

© 2017 Cambridge University Press. This article has been accepted for publication by *Religious Studies*. The published version can be found at: <http://dx.doi.org/10.1017/S0034412516000408>. Please consult published version before citing or quoting.

Bibliographical Reference:

Stewart Clem, "Dropping the Debt: A New Conundrum in Kant's Rational Religion," *Religious Studies* (2017) (in press, DOI: 10.1017/S0034412516000408)

Dropping the debt: a new conundrum in Kant's rational religion

STEWART CLEM

Department of Theology, University of Notre Dame, 130 Malloy Hall, Notre Dame, IN 46556, USA
sclem@nd.edu

Abstract: In this essay, I argue that Immanuel Kant fails to provide a satisfactory account of 'moral debt' in *Religion Within the Boundaries of Mere Reason*. More precisely, he fails to answer the question of why we should assume that a debt exists in the first place. In light of recent scholarship on this area of his thought, I sketch some possible readings of Kant on the nature of moral transformation that suggest how he might account for this debt. I then argue that these accounts fail to justify its existence within Kant's project.

Introduction

A prominent theme in Immanuel Kant's *Religion Within the Boundaries of Mere Reason* is the possibility of moral transformation. For Kant, transformation is a 'change of heart', which involves 'the recovery of the *purity* of the law, as the supreme ground of all our maxims' (6:46).¹ The one who undergoes such conversion is 'upon the road of endless progress toward holiness' (6:47). The transition from 'evil human being' to 'good human being' is thus marked by a 'change of the supreme inner ground of the adoption of all the human being's maxims in accordance with the ethical law, so far as this new ground (the new heart) is itself now unchangeable' (6:51). This change of heart must be possible, because it is a duty required of us (6:67). Yet, Kant is well aware of the difficulties that result from such an explanation.

In the first section of Part II, Kant addresses the idea of a humanity pleasing to God,

which includes not only the notion of a pure will but also freedom from guilt (6:61). The problem facing all human beings is that we possess a propensity to evil (*Hang zum Bösen*),² and thus our starting point is not one of moral purity. But even if conversion is possible, what can we do about the fact that our lives, when taken as a whole, fall short of being pleasing to God? ‘The distance between the goodness which we ought to effect in ourselves and the evil from which we start is, however, infinite, and, so far as the deed is concerned – i.e. the conformity of the conduct of one’s life to the holiness of the law – it is not exhaustible in any time’ (6:66). What results from this is that we carry a moral ‘debt’ before God, and it is not immediately obvious how this debt could ever be removed.

Kant states that the problem of moral debt is ‘apparently the greatest difficulty’ for the notion of a humanity pleasing to God (6:71–72). But the word ‘apparently’ is telling – he thinks that this difficulty can be resolved. In this essay, I critically engage the contemporary scholarship on Kant’s treatment of moral debt. My ultimate purpose, however, is to argue that Kant has failed to answer an even more fundamental question: Why should we assume that a ‘debt’ exists in the first place? In light of the recent scholarship on this area of his thought, I sketch some possible readings of Kant on the nature of moral transformation. I then argue that these accounts fail to justify the role that debt is supposed to play in the process of transformation.

Kant on moral debt: the standard picture(s)

The purpose of the following two sections is to lay out what Kant *does* say regarding the debt, with the aim of showing (in a subsequent section) that he in fact neglects to establish the existence of such a debt. In Part II, Section 1, Division C of *Religion*, he articulates the problem thusly:

Whatever his state in the acquisition of a good disposition, and, indeed, however steadfastly a human being may have persevered in such a disposition in a life conduct conformable to it, *he nevertheless started from evil*, and this is a debt which is impossible for him to wipe out [*diese Verschuldung ist ihm nie auszulöschen möglich*]. He cannot regard the fact that, after his change of heart, he has not incurred new debts as equivalent to his having paid off the old ones. (*Religion* 6:72)

As he also goes on to explain,

The emergence from the corrupted disposition into the good is in itself already sacrifice (as ‘the death of the old man,’ ‘the crucifying of the flesh,’) and entrance into a long train of life’s ills which the new human being undertakes in the disposition of the Son of God, that is, simply for the sake of the good, yet are still fitting *punishment* [*Strafe*] for someone else, namely the old human being (who, morally, is another human being). (*Religion* 6:74)³

Thus, for Kant, moral transformation in some sense involves a transition into an entirely new identity. What remains unclear from these passages alone is just the extent to which conversion results in a ‘new’ human being. In any event, he is unambiguous about the fact that an evil disposition (from which every human being begins) creates a debt that must be accounted for in some way.

To set this problem in context, I have reconstructed the following set of propositions, taken from Parts I and II of *Religion*, as comprising the ‘standard picture’ of Kant on the nature of moral debt. These propositions are not premises of an argument, nor do they generate a specific conclusion when taken together. Nevertheless, they logically hang together, and they present a problem within Kant’s account (which he acknowledges) that requires an additional premise (or set of premises) in order to be resolved. The propositions are as follows:

(a) Human beings possess an innate propensity to evil, resulting in a *Gesinnung* that governs all specific maxims (6:29).

(b) Considered under the rubric of ‘religion,’ moral evil can be categorized as ‘sin’ (i.e. as a violation of God’s commands) (6:72; cf. 6:153).

(c) Our sinful stance carries with it a *debt* that is owed to God (6:72). Hereafter I refer to this as the ‘debt thesis’.

(d) This debt is infinite, because moral evil brings with it an infinity of violations of the law (6:72).

(e) Moral transformation is possible, through a change of *Gesinnung*, after which one becomes, morally, ‘a new human being’ (6:74). Hereafter I refer to this as the ‘new human thesis’.

(f) We carry our debt with us even after this conversion (6:74).

(g) There must be some payment for this debt, some satisfaction of Supreme Justice (6:73).

(h) Atonement cannot be vicarious – no-one can pay another person’s debt (6:72).

The problem that lingers is: what becomes of the debt, then? Or, perhaps we should ask: what *can* be done to erase or remove this debt?

There are various (and conflicting) interpretations of Kant’s resolution to this problem. Some argue that Kant does in fact hold to a vicarious atonement through Christ. Others argue that the ‘new man’ cancels the debt of the ‘old man’ with the assistance of divine grace, but only because the person has first made herself worthy of such assistance. Those who take the latter approach also argue that this creates additional problems for Kant’s account. One of the better-known examples of the former approach is that offered by Firestone & Jacobs (2008).⁴ Their first interpretive move is to make a distinction between the debt of the ‘old man’ and the debt of the ‘new man’. They would modify proposition *f* (see above) to say something like, ‘Our previous debt (that of the ‘old man’) is cancelled in the process of conversion, but we accrue a new debt through our moral wrongdoings that occur after conversion’. According to Jacobs and Firestone, these are two *kinds* of debt, and they are dealt with differently. The debt that exists before conversion is of an infinite nature; the post-conversion debt is finite. They rightly recognize that Kant rejects the possibility of vicarious atonement, but they maintain that this dictum only applies to pre-conversion debt. The problem is not that vicarious payment for debt is vicarious;

rather, the problem is that the debt is infinite. The passing of an infinite debt from one person to another does not objectively do away with the debt itself.

Thus, according to Firestone and Jacobs, the debt that results from our wrongful actions that occur after a change of heart presents a different kind of problem. They note (correctly, in my view) that the original debt is wiped out in the process of conversion (Firestone & Jacobs (2008), 176).⁵ Yet this reasoning somehow leads them to claim, ‘it is entirely possible that Kant thinks post-conversion failings, which bring non-dispositional (and thus finite) guilt, are debts for which atonement can take place’ (*ibid.*, 177). And by ‘atonement’, they mean vicarious atonement. They would therefore modify proposition *h* (see above) to read, ‘Vicarious atonement is not possible for *infinite* debts.’ Furthermore, they argue not only that atonement is possible for Kant, but also that it is possible to read him as affirming an actual atonement by the Son of God.⁶

The problem with this reading of Kant is that it is simply not grounded in the text. The authors rely on a distinction between ‘old’ and ‘new’ debts that Kant himself does not make. For example, they cite *Religion* 6:72, where Kant writes, ‘this **original debt**... cannot be erased by somebody else. For it is not a transmissible liability.’⁷ But Kant is not making a distinction here between debts that exist prior to or post-conversion. The context suggests that by ‘original’, Kant means something like ‘fundamental’, in the sense that none of our good actions can take away the debt that has already accrued. Kant’s word *vorhergehende* should be understood as ‘foregoing,’ or ‘preceding’; in other words, the debt always precedes our good actions, such that it is foolish to think that we can pay an infinite sum with something that is of finite worth. Kant’s words – omitted by Firestone and Jacobs with an ellipsis – reveal as much when he describes this ‘original debt’ as that which ‘precedes whatever good a human being may ever do’ (6:72).

Moreover, Kant never implies that a debt accrues to us after a change of heart. Indeed, he seems to suggest the opposite when he writes, '[A person] cannot regard the fact that, after his change of heart, he has not incurred new debts as equivalent to his having paid off the old ones' (6:72). The authors also fail to account for the noumenal-phenomenal distinction as it applies to moral transformation.⁸ We must keep in mind that, for Kant, our change of disposition is not transparent to us. There is not an observable, one-to-one correspondence between our actions and a change of heart; we cannot confidently draw a temporal line marking the 'before' and 'after' of conversion. Yet, Firestone and Jacobs write, 'If then, failings can occur after the dispositional revolution (which Kant's dispositional philosophy presumes is possible), these debts do not carry the infinite guilt of the disposition, which persists throughout time. They are *finite* debts' (Firestone & Jacobs (2008), 177). What is wrong with this assessment is that Kant does not use the language of debt in his description of post-conversion moral failings, and this basic failure leads the authors to develop a faulty interpretation of Kant on atonement.

Jacqueline Mariña similarly interprets Kant as holding to vicarious atonement (Mariña (1997), 379-400). Yet, going beyond Firestone and Jacobs, she claims that the prototypical 'Son of God' (Kant (1998), 6:74) is to be identified with the historical figure of Jesus Christ.⁹ She cites a cryptic passage in *Religion* in which Kant declares the good principle 'appeared in an actual human being', and 'that human being opened the doors of freedom to all who, like him, choose to die to everything that holds them fettered to earthly life to the detriment of morality' (*ibid.*, 6:82). She infers from this passage, 'Here Kant moves away from the idea that the moral idea lying in our reason *alone* contains both the necessary and sufficient conditions for humanity's successful attainment of the moral goal' (Mariña (1997), 395). This is not because Kant has unwittingly slipped into fideism, but rather because '[s]uch an atonement is necessary

because we cannot possibly repay the debt of sin' (*ibid.*, 390). She reads Kant's analysis of debt and atonement as an antinomy (*ibid.*, 389),¹⁰ in which he ultimately favours the necessity of a vicarious atonement. While I do not have the space to engage Mariña's analysis at length, I would only suggest that she has put too much stock in Kant's imagery, which he ultimately intends to strip away in the interests of pure rational faith.¹¹ While Mariña does not rely on Firestone and Jacob's distinction between pre- and post-conversion debt, all three authors share a commitment to an actual, vicarious atonement as Kant's solution to the debt problem.

Other scholars reject this interpretation of Kant in favour of a divine assistance model. That is, they read Kant as positing the need for external assistance in order to satisfy the debt of the 'old man'. But this assistance does not come in the form of a vicarious atonement. Rather, assistance comes as a capacity to atone for one's own previous failings. Representatives of this reading include John Hare (Hare (1996), 56–60), Gordon Michalson (Michalson (1990), 107–124), Philip Quinn (Quinn (1986), 440–462) and Nicholas Wolterstorff (Wolterstorff (2010), 56–67).¹² I do not wish to obscure the differences in each of their interpretations – they are not all claiming the same thing in the same way. Nevertheless, they recognize that Kant must resolve the debt problem somehow, and their readings share the common feature of affirming my propositions *f*, *g*, and *h* above. Michalson articulates one way to understand this position when he explains, 'Christ *does* atone for the sin of radical evil, but his atoning function turns out to reside in a rational feature that we can make our own: to appropriate Christ as the one who atones is in fact to make manifest and gain access to a feature of our own moral structure' (Michalson (1990), 114). Thus, with divine assistance, the 'new man' atones for the sins of the 'old man', not by paying off the debt with good deeds (which Kant maintains is impossible), but by enduring the consequences of the old man's actions even after conversion.

The point of these analyses has been to illustrate the standard picture of Kantian moral debt. While I have outlined two different approaches of arriving at this picture, all of the authors I have mentioned in this section share a similar set of interpretive commitments when it comes to this area of Kant's thought. All of them affirm the propositions I outlined earlier in this section, with slight modification in some cases. In the following section, I turn to a recent account that challenges this picture.

A challenge to the standard picture

In his recent article, 'Kant on the Debt of Sin', Lawrence Pasternack offers a new reading of Kant on the debt problem (Pasternack (2012)). Through a careful exposition of several key passages, he demonstrates that Kant in fact does not claim that the debt of sin remains after conversion. In short, he denies proposition *f*, above. He argues that the debt is neither 'forgiven', nor 'paid' by anyone else – not even by the 'new man' after conversion. In my estimation, his explication presents a serious challenge to both sets of interpreters – the 'vicarious atonement' group and the 'divine assistance' group – on their readings of proposition *h*. Specifically, Pasternack's reading takes seriously Kant's claim that atonement cannot be vicarious (*pace* Mariña, Jacobs, and Firestone). Moreover, contra the 'divine assistance' interpreters, his account shows that the idea of the 'new man' atoning for the 'old man's' debt would constitute a vicarious atonement, and would thus be a violation of Kant's own principle.

According to Pasternack, Kant does not ascribe debt to the new man. 'However, it may be argued that the old man's debt lingers within him, and so divine grace is still needed. But this interpretation, I do not think, is one that the text forces upon us; and [...] it leads Kant into a

conflict between the need for a forgiving grace and his views on divine justice' (*ibid.*, 45). Instead, Pasternack proposes that the new man is not 'forgiven' of his debt, but rather 'relieved' of it, because it can no longer rightly be attributed to him. This gives rise to a new resolution of the debt problem: '[S]ince the moral characteristics which put the old man into the state of sin are not present in the new man, the new man does not bear the debt. In the event of a moral transformation, the debt just gets wiped off the books, so to speak' (*ibid.*). There is no transfer of debt, either to the new man or to a vicarious substitute.

Pasternack supports this claim by illustrating what Kant means in stating that the new man is 'relieved [*entschlagen*] of all responsibility... though fully in accord with divine justice' (*Religion* 6:76). Kant does not use the word 'forgiven' (*vergeben*) to describe what happens to the debt. It is as if the 'old man' has literally died, and there is no-one else who bears the responsibility of the debt. This is what Kant means when he says, 'the punishment must be thought as adequately executed in the conversion itself' (6:73). Whatever relation exists between the old man and the new man, it is not one of moral responsibility. Nonetheless, Pasternack speculates that 'the new man still serves as a proxy, trying to right the wrongs committed by the old man. Though from the divine perspective, the debt of sin is gone, from the phenomenal perspective, there is much the new man takes up on behalf of the old' (*ibid.*, 50). But none of this should obscure the 'new human thesis', i.e. Kant's claim that the new man is 'morally, a new human being' (*Religion* 6:74).

While some may find this reading perplexing, given that there are places where Kant appears to suggest otherwise, it is consistent with Kantian principles. Where Kant appears to suggest that the new man bears the old man's punishment (e.g. the 'long train of life's ills' in *Religion* 6:74), for example, he is really referring to the enduring *effects* of the old man in the

phenomenal realm. In my judgement, Pasternack's reading offers a compelling alternative to the standard picture, and thereby challenges a significant portion of contemporary Kant scholarship. He modifies the standard picture, first by rejecting proposition *f*, and then by resolving what remains with the additional premise that the moral debt is stricken (*entschlagen*) from the books. I think that Pasternack is correct insofar as this is what Kant himself wants to argue.

A new conundrum

While I endorse Pasternack's account as an accurate reading of Kant, I propose that it gives rise to new difficulties. That is, if this reading is correct (and I think it is), then Kant faces a bigger problem than has been previously recognized. For my purposes, I am interested in the question of why Kant assumes that sin creates a 'debt' in the first place. In the course of analysing this question, I wish to demonstrate that the 'debt thesis' (proposition *c*) is interdependent with the 'new human thesis' (proposition *e*), which in turn threatens the overall coherence of Kant's account.

There appear to be two possible ways forward, in light of the revised picture we now have before us. The first is that we could downplay Kant's language of 'another human being' resulting from conversion. Even after adopting a new, good disposition, the person is still numerically the same. But then it is difficult to make sense of the debt being relieved, or annihilated, along with the annihilation of the old man. Alternatively, we could take Kant's distinction between 'old man' and 'new man' literally. That is, while we are never granted epistemic certainty that we have undergone a genuine change of heart, the person who has adopted a new *Gesinnung* is a numerically distinct being (noumenally, in the eyes of God). But

what could it mean to become a new, numerically distinct person? How could this fit into Kant's philosophical framework? Would the notion of 'conversion' even remain intelligible?

Others have noted that Kant's language of becoming 'another human being' is imprecise. Therefore we should not read too much into it, nor can we criticize it directly. In order to analyse this notion, we would first need to speculate as to what Kant means by 'another human being', but then we would be criticizing something that Kant never actually claimed (although the possibility still remains that the notion is ultimately an empty one). I believe that we *should* be critical, however, if Kant were to put this notion to work by using it to explain some other aspect of his project. And it is clear that he does.

Whatever Kant means by the person's transformation into a new human being, it is significant enough such that the debt of sin is completely 'stricken' from the record. We should be mindful of the qualifier 'morally' when Kant says that the 'new man' is *morally* another human being (6:74). This suggests that the person is not a numerically distinct, noumenal being in the fullest sense. It most likely means that God, who has access to our true, noumenal selves, discerns a change of heart, a new *Gesinnung*, which grants a person a new moral status. Yet, the language of 'debt', while also imprecise, seems to carry some metaphysical baggage. The very notion of a debt is something that one carries around with him – it lingers, it has a residue.¹³ But Kant wants to say that this debt is annihilated, because the 'old man' who carried this debt is annihilated. This interpretive difficulty is one that Pasternack acknowledges but does not wholly resolve; that is, Kant appears to claim, on the one hand, that the new man is numerically distinct, and, on the other, that the new man simply *is* the old man albeit with a new moral status. Pasternack writes, 'The *moral status* of the old and new man's *Gesinnungen* are fundamentally different and it seems that by virtue of this qualitative difference, in both *Religion* and in *The*

Conflict of the Faculties, Kant presents the old man and the new as *numerically distinct* (Pasternack (2012), 48, emphasis mine).¹⁴ This is not to suggest that ‘new moral status’ and ‘numerically distinct’ are interchangeable terms. But Kant’s presentation unhelpfully lends itself to such an interpretation, which only creates more problems than it solves.

The language of debt creates an unnecessary tension in Kant’s account – it puts too much pressure on the concept of a ‘new man’ existing after conversion, to the point that we must posit the creation of a numerically distinct being if the moral debt is to be truly erased. I think it is safe to assume this is not something Kant really wants to affirm. If he does, then this would raise a new set of interpretive (and evaluative) problems. On the other hand, if we want to interpret Kant’s ‘morally’ new human being in the sense I’ve articulated above, then it becomes unclear *what* the moral debt could possibly be.

Pasternack mentions the problem I am addressing here, but only in passing. He observes, ‘Kant takes it for granted that sin [*Sünde*] carries a debt to God. He never explicitly argues for it but rather assumes that through sin, we incur a debt that needs to be repaid’ (*ibid.*, 33 fn. 9). Moreover, ‘Although we can glean from Kant an argument as to why the debt of sin is infinite, there is little to be said about why sin involves debt *simpliciter*. My suspicion is that he regarded their relationship as analytic and is merely following a long-standing convention’ (*ibid.*). This may be so, but I do not think that Kant should be let off the hook so easily. Pasternack’s reading on this point seems plausible enough, given that no other alternative readily presents itself. But if Kant did in fact harbor this assumption, then should we not question whether it is warranted on his own terms?

Where could the assumption come from that ‘sin’ analytically contains the notion of ‘debt’, if not from the Judeo-Christian tradition? Pasternack readily acknowledges this

connection. He cites Gary Anderson's book, *Sin: A History* (Anderson (2009)) as reliable source for tracing the development of 'debt' language to describe sin in Second-Temple Judaism and in portions of the New Testament. Over time, this language became firmly ensconced in the Christian tradition, but as Anderson notes, there is no *necessary* connection between sin and debt. Thus, while Pasternack offers this insight as a 'philological explanation' for why Kant assumed that sin creates a debt, I would argue that this is precisely where we should hold Kant to his own standards and ask why, on the principles of *rational* religion, we should assume that there is a 'debt of sins' (6:72) in the first place. In my judgement, Kant is unreflectively stipulating a feature of historical faith and confusing it with a principle of rational faith.

Kantian reasons for the existence of moral debt?

One might reasonably object, of course, that even if I am right that Kant has failed to articulate an analytic relationship between sin and debt, there may still be good reasons for affirming such a relationship – perhaps even good Kantian reasons for doing so. While it is possible that a case can be made for the concept of debt as inherent to the notion of sin, i.e. in a sense that is *compatible* with Kant's arguments, my aim in the present section is to show that such strategies are dubious and, more importantly, do not diminish the force of my central argument, namely that Kant has failed to establish the existence of moral debt and that its inclusion in his account of transformation ultimately violates his own principles of rational religion.

To be fair, we must acknowledge that Kant does understand that the nature of moral debt is different from other types of debt, such as financial debt. It is likely that 'moral debt' is to be understood as a metaphor. At the very least, the language of moral debt does capture certain intuitions about wrongdoing, including the notion that we often 'owe' something to the one

whom we have wronged. Kant is consistent in this regard in that he does not allow for the possibility of ‘transferring’ moral debts (unlike financial debts), thus proscribing the possibility of vicarious atonement for another’s sins. Yet there is no doubt that for Kant this is not *merely* a metaphor. The infinite debt of sin carries with it an infinite guilt, by virtue of which ‘every human being has to expect *infinite* punishment and exclusion from the Kingdom of God’ (6:72). The notion of debt cannot simply be another way of saying that we ‘owe’ something to a person. Our specific moral obligations to other persons are contingent upon circumstances, and when we fail to meet those obligations we fail to give something that was owed to them. But this does not necessarily mean that the failed obligation ‘creates’ a debt to the person (i.e. some additional *thing*) and it cannot be translated *simpliciter* into a transactional account of moral obligation.

One might be tempted to find a Kantian source for the idea of moral debt in relation to the highest good and the *sui generis* duty of human beings to promote this good (6:97). Echoing St. Anselm’s famous argument in *Cur deus homo*, perhaps the Kantian promotion of the highest good could be read as analogous to the Anselmian promotion of God’s honor. On such an account, the infinity of goodness, impinged upon by human wrongdoing, is the cause of the debt (and of the debt’s infinite nature). One might even appeal to Kant’s affirmation of eternal punishment in *The End of All Things*, in which he speculates that ‘for eternity we would have to anticipate for ourselves the consequences suiting that merit or guilt under the dominion of the good or evil principle’ (Kant (1996b), 8:330).¹⁵ But Kant explicitly states that it is *not* the infinity of the highest lawgiver (or the highest good) that creates infinite guilt, but rather ‘because the evil is in the *disposition* [of the person] and the maxims in general (in the manner of *universal principles* as contrasted with individual transgressions)’ (6:72). Thus, the infinity of the debt is properly a reflection of the universality of one’s maxims, as a feature of one’s noumenal self. But

to whom is the debt owed, then? The moral debt cannot be construed as a debt owed to God (either literally or by analogy) on strictly *analytical* principles.

One obvious place to turn, of course, is the Bible. While much of the debt language used to describe sin and guilt Western Christianity was not developed until the Middle Ages, there are biblical allusions to this notion scattered throughout the Old and New Testaments.¹⁶ For example, in his epistle to the Colossians, St. Paul refers to humanity's 'debt' against God being 'wiped' or 'blotted out' (ἐξάλειψας) by Christ's sacrificial death (Col. 2:14).¹⁷ Thus, while Kant would reject a vicarious atonement, in both the New Testament and in Kant there is a change in relation that allows one's debt to be wiped out. For Kant, this change is brought about by a restoration of each individual's relationship to God, instigated by the Change of Heart rather than through the actions of the literal person of Christ. These considerations are interesting from a historical and textual standpoint, but they cannot help us in constructing a possible Kantian account of the analytic relationship between wrongdoing and debt. If such an account is grounded primarily in biblical revelation, then it must be ruled out on Kant's own principles.

'Debt' or 'guilt'? A tension in *Religion* 6:72

Some may wish to defend Kant by observing that the speculative connections I have considered above are not the only resources for establishing a link between sin and debt. Even if these speculations are ultimately unsatisfying, and even if the link between sin and debt cannot be established analytically, one might argue that their relationship can be established through a synthetic a priori judgement. Kant, of course, does not claim this, but perhaps such a move is implicit in his argument. It is unclear to me, however, the criteria by which we might establish a

synthetic a priori connection between sin and debt. Regardless, I believe that there is sufficient reason to deny such a connection, given the way Kant describes moral debt in *Religion* 6:72.

Rather than supposing that there is a synthetic a priori relationship between sin and debt, I suggest that Kant in fact wishes to decouple these concepts. More precisely, Kant is plagued by the valence of the German word *Schuld*, which can connote both ‘guilt’ and ‘debt’. On my reading of *Religion* 6:72, Kant ultimately wishes to underscore the notion of *guilt* that attaches itself to moral transgression, even if he is inconsistent in his use of the term *Schuld*. More importantly for the purposes of this essay, by demonstrating that guilt (as opposed to debt) is Kant’s ultimate concern in 6:72, this reading supports the thesis that there is no inherent connection (either analytic or synthetic a priori) between sin and debt, which makes the appearance of ‘debt’ language in *Religion* all the more peculiar.

One locus of textual support for my interpretation can be found in Kant’s rejection of the transference of *Schuld*. Kant is adamant that the ‘debt’ of sin cannot be transferred, yet he never states why. Perhaps this is supposed to be intuitive, since it is far from obvious how such a debt could be transferred in the first place; yet we know empirically that financial debts are transferred quite regularly. Yet, if this goes to show just how *disanalogous* these two senses of debt really are, then perhaps it is best to read Kant as favouring another sense of *Schuld*: guilt. If so, then we can more readily understand why Kant would assume (rather than explain) the fact that *Schuld* cannot be transferred. It is unintelligible to speak of ‘guilt’ being transferred, especially given the way Kant deploys the concept in *Religion*.¹⁸ This is why the *Schuld* is ‘the most personal of all liabilities, namely a debt of sins which only the culprit, not the innocent can bear, however magnanimous the innocent might be in wanting to take the debt upon himself for the other’ (6:72), as opposed to financial debts, which are transmissible.

We can gain further insight into Kant's conception of *Schuld* by considering what makes it infinite. Kant explains that moral transgressions produce 'an infinity of guilt'.¹⁹ As mentioned above, he is explicit in his rejection of the notion that this infinite guilt is a direct result of the infinity of the highest lawgiver; the cause is rather that 'the evil is in the *disposition* and the maxims in general' (6:72). If Kant had wanted to emphasize the debt-like nature of *Schuld*, then we might expect him to ground the infinity of moral transgression in the infinity of the highest lawgiver. The fact that he rejects this connection supports my reading of Kant as wanting to emphasize the guilt-like nature of *Schuld*. If the essence of wrongdoing lies in the offense to a higher authority, then it makes sense to describe it as a 'debt' that is owed to someone. 'Guilt', however, is more subjective, i.e. it describes a feature of the moral agent in and of himself, as opposed to the interpersonal nature of a debt.²⁰

Almost as an afterthought, Kant writes, 'Consequently, every human being has to expect *infinite* punishment and exclusion from the kingdom of God' (6:72). Kant's reasons for maintaining a conception of afterlife have been much discussed, and I will not rehearse that discussion here. What is important to note is the significance of *punishment*. If Kant's *Schuld* is to be read as favouring its debt-like nature, then it seems strange for him to speak of 'punishment', as opposed to, say, infinite 'payment'. Obviously, such terms are being used here in the abstract, and Kant does not describe what infinite punishment (nor infinite debt repayment) might look like. Nonetheless, his brief remark that the evil in one's disposition warrants infinite *punishment* suggests that Kant favours *Schuld* as guilt rather than debt.

Given these observations in the preceding section, we can conclude that Kant wishes to frame the sin-*Schuld* relationship more narrowly as a sin-guilt relationship. I should also clarify what I do *not* intend to claim, namely, that Kant has simply been misread on this point. I am not

suggesting that he has unwaveringly described *Schuld* in guilt-like terms – far from it, in fact. There is no denying that he uses unambiguous debt language in places, for example, when he writes that the New Man ‘cannot regard the fact that, after his change of heart, he has not incurred new debts [*neue Schulden*] as equivalent to his having paid off [*bezahlt*] the old ones’ (6:72). What I am suggesting is that Kant has inherited a moral category that creates a tension when placed within the context of his unique philosophical project. This tension is evident in the tradition of *Schuld* itself: from the multivalence of the German word and from the connotations it has inherited through Christian theology and biblical interpretation. Kant never resolves this tension – if he is even aware of it. He vacillates between the different senses of the word, although, as I have suggested in this essay, there is only one that is consistent with his overall project in *Religion*. As I have also suggested, it is this ‘guilt’ sense that Kant ultimately favours, if only implicitly.

Conclusion: dropping the debt

In light of the foregoing analysis, it seems that the only way for Kant to maintain consistency is to drop the notion of debt altogether. It remains unclear what work it is doing other than serving as a metaphor for the evil *Gesinnung* itself. If so, it is a terribly unhelpful metaphor. But the fact that Kant dedicates an entire ‘difficulty’ to the notion of debt (stemming directly from issues developed in the ‘first difficulty’) suggests that it is meant to play a more constructive role. Moreover, on my reading, the language of the ‘new human thesis’ only emerges as a result of the ‘debt thesis’ (i.e. the latter necessitates the former).²¹ But dropping the ‘debt thesis’ would help to disambiguate the ‘new human thesis’ by making it less mysterious, namely, by clarifying that it is something much softer: a change in moral status.

While one cannot simply read Kant's mind on this point, the appearance of debt language does seem to be an instance of his smuggling a Christian theological concept into his schema of rational religion. Goethe once famously declared in his personal correspondence, 'It took a lifetime for Kant to wash his philosophical cloak of many stains and prejudices. But now [in his *Religion* book] he has wantonly slobbered on it with the stain of radical evil, just so that Christians, too, might be enticed to kiss its hem.'²² Perhaps Goethe was right, at least on this particular point. If we drop the notion of debt from Kant's account, the only significant result would be one less item overlapping the 'two circles' (historical religion and pure rational religion) articulated in the Second Preface of *Religion*.

I am in agreement with Pasternack's rejection of faulty readings of Kant on moral debt. I am also in agreement with Hare, Wolterstorff, et al., that Kant's notion of atonement contains a conundrum. The contribution of the foregoing analysis, however, is to show that Kant's conundrum lies at a far more fundamental level than has been previously recognized. Gordon Michalson once wrote, 'Strictly speaking, there is no reason whatsoever why a christology should appear in the *Religion*, since the book is represented as an inquiry into the moral content of historical religions in general, rather than as a moral interpretation of Christianity alone' (Michalson (1990), 109). While Michalson's assessment concerns Kant's Christology *per se*, I have made a stronger claim concerning Kant's entire notion of moral debt. I contend that not only is there no reason for the notion of debt to appear in Kant's *Religion*, but that this addition is unwarranted on his own grounds and only serves to problematize Kant's account of moral transformation.²³

References

- ANDERSON, GARY (2009) *Sin: A History* (New Haven, CT: Yale University Press).
- CHIGNELL, ANDREW (2011) 'The Devil, the Virgin, and the Envoy: Symbols of Moral Struggle in *Religion*, Part Two, Section Two', in Otfried Höffe (ed.) *Die Religion innerhalb der Grenzen der Bloßen Vernunft (Klassiker Auslegen)* (Berlin: Akademie Verlag), 111-129.
- CHIGNELL, ANDREW (2012) 'On Defending Kant at the AAR', *Faith and Philosophy*, **29**, 144-150.
- FIRESTONE, CHRIS L. & NATHAN JACOBS (eds) (2008) *In Defense of Kant's Religion* (Bloomington, IN: Indiana University Press).
- HARE, JOHN (1996) *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance* (New York: Oxford University Press).
- KANT, IMMANUEL (1900 –) *Kant's gesammelte Schriften*, 29 vols (Berlin: Walter de Gruyter).
- KANT, IMMANUEL (1996a) *Practical Philosophy*. Mary J. Gregor (ed.) (Cambridge: Cambridge University Press).
- KANT, IMMANUEL (1996b) *Religion and Rational Theology*, Allen W. Wood (ed.) (Cambridge: Cambridge University Press).
- KANT, IMMANUEL (1998) *Religion within the Boundaries of Mere Reason and Other Writings*, Allen Wood and George di Giovanni (eds) (Cambridge: Cambridge University Press).
- KUEHN, MANFRED (2014) 'Kant's Jesus', in Gordon Michalson (ed.) *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide* (Cambridge: Cambridge University Press), 156–174.
- MARIÑA, JACQUELINE (1997) 'Kant on Grace: A Reply to His Critics', *Religious Studies*, **33**, 379-400.
- MICHALSON, GORDON (1990) *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge: Cambridge University Press).
- MUCHNIK, PABLO (2009) *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the Apriority of History* (Lanham, MD: Lexington Books).
- PASTERNAK, LAWRENCE (2012) 'Kant on the Debt of Sin', *Faith and Philosophy*, **29**, 30-52.
- QUINN, PHILIP (1986) 'Christian Atonement and Kantian Justification', *Faith and Philosophy*, **3**, 440-462.
- WOLTERSTORFF, NICHOLAS (2010) 'Conundrums in Kant's Rational Religion', in Terence Cuneo (ed.) *Enquiring About God: Selected Essays, Volume 1* (Cambridge: Cambridge University Press), 56-67.

¹ All English quotations of Kant's works are from the *The Cambridge Edition of the Works of Immanuel Kant* (Kant (1996a), (1996b), and (1998)). References to the German text are from *Kant's gesammelte Schriften* (Kant (1900 –)).

² See 6:29ff; see also *Groundwork of The Metaphysics of Morals* 4:405 (in Kant 1996a). On Kant's notion of the propensity to evil, I have been aided by Muchnik (2009), especially Chapter 3, 'Radical Evil, Inscrutability, and Moral Self-Constitution'. One of Muchnik's key interpretive

tools is to distinguish clearly between the notions of *Hang zum Bösen* (which describes humanity as a species) and *böse Gesinnung* (which describes the individual agent).

³ See also *The Conflict of the Faculties* 7:54 (in Kant 1996b), where Kant states, ‘[T]he end of religious instruction must be to make us *other* human beings and not merely better human beings.’

⁴ This book has attracted a great deal of attention in Kant scholarship, much of it negative. For a collection of critical engagements, see the symposium in *Faith and Philosophy*, **29**, especially Chignell (2012). I have not come across a critical engagement specifically with Firestone and Jacob’s interpretation of Kantian moral debt, which is my concern in this section.

⁵ In Kant’s own words, ‘the punishment [for the original debt] must be brought about as adequately executed in the situation of conversion itself’ (*Religion* 6:73, in (Kant 1998)). This aspect of Firestone and Jacob’s account actually comes quite close to the view of Lawrence Pasternack, which I discuss below.

⁶ ‘[I]f Kant’s prohibition on vicarious atonement applies only to infinite dispositional guilt, then Kant’s language of vicarious atonement in this passage may refer to an actual atonement performed by the prototype,’ in *ibid.*, 178.

⁷ Quoted in *ibid.*, 177 (emphasis added by the authors).

⁸ See, among other places, 6:41, 6:48, and 6:75 where Kant discusses the atemporality of our dispositions. See also Hare (1996), 58-59.

⁹ For a recent and helpful study on the historical figure of Jesus in relation the figure of Christ in Kant’s *Religion*, see Kuehn (2014).

¹⁰ Kant does use the term ‘antinomy’ in his discussion of saving faith in *Religion* 6:116-117 (incorrectly cited by Mariña as 6:106-107).

¹¹ For an excellent analysis of such imagery, see Chignell (2011).

¹² The title of this article is drawn from Wolterstorff’s usage of ‘conundrum’ here.

¹³ Note the language Kant uses to describe it: ‘[This debt is] the most personal of all liabilities, namely a debt of sins which only the culprit, not the innocent, can bear, however magnanimous the innocent might be in wanting to take the debt upon himself for the other’ (*Religion* 6:72).

¹⁴ I should underscore the fact that Pasternack is not granting that the moral shift to the ‘new man’ entails a shift in numerical identity; he is only acknowledging that Kant’s presentation appears to suggest this. Pasternack’s primary concern here is to show that the moral shift (however understood) is sufficient to understand why the debt of sin is relieved.

¹⁵ It should also be noted that Kant shows his typical restraint in restricting the value of his speculations on the afterlife to *practical* considerations rather than metaphysical certainties. See 8:329 and 8:332, *Note*.

¹⁶ See Anderson (2009) for examples and analysis.

¹⁷ In the Luther Bible, the most widely known German translation that would have been familiar to Kant, the Colossians text uses the word *ausgetilgt* for ‘wiped out’ rather than Kant’s preferred *entschlagen*. The words share an overlapping valence, of course, and it is also possible that some other translation (or some other theological text) familiar to Kant uses *entschlagen* in a biblical context. I am grateful to Lawrence Pasternack for drawing my attention to these biblical connections.

¹⁸ Contemporary psychologists speak of ‘guilt transference’ and ‘blame shifting’ as forms of psychological projection, but the meanings of these terms derive from unrelated contexts. Regardless, such concepts would have been foreign to Kant and his contemporaries.

¹⁹ George di Giovanni translates *Schuld* as ‘guilt’ in this passage, but he translates it as ‘debt’ elsewhere. The German text reads, ‘Da nun das Sittlich=Böse [...] eine Unendlichkeit von Verletzungen des Gesetzes, mithin der Schuld bei sich führt’ (6:72).

²⁰ We often find a similar distinction in legal settings: another party can assume someone else’s debt, but not another person’s *guilt*. See, for example, 11 U.S. Code § 101.

²¹ Even when Kant uses the language of a ‘new’ human being in *The Conflict of the Faculties* 7:54 (a later work), he seems to have in mind a change of supreme maxim, as opposed to a change of mores – a distinction developed in *Religion*.

²² Johann Wolfgang von Goethe, letter to Johann Herder of June 7, 1793 (quoted in Chignell (2012), 144).

²³ I am grateful to Gerald McKenny, Lawrence Pasternack, Jean Porter, and an anonymous reviewer for their insightful criticism and comments on earlier versions of this article.