

# Kant's Retributive Theory of Remorse, and a Non-Retributive Kantian Alternative

Benjamin Vilhauer  
Professor, CUNY City College & Graduate Center

## 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 that 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 both Kant's texts and recent work in empirical psychology, I argue that experiences of remorse which conform to the sympathetic account may fulfill Kant's forward-looking goals better than those conforming to his own account.

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

## I. 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 behavior 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 intuitive distinction between them just mentioned.<sup>1</sup>

---

<sup>1</sup> For embarrassment [*Verlegenheit*], see e.g. AP 7:121, 132; for shame [*Scham*], see e.g. 2C 5:88, OB 2:218.

The difference between remorse and guilt is somewhat complex. “Guilt” is used both to refer to painful moral feeling and also to a state of culpability which can be determined by conscience, 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 guilty feeling, though he does so in at least one place (*schuldig zu fühlen*, RR 6:38)<sup>2</sup>.

---

<sup>2</sup> Abbreviations and translations for Kant’s texts are from the following books unless otherwise noted. “<sup>t</sup>” in passages from Kant indicates that I have altered the Cambridge translation to reflect my own translation. Pagination is by *Akademie* edition, which is included in texts cited, unless otherwise noted. “<sup>NA</sup>” at the end of entries in the following list in this note indicates texts not included in the *Akademie* edition; quotes from these are paginated according to the volume cited. Translations from German-language volumes in this list are my own. **A/B**: *Critique of Pure Reason*, trans. and ed. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1998. **2C**: *Critique of Practical Reason*, in *Practical Philosophy*, trans. and ed. Mary J. Gregor, 137-271. Cambridge University Press, 1996. **3C**: *Critique of the Power of Judgment*, trans. Paul Guyer and Eric Matthews, ed. Paul Guyer, Cambridge University Press, 2000. **AP**: *Anthropology from a Pragmatic Point of View*, trans. Robert B. Louden. In *Anthropology, History and Education*, ed. Günther Zöller and Robert B. Louden, 231-429. Cambridge: Cambridge University Press, 2007. **CA**: Collins notes from Kant’s Anthropology lectures in *Lectures on Anthropology*, trans. Robert B. Clewis, Robert B. Louden, G. Felicitas Munzel, and Allen W. Wood. Ed. Allen W. Wood and Robert B. Louden, Cambridge: Cambridge University Press, 2012, 11-26. **CO**: *Immanuel Kant: Correspondence*, trans. and ed. Arnulf Zweig, Cambridge: Cambridge University Press, 1999. **CE**: Collins notes from Kant’s Ethics lectures, in *Lectures on Ethics*, trans. Peter Heath, ed. Peter Heath and J.B. Schneewind, Cambridge: Cambridge University Press, 1997, 37-222. **CF**: *The Conflict of the Faculties*, trans. Mary J. Gregor and Robert Anchor, in *Religion and Rational Theology*, trans. and ed. Allen W. Wood and George di Giovanni, Cambridge: Cambridge University Press, 1996, 237-327. **ET**: *The End of All Things*, trans. Allen W. Wood, in *Religion and Rational Theology*, ed. Allen W. Wood and George di Giovanni, 221-31. Cambridge: Cambridge University Press, 1996. **FA**: Notes from Kant’s Anthropology lectures by Michael Friedländer, in *Lectures on Anthropology* (ibid.), 37-255. **G**: *Groundwork of the Metaphysics of Morals*, in *Practical Philosophy* (ibid.), 41-108. **HE**: Notes from Kant’s Ethics lectures by J.G. Herder, in *Lectures on Ethics* (ibid.), 1-36. **HM**: Notes from Kant’s Metaphysics lectures by J.G. Herder, in *Kant’s gesammelte Schriften: Vorlesungen über Metaphysik und Rationaltheologie (Band 28, Erste Hälfte)*, ed. Gerhard Lehmann, Berlin: Walter de Gruyter, 1968. **HN**: *Kant’s gesammelte Schriften: Band 19, Dritte Abtheilung: handschriftlicher Nachlass, Sechster Band*, ed. Preußische Akademie der Wissenschaften, Berlin and Leipzig: Walter de Gruyter, 1934. **LP**: *Lectures on Pedagogy* (Kant’s own lecture notes),

The English word “remorse” has a helpful simplicity in its exclusive reference to painful moral feeling, and I will 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.<sup>3</sup> “*Reue*” has “rue” as a close cognate in English, and both can 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*” (CE 27:353) and “*wahre* [true] *Reue*” (VE 27:464). In the Collins Ethics lecture notes, Kant identifies “*moralische Reue*” as *Reue* for behavior “in regard to morality “[*in Ansehung der Moralität*]” and distinguishes it from *Reue* because one has acted “imprudently [*unklug*]” (CE 27:353). Similarly, in his 1792 letter to Maria von Herbert (CO 11:333), he distinguishes *Reue* over “imprudence [*Unklugheit*]” from *Reue* “grounded in a purely

---

trans. Robert B. Loudon, in *Anthropology, History, and Education* (ibid.), 437-85. **MA:** Notes from Kant’s Anthropology lectures by Christian Coelestin Mrongovius, in *Lectures on Anthropology* (ibid.), 335-509. **MM:** *The Metaphysics of Morals*, in *Practical Philosophy* (ibid.), 363-602. **MT:** *On the Miscarriage of All Philosophical Trials in Theodicy*, trans. George di Giovanni, in *Religion and Rational Theology* (ibid.), 19-38. **FA:** Notes from Kant’s Anthropology lectures by Michael Friedländer, in *Lectures on Anthropology* (ibid.), 261-279. **RR:** *Religion within the Boundaries of Mere Reason*, trans. George di Giovanni, in *Religion and Rational Theology*, (ibid.), 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 *Practical Philosophy* (ibid.), 1-10. **VE:** Notes from Kant’s Ethics lectures by Johann Friedrich Vigilantius in *Lectures on Ethics*, (ibid.), 249-452.

<sup>3</sup> Other words rendered as “remorse” and derivatives in the Cambridge translation include *Zerknirschung* (CF 7:10, 7:55), *zerknirschten* (3C 5:263) *Kummers* (VE 27:642), and *Verweis* (2C 5:38).

moral judgment [*auf bloßer sittlicher Beurtheilung ...Verhaltens gründet*]" about one's behavior.<sup>4</sup>

I will therefore use "remorse" to translate Kant's *wahre, moralische Reue*.

The main questions I wish to address in this paper 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? 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[.] (CE 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"

---

<sup>4</sup> The Cambridge edition often translates "*Reue*" as "repentance", which can evoke a theological context, and this context is often part of Kant's discussions, but is peripheral to my interests here. However, 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 phenomenologically and morally isolated from the feelings specifically about God which are also involved in repentance, such as fear of divine punishment. Kant is careful to distinguish the "inner sorrow [*innere Traurigkeit*]" of "*wahre* [true] *Reue*" from the sorrow of *Buße*, which the Cambridge edition also translates as "repentance", but can also be rendered as "penance" or "penitence". Kant remarks that "*Buße*" is "not a good 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 (CE 27:464). Therefore, in the following I will typically replace appearances of "repentance" and its derivatives in the Cambridge translation with "remorse" or derivatives in cases where it seems clear that Kant is referring to *wahre, moralische Reue*.

(CE 27:356). 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*]" (CO 11:333), and an "excess of remorse [*Kummers*] over ...transgressions of duty" can prompt suicide (VE 27:642). Shaping conscience in these ways requires judgment 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). 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.<sup>5</sup> Retributive justifications are often called *backward-looking*, because they refer only to past wrongs. Another justification is *forward-looking*: our wrongs should pain us because this motivates us to act better in the future.<sup>6</sup> 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

---

<sup>5</sup> For retributive justifications of remorse, see Freud 1989: 83-96; Walker 1980: 129; Murphy 2012: 122-3, 138; Smith 2016: 356-357. Freud is the most influential advocate of a retributive conception of painful moral feelings. He conceptualizes painful moral feelings as the result of internalizing a vengeful parent intent on castration. Characterized abstractly, in Freudian internalization, an aspect of one's own mind comes to represent another agent, such that one's own action on oneself comes to represent action on oneself by the other agent. Kant's model of conscience also involves internalization in this sense, though for Kant the internalized agent is God.

<sup>6</sup> Mill is probably the most influential advocate of a forward-looking justification of remorse: see *Utilitarianism* (2001: 28-9). For a contemporary discussion, see Proeve and Tudor 2016: 117-120

our wrongs cause.<sup>7</sup> This third justification has not been discussed as widely as the first two, but it does not rely on retributivism and it is not purely forward-looking, as I will explain below. Kant does not speak in terms of the value of care, but it is implicit in his theories of sympathy and friendship. I will argue that Kant's account of our reasons for remorse is a hybrid of the retributive and forward-looking justifications. I will go on to offer a revisionist but textually-grounded care-based Kantian alternative.

## II. 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<sup>t</sup> him [*ihn sein Gewissen nachher peinigen kann*][.] (MM 6:394)

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.<sup>8</sup>

---

<sup>7</sup> A sympathy-based justification of remorse is discussed in Pereboom 2014: 186-7.

<sup>8</sup> Kant's account of remorse has received little discussion in the literature. It has been discussed indirectly in relation to Kant's account of conscience, by way of a question about whether conscience punishes, in the context of a critique of Thomas Hill's interpretation by Allen Wood. Hill holds that conscience punishes, while Wood holds that conscience does not punish (Hill 2002: 301, 352-3; Wood 2008: 187-8). Wood rightly highlights passages in which Kant criticizes penitent self-torture (e.g. MM 6:485, CE 27:464, CF 7:55-6). However, close analysis shows that what Kant is criticizing is the penitence of *Buße*, which Kant thinks is founded in *Reue* over imprudence, rather than the self-retribution of *wahre, moralische Reue*. See note 4 above. I discuss this issue in more detail in [author's paper 1, under review]. There is a passage which neither cites which supports Hill's interpretation, at MT 8:260, where Kant describes the conscientiousness of the virtuous man as "a conscientiousness [*Gewissenhaftigkeit*] in all its severity which, the more virtuous a human being is, all the more harshly punishes [*bestraft*] him

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. The "monkish ascetics" passage discussed earlier states that *moralische zu bereuen* requires a "view to improving" (MM 6:485). In the *Religion*, Kant criticizes "remorseful self-inflicted torments [*reueige Selbstpeinigungen*]" that do not...originate in any genuine disposition toward improvement" (RR 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" (RR 6:77n). The *Anthropology* warns against regarding our "record of guilt as...simply wiped out (through remorse<sup>1</sup> [*Reue*]), so that [we are] spared the effort toward improvement" (AP 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" (MA 25:1335, my boldface). 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

---

because of the slightest indiscretion frowned upon by the moral law in him." We should assume that virtuous people have consciences which function properly, so if their consciences punish, then proper function of conscience is to punish. However, this point does not give us a fundamental justification of the pain of remorse which conscience inflicts, because punishment itself requires further justification.

on acting better in the future. But retributive justifications of remorse essentially involve references to past wrongs, and so the MA 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 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.<sup>9</sup> Forward-looking reasoning is crucial for Kant in contexts as varied as punishment by the state (MM 6:331, 6:336; CE 27:286), education (2C 5:152; LP 9:451-2), and the prudent permissible pursuit of happiness (G 4:399; 2C 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 behavior 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

---

<sup>9</sup> I am not claiming that duties to persons as ends can be fully explained in terms of the forward/backward dichotomy, and I do not think retributivism is necessary to account for such duties, but I lack space to address this matter in detail.



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, Kant attributes various theses to Schulz which Kant does not accept. One of 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 retribution. 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" (HM 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<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" (CE 27:354). Kant's death-bed case partly overlaps with the solitary confinement case sketched above, in that Kant thinks the imminence of death means that 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 channeled into improved behavior 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. We have seen this in Kant's warnings that remorse can lead to distraction, brooding which makes life useless, and even suicide.

### III. Concerns about Self-Retribution

In this section I will explain 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. It is based on a critique of Kant which I can only sketch here, though I provide detailed arguments for it elsewhere.<sup>10</sup> 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. This intuition is part of why so many endorse the view that arguments in the criminal court must be proven beyond reasonable doubt.<sup>11</sup> Kant himself addresses this intuition in a discussion of imputation [*Imputation*] of crimes in the *Vigilantius Ethics* notes (VE 27:558-27:573). "Imputation" is Kant's term for assigning moral responsibility for actions, so imputation is a precondition for judgments about desert, and thus for retributive justification. He describes the justificatory standard we must meet in imputing crimes as "the greatest possible" [*größtmögliche*] 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" (VE 27:567)) and the nature of the "motive to the action" (VE

---

<sup>10</sup> Author's paper 2, in print.

<sup>11</sup> Pereboom 2006 points out the significance of this intuition for Kant's account of transcendental freedom, though not in the context of remorse.

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” (ibid.). 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 VE 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 transcendently 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 transcendently free (see e.g. 2C 5:30). Kant may mean to assert this supposed knowledge as an ungrounded or self-grounding “fact of pure reason” (2C 5:31). He may also think it has a kind of phenomenological grounding in our moral feelings, and in respect for moral law in particular, which he thinks we should represent 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 (2C 5:116), so it unclear how the phenomenology can play a grounding role. It seems fair to claim that Kant’s theoretical argument against knowledge that we are transcendently free makes it prudent to be cautious about being easily persuaded by his practical argument in favor 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 transcendently free in the first *Critique*, give us ample space for a postulate that we are transcendently free.<sup>12</sup> But to postulate is not to know, and without knowledge, we cannot meet the justificatory standard retribution requires. We cannot justly inflict suffering on people by postulating that they deserve to suffer.

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 I think 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.<sup>13</sup>

---

<sup>12</sup>Kant himself of course assigns this status to the belief in transcendental freedom at 2C 5:132, though it is not clear whether this is an alternative way he sometimes thought about this belief or a slip of the pen.

<sup>13</sup> I defend an alternative-possibilities account of transcendental freedom in [author’s paper 3, in print].

But when we make *judgments 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 judgment. There is a kind of flexibility available to us in deciding what it means to regard completed actions under the idea of freedom, which justice requires us to take 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 (ME 6:228; CE 27:291; VE 27:567) implies that we can assign responsibility in different degrees in different cases. Herder's *Metaphysics* notes suggest that the appropriate degree of imputation is sometimes "vanishingly small" (HM 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 accordingly diminish the degree of imputation, 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.

Now I will turn to a second concern about the retributive component of Kant's account, which is practical rather than metaphysical or epistemological. That is, 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 suicide are all obviously things which obstruct improvement, and these would all seem to follow naturally from the belief that we deserve to suffer. If inflicting suffering on ourselves because we believe 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" (VE 27:575), as Kant claims, 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" (VE 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 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, 160-164). The degree to which the pain at issue in the DSM is principally motivated by a 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.

We may also have an example of suicide motivated by self-retribution in Kant's own 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 previously 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 (CO 11:273-4). Kant writes back in 1792 encouraging her not to kill herself, counseling "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" (CO 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 behavior (CO 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 skepticism about ever really being sure about how pure one's disposition is (G 4:407, MM 6:392-3).<sup>14</sup> She committed suicide in 1803. It is 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.

---

<sup>14</sup> Ware 2009 argues that the role of God's judgment in conscience resolves this concern. In [author's paper 1, under review] I argue that this resolution is unstable because the God of conscience is an aspect of one's own self-judgment.

#### IV. 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 will explain how it works in a general way, and next, 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 made it the case that I *deserve* to suffer when she does, but because sympathy is *part of caring*. It would be absurd to suppose that by befriending someone, I have gotten myself into a situation such that I deserve to suffer when she does—rather, we think that sympathetic suffering is part of the *nature* of friendship, because it is part of the nature of *care*.

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 everyday moral common sense that I *ought* to believe that I deserve to suffer. But even if I am skeptical 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. Sungwoo Um (2020) calls this focused aspect of care "particularized care".

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.<sup>15</sup> Certainly 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 particular 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* [*Bedürfnis*]. Some philosophers may wish to model such needs in terms of Strawsonian reactive attitudes, which can be understood as 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.

---

<sup>15</sup> I set out this view in a way that is independent of Kantian ethics in [author's paper 4, in print].

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.<sup>16</sup> 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” (G 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 überhaupt Pflicht*]” (MM 6:456), and that “it is a duty [*Pflicht*] to actively sympathize [*thätige Theilnehmung*] in [others'] fate” (MM 6:457). If it 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

---

<sup>16</sup> Others who have discussed Kant and care include Halpern (2001:101-127), Wood (2008: 176-7), and Paytas (2015).

*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 will refer to as "natural sympathy", and the other is an active (*thätige*) sympathy guided by practical reason, which I will refer to as "rational sympathy". Kant draws this distinction in at least five places over a period of at least 20 years. To keep things concise, I will label the terms for rational sympathy with "(a)" and the terms for natural sympathy with "(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*". FA 25:607-11 distinguishes (a) "*reason's sympathy*" and (b) "*physical sympathy*". AP 7:235 and MA 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 be affected, even against our will, by sympathy (Theilnehmung) for others' condition who...play at will on the organ of the sentimentalist*" (AP 7:235). VE 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 interfere with self-governance and can (for example) prevent us from helping effectively even when we discover means to do so (FA 25:589, MM 6:407), and dispose

us to violations of the law when they conflict with duty (RR 6:30 and FA 25:611). In (b) the agent allows feelings to flow passively, without exercising the discipline necessary for (a)-type sympathy.<sup>17</sup>

Sympathy is an activity of the imagination (MM 6:321n, 6:457; AP 7:179, 7:238), 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" (AP 7:153). Sympathy is best understood as an activity of the *a posteriori* productive imagination, which can function both involuntarily and voluntarily (AP 7:174; MA 25:1257). Kant calls involuntary productive imagination "fantasy" (*Phantasie*) (AP 7:167, 7:175, also see MA 25:1258, M 29:884-5), and makes an explicit connection between *Phantasie* and *Empfindelei* (sentimentality) at 3C 5:273, which I argued above is a term for natural sympathy. Kant contrasts fantasy with a voluntary, rationally-ordered counterpart called "disciplined fantasy <*phantasia subacta*>" at M 29:885. This suggests that natural sympathy is an aspect of involuntary productive imagination, and rational sympathy is an aspect of voluntary productive imagination.

Kant says that the sympathetic imagination puts us "in the other's place" (FA 25:575, also see MM 6:321n; HE 27:58, 27:65; FA 25:575, 25:607, for similar language).<sup>18</sup> The voluntary exercise of this capacity, which we might call *projective imagination*, is a *skill* which enables rational sympathy:

the 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<sup>t</sup> (*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*).

---

<sup>17</sup> I give this exegesis in more detail in [author's papers 5 and 6 (in print) and 7 (under review)].

<sup>18</sup> Timmermann (unpublished book manuscript) discusses the idea that imagination puts us in the other's place, but does not see this as a necessary condition of adopting others' ends.

(FA 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” (HE 27:58), and “[w]e are sensible of this sympathizing feeling in our entire soul” (FA 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<sup>t</sup> (*Theilnehmung*)” (FA 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 et al., 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” (Decety et al., 2007: 254). *Empathic distress*, by contrast, is a feeling which Decety et al. (2007: 254) call “emotional contagion”. Hodges et al. (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 et al., 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 et al., 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” (CE 27:430).<sup>19</sup> Friendship is an “ideal of each sympathizing and communicating<sup>t</sup> [*Ideal der Theilnehmung und Mittheilung*] about the other’s wellbeing” which guides us toward a “maximum” (MM 6:469) in which “each mutually sympathizes<sup>t</sup> (*teilnehmen*) with every situation of the other, as if it were encountered by himself” (VE 27:677). This ideal gives us reasons to establish strong particularized sympathetic bonds with our friends. I think this is a manifestation in Kant’s ethics of the value of particularized care.

My claim is that we ought to care about the people we have wronged in this way too. Kant nearly suggests this account of remorse in a discussion of sympathy and the “oppression” of people “subordinate to the aristocracy” (FA 25:606). Kant 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” (FA 25:607). “If the ills are natural, for

---

<sup>19</sup> See Baron and Fahmy (2009: 222) for a discussion of this point.

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” (FA 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” (FA 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” (FA 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” 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 behavior* 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 behavior generally so as to not contribute to future oppression. This next step would establish the basic operations of the sympathetic Kantian remorse I am advocating here.

I argued earlier that suffering in sympathy with people we care about is valuable even when we cannot help, and I think 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 some 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). Kant 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 here 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 word 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.<sup>20</sup>

---

<sup>20</sup> I make this argument in more detail in [author's paper 7, under review]. Also see Denis 2000.



Further, I think that 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 fulfill the imperfect duty to make others' permissible ends "as far as possible...also *my* ends" (G 4:430). This implies that the duty of sympathy is grounded in the duty to make others' permissible ends my own.<sup>21</sup> The argument for this view turns on a distinction between *adopting* and *promoting* others' ends.<sup>22</sup> 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 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*—that is, 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 auto meridian sensory response in me, or because I desire to improve my reputation, or because I have a rational desire to fulfill 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 supports

---

<sup>21</sup> I make this argument in detail in [author's paper 6, in print].

<sup>22</sup> Fahmy (2010: 314-27) also draws on the distinction between adopting and promoting others' ends, but interprets it differently.

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 and metaphysical 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 channeled into improved behavior. It is *prima facie* plausible to suppose that sympathetic remorse would fulfill 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 behavior: as noted earlier, rational sympathy is similar (and plausibly identical) to what is called "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 et al., 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 can 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. give evidence that, while empathic *distress* is *positively* correlated with depression, empathic *concern* is *negatively* correlated with depression (2003: 438). O'Connor et al. explain that “[the] empathic reaction in depressives often leads to great distress because they tend to unrealistically blame themselves for pain felt by others” (2007: 49), and use an explicitly retributive model of self-blame, describing “self punishment” in depression as “meted out...while thinking ‘I deserve this’ ” (ibid., 67). They argue that it is just this unwarranted self-retribution which “transform[s] empathic concern into empathic distress” in the case of depression, and therefore advocate depression therapy which targets unwarranted self-retribution (ibid., 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.

## V. Potential Objections and Replies

In this section I will address 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 by proposing to ground remorse in the duties of sympathy and friendship, I am associating it with *imperfect* duties which grant us *latitude* (see e.g. MM

6:392, 6:411) 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*]" (HN 19:178, *Reflexionen* 6848). 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 in such a way that it "[holds their] duty before [them] for [their] acquittal or condemnation in every case that comes under a law" (MM 6:400).<sup>23</sup> 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 not all wrongs cause pain. 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 basis for remorse in this case on the model proposed here. Such cases are of course challenging

---

<sup>23</sup> See Timmermann (2006) for a helpful discussion of conscience and the distinction between perfect and imperfect duty.

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 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:

if 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[.] (FA 25:606)

Suppose I treat someone disrespectfully who has been disrespected for so long that he has become inured to it. Where is the pain with which I should sympathize? Kant's claim 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 my treatment of him, and by the way I contributed to denying him access to the kinds of experiences he could have as a member of a society free from oppression.

This strategy of sympathy with normatively-adjusted possible versions of actual others can be generalized to all wrongs which do not actually cause pain. 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. Such care for this version of herself seems more likely to help her persist than a new infliction of the self-retribution Kant's own account of remorse prescribes.

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 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. This is a view Kant himself should support. He holds that we have a duty of forgiveness (MM 6:460-1), and adopting this attitude toward others is a way of being forgiving. When we put ourselves into the position of others who adopt this attitude, this is a way of extending forgiveness to ourselves in turn.

This is the first step toward extending the attitude toward oneself advocated in this paper into a broader Kantian moral psychology which is non-retributive interpersonally as well as intrapersonally. To be sure, Kant himself wants to offer an interpersonal moral psychology which is non-retributive in important ways. We can see this in his critique of antipathy (*Antipathie*) (FA 607-611), which he also calls *Schadenfreude* (MM 6:459, CE 27:440, VE 27:695). *Schadenfreude* is an activity of the imagination (MM 6:460) which inverts the function of sympathetic participation such that “one suffers pain because the other rejoices” and “one has a sensation of joy because the other has pain” (FA 25:607). Kant calls it “diabolical”, and says that in its most extreme form, it is “an ideal, or a maximum of moral evil” (FA 25:608).

Kant's discussion of *Schadenfreude* in the *Metaphysics of Morals* comes directly after his discussion of sympathy. He says that *Schadenfreude* is “directly opposed to one's duty in accordance with the principle of sympathy” (MM 6:459-60). He specifies that “[t]he sweetest form of *Schadenfreude* is the desire for revenge” (MM 6:460), and goes on to argue that “[i]t is

therefore a duty of human beings to be *forgiving*” (MM 6:461).<sup>24</sup> The course of this argument implies that the duty of sympathy extends to people who have wronged us, and that it is (at least in part) because of this that we have duty to be forgiving rather than vengeful. But as we have seen, on Kant’s account, remorse has a self-retributive dimension, and there is reason to think that this self-retributive dimension is well-understood as involving vengeance upon oneself. Kant thinks that conscience includes an aspect of ourselves which we represent as God. Kant holds that God, in contrast to human beings, *justly* punishes out of vengeance (MM 6:460). This suggests that we *justly punish ourselves out of vengeance* on his model. The references to deserved self-torture we saw earlier at MM 6:394 and CO 11:334 also support this interpretation.

If this is right, then to demand that wrongdoers feel remorse on Kant’s own model of remorse is to demand that wrongdoers take vengeance upon themselves. That it is a mediated demand for vengeance does not make it any less a demand for vengeance, and it should therefore be recognized as a violation of the duties of sympathy and forgiveness by Kant’s own lights. Presumably demands for vengeance are such that the wronged are pleased by their satisfaction, and such pleasure is therefore an instance of *Schadenfreude*. If the wronged instead demand rationally sympathetic remorse, then while this remains a demand that the wrongdoer feel pain, the pain demanded is merely the pain necessary to fulfill the wrongdoer’s own duty of sympathy, rather than a pain involving vengeance. This takes us some distance toward fulfilling the duty of forgiveness. However, if the demand for the wrongdoer’s painful sympathy remains one whose satisfaction brings pleasure to the wronged, then it remains *Schadenfreude* even if it is no longer vengeance. The conclusion to which we are drawn is that the only way to be *truly* forgiving is

---

<sup>24</sup> Kant thinks that there is not only a duty to sympathize but a need to sympathize (2C 5:34), and the same thing quite plausibly holds with respect to the duty to forgive—that is, while Kant does not comment on a need to forgive, contemporary psychology offers evidence which supports the view that such a need exists (Worthington 2006: 61-68).



for the wronged to be sympathetically pained by the wrongdoers' sympathetic pain. This reciprocity of sympathy would motivate the wronged to offer wrongdoers opportunities to make amends. It would establish a harmony of wills which includes the wills of remorseful wrongdoers, and would contribute to reunification of the moral community. It would take great efforts of conscience-cultivation for real people to conform to this moral psychology, but I think it has attractions which extend well beyond Kantian ethics.

## VI. Conclusion

I have presented a non-retributive account of remorse as the first 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 agents who have acted wrongly 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 effort.

## Bibliography

- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA: American Psychiatric Association Publishing.
- Baron, Marcia and Fahmy, Melissa Seymour. 2009. "Beneficence and Other Duties of Love". In Hill, Thomas E., ed. *The Blackwell Guide to Kant's Ethics*. Oxford: Wiley-Blackwell. 211-228.
- Batson, C. D., Early, S., Salvarani, G. (1997). "Perspective taking: imagining how another feels versus imagining how you would feel." *Personality and Social Psychology Bulletin*, 23, 751-758.

- Decety, Jean, Jackson, Philip L., Brunet, Eric. "The Cognitive Neuropsychology of Empathy". In Farrow, Tom F.D. and Woodruff, Peter W.R., eds. *Empathy and Mental Illness*. Cambridge: Cambridge University Press, 240-260.
- Denis, Lara. 2000. "Kant's Cold Sage and the Sublimity of Apathy". *Kantian Review*, 4:48-73.
- Fahmy, Seymour Melissa. 2009. Active Sympathetic Participation: Reconsidering Kant's Duty of Sympathy. *Kantian Review* 14 (1):31-52.
- . 2010. "Kantian Practical Love". *Pacific Philosophical Quarterly* 91 (3): 313-331.
- Freud, Sigmund. 1989. *Civilization and its Discontents*. New York: Norton, 1989.
- Ghorbani, Nima; Bing, Mark N.; Watson, P.J.; Davison, H. Kristl; LeBreton, Daniel L. 2003. "Individualist and collectivist values: evidence of compatibility in Iran and the United States". *Personality and Individual Differences* 35 (2003):431-447.
- Guyer, Paul. 2010. "Moral feelings in the Metaphysics of Morals". In Denis, Lara, ed. *Kant's Metaphysics of Morals: A Critical Guide*. New York: Cambridge University Press.
- Halpern, Jodi. 2001. *From Detached Concern to Empathy: Humanizing Medical Practice*. Oxford: Oxford University Press.
- Hill, Thomas E. 2002. *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford University Press.
- Hodges, Sara D. and Biswas-Diener, Robert. "Balancing the Empathy Expense Account: Strategies for Regulating Empathic Response". In Farrow, Tom F.D. and Woodruff, Peter W.R., eds. *Empathy and Mental Illness*. Cambridge: Cambridge University Press, 390-407.
- Mill, John Stuart. 2001 (1863). *Utilitarianism*. Ed. George Sher. Indianapolis: Hackett.
- Murphy, Jeffrie G. 2012. *Punishment and the Moral Emotions: Essays in Law, Morality, and Religion*. New York: Oxford University Press.
- O'Connor, Lynn; Berry, Jack W.; Lewis, Thomas; Mulherin, Kathleen; Cristosomo, Patrice S. 2007. "Empathy and Depression: The Moral System on Overdrive". In Farrow, Tom F.D. and Woodruff, Peter W.R., eds. *Empathy and Mental Illness*. Cambridge: Cambridge University Press, 49-75.
- Paytas, Tyler. 2015. "Rational Beings with Emotional Needs: The Patient-Centered Grounds of Kant's Duty of Humanity". *History of Philosophy Quarterly* 32 (4):353-376.
- Pereboom, Derk (2006). "Kant on Transcendental Freedom". *Philosophy and Phenomenological Research* 73 (3):537-567.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Proeve, Michael and Tudor, Steven. 2016. *Remorse: Psychological and Jurisprudential Perspectives*. New York: Routledge.
- Smith, Nick. 2016. "Dialectical Retributivism: Why Apologetic Offenders Deserve Reductions in Punishment Even Under Retributive Theories". *Philosophia* 44:343-360

- Tangney, J. P. 1991. "Moral affect: The good, the bad, and the ugly". *Journal of Personality and Social Psychology*, 61, 598– 607.
- Thomason, Krista K. 2017. "A Good Enough Heart: Kant and the Cultivation of Emotions". *Kantian Review* 22 (3):441-462.
- Timmermann, Jens. 2006. "Kant on Conscience, 'Indirect' Duty, and Moral Error." *International Philosophical Quarterly* 46, 293–308.
- Um, Sungwoo. 2020. "Gratitude for Being". *Australasian Journal of Philosophy* 98 (2):222-233.
- Walker, Nigel. 1980. *Punishment, Danger, and Stigma: The Morality of Criminal Justice*. Oxford: Blackwell.
- Ware, Owen. 2009. The Duty of Self-Knowledge. *Philosophy and Phenomenological Research* 79 (3): 671-698.
- Wood, Allen. 2008. *Kantian Ethics*. New York: Cambridge University Press.
- Worthington, Everett L. J. *Forgiveness and Reconciliation : Theory and Application*. New York: Routledge, 2006.