such that the *only* reasons to feel remorse are forward-looking. In the Mrongovius *Anthropology* lecture notes (1784–5), Kant criticizes 'idle desires' connected with the 'wish that something would not have happened which, however, now is impossible' as a 'distraction' which is 'senseless and harmful'. He gives 'remorse' (*Reue*) as an example of such desires, and says that it 'is good **merely insofar** as it impels us to cancel the consequences thereof and to act better in the sequel' (*Anth-Mr*, 25: 1335, boldface added). His point seems to be that when we feel remorse, we should try to eliminate painful thoughts connected with the wish that we had not acted badly and focus on acting better. But retributive (backward-looking) justifications of remorse would seem to be intrinsically bound up with the thought that past bad actions should not have happened, and so the *Anth-Mr*, 25: 1335 remarks suggest that there is no role for retribution. The *Starke 2 Anthropology* lecture notes (1790–1791) express a similar thought: 'fruitless remorse [Reue] is useless, because it attacks our powers too much' (*Anth-S2*, 34). The strongest suggestion of a purely forward-looking account comes in a pair of passages in the *Menschenkunde Anthropology* lecture notes (likely 1780–1):

Remorse [Reue] for crimes committed, as soon as it is not connected with the endeavour to make amends for the crime, is an empty delusion; for no other man is served by the fact that one plagues oneself with a torment, so one must also make an effort to help others again.<sup>t</sup> (*Anth-Me*, 25: 934)

Remorse [*Reue*] alone, as such, has no value, except insofar as it is the motive force of improvement. Remorse is serious [*ernstlich*] when it passes over as quickly as possible to good deeds. He who believes that remorse has a value in itself is very much mistaken. (Anth-Me, 25: 1086)

If remorse is a *delusion with no value in itself* if it does not motivate us to improve or make amends, then the justification of remorse is purely forward-looking, and there is no role for retribution. If we attribute this view to Kant, then we have a clear conflict with the view suggested at *MM*, 6: 394 and 6: 460 (quoted above).

There are, however, remarks in which Kant indicates that there are multiple reasons for remorse, and this would imply a hybrid theory. In Kant's critique of Johann Schulz's moral theory, he attributes a number of theses to Schulz with the clear implication that Kant does not accept them. One of these theses is that 'Remorse [*Reue*] is merely a misunderstood representation of how one could *act better in the future*, and in fact, nature has no other purpose in it than the end of improvement' (*RS*, 8: 110). If Kant rejects this, he thinks that remorse has *another purpose distinct* from improvement, and it is reasonable to assume that this is retribution, in light of the foregoing. Support for this view is also offered by Herder's *Metaphysics* notes, which attribute to Kant the statement that '[i]f remorse [*Reue*] about the past prevents all attention to the future, it is absurd' (*Met-Her*, 28: 90). This suggests that it is appropriate for remorse to be both forward-looking but also backward-looking to some degree. A final and definitive text supporting a hybrid justification appears in the Collins *Ethics* lecture notes. There Kant writes that preachers attending the dying 'must . . . see to it, that people do indeed feel remorse<sup>t</sup> for [*bereuen*] the transgression of self-regarding duties, since these can no longer be remedied, but that if they have wronged another, they genuinely try to make amends' (*Eth-C*, 27: 354). Kant here gives us a case where he thinks the imminence of death means that there is a forward-looking way to respond to some but not all of one's past wrongs. He states clearly that we should feel remorse for the actions to which we *cannot* respond in a forward-looking way, and based on *MM*, 6: 394 and 460, it is fair to assume that we should feel such remorse because we deserve it. His point about the actions to which we *can* respond in a forward-looking way seems to be that we should feel remorse about them but channel it into improvement and making amends.

Though there are tensions among the passages we have considered, together they give us sufficient evidence to attribute the following view to Kant. We must retributively inflict remorse on ourselves for past wrongs. But this self-retribution has two forward-looking constraints. First, remorse should be channelled into improved behaviour: it should lead us to act better in general (which we have seen at *Rel*, 6: 77, *MM*, 6: 485, *RS*, 8: 110, and *Anth-Me*, 25: 934 and 1086), and it should prompt us to make amends to the particular people we have wronged (which we have seen at *Eth-C*, 27: 354 and *Anth-Me*, 25: 934). Second, remorse must be moderated insofar as moderation is required to avoid damage to our rational agency. We saw this recently in his admonitions against remorse's 'distraction' (*Anth-Mr*, 25: 1335) and 'attacks [on] our powers' (*Anth-S2*, 34). We have a duty to cultivate conscience, which implies that we have an ability to rationally shape conscience (*MM*, 6: 401), and the reasons Kant gives us for feeling, channelling, and moderating remorse give us principles for shaping it. We should be careful about which actions we cause ourselves pain over,

to avoid a ‘micrological’ conscience ‘burdened with many small scruples on matters of indifference’, and a ‘morbid conscience’ which ‘seeks to impute evil in [one’s] actions, when there is really no ground for it’ (Eth-C, 27: 356). We are obligated to ‘sharpen’ conscience if it is too dull (MM, 6: 401), but we must not make it too sharp: to brood over remorse (*über Reue zu brüten*) can ‘make one’s whole life useless by continuous self-reproach (*Vorwürfe*)’ (Corr, 11: 333), and an ‘excess of remorse (*Kummers*) over . . . transgressions of duty’ can prompt suicide (Eth-V, 27: 642).

This section has set out Kant’s basic moral reasons for requiring and constraining remorse. We might think of these reasons as outlining the abstract structure of his theory of remorse. But we need more concrete details to answer some important questions that remain. How much remorse does Kant think we should feel, and for how long? And what is the mechanism by which we channel remorse into self-improvement? Kant engages these questions most directly in his discussions of the ‘descent into the hell of self-cognition’, which has connections to his concept of conversion.

### 3. The descent into the hell of self-cognition

Kant says he borrows the phrase *Höllenfahrt des Selbsterkenntnisses* from his acquaintance Johann Hamann (CF, 7: 55).<sup>9</sup> A version of this phrase appears in the Collins Anthropology notes (dated to 1772–3), where Kant says that one of the reasons the science of anthropology is neglected, despite being very interesting, is that ‘one conjectures he would not find much to rejoice at if he were to undertake the difficult descent into Hell toward the knowledge of himself [*die schwierige Höllenfahrt zur Erkenntniß seiner selbst*]’ (Anth-C, 25: 7). Kant uses it again in a famous passage in the Doctrine of Virtue:

For . . . a human being, the ultimate wisdom, which consists in the harmony of a being’s will with its final end, requires him first to remove the obstacle within (an evil will actually present in him) and then to develop the original predisposition to a good will within him, which can never be lost. (Only the descent into the hell of self-cognition [*Höllenfahrt des Selbsterkenntnisses*] can pave the way to godliness.) (MM, 6: 441)

Following a number of commentators, I think Kant’s reference to *removing an evil will, and developing the predisposition to good* means that we should link this passage to the conversion in the *Religion* (Rel, 6: 73), which Kant also calls a ‘justification [*Rechtfertigung*] of a human being who is indeed guilty but has passed into a disposition well-pleasing to God’ (6: 76).<sup>10</sup> Kant does not explicitly discuss remorse in the interpretatively challenging discussion of the conversion in the *Religion*, but I will argue that we can connect that discussion to a crucial discussion of the *Höllenfahrt* in *Conflict of the Faculties* (CF, 7: 10, 54–9) in a way that sheds light on both these texts and on Kant’s theory of remorse.

In Kant’s view, we all begin our lives as *radically evil*, which means that we are willing to comply with the moral law only insofar as we are also able to satisfy the incentives of self-love (Rel, 6: 36). This is a way of incorporating evil into our maxims. Kant claims that since this evil is in our ‘maxims in general (in the manner of *universal*



*principles* as contrasted with individual transgressions), it entails ‘an *infinity* of violations of the law’ and thus an ‘*infinity* of guilt [*Schuld*]’ for which we must expect ‘*infinite* punishment [*Strafe*] and exclusion from the Kingdom of God’ (*Rel*, 6: 72).<sup>11</sup> He holds that elimination of radical evil requires us to undergo a conversion in which we transform our disposition so that we choose to satisfy self-love only on the condition that we do our duty. But he thinks that the mere fact of becoming a better person through conversion does not negate our desert of punishment (*Rel*, 6: 72). Instead, he thinks conversion introduces a puzzle about *when* we can appropriately be punished. He thinks that it should not be thought of as ‘fully exacted before’ the conversion (*Rel*, 6: 73). On the other hand, ‘*after his conversion* . . . the punishment cannot be considered appropriate to his new quality (of thus being a human being well-pleasing to God)’ (*Rel*, 6: 73). Kant’s proposed solution is as follows:

[S]ince neither before nor after conversion is the punishment in accordance with divine wisdom but is nevertheless necessary, the punishment must be thought as adequately executed in the situation of conversion itself . . . Now conversion is an exit from evil and an entry into goodness, ‘the putting off of the old man and the putting on of the new’, since the subject dies unto sin (and thereby also the subject of all inclinations that lead to sin) in order to live unto justice. As an intellectual determination, however, this conversion is not two moral acts separated by a temporal interval but is rather a single act, since the abandonment of evil is possible only through the good disposition that effects the entrance into goodness, and vice-versa. The good principle is present, therefore, just as much in the abandonment of the evil as in the adoption of the good disposition, and the pain that by rights accompanies the first derives entirely from the second. The emergence from the corrupted disposition into the good is in itself already sacrifice (as ‘the death of the old man,’ ‘the crucifying of the flesh’) and entrance into a long train of life’s ills which the new human being undertakes in the disposition of the Son of God[.] (*Rel*, 6: 73–4)

While it is important to emphasize that Kant is cautious about using the historical story of Christ as an epistemological foundation in theology throughout his corpus and seems to rely on it only as metaphor, it is clear that he is telling us that the conversion involves a pain that is appropriately understood as excruciating and is like the pain of a mortal wound in some (perhaps symbolic) sense. And since Kant presents this pain as answering to the demand for the infinite punishment he claims we deserve, it is also presumably meant to be understood as infinite. The sense in which it is infinite is unclear, however.<sup>12</sup>

Is this pain to be understood as infinite in intensity? It is not clear that Kant’s account of the intensive magnitude of sensation permits infinitely intense sensations. At A176/B217, Kant says that the intensive magnitude of a sensation ‘can be raised from 0 up to any greater degree’, but this presumably does not entail that we can arrive at an infinite degree. And even if Kant’s theory of sensation allows for the possibility of infinitely intense pain, human physiology and psychology probably set limits. But despite these limits, Kant is clearly prescribing a very intense pain, and perhaps its intensity is meant to symbolically represent an infinite intensity.

Is this pain to be understood as infinite in duration? This passage is not clear about how long the pain Kant prescribes should endure, but he goes on to describe the ‘the suffering which the new human being must endure while dying to the *old* human being’ as occurring ‘throughout his life’ (*Rel*, 6: 74), and this makes it clear that at least some aspect of conversion pain is experienced as long as we live. Further, Kant’s appeal to the claim that we can only make ‘endless progress’ toward complete conformity with the moral law in his argument for immortality as a postulate of pure practical reason (*CPrR*, 5: 122; also see *Rel* 6: 75) might lead us to speculate that at least some of this suffering must continue forever, even beyond earthly life, and thus must be eternal.<sup>13</sup>

One way to get clearer about the nature of conversion pain is to attend to the distinction Kant draws toward the end of the passage above between the sacrifice ‘already involved in the emergence from the corrupt disposition’ and the ‘long train of life’s ills’ to which this ‘emergence’ is an ‘entrance’. This suggests that Kant is prescribing some kind of *first* stage of pain of finite duration, followed by a *second* stage that involves ‘life’s ills’, which in connection with the remark about suffering throughout life (*Rel*, 6: 74, quoted above) can be understood to continue as long as one’s life continues.<sup>14</sup> Perhaps pain in the first stage symbolically represents infinitely intense pain, and pain in the second stage is literal eternal pain.

In light of the discussions earlier in this article, it seems natural to think that remorse for our past evils would fit somewhere into the pain of conversion, but Kant does not explicitly address the pain of remorse in this discussion of the conversion. His clearest idea in this passage is about pain that results from resistance to the inclinations which ‘lead to sin’, a resistance which we must demand of ourselves in a new way after the conversion.<sup>15</sup> That is, while our inclinations persist after the conversion, we have rejected the evil maxim of conditioning compliance with law on satisfaction of our inclinations and have adopted a new maxim of conditioning satisfaction of inclinations on compliance with law, and resisting inclinations can be painful. Since having inclinations which do not necessarily conform to law is simply part of *what it is* to be a sensibly conditioned rational agent, we must assume that our inclinations will persist as long as we are such agents, and so it makes sense to think that the pain deriving from resisting those that dispose us to act immorally would continue as long as we continue as such agents. This kind of pain is thus a good candidate for what I claim is a *second* stage of conversion pain.

However, Kant’s focus on the pain of resisting inclinations creates an interpretative conundrum. Kant thinks that to understand the infinite punishment we must receive for past evils as ‘adequately executed in the situation of conversion itself’ (*Rel*, 6:73, quoted above), we must see suffering in this life as punishment for our past evils. But the pain of resisting inclinations is a puzzling thing to regard as punishment for past evils, for three reasons. First, since having inclinations which do not necessarily conform to law is simply part of being a sensibly conditioned rational agent, dealing with the pain of resisting inclinations is presumably part of the daily affairs of such agents. How can an ordinary kind of unpleasantness required to maintain rational agency be regarded as punishment? Second, it is pain we must grapple with in order to do the right thing *in the present*, so it seems to lack the backward-looking reference which seems essential to regard pain as deserved in light of past wrongs. The third reason takes longer to explain but is the most important.

Consider some features of Kant's view of the ontological and moral status of inclinations. The 'natural inclinations do not have us for their author' – they are 'conatural to us' and we 'cannot presume ourselves responsible for them' (*Rel*, 6: 35). They themselves are 'innocent' and bear 'no direct relation to evil' (*ibid.*). They only lead to sin if we adopt a fundamental maxim of conditioning our compliance with law on satisfaction of our inclinations and that fundamental maxim does not have a 'ground . . . in the natural inclinations' (*Rel*, 6: 34). The 'inclinations only make more difficult the *execution* of the good maxims opposing them' (*Rel*, 6: 57n). Now consider two sensibly conditioned rational agents: agent 1, a hypothetical agent who begins life in the *right* way, by prioritizing duty over inclination, and thus does *not* deserve to suffer, and agent 2, who begins life in the *wrong* way, by inverting that priority, and thus *does* deserve to suffer. Kant's view of the ontological and moral status of inclinations would seem to imply that both 1 and 2 would have to contend with the pain of resisting inclinations. If God or Reason say that it is *just* for 2 to have this pain *because of his past evils*, then 1 can reasonably object that it is *unjust* for him to *have this pain too*. We might suppose that there is an implicit background metaphysics which respects the deserts of 1-type agents by ensuring that they are never incarnated as sensibly conditioned agents, but that would seem to imply that we *are* in some sense responsible for being assailed by inclinations, and (as explained) Kant appears to reject this.<sup>16</sup>

Kant's idea that the pain of conversion is justified by our past evils would thus make more sense if he had a more plausible locus of punitive pain. I think what I claimed above to be the *first* stage of conversion pain can provide such a locus if the pain in that stage is remorse. Evidence to support this idea appears in *Conflict of the Faculties*, where Kant explicitly discusses the concepts of conversion, the *Höllenfahrt*, and remorse. Care is required to put this discussion forward as evidence about Kant's own view, because he offers it in the context of a critique of the Pietist conception of conversion.<sup>17</sup> In the Preface to this work, he says that Pietists hold that forgiveness requires 'an overwhelming remorse [*Zerknirschung*]', which he also calls 'a deep remorseful sorrow' (*ein tiefer reuiger Gram*) (*CF*, 7: 10). However, he faults their view *not* because of the deep remorse it prescribes, but because they hold that the 'human being' cannot 'attain this by himself', and think 'the remorseful' [*reuivolle*] sinner must especially beg this remorse [*Reue*] from heaven' (*ibid.*). He goes on to argue that it is 'obvious that anyone who still has to beg for this remorse' [*Reue*] (for his transgressions) does not really *feel remorse for* [*wirklich nicht reuet*] his deeds' (*ibid.*). So he is not denying the claim that we ought to feel deep remorseful sorrow – he is rather asserting (*contra* Pietism) that we can attain it on our own. Later, in the 'General Remark: On Religious Sects' (*CF*, 7: 48–61), he explicitly states that this remorse is part of conversion. At 7: 54, he asks 'how is rebirth (resulting from a conversion by which one becomes another, new man) possible[?]' and writes that

According to the Pietist hypothesis, the operation that separates good from evil (of which human nature is compounded) is a supernatural one – a feeling of remorse<sup>t</sup> [*Zerknirschung*] and crushing [*Zermalmung*] of the heart in *repentance* [*Buße*], a sorrow . . . bordering on despair which can, however, reach the necessary intensity only by the influence of a heavenly spirit. The human being must himself beg for this grief, while grieving over the fact that

his sorrow is not great enough (to drive the pain [*Leidsein*] completely from his heart). Now as the late Hamann says: ‘This descent into the hell of self-knowledge [*Höllenfahrt des Selbsterkenntnisses*] paves the way to deification’. In other words, when the fire of repentance [*Buße*] has reached its height, the amalgam of good and evil *breaks up* and the purer metal of the *reborn* gleams through the dross, which surrounds but does not contaminate it, ready for service pleasing to God in good conduct . . . But even in the highest flight of a mystically inclined imagination, one cannot exempt man from doing anything himself, without making him a mere machine[.] (CF, 7:55–6)

This passage clarifies Kant’s view of remorse in five ways. First, it appears to have the same basic structure as the critique of Pietism in the preface: it is obviously critical, but the target of criticism is not the view that we ought to experience profound remorse, but is instead the view that we cannot attain it on our own. Second, it provides support for the idea that Kant himself endorses the view that we ought to experience such remorse, because he clearly indicates that such remorse is (or is part of) the *Höllenfahrt des Selbsterkenntnisses*, which is established as an element of Kant’s own moral psychology by the way he refers to it at *MM*, 6: 441 (quoted above). Third, Kant’s reference to the kind of remorse which Pietists might acquire by begging God for it as *Buße* rather than *Reue* is in line with the distinction we observed earlier, such that *Buße* is not *wahre, moralische Reue*. Fourth, if remorse is self-retribution inflicted by a part of ourselves which we must also represent as God, then it is comprehensible to regard it as punishment that we deserve for our past evils, which I argued above is a problematic way to regard the pain of resisting inclinations. Fifth, if we suppose that Kant’s own view is that we should experience remorse like this under our own power (presumably through the faculty of conscience), then Kant’s smelting metaphor gives us a detailed phenomenological picture of the mechanism he thinks we use to improve ourselves with remorse, and of how the experience of remorse comes to an end.<sup>18</sup> We force the pain of remorse to rise to a great intensity, which leads to a kind of heartbreak, and with it, a conversion in which we adopt the disposition to satisfy self-love only on the condition that we do our duty, at which point the remorse abates and is driven ‘completely from [our] heart[s]’ (CF, 7: 55, quoted above). Kant presumably sees the ability to drive remorse from our hearts when we complete our conversions as (at least part of) the solution to the problem about excess remorse which he warns us about at *Eth-V*, 27: 642.

I want to be cautious about suggesting that Kant thinks there is a *causal* process here, such that remorse causes the conversion, and the conversion causes remorse to cease, since it is clear that Kant thinks of this phenomenology as the manifestation of noumenal freedom. However, if we have an empirical character which is governed by causal laws, along with a noumenal character which is *not* governed by causal laws (see e.g. A539/B567), we can entertain the possibility that there are empirical-psychological laws according to which the phenomenal experience of remorse sometimes causes our phenomenal experience of conversion without suggesting that remorse and conversion are not manifestations of noumenal freedom. We must, however, assume that there is enough diversity in such laws, or in the antecedent conditions determining their instantiation, to accommodate experiences of remorse which do not conform to the ideal Kant seems to prescribe here, since actual

experiences of remorse clearly differ widely, and it seems safe to say that many do not conform to Kant's ideal. This is in one way unproblematic, since Kant regularly emphasizes that his moral theory involves ideals which are (at best) challenging to actualize. It is problematic in other ways, however.

#### 4. The hazard of suicide

Kant's remarks on conversion and the ideal experience of remorse are apparently meant to describe something he thinks it is *possible* for us to experience, but also something we *deserve* to experience and that we *must* experience to purge ourselves of radical evil. (See e.g. *MM*, 6: 441, quoted above.) This suggests that he is asking us to crush our hearts with a remorse whose intensity symbolizes crucifixion, or a mortal wound, *until* we complete our conversions. What does this mean for remorseful people who have not had the conversion experience, or are not confident they have? Kant's theory would seem to imply that they must continue to punish themselves with ever-more intense remorse, until they are confident that they have made the *Höllenfahrt* and converted themselves. But it seems safe to say that many might seek to follow this advice without ever having an experience that gave them confidence that they were converted and could therefore justifiably give themselves a respite from remorse. We have seen a number of remarks from Kant in which he acknowledges that remorse which is protracted or intense can damage rational agency, and even dispose people to suicide. People who strive for the *Höllenfahrt* without ever becoming confident they have achieved it are likely to experience remorse that is protracted *and* intense.

The concern that we might lack confidence about whether we are converted is a serious one, at least in the context of Kant's moral epistemology. There are two puzzles here. First, there are (perhaps surprisingly) no behavioural criteria for conversion. Kant thinks that conversion does not guarantee subsequent moral perfection – conversion only gives us 'hope ... to find [ourselves] upon the good (though narrow) path of constant progress from bad to better' (*Rel*, 6: 48). I think Kant does assume that conversion entails that one's actions will improve in the long run, but it may be a very long run, since reason requires us to postulate immortality to ensure that we have enough time to make the necessary improvement (*CPrR*, 5: 122). Years or decades of actions that violate duty after an experience that *seemed* like a conversion do not entail that that it was *not* a conversion. If it *was* a conversion, then perhaps it is sufficient to respond to modest moral failures with moderate experiences of remorse, rather than the intense remorse of the *Höllenfahrt*, because we can hope that we are already on the path from bad to better. But if it was in fact *not* a conversion, then intense remorse is demanded. How can we tell? The second puzzle is that even if, after a seeming conversion, we seem to be fulfilling *all* our duties, and fulfilling them for the *right reasons*, we still cannot be confident that we are converted. This is because of Kant's scepticism about ever really being certain about our underlying motives:

[A] human being cannot see into the depths of his own heart so as to be quite certain, in even a *single* action, of the purity of his moral intention and the sincerity of his disposition, even when he has no doubt about the legality of

the action . . . [H]ow many people who have lived long and guiltless lives may not be merely *fortunate* in having escaped so many temptations? In the case of any deed it remains hidden from the agent himself how much pure moral content there has been in his disposition. (MM, 6: 392–3, also see G, 4: 407)

In the context of conversion, this passage implies that even if we seem to have had a conversion experience, we may still be radically evil, and we may need to continue to strive for intense remorse.

The concern that some may be driven to suicide by remorse is also a serious one, and it is recognized not only by Kant, but also in contemporary clinical psychology.<sup>19</sup> There may even be a kind of tragic rationality in responding to the belief that one deserves infinite punishment by excising oneself from the world. Perhaps this is why Kant makes the troubling remark that suicide from ‘excess of remorse [*Kummers*] over [one’s] transgressions of duty’ is not a ‘crude’ kind of suicide ‘which should be an object of general hatred’, but is rather a suicide which ‘could betray a worth of the soul’, like suicide for ‘the conservation of [one’s] honour’ (Eth-V, 27: 642), despite the fact that he sees suicide as a grave violation of duty.

We may even have an example of suicide motivated by remorse in Kant’s own correspondent Maria von Herbert. In a 1791 letter to Kant, she writes that when she revealed a protracted but harmless lie (apparently connected with a previous relationship) to the man she loved, ‘his love . . . vanished’ (Corr, 11: 273). She asks Kant for ‘solace, or for counsel to prepare [her] for death’, making it clear that she has been contemplating ‘tak[ing her] own life’ (ibid.). In 1792, Kant writes back to encourage her not to commit suicide, recommending ‘composure’, and advising that ‘life, insofar as it is cherished for the good that we can do, deserves the highest respect and the greatest solicitude in preserving it and cheerfully using it for good ends’ (Corr, 11: 334). But he denies her a ‘moral sedative’: he remarks that even a harmless lie is ‘a serious violation of duty to oneself and one for which there can be no remission’, and lectures her about some of the details of his theory of remorse discussed above, sternly advising that her ‘bitter self-reproach’ for the lie should not be *Reue* over ‘imprudence’ (*Unklugheit*) but *Reue* ‘grounded in a purely moral judgement’ (*auf bloßer sittlicher Beurtheilung . . . Verhaltens gründet*) about her behaviour (Corr, 11: 331–3). He goes on to attempt to temper his admonition with a remark part of which we have already considered:

But to brood over one’s remorse [*über . . . Reue zu brüten*] and then, when one has already caught on to a different set of attitudes, to make one’s whole life useless by continuous self-reproach [*Vorwürfe*] on account of something that happened once upon a time and cannot be anymore – that would be a fantastic notion of deserved self-torture [*verdienstlicher Selbsteinigung*] (assuming that one is sure of having reformed). (Corr, 11: 333–4)

This remark obviously leaves open the possibility that we have a reason to think continued self-torture is deserved if we are *not* sure of having reformed, that is, if we are not sure our conversion is complete. Von Herbert appears to have been quite familiar with Kant’s work, and thus may well have been familiar with his scepticism about ever really being certain about our underlying motives. She wrote to Kant again

in 1793, expressing a ‘sense of constantly reproaching [*vorwerfen*] herself’ (*Corr*, 11: 401), and a continued desire to ‘shorten [*her*] so useless life’ (*Corr*, 11: 402), and once again in 1794, stating that while she still wished to die, she felt that ‘if people take morality and friends into account [*they can*] with the greatest desire to die still wish for life and try to preserve it no matter what’ (*Corr*, 11: 486). Kant did not reply to either of these letters, and von Herbert committed suicide in 1803.

Owen Ware argues that in the *Religion and Metaphysics of Morals*, Kant struggles against his scepticism about the knowledge of our underlying motivations which could demonstrate the completeness of our conversion. Ware thinks that while Kant holds that we lack knowledge that our conversion is complete, he arrives at ‘subjectively sufficient grounds’ for the belief that our conversion is complete, grounds which yield what Kant calls ‘conviction’ (Ware 2009: 694, citing A822/B580). Ware argues that a key threat to knowledge of our underlying motivations is that when we evaluate ourselves, we are always inclined to see ourselves as better than we really are, or as Kant puts it in the *Groundwork*, that we like to ‘flatter ourselves by falsely attributing to ourselves a nobler motive’ (*G*, 4: 407). He argues that Kant circumvents this problem in his account of conscience by appointing God as an authoritative and distinct person within us: representing the verdict of conscience as handed down by God provides the verdict with an authority that self-evaluation lacks. Ware thinks this implies that if we are diligent about submitting not only our particular actions but also our maxims for the scrutiny of conscience, and conscience does not hand down a verdict of ‘guilty’, we can have conviction that our conversion is complete. This is a plausible and interesting account of some of Kant’s aims in his theory of conscience. But as discussed earlier (and as Ware acknowledges), we must recognize that what we represent as God in conscience is also an aspect of ourselves which we ourselves *set up* as authoritative, however rationally rarefied it may be. As Kant puts it, ‘[a]lthough it certainly sounds questionable, it is in no way reprehensible to say that every human being *makes a God* for himself’ (*Rel*, 6: 169n). Would it be unreasonable for someone uncertain about the completeness of his conversion to worry that he might have unwittingly incorporated a bit of self-flattery into the God he has made and that the verdict of his conscience therefore lacked authority, so that even if his conscience did not demand continued self-retribution, he might deserve it anyway? This objection might be seen as asking more of Kant’s ethics than it can reasonably provide, were it not for the serious damage to rational agency that threatens us if the authority of Kantian conscience is unstable in rational reflection. It seems to me that it is unstable and that its authority depends as much on a kind of faith as it does on reason, a faith which is least likely to be found in the people who need it most. Maria von Herbert may be an example of such a person, and her brother, Baron Franz Paul von Herbert, may be another.

The historical evidence that Franz Paul had expertise in Kant’s philosophy is as strong as it is with respect to Maria. The von Herbert family were lead paint manufacturers in Klagenfurt, and Franz Paul left the family business to study Kant’s philosophy in Jena and Weimar, making the acquaintance of a number of Kantian philosophers including Karl Leonard Reinhold. According to Wilhelm Baum, Reinhold and von Herbert had many conversations, and Reinhold wrote in a letter that von Herbert was his ‘house and table companion’ for four weeks (Baum 1996: 489). Upon Franz Paul’s return to Klagenfurt, he established a discussion circle to debate Kant’s

philosophy, despite strong local political opposition (Langton 1992: 481, Baum 1996: 499–504). Franz Paul committed suicide in 1811, after leaving this gnomic but sadly evocative passage in his will in 1810:

[M]y children, may you take the truth deeply into your hearts that an upright way of life is impossible without the faith in the divine judge announced by the conscience, without this faith the consequent man owes himself [*bleibt sich . . . schuldig*] for the last answer to the last question. (Baum 1996: 513)

This passage is of course too oblique and brief to count as a critique of Kant's idea of conscience, but it seems fair to say that it manifests concern about the reflective instability of that idea. It would probably be unfair to suggest that Kant's philosophy played a role in the death of either of the von Herbert siblings. People commit suicide for a variety of reasons, and the von Herberts' consciences were not the only painful things in their lives. Maria had lost her love, and Franz Paul's activities were the subject not only of political opposition but also active police scrutiny (Baum 1996: 499–504). Further, for all we know, the von Herberts may have had a genetic predisposition to major depressive disorder and may have suffered from chronic lead exposure, which is also positively correlated with major depression. But it does seem reasonable to think that we have textual evidence that remorse and anxieties about conscience *may* have played a role in their deaths and that neither found anything in Kant's philosophy to keep them alive. It is probably unreasonable to hope that any philosophy can prevent suicide. But it does not seem unreasonable to expect an account of how to moderate remorse from a philosopher who advocates heart-crushing remorse despite acknowledging that it can prompt suicide.

It may be objected that Kant *does* address the problem of moderating extreme feelings in his account of rationally regulating feeling to avoid what he calls affect (e.g. *MM*, 6: 407–9). 'Affect' refers to feeling that we allow to develop without moderation by reason, which can be overwhelming in a way that is damaging to agency, because it makes 'reflection . . . impossible or more difficult' (6: 407). Kant holds that we have what he calls a 'duty of *apathy*' to suppress affect, which we can fulfil through our 'inner freedom . . . to bring all [our] capacities and inclinations under [our] control and so to rule over [ourselves]', and avoid being 'governed by [our] feelings' (6: 408). Perhaps we should regard remorse so intense that it might overwhelm us and drive us to suicide as affect, and expect ourselves to draw on this inner freedom to diminish its intensity. But it would seem that we cannot regard even terribly intense remorse as affect, because of the role Kant gives such remorse in conversion – such remorse is an aspect of the infinite punishment we deserve, and is part of the descent into the hell of self-cognition which is necessary for conversion.

## 5. Conclusion

The conclusion I draw is that, while Kant has a detailed and interesting account of remorse, he does not provide an adequate explanation of how we can prevent remorse from damaging our rational agency, so we cannot be confident that we ought to cultivate our consciences in the way he demands. As far as my argument here is concerned, it may yet be that Kant is right about how we should experience remorse



and that it therefore falls to Kant's successors in the ethical tradition he inaugurated to explain how we can experience such remorse without damage to rational agency. But another path open to Kantian ethicists is to seek alternative models of remorse which can be developed within Kantian ethics, in a way that is revisionist but nonetheless grounded in Kant's philosophy.

One alternative model might retain the abstract structure of Kant's theory of remorse – the hybrid justification with a retributive basis and forward-looking constraints – and seek a new account of the mechanism we should use to improve ourselves with remorse which dispenses with the idea of heart-crushing pain and the hazards it presents. Perhaps more moderate self-retribution inflicted in response to particular misdeeds, instead of a general effort to hurt ourselves until we are transformed, would suffice for moral progress and present fewer hazards.

Another alternative model departs farther from Kant's own model: it rejects retributivism about remorse and instead grounds the value of remorse in the duty of sympathy (e.g. *MM*, 6: 456–8). The duty of sympathy gives us a reason to be pained by the pain we cause people we have wronged, and this is a reason to be pained by our wrongs which cannot be reduced to a forward-looking reason. Thus a Kantian sympathy-based model is, like Kant's own model, not purely forward-looking. The duty of sympathy arguably derives from the duty to take others' ends as one's own (Vilhauer 2022a) and is at least as fundamental in Kant's moral theory as retributivism.<sup>20</sup> Sympathy-based remorse is arguably less damaging to rational agency than self-retributive remorse. But the details of this argument must be presented elsewhere (Vilhauer 2022b).

**Acknowledgements.** Thanks to the editors and reviewers at *Kantian Review*, as well as Matthew Altman, Jeffrey Blustein, and Derk Pereboom, for helpful communications at various points in the history of this project.

## Notes

1 Some aspects of the interpretation presented here are also presented in Vilhauer 2022b, from which the present article draws some text. The earlier article's purpose is to contrast Kant's own account of remorse with an alternative account that I argue is preferable and available in his corpus (briefly outlined in the conclusion of the present article).

2 Abbreviations and translations for Kant's texts are as follows, unless otherwise noted. 't' within quoted passages indicates my modification of the Cambridge translation. Pagination is by *Akademie* edition if available and by volume cited if not. Translations from volumes in German are my own. A/B=*Critique of Pure Reason* (Kant 1999). Anth=*Anthropology from a Pragmatic Point of View*, trans. Robert B. Loudon, in Kant 2007: 231–429; Anth-C, Anth-F=Collins and Friedländer notes from Kant's Anthropology lectures, in Kant 2012: 11–26, 37–255; Anth-Me=Menschenkunde notes from Kant's anthropology lectures, in Kant 1997b: 849–1203; Anth-Mr=Mrongovius notes from Kant's anthropology lectures, in Kant 2012: 335–509. Anth-S2=Starke 2 notes from Kant's Anthropology lectures, in Kant 1831 (not included in Akademie edition); CF=*The Conflict of the Faculties*, trans. Mary J. Gregor and Robert Anchor, in Kant 1996b: 237–327; Corr=*Correspondence* (Kant 1999); CPR=*Critique of Practical Reason*, in Kant 1996a: 137–271; EAT=*The End of All Things*, in Kant 1996b: 217–31; Eth-C, Eth-H, Eth-V=Collins, Herder and Vigilantius notes from Kant's ethics lectures, in Kant 1997a: 37–222, 1–36, 249–452; G=*Groundwork of the Metaphysics of Morals*, in Kant 1996a: 41–108; Met-Her=Herder notes from Kant's metaphysics lectures in Kant 1968: 5–166. L-Th=*Lectures on the Philosophical Doctrine of Religion*, in Kant 1996b: 339–451; MM=*The Metaphysics of Morals*, in Kant 1996a: 363–602; MPT=*On the Miscarriage of All Philosophical Trials in Theodicy*, trans. George di Giovanni, in Kant 1996b: 19–38; OFBS=*Observations on the Feeling of the Beautiful and Sublime*, trans. Paul Guyer, in Kant 2007: 23–62; Rel=*Religion within the Boundaries of Mere Reason*, trans. George di

Giovanni, in Kant 1996b: 39–216; RS=Review of Schulz's *Attempt at an Introduction to a Doctrine of Morals for All Human Beings Regardless of Different Religions*, in Kant 1996a: 1–10.

3 See Clewis (2009: 114–16) for a discussion of disgust in the context of Kant's aesthetics.

4 Many Kant passages in this paper are drawn from lecture notes taken by students or professional note-takers. These notes must be considered less authoritative than Kant's published works. But the notes typically correspond closely to Kant's published remarks, and they often augment Kant's published remarks in ways that add depth and detail, so they are nonetheless crucial sources.

5 See Cohen (2009: 44–5) for a discussion of how embarrassment thwarts anthropological observation.

6 See Sussman (2008) for a discussion of Kant's account of shame.

7 See Timmermann (2014) for a helpful discussion of the perspectival structure of Kant's ethics.

8 Kant does not think that God is merely a projection of an aspect of ourselves or a construction of reason. He thinks that reason requires us to postulate that God exists in a quite traditional sense (*CPPrR*, 5: 124–32). In his theory of conscience, however, he holds that conscience involves representing an aspect of ourselves as God – he does not suppose that conscience gives us direct access to God.

9 See Kuehn (2001: 118–35) for an account of their relationship.

10 E.g. Muchnik (2014: 240), Sweet (2013: 99), Ware (2009: 684–90).

11 Kant's view that everyone deserves infinite punishment may imply that people who commit horrible crimes deserve no more punishment than people who do not, and this on its own could be seen as a fatal flaw in his retributivism. But I lack space to consider this problem here.

12 A concern about quantifying retribution may seem puzzling, because it may seem to be the sort of thing that cannot be measured with the kind of precision necessary for quantification. Kant has detailed views on quantifying retribution, however. He holds that we deserve (a) *infinite* self-inflicted retribution when our consciences impute wrongs, (b) retribution of a severity *equal* to the severity of crimes of which we are convicted in the context of public law, and (c) non-violent retribution (such as ostracism) which comes in *degrees* of severity when others impute wrongs to us in the context of private relations, in such a way that the degree is reduced when sensible incentives to act wrongly were forceful (Vilhauer 2024).

13 It is natural to wonder whether Kant thinks of conversion pain as partly analogous to the pain of the damned in Hell, both because of its role as punishment required by God and because of its connection to the *Höllenfahrt*. If Kant does have an analogy along these lines in mind, then we might hope that his views on pain in Hell could cast light on his remarks on conversion pain. But Kant's remarks on pain in Hell are as interpretatively challenging as his remarks on conversion pain. The most direct connection he draws between Hell and pain may be at *Eth-V*, 27: 691, where he says we picture Hell as a 'state containing nothing but evil and involving a total loss of consolation and the utmost pain', but he emphasizes that this picture derives from an 'unattainable ideal; a notion of the uttermost degree thinkable, that was already to hand before it took on symbolic form . . .'. This text would seem to give us dubious grounds for reading Kant as asserting a view about the nature of pain in Hell (but see Mahon [2015: 118] for an account that takes Kant to be asserting a view). Even if we suppose he is asserting a view, the notion of *utmost pain* is no clearer than the notion of *infinite pain* in the *Religion*. *The End of All Things* contrasts 'two systems pertaining to the future eternity': 'that of the unitists, awarding eternal blessedness to all human beings (after they have been purified by a longer or shorter penance)' and that of the 'dualists', 'which awards blessedness to some who have been elected, but eternal damnation to all the rest' (*EAT*, 8: 328–9; in this passage he does not consider a system in which all are awarded either eternal blessedness or damnation based on their free actions and maxim determinations rather than election). He says that the 'unitistic system appears to lull us too much into an indifferent sense of security', so 'it is wise to act as if another life . . . is unalterable' and thus from 'a practical point of view, the system to be assumed will have to be the dualistic one' (8: 330). But this sort of reason for believing that damnation is eternal would not seem to yield practical knowledge that damnation is eternal. Since an eternity in hell would seem to deny the damned the opportunity for endless progress toward complete conformity with the moral law essential in Kant's argument for immortality (see main text above), we have reason to doubt that Kant's considered view is that damnation is eternal.

14 While practical reason requires us to see the phenomenology of conversion as the manifestation of noumenally timeless action, transcendental idealism implies that this need not be in tension with the claim that the phenomenology of conversion is a process with a determinate temporal structure. (But see Loncar [2013: 360] for an argument which disputes the notion that conversion has 'empirical

manifestations' with a real temporal 'before and after', and Wood [2020: 110–114] for an argument that the conversion is not a datable event in time.)

15 A complete analysis of Kant's notion of punitive conversion pain would have to grapple with his apparent inclusion of all pain, including the pain of childbirth (*Rel*, 6: 74n). But it is in significant part because so much of the pain in this world is distributed entirely without regard to the virtue of the afflicted that Kant thinks reason requires us to postulate the existence of God and immortality as necessary for the highest good (see e.g. *CPrR* 5: 129). Kant's inclusion of all pain appears to be a regrettable effort to incorporate post-Edenic punishment in a way that tends to undermine his moral system. For this reason, it is of dubious significance and is not addressed in the main text. Pain deriving from resisting inclinations that dispose us to act immorally is obviously related to virtue (even if the relationship is complex) and is thus discussed in the main text.

16 Two caveats are needed here. First, part of the pain of resisting inclinations that dispose us to act immorally is plausibly the result of our free cultivation of immoral inclinations earlier in our lives. This part of the pain of resisting inclinations could plausibly be regarded as punitive. Second, Kant may also mean to refer to pain we deserve to suffer as a result of continuing to *fail* to resist inclinations eternally, even after the conversion, because he thinks conversion does not establish moral perfection, but only gives one hope that one has now found oneself 'upon the good (though narrow) path of constant progress from bad to better' (*Rel*, 6: 48). According to the interpretation presented here, the sort of pain Kant ought to demand in that context is remorse, and it would make sense to see this pain as punitive too. But the quotes just considered in the main text indicate that punitive pain can only be part of the pain involved in dealing with inclinations.

17 Wood also discusses the relationship between Kant's conception of conversion and the Pietist conception (2020: 110–4), but does not discuss the issues about remorse introduced in *The Conflict of the Faculties*.

18 Smelting as a metaphor for moral purification is also common in the Bible. See e.g. Isaiah 1: 25 (English Standard Version): 'I will turn my hand against you and will smelt away your dross as with lye and remove all your alloy.'

19 See e.g. the *Diagnostic and Statistical Manual of Mental Disorders* entry for major depressive disorder (American Psychiatric Association 2013: 160–4).

20 See Wood (2008: 206–23) for an argument that Kant's retributivism is not well supported by his basic moral theory and is in tension with some of his fundamental doctrines.

## References

- American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) Arlington, VA: American Psychiatric Association Publishing.
- Baum, Wilhelm (1996) 'Der Klagenfurter Herbert-Kreis zwischen Aufklärung und Romantik'. *Revue Internationale de Philosophie*, 50/197(3), 483–514.
- Clewis, Robert R. (2009) *The Kantian Sublime and the Revelation of Freedom*. Cambridge: Cambridge University Press.
- Cohen, Alix (2009) *Kant and the Human Sciences: Biology, Anthropology, and History*. New York: Palgrave Macmillan.
- Hill, Thomas E. (2002) *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Oxford University Press.
- Kant, Immanuel (1831) *Immanuel Kant's Anweisung zur Menschen- und Weltkenntniß. Nach dessen Vorlesungen im Winterhalbjahre von 1790/91*. Ed. Fr. Ch. Starke, Leipzig: Die Expedition des europäischen Aufsehers.
- (1968) *Kant's gesammelte Schriften, Bd. 28.1: Vorlesungen über Metaphysik*. Ed. Gerhard Lehmann, Berlin: Walter de Gruyter (Göttingen Academy of Sciences).
- (1996a) *Practical Philosophy*. Trans. and ed. Mary J. Gregor. Cambridge: Cambridge University Press.
- (1996b) *Religion and Rational Theology*. Trans. and ed. Allen W. Wood and George di Giovanni. Cambridge: Cambridge University Press.
- (1997a) *Lectures on Ethics*. Trans. Peter Heath, ed. Peter Heath and J. B. Schneewind. Cambridge: Cambridge University Press.

- (1997b) *Kant's gesammelte Schriften: Band 25, Vorlesungen über Anthropologie*. Ed. Reinhard Brandt and Werner Stark, Berlin: Walter de Gruyter (Berlin-Brandenburg Academy of Sciences).
- (1999) *Correspondence*. Trans. and ed. Arnulf Zweig, Cambridge: Cambridge University Press.
- (2007) *Anthropology, History and Education*. Ed. Günther Zöllner and Robert B. Loudon. Cambridge: Cambridge University Press.
- (2012) *Lectures on Anthropology*. Ed. Allen W. Wood and Robert B. Loudon. Cambridge: Cambridge University Press.
- Kuehn, Manfred (2001) *Kant: A Biography*. Cambridge: Cambridge University Press.
- Langton, Rae (1992) 'Duty and desolation'. *Philosophy*, 67(262), 481–505.
- Loncar, Samuel (2013) 'Converting the Kantian self: radical evil, agency, and conversion in Kant's *Religion within the Boundaries of Mere Reason*'. *Kant-Studien*, 104(3), 346–66.
- Mahon, James Edwin (2015) 'Kant, morality, and hell'. In Robert Arp and Ben McCraw (eds.), *The Concept of Hell* (New York: Palgrave-Macmillan), 113–26.
- Muchnik, Pablo (2014) 'The heart as locus of moral struggle in the Religion'. In Alix Cohen (ed.), *Kant on Emotion and Value* (New York: Palgrave Macmillan), 224–44.
- Sussman, David (2008) 'Shame and punishment in Kant's "Doctrine of Right"'. *The Philosophical Quarterly*, 58(231), 299–317.
- Sweet, Kristi E. (2013) *Kant on Practical Life: From Duty to History*. Cambridge: Cambridge University Press.
- Timmermann, Jens (2014) 'Kant and the second-person standpoint'. *Grazer Philosophische Studien*, 90, 131–47.
- Vilhauer, Benjamin (2022a) "'Reason's sympathy" and others' ends in Kant'. *European Journal of Philosophy*, 30(1), 96–112.
- (2022b) 'Kantian remorse with and without self-retribution'. *Kantian Review*, 27(3), 421–41.
- (2024) 'Five perspectives on holding wrongdoers responsible in Kant', *British Journal for the History of Philosophy*, 32(1), 100–25.
- Ware, Owen (2009) 'The duty of self-knowledge'. *Philosophy and Phenomenological Research*, 79(3), 671–98.
- Wood, Allen (2008) *Kantian Ethics*. New York: Cambridge University Press.
- (2020) *Kant and Religion*. New York: Cambridge University Press.