CrossMark

Kantian Review (2022), **27**, 421–441
doi:10.1017/S1369415422000140

KANTIAN REVIEW

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Kantian Remorse with and without Self-Retribution

Benjamin Vilhauer

City College and Graduate Center, City University of New York, USA
Email: bvilhauer@ccny.cuny.edu

Abstract

Kant's account of the pain of remorse involves a hybrid justification based on self-retribution, but constrained by forward-looking principles which say we must channel remorse into improvement and moderate its pain to avoid damaging our rational agency. Kant's corpus also offers material for a revisionist but textually grounded alternative account based on wrongdoers' sympathy for the pain they cause. This account is based on the value of care, and has forward-looking constraints much like Kant's own account. Drawing on Kant's texts and recent work in empirical psychology, I argue that sympathetic remorse may fulfil Kant's forward-looking goals better than self-retributive remorse.

Keywords: remorse; retributivism; conscience; care; sympathy; transcendental freedom

1. Introduction

We can react to the belief that we have acted wrongly with a variety of painful feelings, such as embarrassment, shame, remorse and guilt. However, embarrassment and shame are different from remorse and guilt because they can be responses to behaviour which is not immoral but merely prompts undue attention, or anxiety about mockery, such as spilling a plate of messy food on oneself at a conference. Kant comments on all these feelings, and his remarks reflect the distinction just mentioned.[1]

The difference between remorse and guilt is somewhat complex. 'Guilt' is used to refer both to painful moral feeling and also to a state of culpability which can be determined by God or a court, which may have painful feeling as a component, but need not. Courts can 'find' people guilty even if they do not *feel* guilty. The German word Kant uses which translates as 'guilt' is *Schuld*, which can also mean debt or obligation. Kant seems rarely to write about feeling guilty, though he does so in at least one place (*schuldig zu fühlen*; *Rel*, 6: 38).

The English word 'remorse' has a helpful simplicity in its exclusive reference to painful moral feeling, and I primarily rely on this term below. With a bit of work, 'remorse' can be directly matched up to Kant's German terminology. Kant uses multiple terms which translators render as remorse, but *Reue* and the related verb *bereuen* are by far the most common.[2] *Reue* has 'rue' as a close cognate in English, and both can

© The Author(s), 2022. Published by Cambridge University Press on behalf of Kantian Review.

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

mean painful regret for my actions either because they were immoral, or because they were imprudent and brought negative consequences upon me. Only the former meaning fits that of 'remorse'. However, Kant draws distinctions which mark out a kind of *Reue* which fits the former meaning, which he calls *moralische* (moral) *Reue* (Eth-C, 27: 353) and *wahre* (true) *Reue* (Eth-V, 27: 464). In the Collins ethics lecture notes, Kant 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 1792 letter to Maria von Herbert (Corr, 11: 333), he distinguishes *Reue* over 'imprudence' (*Unklugheit*) from *Reue* 'grounded in a purely moral judgment' (*auf bloßer sittlicher Beurtheilung … Verhaltens gründet*) about one's behaviour.[3] I will therefore use 'remorse' to translate Kant's *wahre, moralische Reue*.

The main questions I wish to address in this article are the following: what are our reasons for feeling remorse according to Kantian moral psychology, and how should we experience remorse based on those reasons?[4] I do not mean to claim that we typically deliberate about how we should experience such pain, on the basis of explicit justifications for it, or that Kant thinks we do. Kant holds that we have 'an instinct (*Instinkt*), an involuntary and irresistible drive in our nature, which compels us to judge with the force of law concerning our actions, in such a way that it conveys to us an inner pain at evil actions' (Eth-C, 27: 296–7). Kant does not explain the sense in which this is instinctive, but it seems right to think that remorse often has an immediacy which makes it prior to deliberation about reasons for feelings. However, the faculty that prompts remorse is conscience, and we have a duty to cultivate conscience, which entails an ability to rationally shape conscience (*MM,* 6: 401). We must be reflective 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).[5] We must '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). Shaping conscience in these ways requires judgement about *when* and *how* we should feel remorse, and this requires reflection on *why* we should feel remorse.

Suppose that we could alter our reactions to our own wrongs so that we felt no pain. Perhaps we could take pills serving the function of the 'moral sedative' Kant refuses to offer his conscience-stricken correspondent Maria von Herbert (whose story I will return to later).[6] Most philosophers would think we ought not take such pills. Why? There are at least three independent ways of justifying the pain of remorse. One is *retributive*: we should have painful feelings in response to past wrongful actions because we *deserve* them.[7] Retributive justifications are often called *backward-looking*, because they refer only to past wrongs.[8] Another justification is *forward-looking*: our wrongs should pain us because this motivates us to act better in the future.[9] A third justification is based on the value of *care*: we should feel pain because we should care about the people we wrong, and this requires sympathizing with the pain our wrongs cause.[10] The care justification does not rely on retributivism and it is not purely forward-looking, as I explain below. Kant does not speak in terms of the

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

value of care, but it is implicit in his theories of sympathy and friendship. I argue that Kant's account of our reasons for remorse is a hybrid of the retributive and forward-looking justifications. I go on to offer a revisionist but textually grounded Kantian care alternative.

2. Kant's account of reasons for remorse

The most direct evidence for a retributive component in remorse appears in Kant's discussion of our negative duty to promote others' 'moral well-being':

> [T]he *pain* one feels from the pangs of conscience (*Der Schmerz, den ein Mensch von Gewissensbissen fühlt*) 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[t] him (*ihn sein Gewissen nachher peinigen kann*)[.] (*MM*, 6: 394)[11]

This seems to imply the general view that to experience pain from the pangs of conscience – in other words, to suffer remorse – is to *deservedly* suffer inner reproach.[12]

Elsewhere Kant indicates that forward-looking considerations play a role in justifying remorse, because it motivates us to improve, by acting better in general and by making amends to the people we have wronged. In the *Metaphysics of Morals,* Kant states that *moralisch zu bereuen* requires a 'view to improving' (*MM*, 6: 485). In the *Religion*, he criticizes 'remorseful self-tortures[t] (*reuige Selbstpeinigungen*) that do not ... originate in any genuine disposition toward improvement' (*Rel*, 6: 77), and argues that at the end of life 'conscience ought rather to be *stirred up* and *sharpened*, in order that whatever good yet to be done, or whatever consequences of past evil still left to be undone (repaired for), will not be neglected' (6: 77n.). The *Anthropology* warns against regarding our 'record of guilt as ... simply wiped out (through remorse[t] [*Reue*]), so that [we are] spared the effort toward improvement' (*Anth*, 7: 236).

Kant sometimes suggests such a complete reliance on forward-looking reasons that he can seem to advocate a *purely* forward-looking account, such that the *only* reasons to feel remorse are forward-looking. In the Mrongovius anthropology lecture notes (1784–5), Kant critiques 'idle desires, *pia desideria*' connected with the 'wish that something would not have happened which, however, now is impossible', which is 'senseless and harmful' and leads to 'distraction'. 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, my emphasis). His point here 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 in the future. But retributive justifications of remorse essentially involve references to past wrongs, and so the Anth-Mr, 25: 1335, remarks suggest that there is no role for retributivism.

Some philosophers think of Kant as aiming to oppose consequentialist reasoning in all things, and may thus resist the notion that Kant endorses forward-looking reasons for remorse. But the notion of doing ethics without *any* forward-looking reasoning is

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

surely absurd – the key idea for Kant is that forward-looking reasoning is not always the right way to think about moral matters, and must always be regulated by duties to persons as ends which constrain forward-looking reasoning. Forward-looking reasoning is crucial for Kant in contexts as varied as punishment by the state (*MM*, 6: 331, 336; Eth-C, 27: 286), education (*CPrR*, 5: 152; Ped, 9: 451–2) and the prudent permissible pursuit of happiness (*G*, 4: 399; *CPrR*, 5: 25). So forward-looking reasons in the justification of remorse need not conflict with Kantian non-consequentialism.

On the other hand, a *purely* forward-looking justification of remorse conflicts with deep intuitions about the significance of painful moral emotions which it is natural for Kantians to wish to preserve. Suppose that I am imprisoned in solitary confinement for a series of assaults that left my victims disabled and in pain, and I am certain to die before I am released, and the conditions of my imprisonment mean I can do nothing to make amends to my victims or improve my behaviour toward people in general. I might conclude that I have no forward-looking reasons for remorse, and if these are the only reasons for remorse I endorse, it would be rational to make an effort to free myself from remorse altogether. Utilitarians need not object to this effort, but Kantians may regard it as trivialization of wrongdoing which privileges my happiness over my appreciation of the gravity of my wrongs.

There are, however, remarks which suggest Kant has a hybrid theory with both retributivist and forward-looking components. In Kant's critique of Johann Schulz's moral theory, he attributes various theses to Schulz which Kant does not accept. One 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). This suggests that Kant thinks remorse has another purpose in addition to improvement, and *MM*, 6: 394 (quoted above) suggests this is experiencing pain we deserve. This idea is also supported by Herder's metaphysics notes, which state that '[i]f remorse (*Reue*) about the past prevents all attention to the future, it is absurd' (Met-Her, 28: 90). This suggests that remorse is *not* absurd if it looks backward *and* forward. Another text supporting a hybrid theory appears in the Collins ethics lecture notes, which state that preachers attending the dying 'must ... see to it, that people do indeed feel remorse[t] 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's death-bed case partly overlaps with the solitary confinement case sketched above: he thinks the imminence of death means there is a forward-looking way to respond to some but not all of one's past bad actions. He states that we should feel remorse for the actions to which we *cannot* respond in a forward-looking way, and *MM*, 6: 394, makes it reasonable to assume we should feel such remorse simply because we deserve it.

Overall, the evidence surveyed means that we should attribute the following view to Kant. We must retributively inflict remorse on ourselves for past wrongs, under two forward-looking constraints: (1) remorse should be channelled into improved behaviour when possible, so that it prompts us to act more morally toward others in general and to make amends to the particular people we have wronged; (2) remorse must be moderated insofar as that is necessary to go on with our lives as effective moral agents.[13] We have seen this in Kant's warnings that remorse can lead to distraction, brooding which makes life useless, and even suicide.

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

3. Concerns about self-retribution

This section explains two concerns about the retributive component of Kant's account. The first concern is about the epistemology of transcendental freedom and its implications for retribution (Vilhauer 2017, forthcoming b). We have a strong intuition that justifications for retribution must meet the highest possible practical justificatory standard, since retribution is about the intentional infliction of harm which is purportedly deserved even if it has no forward-looking justification (Vilhauer 2015). This intuition is part of why so many endorse the view that arguments in the criminal court must be proven beyond reasonable doubt.[14] Kant himself addresses this intuition in a discussion of imputation (*Imputation*) of crimes in the Vigilantius ethics notes (Eth-V, 27: 558–73). 'Imputation' is Kant's term for assigning moral responsibility for actions, so imputation is a precondition for judgements about desert, and thus for retributive justification. He describes the justificatory standard we must meet in imputing crimes as 'the greatest possible[t] (*größtmöglichste*) moral and logical certainty (*Gewißheit*)', and states that it extends not only to questions of whether the deed to be imputed was actually done by the agent at issue (whether 'the man did it': Eth-V, 27: 567) and the nature of the 'motive to the action' (27: 559), but also that it is 'absolutely necessary in addition, that he act with freedom, indeed it is only when considered as a free being that he can be accountable' (27: 559). This standard of *greatest possible logical and moral certainty* is relevant not only for the courts constructed in our legal institutions, but also for conscience, because Kant thinks of conscience itself as a kind of court which adjudicates 'the internal *imputation* of a *deed*' (*MM*, 6: 438).

There is reason to doubt that this highest possible justificatory standard can really be met in Kantian ethics. While Kant does not specify at Eth-V, 27: 559 that the certainty we require about agents' freedom is certainty about transcendental freedom, I take that to be the default interpretation. I take it to be Kant's view that we can only deserve to suffer if we have the radical independence from natural causation which transcendental freedom affords. I think he is right to think this. But the first *Critique*'s argument that we cannot have theoretical knowledge of transcendental freedom entails that we cannot meet this standard through theoretical reasoning. Kant of course advocates a practical epistemology of transcendental freedom in the second *Critique* and afterwards which he thinks delivers practical knowledge that we are transcendentally free, and he seems to think we meet the standard of certainty for retribution in this way. In his practical epistemology, he appeals to the 'ought implies can' principle to argue from the claim that we know we *ought* to act in certain ways to the claim that we know we *can* act in those ways, which he claims to entail practical certainty that we are transcendentally free (see e.g. *CPrR*, 5: 30). Kant may mean to assert this supposed knowledge as an ungrounded or self-grounding 'fact of pure reason' (5: 31). He may also think it has a phenomenological grounding in moral feeling. Kant thinks we should represent respect for moral law as determined in us through moral law with a self-wrought spontaneity fundamentally distinct from causation according to natural law. But he acknowledges that 'this determination has exactly the same inward effect, that of an impulse to activity', as sensible incentives which have no such special origin (5: 116). Paula Satne (2021) argues that *CPrR*, 5: 98–9 shows that Kant includes *Reue* itself as part of a phenomenological grounding for practical

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

knowledge of transcendental freedom.[15] But the first *Critique* is clear that we must regard all phenomena, including all feelings, as deterministically caused by the past and empirical laws, so it is not clear how the phenomenology can play a grounding role. It seems fair to claim that Kant's theoretical argument against knowledge that we are transcendentally free makes it prudent to be cautious about being easily persuaded by his practical argument in favour of such knowledge. The fact that retributivism relies crucially on transcendental freedom, and that justifications for retribution demand the *highest possible* justificatory standard, imply that if we have doubts about Kant's practical epistemology in *any* context, we should take those doubts most seriously in the context of justifications for retribution. I think that Kant's practical epistemology, in combination with his argument that it is possible that we are transcendentally free in the first *Critique*, give us room for a postulate that we are transcendentally free.[16] But to postulate is not to know.

It may be objected that the entire edifice of Kantian ethics collapses if we put practical knowledge of transcendental freedom in doubt in the context of retribution, but I think this is a mistake. A postulate of transcendental freedom is sufficient for regarding ourselves '*under the idea of freedom*' and bound by moral law, and Kant's argument that we must regard ourselves in this way when we *deliberate about how to act* in *Groundwork* III has merit, because it is plausible that deliberation requires us to postulate that we have the kind of control afforded by transcendental freedom over the alternative courses of action among which we deliberate.[17] But when we make *judgements about whether to impute actions to human beings*, ourselves and others, there are both prospective and retrospective elements to consider: we deliberate about alternative ways we can act with respect to a completed action which now presents itself as an object for judgement. There is a kind of flexibility available to us when we consider completed actions under the idea of freedom, and justice demands that we take it seriously. Kant holds that 'a person is a subject whose actions can be imputed to him' (*MM*, 6: 223), but his doctrine of degrees of imputation (*MM*, 6: 228; Eth-C, 27: 291; Eth-V, 27: 567) implies that we can assign responsibility in different degrees in different cases.[18] Herder's metaphysics notes suggest that the appropriate degree of imputation is sometimes 'vanishingly small' (Met-Her, 28: 41). I think these points imply that we can allow the (as it were) local diminution of degrees of imputation in cases where we confront especially high justificatory standards for imputation, and when we confront the highest possible justificatory standard, as we do in justifications for retribution, we should diminish the degree of imputation accordingly, so that the role which is played by retribution in justifying suffering diminishes in a corresponding way. We can still impute actions to agents, as wrongs rather than mere effects of things, but we can do so in a degree appropriate to the requisite standard of justification. Kant himself does not draw on his doctrine of degrees of imputation in this way. He thinks, for example, that the state must punish retributively according to the *lex talionis*, which he thinks entails a 'principle of equality' commanding execution of murderers and enslavement of thieves (*MM*, 6: 333). But Kantians can adopt this approach without fear of undermining Kantian ethics as a whole.

The second concern about the retributive component of Kant's account is practical rather than epistemological. Self-retribution appears to be responsible for many of the practical hazards of remorse. As discussed above, we are meant to channel remorse into improvement. Distraction, brooding which makes life useless, and

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

suicide are all obviously things which obstruct improvement, and they can all follow naturally from the belief that we deserve to suffer. If inflicting suffering on ourselves because we deserve it has a value independent from the value of treating people better in general and making amends, then it makes sense to attend to inflicting that suffering in a way that is independent of the attention we invest in acting better, and given the finitude of attention, this inevitably distracts us from acting better. Protracted distraction results in brooding. Further, if one 'connects the transgression or violation of his conscience with the idea of losing his entire moral worth', as Kant claims (Eth-V, 27: 575), there is a kind of tragic rationality in thinking that blotting oneself out through suicide is appropriate self-retribution, though this violates duty and permanently forecloses the possibility of improvement. This tragic rationality may explain Kant's distressing remark that suicide from excess remorse 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' (27: 642).

The idea that excess remorse can lead to suicide is borne out in contemporary clinical psychology. The most widely referenced text in clinical psychology, the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), uses the term 'guilt' to refer just to feeling, independent of the associations with the legal culpability one may have even if one does not *feel* guilty as noted earlier, and it thus uses 'guilt' in a way I take to be coreferential with 'remorse' as used here. It lists '[f]eelings of worthlessness or excessive or inappropriate guilt' among the diagnostic criteria for major depressive disorder, in which 'suicidal ideation' and 'suicide attempts' are common (American Psychological Association 2013: 160–4). The degree to which the pain at issue in the *DSM* is motivated by the desire for self-retribution is an empirical question with which psychology is still grappling, but there is good reason to think that many psychologists have seen it this way, given the influence of the Freudian idea that guilt is essentially a matter of exacting vengeance upon ourselves.[19]

We may also have an example of suicide motivated by self-retribution in Kant's correspondent Maria von Herbert. In a 1791 letter to Kant, she writes that a man she loved had fallen out of love with her when she revealed a protracted but harmless lie, apparently connected with the fact that she had had a previous relationship. She asks Kant for 'solace, or for counsel to prepare [her] for death', proceeding to make it clear that she meant she was contemplating suicide (Corr, 11: 273–4). Kant writes back in 1792 encouraging her not to kill herself, counselling 'composure', and remarking 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' (11: 334). But he refuses to provide a 'moral sedative': he tells her that even a harmless lie is 'a serious violation of duty to oneself and one for which there can be no remission', and (as mentioned earlier) instructs that her 'bitter self-reproach' for her lie should not be *Reue* over 'imprudence' (*Unklugheit*) but *Reue* 'grounded in a purely moral judgment' (*auf bloßer sittlicher Beurtheilung … Verhaltens gründet*) of her behaviour (11: 331–3). He notes that 'self-torture' (*Selbstpeinigung*) is not 'deserved' (*verdienstlicher*) if one is 'sure of having reformed', but von Herbert clearly had read Kant's work in enough detail to be familiar with his denial of knowledge about the purity of our dispositions (*G*, 4: 407; *MM*, 6: 392–3). She committed suicide in 1803. It is

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

impossible to know how great a role self-retribution played in her death, but the texts make it reasonable to worry that it played some role.

4. A sympathy-based Kantian account of reasons for remorse

The non-retributive proposal turns on the idea that we ought to sympathize with the pain our wrongs cause. First, I explain how it works in a general way, and then how it can be grounded in Kant's texts.

It is in the nature of care that, when I care about someone, I sympathize with her joy and also her pain. When I care about someone (for example, because I have befriended her) I sympathize with her pain *not* because by establishing a connection of care I have acted in a way which makes me *deserve* to suffer when she does, but because sympathy is *part of caring*. It would be absurd to suppose that I deserve to suffer because of my actions in making a friend. Instead, sympathetic suffering is part of the nature of friendship because it is part of the nature of care.[20]

When I have *wronged* someone I care about, and caused them pain, I may well believe that I deserve to suffer, and it may seem to be a matter of moral common sense that I *ought* to believe that I deserve to suffer. But even if I am sceptical about the notion that anyone can deserve to suffer (perhaps because the first *Critique* places it in doubt, and perhaps for independent moral reasons) the fact that someone I care about is in pain gives me a reason to be pained. Sympathetic pain gives us a reason to remove the cause of the other's pain, and when the cause is *our own actions*, it gives us a reason to be *pained by those actions*, and to improve, both by acting better in general, and by making amends.

Grounding the value of sympathy in the value of care means this is not just a hedonistic calculus: if sympathy is a manifestation of care, then guidance by sympathy is guidance by care. This grounding also steers us away from the utilitarian thought that we should sympathize equally with everyone, so that we are motivated to maximize overall happiness. Once we dispense with the utilitarian construal of care, it is intuitive to think that the value of care gives us reasons to care about everyone to some degree, but to care in a focused and heightened way about some particular others, such as our friends.

The key idea in the sympathetic justification of remorse is that we should have such particularized care for the people we have wronged – that wronging people gives us a reason to care about them which is virtuous in a way that parallels the virtuousness of making friends, and which is independent of reasons of desert. Perfectly virtuous agents like the Kantian sage would never wrong people in the first place, but a theory of remorse is necessarily a theory that applies to imperfectly virtuous agents. The idea is that in wronging another, the wrongdoer establishes a particularized moral connection with the person wronged which demands care in a way which is not grounded on desert.

Human nature as it is empirically given to us is such that when someone hurts us in a way that violates morality, we have a desire for the wrongdoer not only to make amends, but also to understand what he has done in a way that is not just cognitive but also involves painful emotions. This desire is often strong enough to constitute the kind of disposition Kant calls a *need* (*Bedürfniß*). Some may wish to model such needs in terms of Strawsonian reactive attitudes, which can be understood as

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

essentially involving desires for the wrongdoer to experience deserved suffering. But it artificially circumscribes such needs to assume that they are always retributive. Wrongdoers' sympathetic pain sometimes satisfies such needs, and since sympathetic pain does not have to be understood in terms of deserved suffering, such needs do not always have to be understood as retributive.

While remorse based on sympathy gives us forward-looking reasons, the sympathetic justification cannot be *reduced* to a forward-looking justification, because care is not valuable just because of its consequences. Care as understood here entails sympathizing with suffering even when there is nothing we can do to help. If I am trapped on a desert island and receive a message in a bottle informing me that someone I claim to care about is suffering, and I do not suffer sympathetically just because I cannot help, this is a strong indication that my claim to care is false. The same thing holds when we care about people we have wronged – sympathy motivates us to make amends if we can, but if we cannot, we still sympathize, because we care.

Kant does not talk about the value of care in terms that lend themselves to easy linkage with contemporary care ethics.[21] However, his accounts of sympathy and friendship have some parallel implications. We can draw ideas from these accounts to develop a Kantian account of sympathetic remorse which is revisionist but textually grounded. Kant's best-known remarks on sympathy appear in the *Groundwork*, and can appear to reject any role for sympathy in his moral psychology (*G*, 4: 398–9). He says that while sympathy is 'amiable' and a disposition to be encouraged, it is 'on the same footing with other inclinations', and we can have a 'far higher worth than what a mere good-natured temperament' confers 'even if we are cold and indifferent to the sufferings of others' (4: 398). But in Kant's more detailed account of sympathy in the *Metaphysics of Morals,* he explains that 'Sympathetic Feeling is Generally a Duty' (*Theilnehmende Empfindung ist uberhaupt Pflicht*) (*MM*, 6: 456), and that 'it is a duty (*Pflicht*) to actively sympathize (*thätige Theilnehmung*) in [others'] fate' (6: 457). If there is a duty to sympathize, then sympathy cannot be merely a matter of inclination – it must count as a moral feeling and must be capable of guidance by practical reason. The apparent conflict between the *Groundwork* and the *Metaphysics of Morals* arises from the *Groundwork*'s elision of a distinction between two ways of sympathizing which can involve qualitatively identical sympathetic joys and pains but differ in their relation to practical rationality. One is a passive, inclination-driven sympathy, which I refer to as 'natural sympathy', and the other is an active (*thätige*) sympathy guided by practical reason, which I refer to as 'rational sympathy'. Kant draws this distinction in at least five places over a period of at least twenty years. (The terms for rational sympathy are labelled '(a)', and the terms for natural sympathy '(b)'.) *MM*, 6: 456, distinguishes (a) *humanitas practica*, the '*capacity* and the *will* to *share in others' feelings*', which is '*free*', and based on 'practical reason', and (b) *humanitas aesthetica*, 'the receptivity, given by nature itself, to the feeling of joy and sadness in common with others', which 'can be called *communicable* … like receptivity to warmth or contagious (*ansteckender*) diseases … since it spreads naturally'. Anth-F, 25: 607–11 distinguishes (a) 'reason's sympathy' and (b) 'physical sympathy'. *Anth*, 7: 235, and Anth-Mr, 25: 1320–1, distinguish (a) 'sensitivity' (*Empfindsamkeit*) and (b) 'sentimentality' (*Empfindelei*). Sensitivity 'possesses choice' and 'permits or prevents both the state of pleasure as well as displeasure from entering the mind' in a way that allows us to 'judge [others'] sensation', while sentimentality 'is a weakness by which we can

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

be affected, even against our will, by sympathy (*Theilnehmung*) for others' condition who ... play at will on the organ of the sentimentalist' (*Anth*, 7: 235). Eth-V, 27: 677–8, distinguishes (a) 'moral' sympathy and (b) 'instinctual' sympathy. These discussions all support a distinction between (a) sympathy which is voluntary and guided by reason, and (b) sympathy which is passive and irrational. In (a) the feelings are regulated so that they do not rise to 'affects' – feelings which may interfere with self-governance, prevent us from helping effectively even when we discover means to do so (Anth-F, 25: 589; *MM*, 6: 407) and dispose us to violations of the law when they conflict with duty (*Rel*, 6: 30 and Anth-F, 25: 611). In (b), agents allow feelings to flow passively, without exercising the discipline necessary for (a)-type sympathy.[22]

Sympathy is an activity of the imagination (*MM*, 6: 321n., 457; *Anth*, 7: 179, 238), and the difference between rational and natural sympathy is a difference between active and passive ways in which this activity can proceed. Imagination is a fundamental power in Kant's theory of mind, one of two 'parts' of sensibility, the other of which is 'sense' (*Anth*, 7: 153). Sympathy is best understood as an activity of the *a posteriori* productive imagination, which can function both involuntarily and voluntarily (*Anth*, 7: 174; Anth-Mr, 25: 1257). Kant calls involuntary productive imagination 'fantasy' (*Phantasie*) (*Anth*, 7: 167, 175),[23] and makes an explicit connection between *Phantasie* and *Empfindelei* (sentimentality) at *CPJ*, 5: 273, which I argued above is a term for natural sympathy. Kant contrasts fantasy with a voluntary, rationally ordered counterpart called 'disciplined fantasy' (Met-Mr, 29: 885), and rational sympathy is well-understood as a kind of disciplined fantasy.

Kant says that the sympathetic imagination puts us 'in the other's place' (Anth-F, 25: 575).[24] The voluntary exercise of this capacity, which we might call *projective imagination*, is a *skill* which enables rational sympathy:

> [T]he power to transpose the I is necessary, and to put oneself in the point of view and place of the other, so that one thinks with him, and has sympathy with him[t] (*sich in ihm fühlt*) ... To take a point of view is a skill (*Geschicklichkeit*) which one can acquire by practice (*sich durch Uebung erwerben kann*). (Anth-F, 25: 475, also see 25: 606-7)

Kant thinks the sympathy this enables is very vivid: 'we really feel ourselves to be in his place' (Eth-H, 27: 58), and '[w]e are sensible of this sympathizing feeling in our entire soul' (Anth-F, 25: 606). Natural sympathy happens when this occurs involuntarily. We draw on the skill of projective imagination when we sympathize not only with actual others, but also with possible others. In the Friedländer anthropology notes, Kant says that '[w]hen we read something, a history or a novel, we always put ourselves in the other's place and this is sympathy[t] (*Theilnehmung*)' (25: 476).

Kant's distinction between rational sympathy, on the one hand, and episodes of natural sympathy which prompt agency-disrupting affect, on the other, corresponds closely (and is plausibly identical) to a distinction drawn in contemporary empirical psychology between *empathic concern* and *empathic distress* (Tangney 1991: 599). *Empathic concern* involves 'feelings of compassion and warmth felt for the target of empathy' (Hodges and Biswas-Diener 2007: 390). It is an 'intentional capacity' which involves 'emotion-regulation'– it 'involves an explicit representation of the subjectivity of the other' rather than 'a simple resonance of affect between the self and other'

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

(Decety et al. 2007: 254). *Empathic distress*, by contrast, is a feeling which Decety et al. (ibid.) call 'emotional contagion'. Hodges and Biswas-Diener (2007: 402) say that it 'occurs when people fail to rein in emotional empathy', and note that '[t]he quintessential example of this phenomenon is the bystander who witnesses a gruesome accident and can only stand by, gasping and shrieking, rather than comforting the victim or going for help'. Psychologists think that it is the development of regulatory processes which allows us to feel empathic concern rather than empathic distress. It appears that some of this regulation is unconscious, but there is evidence that conscious perspective-taking plays a role in this regulation too. To 'imagine things from the empathy target's point of view consistently increases empathic concern' (Hodges and Biswas-Diener 2007: 393; also see Batson et al. 1997). On the other hand, imagining things from the other's perspective *too vividly* can prompt empathic distress, and we can modify *how* we frame our engagement with the other's position to moderate our empathic feelings (Hodges and Biswas-Diener 2007: 393).

Universalizing our maxims might seem to require us to sympathize equally with everyone, and this might prohibit us from cultivating especially strong sympathy for particular others. But Kant does not advocate this. We have a duty of friendship (*MM*, 6: 469), and while we ought to have 'general good will toward everyone', 'to be everybody's friend will not do, for he who is a friend to all has no particular friend; but friendship is a particular bond' (Eth-C, 27: 430).[25] Friendship is an 'ideal of each sympathizing and communicating[t] (*Ideal der Theilnehmung und Mittheilung*) about the other's wellbeing' which guides us toward a 'maximum' (*MM*, 6: 469) in which 'each mutually sympathizes[t] (*teilnehmen*) with every situation of the other, as if it were encountered by himself' (Eth-V, 27: 677). This ideal gives us reasons to establish strong particularized sympathetic connections with our friends.

My claim is that we should also establish such connections with people we have wronged. Kant himself nearly suggests this in a discussion of sympathy and the 'oppression' of people 'subordinate to the aristocracy' (Anth-F, 25: 606). He says that 'a humble person can easily put himself in the position of the higher one and assume greater dispositions. However, the distinguished one cannot assume the state of the humble one, hence he also does not sympathize (*sympathesirt*) with his misfortune' (25: 607). 'If the ills are natural, for example, famine, then the distinguished person sympathizes with the humble one just as well as the latter with him, but in the case of ... ideal ills, the distinguished one does not sympathize (*sympathesirt*) with the humble one, but the latter does in fact sympathize (*sympathesirt*) with the former' (25: 606–7). The distinguished one 'thinks that the one who is thus not accustomed to the refined life is indeed just a humble man, hence he always gets on [in life], if he can just live', and does 'not become as aware' of the 'distance' of the humble man's 'social standing from the civic one in general' (25: 607). Kant says that while a commoner 'has compassion (*Mitleiden*) for an unfortunate king', the 'unfortunate thing with kings' is that they 'have no inclination' to 'imagine the misfortune of their subjects' (25: 607). Kant's implicit point here is that when the 'distinguished' sympathize naturally, their inclinations may dispose them to imagine what it is like for the 'humble' to be hungry or in pain, but not to imagine their 'ideal' misfortunes – in particular, they do not imagine that the 'humble' have ideas of happiness which include more than just living, and are pained by the way their social standing makes it hard to do more than just live – and that the 'distinguished'

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

should resist their inclinations and sympathize rationally, putting themselves in the place of the 'humble' more accurately, in a way that brings them a greater range of sympathetic feelings, including sympathetic pain. An intuitive next step in this line of thought would be for the 'distinguished' to note that, since it is *their own oppressive behaviour* which is the cause of what is wrong in the lives of the 'humble', they can alleviate that pain by making amends to particular people they have wronged, and improve their behaviour generally so as to not contribute to future oppression. This next step would establish the basic operations of sympathetic Kantian remorse.

I argued earlier that suffering in sympathy with people we care about is valuable even when we cannot help, and this means that a sympathy-based Kantian theory of remorse, like Kant's own theory, is not purely forward-looking. As explained earlier, this is crucial because a *purely* forward-looking conception of remorse can strike Kantian sensibilities as trivializing grave wrongs. I think Kant himself is committed to the view that sympathetic suffering is valuable even apart from its good consequences, but there is a textual challenge to this claim which must be addressed. The apparent endorsement of cold indifference we saw in the *Groundwork* recurs within Kant's theory of friendship, and appears to undercut the claim that sympathetic suffering is valuable when it has no good consequences. Kant writes that when the sage 'could not rescue his friend, [he] said to himself "what is it to me?" In other words, he rejected compassion (*Mitleidenschaft*)' (*MM*, 6: 457). He continues in a way that seems to endorse the attitude of the sage:

> In fact, when another suffers and, although I cannot help him, I let myself be infected (*anstecken*) by his pain (through my imagination), then two of us suffer, though the trouble really (in nature) affects only *one.* But there cannot possibly be a duty to increase the ills in the world and so to do good *from compassion* (*Mitleid*). (*MM*, 6: 457)

On initial inspection, it may be natural to read Kant as arguing quite precisely that there is no reason for painful sympathy when we lack forward-looking reasons. But there is good reason to think that what Kant is criticizing here is painful *natural* sympathy, because in objecting to letting oneself be 'infected' (*anstecken*) by another's pain, he uses the same terminology he uses to describe *humanitas aesthetica* just two paragraphs earlier (in a passage mentioned above), and *humanitas aesthetica* is a term for natural sympathy. This makes it reasonable to think that Kant is *not* claiming that rational sympathy is only valuable when it has good consequences.[26]

Further, fundamental features of Kant's intentional teleology commit him to the view that sympathetic suffering is valuable even when it has no good consequences, because sympathy is necessary to fulfil the imperfect duty to make others' permissible ends 'as far as possible ... also *my* ends' (*G*, 4: 430). The argument for reading Kant this way turns on a distinction between *adopting* and *promoting* others' ends.[27] Many of others' permissible ends are subjective ends, that is, ends which they have only because of features of their individual feelings which are contingent from the perspective of rational agency (*G*, 4: 427). Rational sympathy allows me to project myself into others' perspectives and conform contingent features of my own sensibility to theirs, and this disposes me to be sympathetically pleased when they achieve their

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

ends and pained when they do not, and thus disposes me to work toward the achievement of their ends out of motivation from a contingent feeling-basis like their own. This allows me not only to *promote* but also *adopt* their ends. I can *promote* others' ends without sympathizing if I do so as *means to distinct ends* – I may behave in ways which help others achieve their ends even if I am in pursuit of ends which are not their ends. If my friend wants to alleviate his pangs of hunger, I may give him food because the sounds he makes in eating produce an autonomous sensory meridian response in me, or because I desire to improve my reputation, or because I have a rational desire to fulfil my duty of beneficence. While I promote his end in all these ways, and there may be no difference at all in the consequences I produce, I do so as means to ends which are not his end (see *MM*, 6: 388, for an argument which offers support for this claim). To dispense with sympathetic pain for others who are in pain is to dispense with adoption of their ends, and thereby to fail to take their ends as my own in an important way.

This shows that Kant's moral psychology offers materials for a sympathy-based account of remorse which does not rely on desert, and is therefore not vulnerable to the epistemological problems confronting retribution. This account's grounding in the value of care arguably allows it to satisfy Kant's forward-looking requirements for remorse better than his own account. As explained above, the first forward-looking requirement is that remorse should be channelled into improved behaviour. It is *prima facie* plausible to suppose that sympathetic remorse would fulfil this goal better than self-retributive remorse. If I am hurting because I believe that someone I care about is in pain, it is clear what I must do: I must help her. There is also empirical evidence that rationally sympathetic remorse would prompt such behaviour: as noted earlier, rational sympathy is similar (and plausibly identical) to empathic concern in contemporary empirical psychology, and empathic concern is associated with 'altruistic helping behavior' toward the people with whom we empathize (Tangney 1991: 599; also see Hodges and Biswas-Diener 2007: 402). If I am hurting because I believe I deserve to suffer, there is more conceptual and psychological mediation required to arrive at the motivation to help.

Kant's second forward-looking requirement is that remorse must be moderated to avoid distraction, brooding and suicide, so that we remain effective moral agents. Earlier we noted reasons to think that self-retributive remorse may pose special hazards here, and it is intuitive to think that rationally sympathetic remorse would pose fewer hazards. This claim also finds support in contemporary empirical psychology. Earlier we saw that excess guilt (arguably construed on a self-retributive model) is a diagnostic criterion for depression, and that depression can prompt suicide. Empathic feelings are also common in depression, but the distinction between empathic *concern* and empathic *distress* (which is similar and plausibly identical to the distinction between rational and natural sympathy) is crucial in understanding the relationship between empathy and depression. Ghorbani et al. (2003: 438) give evidence that, while empathic *distress* is *positively* correlated with depression, empathic *concern* is *negatively* correlated with depression. O'Connor et al. (2007: 49) explain that '[the] empathic reaction in depressives often leads to great distress because they tend to unrealistically blame themselves for pain felt by others', and use an explicitly retributive model of self-blame, describing 'self punishment' in depression as 'meted out ... while thinking "I deserve this"' (p. 67). They argue that it is just

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

this unwarranted self-retribution which 'transform[s] empathic concern into empathic distress' in depression, and therefore advocate depression therapy which targets unwarranted self-retribution (p. 70). According to the view I advance, we ought to resist self-retribution, not only because of the damage it does, but also because we cannot be confident enough about transcendental freedom to be confident that self-retribution is warranted. This empirical work suggests that agents who succeed in resisting self-retribution can empathize without the threat of depression and the damage it does to rational agency.

5. Potential objections and replies

This section addresses potential objections: first, an objection about perfect and imperfect duties; second, an objection about wrongs which do not cause pain; third, an objection about duties to ourselves; and fourth, an objection about wrongdoers' sympathy with the desires of the wronged for wrongdoers to feel self-retributive remorse.

First, it may be objected that grounding remorse in the duties of sympathy and friendship associates remorse with *imperfect* duties which grant us *latitude* (see e.g. *MM*, 6: 392, 411) and are thus not suited to guide conscience. The worry is that, while we think sympathetic suffering reflects well upon someone's moral character, we may not think its absence in particular cases is a flaw, while remorse is something we *expect* of people who have wronged someone. As Kant puts it, 'I approve of a pain of compassion (*Mitleids*), but demand a pain of remorse (*Reue*)' (Refl 6848, 19: 178). There are puzzles about conscience and the distinction between perfect and imperfect duty which I cannot address here, but Kant is clear that human beings should cultivate conscience such that it '[holds their] duty before [them] for [their] acquittal or condemnation in every case that comes under a law' (*MM*, 6:400). Beyond avoidance of micrological conscience, we ought not pick and choose what we are conscientious about. So it is important to emphasize that the sympathetic remorse proposed here is not meant to have the latitude of the general duty of sympathy. The proposal is that conscience can respond to *all* our wrongs with sympathetic remorse. Since sympathetic remorse involves rational sympathy, it is guided by practical reason just as Kant's self-retributive remorse is, and is thus governed by the same criteria for identifying wrongs.

The second objection to be considered is that some wrongs cause no pain, and there is nothing to prompt sympathetic remorse in such cases. Consider a case in which my friend requests that I kill him because of a painful terminal illness that will afflict him for many more years if he waits to die naturally. Suppose I establish with certainty that the course of his illness is as he says, and his request is the result of thorough and stable reflection, and I kill him. According to the letter of Kantian principles, I have done a grave wrong in acting in a way that subordinates his dignity as a rational agent to his pain. But my action has ended his pain, so there is no actual pain with which to sympathize. It may thus appear that there is no reason for remorse on the model proposed here. Such cases are of course challenging not only for the account of sympathetic remorse presented here, but also for Kantian ethics in general, as many who endorse the conception of Kantian rational dignity in most circumstances think it can sometimes be outweighed by profound suffering. So we might

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

respond to this objection by adopting a moral theory which is Kantian in many respects but holds that we need not feel remorse for such a killing because it is not wrong.

Kant's texts offer material for a more orthodox response, however. As explained earlier, Kantian sympathy is a function of the imagination, and we can imaginatively transpose ourselves into the position of both *actual* and *possible* others. Further, a passage in the Friedländer Anthropology notes mentioned above indicates that rational sympathy sometimes requires us to sympathize with *possible versions of actual persons*, versions who *are* as the actual people *would* be if they had vivid feelings of their dignity as rational agents. The passage reads as follows:

> [I]f people ... subordinate to the aristocracy ... are constantly under oppression, then they lose the idea of the right of humanity, for since they have no examples where justice prevails, then they think it must be so. There we must sympathize with the other's right, but not with the physical ill[.] (Anth-F, 25: 606)

Suppose I oppress someone who has been oppressed for so long that he has become inured to it.[28] Where is the pain with which I should sympathize? Kant's idea is that I must imagine myself into a version of the other's position adjusted in light of how he *would* feel if he had not lost the idea of the right of humanity. He would be pained by the loss of the valuable experiences he could have as a member of a society free from oppression, and by the way my oppressive actions contributed to that loss.

This strategy of sympathy with normatively adjusted possible versions of others can be generalized to all wrongs which do not actually cause pain.[29] This should not seem *ad hoc*, since it is essential to Kant's theory of rational sympathy that it is imaginative activity regulated by the moral law and the concept of rational agency at its basis. Consider euthanasia again. An agent with a vivid sense of Kantian dignity would not wish to be killed, no matter how intense and protracted his pain, and would experience a kind of sublime joy in contemplating victory over his desire to die, motivated by his sense of dignity as a rational agent. Though my friend does not actually feel this joy, it is the loss of this possible joy which provides the feeling-basis for my sympathetic remorse if I kill him. If I fail in an *attempt* to kill him, I can sympathize with the possible joy he would have lost had I been successful. Such sympathetic imaginings quickly become modally complex, but it seems reasonable to assume that the Kantian imagination has the capacity to handle such complexity.

The third objection is that sympathy cannot be the basis of remorse for violating duties to ourselves, because sympathy is something we feel for others. But we can extrapolate from Kant's idea that we project ourselves into possible versions of others and suppose that we can also project into possible versions of ourselves. Imagine that Maria fails in her suicide attempt, and feels nothing but frustration at her failure. Where is the basis for remorse over her failed attempt? As in the euthanasia case, she can sympathize with the feelings she would have if she had a vivid appreciation of her dignity as a rational agent, and the sublime joy she could find in persisting despite her sorrow, and the loss of that possible joy she would have caused herself. Caring for this version of herself seems more likely to help her persist than a renewed infliction of the self-retribution Kant's own account of remorse prescribes.

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

The fourth objection is that part of the pain felt by others I have wronged may be bound up with the desire that I exact self-retributive remorse upon myself. I argued earlier that the need of the wronged for the wrongdoer to have a painful emotional experience of the wrong is not always well-understood as retributive. But what about cases where it is? In such cases, sympathetically putting myself in others' places *too* completely might yield vicarious self-retribution with consequences like those of exacting self-retributive remorse upon myself. But here too we can project ourselves into the perspectives others *ought* to take: if others *ought* not to wish self-retribution upon us, then we should sympathize with the feelings they would have if they *did* not. The reasons that we should experience sympathetic rather than self-retributive remorse ourselves imply a similar attitude toward others: we ought to demand sympathetic rather than self-retributive remorse of others.

While this idea is revisionist relative to Kant's own ethics, like all the claims about sympathetic remorse advanced here, it prompts useful questions about Kant's claim that 'reconciliation[t] (*Versöhnlichkeit*) is a duty of human beings' (*MM*, 6: 461) and points us toward an attractive revisionist but textually grounded interpretation of it. Kant introduces the duty of reconciliation in the context of a critique of *Schadenfreude* and the desire for revenge (*Rachbegierde*) (*MM*, 6: 459–61).[30] *Schadenfreude* is also called 'antipathy' (*Antipathie*) (Anth-F, 25: 607-11). It is an activity of the imagination which inverts sympathetic participation such that 'one suffers pain because the other rejoices' and 'one has a sensation of joy because the other has pain' (Anth-F, 25: 607), and is therefore 'directly opposed to one's duty in accordance with the principle of sympathy' (*MM*, 6: 460). Kant says *Schadenfreude* in its most extreme form is 'an ideal, or a maximum of moral evil' (Anth-F, 25: 608). He describes *Rachbegierde* as '[t]he sweetest form' of *Schadenfreude* (*MM*, 6: 460). His argument that we have a duty of reconciliation appears to go like this: it is a duty of virtue not to be vengeful, and we need (*bedürfen*) forgiveness (*Verzeihung*), therefore we have a duty of reconciliation (6: 460–1).

This brief argument does not make clear exactly what the duty of reconciliation demands of us. But Kant's point that it requires avoiding the vicious inversion of sympathy in *Schadenfreude* implies that it requires regulation of our feelings about wrongdoers' feelings. It may be sufficient to avoid *Schadenfreude* to have no feelings at all about wrongdoers' feelings. Kant calls 'lack of the feeling whereby the state of others affects us' *frigidity* (*Kaltsinnigkeit*) (Eth-C, 27: 420). But it is hard to see how frigidity toward wrongdoers would respond to their need for forgiveness or promote reconciliation. This fact, along with Kant's point that *Schadenfreude* is 'directly opposed' to our duty of sympathy (*MM*, 6: 460, quoted above), seems to imply that the duty of reconciliation requires the wronged to sympathize with wrongdoers.

But this raises further questions. Which feelings of wrongdoers should the wronged sympathize with? Many wrongdoers suffer at many points in life. For example, violent wrongdoers have often suffered abuse in childhood. It is surely appropriate to sympathize with the pain of childhood abuse, but this may not constitute specifically reconciliatory sympathy. The reconciliatory wronged seek reconciliation *over the wrong*, so it is natural to think that reconciliatory sympathy should aim at wrongdoers' feelings *about the wrong*. Sympathizing with remorseless wrongdoers' feelings about the wrong might dispose us to the '*meek toleration* of wrongs' Kant warns against (6: 461). This suggests that wrongdoers must be remorseful if sympathy

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

with their feelings about the wrong is to be virtuous, and that reconciliatory sympathy *is* (or at least *includes*) sympathy with remorse felt by wrongdoers.

Sympathy for self-retributive remorse felt by wrongdoers may well be virtuous. But when we sympathize with remorse we believe to be deserved, we have conflicting reasons about how to respond to wrongdoers' need for forgiveness. Claudia Blöser convincingly argues that this need is for a lightening of the burden of moral failure (2019). This makes it natural to think that our response to this need should help diminish wrongdoers' remorse, and that the duty of reconciliation requires efforts toward this diminution. If we sympathize with remorse we believe wrongdoers deserve, we may think that wrongdoers should continue to suffer, and that we should therefore resist acting on our sympathies in ways that diminish remorse. Retributivists can of course appeal to additional principles which govern when and how it is appropriate to help wrongdoers diminish their remorse. But this adds conceptual and psychological mediation which diminishes the moral significance and motivational efficacy of reconciliatory sympathy.

Suppose instead that the wronged sympathize with wrongdoers' sympathetic remorse. Sympathetically remorseful wrongdoers are pained by the pain they have caused the wronged, and this motivates them to diminish the pain with which they sympathize. If the wronged sympathize with wrongdoers' sympathetic remorse, and harbour no belief that it is deserved, this naturally motivates them to offer wrongdoers opportunities to make amends. Wrongdoers' amends diminish the pain of the sympathetic wronged in a twofold way: by directly working to undo harm to the wronged, and by diminishing the remorse with which the wronged sympathize. In this way, sympathy with wrongdoers' sympathetic remorse is not only responsive to their needs, but also directly conducive to reconciliation, since it avoids the mediation involved in responding to remorse we believe to be deserved. Thus, a society where the wronged demand sympathetic rather than self-retributive remorse would be a society where the wronged were better able to fulfil their duty of reconciliation.

6. Conclusion

I have presented a non-retributive account of remorse as a step toward a revisionist but textually grounded non-retributive Kantian ethics. But most of the ideas offered here can be incorporated into the interpretations of Kantians committed to retributivism. Even if one holds that self-retributive remorse must play a role in Kantian ethics, one can give sympathetic remorse a role too, and suppose that wrongdoers should be sensitive to both. Kantian ethics is often criticized for having a simplistic moral psychology which does not capture the complexity of moral experience, and while previous commentary has already done a lot to blunt this criticism, finding a role in Kantian ethics for a distinctive kind of remorse based on sympathy can contribute to this goal.[31]

Notes

1 For embarrassment (*Verlegenheit*), see *Anth*, 7: 121, 132; for shame (*Scham*), see *CPrR*, 5: 88, OFBS, 2: 218. Abbreviations and translations for Kant's texts are as follows, unless otherwise noted; '†' within quoted passages indicates my modification of the Cambridge translation; translations from volumes in German are my own. *Anth* = *Anthropology from a Pragmatic Point of View*, trans. Robert B. Louden in Kant 2007:

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

231–429; Anth-F, Anth-Mr = Friedländer and Mrongovius notes from Kant's Anthropology lectures, in Kant 2012: 37–255, 335–509; *CF* = *The Conflict of the Faculties*, trans. Mary J. Gregor and Robert Anchor, in Kant 1996b: 237–327; Corr = Correspondence (Kant 1999); *CPrR* = *Critique of Practical Reason,* in Kant 1996a: 137–271; *CPJ* = *Critique of the Power of Judgement* (Kant 2000); 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; L-Th = Lectures on the Philosophical Doctrine of Religion, in Kant 1996b: 339–451; Met-Her = Herder notes from Kant's metaphysics lectures in Kant 1968: 5–166; Met-Mr = Mrongovius notes from Kant's metaphysics lectures, in Kant 1997b: 107–286; *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; Ped = Lectures on Pedagogy (Kant's own lecture notes), trans. Robert B. Louden, in Kant 2007: 437–85; Refl = Reflections on Moral Philosophy, in Kant 1934: 92-317; *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.

2 Other words rendered as 'remorse' and derivatives in the Cambridge translation include *Zerknirschung* (*CF*, 7: 10, 55), *zerknirschten* (*CPJ*, 5: 263), *Kummers* (Eth-V, 27: 642) and *Verweis* (*CPrR*, 5: 38).

3 The Cambridge edition often translates *Reue* as 'repentance', which can evoke a theological context. I take Kant's view to be that the emotional core of sincere repentance even in a theological context is remorse, and that this core can be distinguished from the feelings specifically about God which are also involved in repentance, such as fear of divine punishment. Kant distinguishes the 'inner sorrow' (*innere Traurigkeit*) of '*wahre* (true) *Reue*' from the sorrow of *Buße*, which the Cambridge edition also translates as 'repentance', but can be rendered as 'penance' or 'penitence' (Eth-C, 27: 464). Kant remarks that *Buße* is 'not a good term; it derives from penances and chastisements (*Büßungen, Kasteyungen*)' which we inflict on ourselves when we recognize that we deserve punishment, in the hope that God will not punish us later (Eth-C, 27: 464). (Kant seems to think of *Buße* as a kind of non-moral, prudential *Reue*). Therefore, in the following I will replace 'repentance' in the Cambridge translation with 'remorse[1]' or derivatives when it is clear that Kant is referring to *wahre*, *moralische Reue.*

4 I do not claim that remorse involves feeling but no cognitive activity. Practical reasoning about our wrongs is clearly embedded in the experience of remorse. But my central question is about why we should *hurt*, so my focus is on feeling.

5 The centrality of maxims for Kant may seem to suggest that remorse must be a response to maxims rather than actions. But maxim-adoption is something we *do* (*Rel*, 6: 24) and thus a kind of action, so theories of remorse about actions extend to maxims. However, maxims create puzzling tensions in Kant's theory of remorse and will therefore be bracketed in the main text. Kant thinks that beginning life as radically evil means we all incorporate evil into 'maxims in general (in the manner of *universal principles* as contrasted with individual transgressions)', which entails 'an *infinity* of violations of the law' and an '*infinity* of guilt (*Schuld*)' for which we all must expect '*infinite* punishment (*Strafe*)' (*Rel*, 6: 72). This punishment must be experienced in a conversion, and is naturally understood as involving remorse. Can we experience infinite remorse without continuous self-reproach? Can a murderer deserve the same degree of remorse as a shoplifter or someone who always acts permissibly despite beginning as radically evil?

6 Thomas Hill discusses conscience-pills in his account of the motivational role of conscience (2002: 352).

7 'Desert' in the main text refers to *action-based* desert, which entails free will and moral responsibility (in Kant's view and in the view of most ethicists). See note 20.

8 For retributive accounts of remorse, see Freud (1989: 83–96), Walker (1980: 129), Murphy (2012: 122–3, 138), Smith (2016: 356–7). To justify pain retributively is not necessarily to endorse vengeance – it is simply to claim that suffering is deserved – but some accounts incorporate vengeance. Freud conceptualizes painful moral feeling as the result of internalizing a vengeful parent. In Freudian internalization, an aspect of one's own mind represents another agent, such that one's own action on oneself represents action on oneself by the other. Kantian conscience involves similar internalization, though the internalized agent is God. Kant's God sounds vengeful sometimes (*MM*, 6: 460) but not always (L-Th, 28: 1086). See note 13.

9 Mill is probably the most influential advocate of a forward-looking justification of remorse: see *Utilitarianism* (Mill 2001: 28–9). Also see Proeve and Tudor (2016: 117–120).

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

10 I describe a care justification which is independent from Kantian ethics in Vilhauer (2004). See Pereboom (2021: 52) for discussion.
11 As discussed below in the main text, Kant warns against self-torture. Elsewhere he encourages us to bear remorse in a way that allows cheerful commitment to moral progress (e.g. *Rel,* 6: 24n.). But his view seems to be that we can only hope for genuine cheerfulness by earning it through a conversion naturally understood as involving profound remorse. See note 5.
12 It may be objected that Kant cannot see conscience as retributive because he holds that conscience does not punish (e.g. Wood 2008: 187). But retribution need not entail punishment, and in any case, Kant does think conscience punishes. See Hill (2002: 340–61) for discussion, and MPT, 8: 260.
13 It may be objected that Kant cannot think we can retributively inflict remorse on ourselves, because he claims we cannot control our feelings. But while we cannot directly will feelings to spring forth, we can act on ourselves in ways that prompt feelings. We feel respect when we subordinate our wills to the law (*G,* 4: 402), and this is something we *do.* As discussed below, we can prompt sympathy by putting ourselves in others' places. As discussed above, Kant speaks of *Reue* as self-torture (*Selbstpeinigung*), a notion in which the concept of acting on oneself to cause feeling is implicit. In everyday life, retributively self-inflicted remorse often involves self-directed inner speech. One says to oneself, 'How could you do that, you terrible person?' One may go on to say, 'You deserve to suffer for doing that!', or may leave that part implicit. Human nature is such that, whether such condemnation issues from others or from ourselves, it often prompts painful feeling. On Kant's account of conscience, we must represent the inner condemner as an aspect of ourselves, and also as God (*MM,* 6: 438–40). See note 8. Some self-retribution seems to happen below the level of conscious intention (and is in this sense 'instinctive', as Kant puts it above). But reducing conscious self-retribution may cultivate a reduction in unconscious self-retribution. This is often a goal of psychotherapy.
14 Pereboom (2006) points out the significance of this intuition for Kant's account of transcendental freedom, though not in the context of remorse.
15 Also see Zupančič (2000: 21–42) and Gamberini (2013).
16 Kant himself assigns this status to the belief in transcendental freedom at *CPrR*, 5: 132, though this may be a slip of the pen.
17 See Vilhauer (2010) for an alternative possibilities account of transcendental freedom.
18 See Blöser (2015) for a helpful discussion of degrees of imputation.
19 See note 8.
20 See note 7. Some may think that my friend deserves sympathy *from* me, as her friend, based on a morally important *need* for sympathy from friends. But it is intuitive to think that need is a desert base which is distinct from action and does not entail free will or moral responsibility. Consider the claim that children deserve the love of their parents, based on their need for this love, even when children are too young to have earned love and conceptions could not reasonably be anticipated. Since I lack textual evidence to think that Kant views need as a desert base, this issue is set to the side in the main text, where I use 'desert' just to refer to action-based desert, which entails free will and moral responsibility. See Vilhauer (Forthcoming a) for further discussion.
21 For other discussions of Kant and care, see Baron (1995), Hay (2013), Miller (2012), Varden (2020).
22 I provide this exegesis in more detail in Vilhauer (2021a, 2022).
23 Also see Anth-Mr, 25: 1258; Met-Mr, 29: 884–5.
24 For similar language see *MM*, 6: 321n.; Eth-H, 27: 58, 65; Anth-F, 25: 575, 607. Timmermann (n.d.) discusses the idea that imagination puts us in others' places, but does not think this is necessary for adopting others' ends.
25 See Baron and Fahmy (2009: 222) for a discussion of this point.
26 See Vilhauer (2021b) for a more detailed argument. Also see Denis (2000).
27 See Vilhauer (2022). Fahmy (2010: 314–27) also draws on the distinction between adopting and promoting others' ends, but interprets it differently.
28 It is doubtful that the oppressed ever really become inured to oppression, but Kant's response to this possibility has important implications for other cases.
29 We can also respond to immoral maxims on which we never act by sympathizing with possible others.
30 Also see Eth-C, 27: 440; Eth-V, 27: 695 for *Schadenfreude.*

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

31 Thanks to the editors and reviewers at *Kantian Review,* as well as Matthew Altman, Jeffrey Blustein, Melissa Seymour Fahmy, Derk Pereboom, Jens Timmermann and Allen Wood for helpful advice at various points in the history of this project.

References

American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition. Arlington, VA: American Psychiatric Association Publishing.

Baron, Marcia (1995) *Kantian Ethics Almost without Apology*. Ithaca, NY: Cornell University Press.

Baron, Marcia and Melissa Seymour Fahmy (2009) 'Beneficence and Other Duties of Love'. In Thomas E. Hill (ed.), *The Blackwell Guide to Kant's Ethics* (Oxford: Wiley-Blackwell), 211–28.

Batson, C. Daniel, Shannon Early and Giovanni Salvarani (1997) 'Perspective Taking: Imagining How Another Feels Versus Imagining How You Would Feel'. *Personality and Social Psychology Bulletin*, 23(7), 751–58.

Blöser, Claudia (2015) 'Degrees of Responsibility in Kant's Practical Philosophy'. *Kantian Review*, 20(2), 183–209.

—— (2019) 'Human Fallibility and the Need for Forgiveness'. *Philosophia*, 47(1), 1–19.

Decety, Jean, Philip L. Jackson and Eric Brunet (2007) 'The Cognitive Neuropsychology of Empathy'. In Tom F. D. Farrow and Peter W. R. Woodruff (eds), *Empathy and Mental Illness* (Cambridge: Cambridge University Press), 240–60.

Denis, Lara (2000) 'Kant's Cold Sage and the Sublimity of Apathy'. *Kantian Review*, 4(1), 48–73.

Fahmy, Melissa Seymour (2010) 'Kantian Practical Love'. *Pacific Philosophical Quarterly*, 91(3), 313–31.

Freud, Sigmund (1989) *Civilization and its Discontents*. Ed. Samuel Moyn, trans. James Strachey. New York: Norton.

Gamberini, Paola (2013) 'Guilt and Repentance: Kant on the Experience of Moral Responsibility in the Retrospective Evaluation of Actions'. In Margit Ruffing, Claudio La Rocca, Alfredo Ferrarin and Stefano Bacin (eds), *Kant und die Philosophie in weltbürgerlicher Absicht. (Berlin and* Boston: De Gruyter), 233–44.

Ghorbani, Nima, Mark N. Bing, P. J. Watson, H. Kristl Davison and Daniel L. LeBreton (2003) 'Individualist and Collectivist Values: Evidence of Compatibility in Iran and the United States'. *Personality and Individual Differences*, 35(2), 431–7.

Hay, Carol (2013) *Kantianism, Liberalism, and Feminism: Resisting Oppression*. New York: Palgrave-Macmillan.

Hill, Thomas E. (2002) *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Oxford University Press.

Hodges, Sara D. and Robert Biswas-Diener (2007) 'Balancing the Empathy Expense Account: Strategies for Regulating Empathic Response'. In Tom F. D. Farrow and Peter W. R. Woodruff (eds), *Empathy and Mental Illness* (Cambridge: Cambridge University Press), 390–407.

Kant, Immanuel (1934) *Kant's gesammelte Schriften*, vol. 19: *Moralphilosophie, Rechtsphilosophie und Religionsphilosophie*. Ed. Erich Adickes and Friedrich Berger, Prussian Academy of Sciences. Berlin: Walter de Gruyter.

—— (1968) *Kant's gesammelte Schriften*, vol. 28.1: *Vorlesungen über Metaphysik*. Ed. Gerhard Lehmann, Göttingen Academy of Sciences. Berlin: Walter de Gruyter.

—— (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) *Lectures on Metaphysics*. Trans. and ed. Karl Ameriks and Steve Naragon. Cambridge: Cambridge University Press.

—— (1999) *Correspondence*. Trans. and ed. Arnulf Zweig. Cambridge: Cambridge University Press.

—— (2000) *Critique of the Power of Judgment*. Trans. Paul Guyer and Eric Matthews, ed. Paul Guyer. Cambridge: Cambridge University Press.

—— (2007) *Anthropology, History and Education*. Ed. Günther Zöller and Robert B. Louden. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press

—— (2012) *Lectures on Anthropology*. Ed. Allen W. Wood and Robert B. Louden. Cambridge: Cambridge University Press.
Mill, John Stuart (2001) *Utilitarianism*. Ed. George Sher. Indianapolis: Hackett.
Miller, Sarah Clark (2012) *The Ethics of Need: Agency, Dignity, and Obligation*. New York: Routledge.
Murphy, Jeffrie G. (2012) *Punishment and the Moral Emotions: Essays in Law, Morality, and Religion*. New York: Oxford University Press.
O'Connor, Lynn, Jack W. Berry, Thomas Lewis, Kathleen Mulherin and Patrice S. Cristosomo (2007) 'Empathy and Depression: The Moral System on Overdrive'. In Tom F. D. Farrow and Peter W. R. Woodruff (eds), *Empathy and Mental Illness* (Cambridge: Cambridge University Press), 49–75.
Pereboom, Derk (2006) 'Kant on Transcendental Freedom'. *Philosophy and Phenomenological Research*, 73(3), 537–67.
—— (2021) *Wrongdoing and the Moral Emotions*. Oxford: Oxford University Press.
Proeve, Michael and Steven Tudor (2016) *Remorse: Psychological and Jurisprudential Perspectives*. New York: Routledge.
Satne, Paula (2021) 'Kantian Guilt'. In Beatrix Himmelmann and Camilla Serck-Hanssen (eds), *The Court of Reason* (Berlin and Boston: De Gruyter), 1511–20.
Smith, Nick (2016) 'Dialectical Retributivism: Why Apologetic Offenders Deserve Reductions in Punishment Even Under Retributive Theories'. *Philosophia*, 44(2), 343–60.
Tangney, J. P. (1991) 'Moral Affect: The Good, the Bad, and the Ugly'. *Journal of Personality and Social Psychology*, 61(4), 598–607.
Timmermann, Jens (n.d.) 'Kant's Theory of Sympathy'. Unpublished book manuscript.
Varden, Helga (2020) *Sex, Love, and Gender: A Kantian Theory*. Oxford: Oxford University Press.
Vilhauer, Benjamin (2004) 'Hard Determinism, Remorse, and Virtue Ethics'. *Southern Journal of Philosophy*, 42(4), 547–64.
—— (2010) 'The Scope of Responsibility in Kant's Theory of Free Will'. *British Journal for the History of Philosophy*, 18(1), 45–71.
—— (2015) 'Free Will and the Asymmetrical Justifiability of Holding Morally Responsible'. *Philosophical Quarterly*, 65(261), 772–89.
——(2017) 'Kant's Mature Theory of Punishment and a First *Critique* Ideal Abolitionist Alternative'. In Matthew C. Altman (ed.), *The Palgrave Kant Handbook* (London: Palgrave Macmillan), 617–42.
—— (2021a) '"Reason's Sympathy" and its Foundations in Productive Imagination'. *Kantian Review*, 26(3), 455–74.
—— (2021b) 'Sages, Sympathy, and Suffering in Kant's Theory of Friendship'. *Canadian Journal of Philosophy*, 51(6), 452–67.
—— (2022) '"Reason's Sympathy" and Others' Ends in Kant'. *European Journal of Philosophy*, 30(1), 96–112.
—— (Forthcoming a) 'Free Will Skepticism and Criminals as Ends in Themselves'. In Matthew C. Altman (ed.), *Palgrave Handbook of Punishment*. New York: Palgrave.
—— (Forthcoming b) 'An Asymmetrical Approach to Kant's Theory of Freedom'. In Dai Heide and Evan Tiffany (eds), *The Idea of Freedom: New Essays on the Interpretation and Significance of Kant's Theory of Freedom*. Oxford: Oxford University Press.
Walker, Nigel (1980) *Punishment, Danger, and Stigma: The Morality of Criminal Justice*. Oxford: Blackwell.
Wood, Allen W. (2008) *Kantian Ethics*. Cambridge: Cambridge University Press.
Zupančič, Alenka (2000) *Ethics of the Real: Kant, Lacan*. London and New York: Verso.

Cite this article: Vilhauer, B. (2022). Kantian Remorse with and without Self-Retribution. *Kantian Review* **27**, 421–441. https://doi.org/10.1017/S1369415422000140

https://doi.org/10.1017/S1369415422000140 Published online by Cambridge University Press