

Satne, P. 'Forgiveness and Punishment in Kant's Moral System' in Larry Krasnoff, Nuria Sanchez Madrid and Paula Satne (eds.), *Kant's Doctrine of Right in the 21st Century*, Cardiff: University of Wales Press, 2018. www.uwp.co.uk

11

Forgiveness and Punishment in Kant's Moral System

Paula Satne

University of Manchester

Introduction

Forgiveness as a positive response to wrongdoing is a widespread phenomenon that plays a role in the moral lives of most persons. Surprisingly, Kantⁱ has very little to say on the matter. Although Kant dedicates considerable space to discussing punishment, wrongdoing and grace, he addresses the issues of *human* forgiveness directly only in some short passages in the *Lectures on Ethics* and in one passage of the *Metaphysics of Morals*.ⁱⁱ As noted by Sussman, the TL passage, however, betrays some ambivalence.ⁱⁱⁱ Kant establishes a duty of virtue to be *forgiving* (TL, 6:460), yet he immediately warns against its excess: meek toleration of recurrent wrongs could manifest a lack of self-respect and a violation of a duty to oneself (TL, 6:461). Sussman claims that this ambivalence ultimately arises from the fact that forgiveness sits uncomfortably in Kant's moral thought. First, forgiveness has an 'ineluctably elective aspect'^{iv} that makes it, to a certain extent, arbitrary and dependent on particular features of the forgiver's psychology and, as such, in tension with Kant's central claims that human beings are autonomous agents capable of determining their own moral status. Second, according to Sussman, Kant's moral retributivism, i.e. 'the particular moral position that every moral wrong against another deserves punishment of the wrongdoer'^v seems to be in tension with the possibility of a 'truly redemptive forgiveness'.^{vi} Moreover,

forgiveness also seems to be in tension with a passage of the *Religion* in which Kant argues that the moral guilt from our original evil disposition cannot be understood as a debt or liability that can be compensated, erased, transferred or otherwise wiped out by others (*Rel*, 6:72). Thus, to the extent that forgiveness might be thought to involve the forgoing of moral guilt, it seems incompatible with Kant's views on culpability and punishment.

This chapter seeks to clarify Kant's views on forgiveness in order to show that, although not often appreciated, personal forgiveness plays an important role in the lives of ordinary human agents as understood by Kant. In particular, I aim to show there is a conception of forgiveness available to Kant that is not incompatible with Kant's views of punishment and culpability. In Section 1, I argue that, for Kant, far from being merely 'elective,' forgiveness is, under certain *conditions*, morally required. I provide a brief summary of an interpretation of Kant's theory of forgiveness that I have defended in recently published work^{vii}, in order to argue that Kant's duty to be *forgiving* should be understood as an *imperfect* duty of virtue which is *conditional* on repentance. Kant is not ambivalent about this duty because he maintains that when the relevant conditions are not met, we have a perfect duty to ourselves not to forgive unrepentant wrongdoers. The *TL* passage thus identifies two different duties. In Section 2, I show that forgiveness, as conceptualised by Kant, does not require the forgoing of punishment or the overcoming of moral guilt and that this could, in fact, be seen as an attractive feature of Kant's position. I end by offering a very brief assessment of Kant's views.

1. Kant on forgiveness

My reconstruction of Kant's argument in support of a duty to be *forgiving*^{viii} appeals to various aspects of Kant's philosophy, including his theory of rational agency, his views on radical evil and moral development and the formula of humanity. In the *TL* passage, Kant

provides only two considerations in support of this duty. Kant says that men have this duty ‘partly because a man has enough guilt of his own to be greatly in need of pardon and partly, and indeed specially, because no punishment, not matter from whom it comes, may be inflicted out of hatred’ (*TL*, 6:460-1). Admittedly, these remarks are obscure and certainly do not amount to a fully stated argument. The second remark makes the somewhat modest point that punishment inflicted out of hatred would be a matter of mere revenge and, as such, unjust. The first remark seems to be more substantive. The passage has sometimes been interpreted as implying that we should forgive each other because we all are to a certain extent evil (guilty) and then used to support the claim that the Kantian duty to forgive is unconditional.^{ix} Here, for reasons of space, I will not be able to provide a full discussion of this line of interpretation. There is some textual support for this reading but I believe that ultimately, the passage taken in isolation is not only too cryptic to provide a full argument but also not very plausible. If the point that Kant wishes to make is that we should forgive each other unconditionally because we are all to a certain extent evil, then it strikes me as unconvincing and probably a *non-sequitur*: if we are all evil, we might as well not forgive anyone. What would be the point of forgiving others? Is Kant inviting us to take a resigned attitude towards our morally tainted nature? Perhaps the thought is that we should forgive others with the hope that others will, in turn, forgive us for our failures. The problem with these suggestions is that if we are all evil, then forgiveness could not guarantee such reciprocity. Given the obscurity of the text and the philosophical implausibility of the interpretation under consideration, I believe that the passage should not be read in isolation. Instead, we need to appeal to different strands of Kant’s practical philosophy in order to provide a reconstruction of a recognisable Kantian theory of forgiveness.

The first important aspect that we need to consider refers to Kant’s theory of rational agency. The central claim of this theory is that agents act on self-given subjective principles

called maxims (*G*, 4:422). Rational agents possess a will which Kant equates with practical reason, understood as the faculty of acting through the conception of a principle (*G*, 4:412). Insofar as agents act rationally, they choose their maxims in accordance with imperatives, that is, objective principles that hold for all rational beings and instruct them how they ought to act (*G*, 4:413). Maxims have a certain degree of generality in the sense that they provide reasons for actions in the form of principles that apply to relevantly similar circumstances. Maxims are thus principles that connect some generic description of circumstances (taken broadly to include inclinations and purposes of the agent) with some generic description of an action type that the agent takes these circumstances to warrant. Maxims are principles of action that generate, explain and – crucially – justify behaviour. Agents act by judging that some aspect of their circumstances justifies a certain type of behaviour. The process by which agents adopt maxims could involve different *degrees* of conscious decision: the more rational an agent, the more consciously aware she is of her maxims. But Kant also claims that we are sometimes uncertain about the true content of our own motives (*G*, 4:407; *Rel*, 6:20), which means that we are not always explicitly aware of the maxims that we adopt. Maxims can be adopted tacitly, implicitly and, in many cases, retroactively.^x However, as maxims are a product of our freedom and principles for which we are responsible, we can and should become aware of them through reflection.^{xi} In fact, a sign of rationality and of a good character is a willingness to take responsibility for one's actions by constantly examining and revising one's maxims. Agents should thus reflect on the moral (and prudential) adequacy of their maxims in light of the requirements of categorical (and hypothetical) imperatives. Kant goes as far as saying that the *first command of all duties to oneself* is 'know (scrutinize, fathom) *yourself* ... in terms of your moral perfection in relation to your duty' (*TL*, 6:441). Kantian ethics is, thus, an ethics of principles that recommends self-reflection and ultimately self-reform requiring that we become aware of our maxims and attempt to get rid of those

ones that on reflection we do not fully endorse. The next step in the argument will be to note that this process of reflecting and reforming our maxims constitutes a fundamental aspect of our moral development, given that the default position of the will is a position of ‘radical evil’.

In Kant’s later writings, it becomes clear that maxims can have different levels of generality, implying that agents act not only under maxims but also under a system of maxims that form a hierarchy, with the more particular maxims fitting under the more general ones. Caswell has noted that although higher-order maxims rationally justify lower-order maxims, they do not fully determine them. The only constraint that the more general maxims impose on the lower subordinate maxims is they must be a means to the end that the agent has selected.^{xii} More importantly, it is the whole system of maxims that provides the justification for the agent’s action. In order to avoid regress, Kant explicitly identifies the need for an ultimate, most general maxim, which is itself a product of free practical reason and which provides the ultimate fundamental principle of justification of all actions (*Rel*, 6:21). The concept of the *Gesinnung*, as the agent’s fundamental moral disposition or character, refers to the higher-order maxim that underlies an agent’s choice of more particular maxims.^{xiii} Kant’s ethical rigorism entails that both actions and agents (including their *Gesinnung*) must be characterised as either good or evil (*Rel*, 6:23-4).^{xiv} Considered materially, a good and an evil *Gesinnung* have the same content: they include empirical incentives subsumed under the general principle of happiness (*G*, 4:399) and the rational incentive of respect for the moral law which is the most basic ‘fact of reason’ (*KpV*, 5:29-50). The difference between the two lies in the form of the will, that is, the manner in which the contents are combined, how the two incentives are related, which one is incorporated as the condition of the other. The person with a good character is the person whose fundamental maxim is to make the moral law the supreme condition of all acts, thus subordinating the

demands of happiness to the demands of morality and taking the moral law as the supreme principle of justification of all acts. In contrast, a bad person is committed to the promotion of her own happiness unconditionally, typically acting on merely subjective valid motives and complying with moral requirements only insofar as they do not demand sacrifices.

In *Religion*, Kant famously claims that ‘the human being is by nature evil’ (*Rel*, 6:32). This is usually taken to mean that the *default* position of the human will is in fact evil. Kant claims that we have a ‘propensity’ (*Hang*) to radical evil, i.e. a free tendency of the will (*Willkür*) to give undue weight to nonmoral incentives, a tendency that implies the adoption of an evil fundamental maxim.^{xv} The choice of *Gesinnung* is said to be free (*Rel*, 6:44), but to the extent that it also supposed to be universal, Kant says that the propensity to evil is an aspect of human nature (*Rel*, 6:32). The view that the human species as whole chooses an evil fundamental maxim is difficult and philosophically controversial;^{xvi} however, what is important for our argument is that Kant also says that it is ethically necessary, and therefore must be possible, to overcome radical evil (*Rel*, 6:66-67) by attempting to change one’s fundamental maxim through a ‘revolution of the heart’ (*Rel*, 6:47, 51).

Earlier, I noted that the process of reflecting and reforming our maxims should constitute a fundamental aspect of our rationality. We can now add that this process of self-reform would require that our maxims are changed in light of the requirements of the moral law. Transforming our *Gesinnung* through a revolution of the heart is thus a necessary aspect of the moral development of a person. In the second *Critique*, Kant characterised moral development as requiring a gradual process of moral change (*KpV*, 5:159-160); in the *Metaphysics of Morals*, which was written after the *Religion*, in addition to the need for a gradual change (*TL*, 6:477), Kant also refers to the need for a single moral decision to break away from vice (*TL*, 6:477). It has been suggested that this single moral decision should be identified with the revolution of the heart proposed by Kant as a solution to the problem of

overturning evil in the *Religion*.^{xvii} I would like to further suggest that the revolution of the heart plays a central role in Kant's theory of moral improvement for two main reasons. First, the revolution is a necessary condition for the possibility of acquiring virtue, understood as the strength to overcome obstacles (vices) and make duty the sole incentive of right acts. Virtue is the 'moral strength of a *man*'s will in fulfilling his *duty*' (*TL*, 6:405) and involves a firm resolution to act out of duty no matter how strong the temptation to act wrongly, thus requiring *reliability* of motivation. However, reliability of motivation cannot be achieved unless the agent is unconditionally committed to the moral law, that is, unless the agent possesses a good fundamental maxim.^{xviii} Second, it provides the rational framework that allows a person to abandon her immoral maxims, insofar as her more particular immoral maxims would now lack a higher-order principle of justification. There are some difficulties associated with the possibility of moral transformation^{xix} and Kant himself notes that the transformation is somehow inexplicable (*Rel*, 6:45); however, it is clear that Kant also thinks that the revolution is a clear human possibility (*Rel*, 6:67 and 6:50). Importantly, even after (or during)^{xx} the revolution, there is still more moral progress to be made (*Rel*, 6:47-48).^{xxi} The acquisition of a good *Gesinnung* provides the framework that allows a person to embark on 'the road of endless progress towards holiness' (*Rel*, 6:47) but holiness is ultimately an unattainable ideal (*TL*, 6:409). After the revolution, the agent should still revise her maxims in order to make sure that moral actions are performed out of a pure sense of duty and she will need to continue cultivating a firm resolution of the will in order to live up to her new maxims. Moral development, thus, involves an ongoing and self-imposed intellectual process of self-knowledge, reflection and self-reform.

The next step is to note that commitment to an ethical project of self-improvement would necessarily require repentance of our immoral acts, that is, the agent should come to see the maxims underlying her immoral acts as something that cannot be fully justified to

others. The judgement that her maxim is unjustified and the realisation that she has wronged others would necessarily involve repentance (understood as a commitment to a change of maxim) and would typically involve taking steps towards the reparation of the wrong. Repentance is a necessary component in the moral development of a person insofar as it is a necessary aspect of a commitment to abandon immoral maxims.

These various elements of Kant's practical philosophy strongly suggest that forgiveness should play a role in the lives of ordinary human agents as understood by Kant. The human predicament is a predicament of evil so agents will, on occasion, act wrongly. Yet at the same time, the moral law as the most basic 'fact of reason' and our fundamental freedom mean that a revolution of the heart is always a 'live' option for human beings. Agents who are committed to this revolutionary and ongoing project of self-reform would necessarily strive to change their bad maxims and repent their immoral acts. In order to complete the argument, we need to derive the duty to forgive repentant wrongdoers from Kant's *formula of humanity* (FH),^{xxii} which says 'so act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means' (*G*, 4:429). Kant says that this Imperative is grounded on the recognition that '*rational nature exists as an end in itself*' (*G*, 4:428). To treat something as an end in itself is to treat it as providing considerations for action that we cannot rationally ignore, and this involves mainly avoiding treating it as a means for the satisfaction of some contingent desire-based ends. Persons possess rational nature and by virtue of possessing such a nature, they are ends in themselves and possess an intrinsic value that makes them worthy of 'respect.' Although the correct interpretation of FH is open to some controversy, here, for the sake of brevity, I will adopt the interpretation that understands 'rational nature' as referring to the *capacity for morality* (*G*, 4:435), which is then understood broadly to refer to the capacity to set ends (*Willkür*) plus the capacity to legislate moral principles (*Wille*) and the capacity

for rational self-constraint termed ‘respect for the law’ (*KpV*, 5:76).^{xxiii} It has been argued that we can derive a *perfect* duty to forgive wrongdoers unconditionally^{xxiv} but we should take seriously Kant’s own classification of the duty to be forgiving as an imperfect duty of virtue, i.e. as one of the duties of sympathy, which together with duties of beneficence and gratitude, constitute duties of love to others (*TL*, 6:448-462). Duties of love are duties of virtue that command the adoption of obligatory ends (*TL*, 6:383) and their corresponding maxims. Imperfect duties are derived from FH, which establishes two main obligatory ends: my own perfection and the happiness of others (*G*, 4:430). To fully respect rational nature in ourselves and others requires that we adopt these ends. The duty to be forgiving is one of the duties of sympathy that falls under the duty to promote the happiness of others. As these are imperfect duties, there is some *latitude* for agents to decide in *what way* and to *what extent* to promote these ends (*TL*, 6:390-4). Duties of virtues do not command us to act in specific ways but rather, to adopt certain principles. The duty to promote an end is, thus, primarily a duty to refrain from adopting the maxim of refusing in principle to promote that end.^{xxv} This shows that Kantian ethics does not recommend that we forgive unconditionally, that is, a maxim of always forgiving offences is not obligatory^{xxvi} Instead, Kantian ethics forbids that we adopt a maxim of refusing in principle to develop forgiving practices and attitudes towards wrongdoers. The duty to be forgiving is a duty to adopt a forgiving maxim which involves taking on the task of *cultivating a forgiving character*,^{xxvii} that is, to be prepared to forgive when the circumstances are considered appropriate. Given the arguments developed in this section, it should be clear that the chief consideration that makes forgiveness appropriate is repentance of immoral deeds as evidence of the wrongdoer’s commitment to a project of moral transformation.^{xxviii}

To summarise: forgiveness does not sit uncomfortably in Kant’s system because Kantian ethics can ground an imperfect duty of virtue to be forgiving; thus, forgiveness is not

merely ‘elective’ but actually morally required.^{xxx} Since imperfect duties cannot be enforced and do not correspond to rights,^{xxx} Kant’s theory is particularly well-suited to retain the idea that there is a duty to be *forgiving* while at the same time rejecting the view that wrongdoers have a *right* to be forgiven.

2. Forgiveness and punishment

Here, I clarify Kant’s views on the nature of forgiveness. Kant says that the ‘duty to be *forgiving (placabilitas)*’ involves overcoming the vice of *malice* which is the ‘direct opposite of sympathy’ (*TL*, 6:460). Here is the passage in full:

The sweetest form of malice is the *desire for revenge*. Besides, it might even seem that one has the greatest right, and even the obligation (as a desire for justice), to make it one’s end to harm others without any advantage to oneself.

Every deed that violates a man’s right deserves punishment, the function of which is to *avenge* a crime on the one who has committed it (not merely to make good the harm that was done). But punishment is not an act that the injured part can undertake on his private authority, but rather an act of a court distinct from him, which gives effect to the law of a *supreme authority* over all those subject to it; and when (as we must in ethics) we regard men as in a rightful condition *but in accordance only with laws of reason* (not civil laws), then no one is authorized to inflict punishment and to avenge the wrongs sustained by men except Him who is also the supreme moral lawgiver; and He alone (namely God) can say ‘Vengeance is mine; I will repay’. It is, therefore, a duty of virtue not only to refrain from repaying another’s enmity with hatred out of mere revenge but also not even to call upon the judge of the world for vengeance, partly because a man has enough guilt of his own to

be greatly in need of pardon and partly, and indeed specially, because no punishment, not matter from whom it comes, may be inflicted out of hatred. It is therefore a duty of men to be *forgiving (placabilitas)*. But this must not be confused with *meek toleration of wrongs (mitis iniuriarum patientia)*, renunciation of rigorous means (*rigorosa*) for preventing the recurrence of wrongs by other men; for then a man would be throwing away his rights and letting others trample of them, and so would violate his duty to himself (*TL*, 6:460-1).

The passage sees forgiveness as a response to *wrongdoing*, which involves the overcoming of the vice of malice understood as the hateful desire for revenge. In addition, and perhaps more enigmatically, Kant says ‘not even to call upon the judge of the world for vengeance’, presumably suggesting that forgiveness requires that one does not even desire that the wrongdoer suffers disproportionately (in the sense of calling upon God to inflict the punishment on the wrongdoer that would satisfy one’s excessive vindictive desires). More importantly, forgiveness is also seen a *personal* and *private* response to wrongdoing. Interestingly, Kant endorses moral retributivism in the very same passage in which he argues for a duty to be forgiving: actions that violate the rights of men (i.e. wrongs), Kant says, deserve to be punished.^{xxxii} Yet immediately after making this assertion, Kant clarifies that punishment cannot be inflicted by a private authority, but only by the court of a supreme authority. Forgiveness cannot be understood as the forgoing of punishment because punishment by a private individual or group is never morally or legally appropriate, as Kant had already stated in the Doctrine of Right that an individual is not allowed to enforce his rights generally (*RL* §44, 6:312). Punishment is not something that an individual can forgo because an individual is not authorised to inflict punishment in the first place. The issue of whether or not to punish a wrongdoer arises only for a supreme authority. In the case of

wrongs that are also legal offences, the supreme authority is the state in the form of the courts (*RL* §49E, 6:332). The issue of personal forgiveness might arise for cases in which the wrong involves a violation of positive law and cases in which the wrong does not involve a violation of the law. Insofar as juridical duties do not command us to commit acts that conflict with ‘inner morality’ (*RL* Appendix, 6:371), they are also indirect ethical duties (*RL* Introduction, 6:220). But duties of virtue are not enforceable (*RL* Introduction, 6:220),^{xxxii} so not all moral wrongs constitute legal offences. The law is concerned only with intentional ‘external acts’ that violate enforceable public requirements (*RL* Introduction, 6:230). The state in fact has a right (*RL*, §49E 6:331) and a duty to punish lawbreakers: ‘for if justice goes, there is no longer any value in men’s living on the earth’ (*RL*, §49E 6:332/p. 141). Traditionally, Kant has been read as a legal retributivist who justifies punishment of breaches of positive law in terms of moral desert. Recently, however, the issue has become highly debated, with interpreters suggesting that Kant’s position involves a mixture of deterrence and retribution. Sometimes Kant is presented as providing a deterrence justification for the institution of punishment which is then understood as constituted by retributive practices.^{xxxiii} Others find a distinction between the threat of punishment, which is justified on deterrence grounds, and the execution of punishment, which is justified by considerations of justice.^{xxxiv} Kant’s position is thus more complex and subtle than a simple commitment to legal retributivism.^{xxxv} In any case, what is clear is that for Kant, legal wrongs require punishment by the state. Moreover, in the *TL* passage Kant adds: ‘and when (as we must in ethics) we regard men as in a rightful condition *but in accordance only with laws of reason* (not civil laws), then no one is authorized to inflict punishment and to avenge the wrongs sustained by men except Him who is also the supreme moral lawgiver; and He alone (namely God) can say “Vengeance is mine; I will repay”’. The passage suggests that moral wrongs *qua moral* cannot be punished by either the state or the individual. This is because persons cannot be

coerced to act for morally worthy motives since the adoption of ends always involves a free choice by the individual (*TL*, 6:381). Thus, the courts are not concerned with the moral worth of a person's motives and legal culpability is not equivalent or proportionate to moral unworthiness even though, as noted previously, legal offences are also violations of an indirect duty to comply with the law. The state is not concerned with bringing about punishment in proportion to moral wrongdoing since the law is only concerned with the mere legality of actions.^{xxxvi} The only one in a position to punish moral wrongs would (perhaps) be God who, as the supreme moral authority, can establish the required proportion between suffering and moral wrongdoing. Insofar as punishment is the prerogative of a supreme authority, either the state or God, but not the individual, it is safe to conclude that forgiveness is relegated to the personal, private, sphere: forgiveness and punishment should be considered as two separate issues. An attractive aspect of Kant's position is, thus, that forgiveness does not translate into the political sphere, so those responsible for violations of the law should still be punished by the state even when their victims are prepared to forgive them.

Kant, however, does allow for a right to grant clemency on some occasions. As punishment is the prerogative of the state, the only possibility of remission of punishment arises from the sovereign, who has the right to grant clemency (*RL* §49E, 6:337). Kant, however, shows some ambivalence about this right. German legal scholars usually allowed a right of pardon, but Kant's grounds for the acceptability of this right are far stricter.^{xxxvii} Kant claims that, by granting clemency, the sovereign is in fact 'doing injustice in the highest degree' and states the impermissibility of granting clemency for 'crimes of *subjects* against one another' (*RL* §49E, 6:337). The sovereign cannot be allowed to grant clemency in cases in which a subject has committed a crime against another because that would be a violation of the 'universal principle of right' (*RL* Introduction 6:230) and the equality between individuals embraced by this principle. Such a form of pardon would prioritise the criminal's freedom in

relation to the freedom of his victim and other citizens and, as such, it would be inadmissible. Yet, despite these reservations, Kant does allow for two types of situations in which the right of pardon could be exercised and thus two types of exceptions to his otherwise strict theory of punishment. The first type of situation arises when Kant is discussing the law of retribution (*ius talionis*), that is, the idea that the severity of the punishment should correspond to the severity of the crime (*RL* §49E, 6:332). Kant gives the example of murder, which according to this law, should be punished with the death penalty but he allows for one exception:

If, however, the numbers of accomplices (*correi*) to such a deed [murder] is so great that the state... [might] dissolve [...] then the sovereign must also have it in his power, in this case of necessity (*casus necessitatis*), to assume the role of judge (to represent him) and pronounce a judgement that decrees for the criminals a sentence other than capital punishment, such as deportation, which still preserves the population (*RL* §49E, 6:335).

In this passage, Kant is not recommending a full pardon, but the commuting of the form of punishment, from the death penalty to deportation. Toomas Kotkas^{xxxviii} has suggested that the example refers to a case in which the majority of the population took part in a revolution that ultimately failed so that all the guilty were sentenced to death but in such a situation the state would be *literally* at risk of dissolution. Hence, the deportation of the guilty (presumably to a province) arises ‘out of necessity.’ The second exception says that the right can only be used in cases in which the crime has been committed against the sovereign himself (*crimen laesae maiestatis*) (*RL* §49E, 6:337). Kotkas shows that the term sovereign refers to ‘the ruler as the executive and a private person’ and claims that by Kant’s time, the crime of *lese-majesty* comprised ‘merely the crime of slander (*Beleidigung*) against the

ruler's person'^{xxxix}. Since the crime of slander is a crime committed against the ruler as a private person, pardoning it would not require a violation of the universal principle of right. Moreover, Kant makes clear that this right should not be exercised if the people's security is at risk. Thus, Kotkas' analysis shows that Kant's right to grant clemency is in fact very limited: both exceptions are not inconsistent with Kant's criminal law theory because neither requires a violation of the universal principle of right.

I now consider the specifically moral issue of whether forgiveness can somehow help the wrongdoer to overcome her moral guilt (or pay the moral debt) for her previous wrongful acts. The textual evidence on this issue is ambiguous. In the *TL* passage, Kant says explicitly that we should forgive each other 'because man has enough guilt of his own to be greatly in need of pardon'. In Section I, I suggested that the passage should not be read in isolation as implying that we should all forgive each other because we are all evil. In addition, such a reading would be in tension with a passage of *Religion*, which says that:

Whatever his state in the acquisition of a good disposition...*he nevertheless started from evil*, and this is a debt which is impossible for him to wipe out. 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...[T]his original debt [i.e. radical evil] cannot be erased by someone else. For it is not a *transmissible* liability...in the manner of a financial debt...but 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 (*Rel*, 6:72).

The passage makes it clear that the moral guilt for our previous evil disposition, and the wrongful acts ensuing from it, cannot be wiped out by others, strongly suggesting that other

people's forgiveness cannot help us to overcome our moral debt (guilt). Moreover, Kant is not only concerned with the guilt involved in particular wrongful acts, but also with the moral debt that we incurred insofar as our choice of *Gesinnung* is evil: '...moral evil (transgression of the moral law, called sin when the law is taken as a *divine command*) brings with it an *infinity* of violations of the law, and hence an *infinity* of guilt...because the evil is in the *disposition* and the maxims in general...' (*Rel*, 6:72). Lawrence Pasternack has provided a good explanation of the infinite nature of this moral debt. The infinitude of the debt arises because in choosing an evil *Gesinnung* we are committing to an infinite number of transgressions, not because our transgressions are in fact infinite, but because we are implicitly 'condoning an unlimited neglect of morality.'^{xi} The *Gesinnung* choice is in this respect infinite, so the moral guilt attached to it is equally infinite and, as such, ultimately a debt of sin owed to God. For this reason, some commentators have argued that freedom from this debt would also ultimately require divine assistance in the form of God's forgiveness (*Rel*, 6:76). Thus, according to the 'divine supplement' interpretation,^{xii} although each individual agent should strive towards a change of heart (*Rel* 6:44, 6:52 and 6:171-172), God's forgiveness is required in order to 'supplement' the inevitable shortcomings in our strivings^{xiii} and help us to complete our moral transformation. The problem with idea of a divine supplement is that it is incompatible with Kant's moral retributivism, i.e. the view that moral wrongs require punishment by a supreme authority,^{xiiii} so that ultimately such divine forgiveness would in effect amount to a form of injustice. Furthermore, the idea of the need for a 'divine supplement' is also in tension with the concept of Highest Good, which requires an ideal state of affairs in which happiness is distributed in exact proportion to our moral worth (*KpV*, 5:110).^{xlv} Since this passage makes it clear that human forgiveness cannot help to erase the wrongdoer's moral guilt and an appeal to divine forgiveness is tension with some fundamental commitments of Kant's practical philosophy, I conclude that Kant's conception

of forgiveness does not, and cannot, require the overcoming of moral guilt.^{xlv} Again, I believe that this could be seen as an attractive feature of Kant's position. Repentance involves the recognition of our own guilt but this guilt cannot be erased. Guilt thus plays a role in the moral development of a person, as recognising and repenting our guilt is an essential aspect of our moral transformation.

In summary, Kantian forgiveness is an individual's private response to wrongdoing, which involves overcoming^{xlvi} feelings of hatred and vindictiveness and the forgoing of the desire that the wrongdoer suffers disproportionately. Forgiveness does not (and cannot) involve the forgoing of punishment or the wiping out of moral guilt. The duty to be forgiving is an imperfect duty of virtue and, as such, it commands that we adopt a maxim of forgiving others in circumstances deemed appropriate. Insofar as maxims express an agent's reasons for action, Kant's account is particularly well suited to providing what Pamela Hieronymi has called an 'articulate account of forgiveness'; that is, forgiveness is something that we do on the basis of reasons, and not a mere psychological process that passively happens to us.^{xlvii} However, there are also some limitations to Kant's account. It seems too narrow to only count hatred and vindictiveness among the emotions that forgiveness should overcome. I suggest that one could improve on his account, without compromising Kant's fundamental practical commitments, by allowing for forgiveness to be understood more generally as the overcoming of various negative emotions, also including anger and resentment, which are usually felt by the victim in response to having been wronged.^{xlviii} Furthermore, it has been suggested that overcoming the negative emotions commonly associated with wrongdoing, (although important in many cases) should not be taken as either a necessary or sufficient condition for forgiveness.^{xlix} Instead, forgiveness might involve other practices, such as reintegration into the moral community, restoration of relationships, developing attitudes of love towards the wrongdoer, among others. Geoffrey Scarre is sceptical about the possibility

of defining forgiveness, arguing that it should be understood as a multi-form phenomenon which includes a broad and varied family of practices.¹ Kant's account can accommodate Scarre's suggestion, inasmuch as it commands the adoption of a maxim and maxims can be discharged in many ways, depending on both the particular circumstances of the case and the forger's practical judgment (*TL*, 6:411), arguably allowing for a variety of forgiving practices and attitudes.

Conclusion

There is space in Kant's philosophy for a genuine theory of forgiveness which establishes an imperfect duty of virtue to adopt a maxim of forgiving offenders under certain conditions, that is, provided that there is evidence that the wrongdoer has repented and is committed to a project of moral improvement. However, on the Kantian account, forgiveness does not involve the forgoing of punishment or the wiping out of moral guilt. To the extent that this is an imperfect duty of virtue that commands that we adopt a maxim of forgiving repentant wrongdoers, Kant's account can accommodate a certain degree of flexibility and allow for a variety of forgiving attitudes and practices.

Notes

¹ Parenthetical references are all to Kant; page numbers refer to the edition of the Prussian Academy. Direct quotes will also provide the page number in the following translations:

G *Groundwork of the Metaphysics of Morals* translated by Gregor M. and introduced by Korsgaard M. Cambridge: Cambridge University Press, 1997.

KpV *Critique of Practical Reason*, translated by Pluhar W. and introduction by Engstrom S. Hackett Publishing Company, Inc., Indianapolis/ Cambridge, 2002.

RL Doctrine of Right in *The Metaphysics of Morals* translated and introduced by M. Gregor (Cambridge: Cambridge University Press, 1991).

TL Doctrine of Virtue in *The Metaphysics of Morals* translated and introduced by M. Gregor (Cambridge: Cambridge University Press, 1991).

Rel *Religion within the Boundaries of Mere Reason* in *Religion within the Boundaries of Mere Reason and Other Writings* translated and edited by A. W. Wood and G. di Giovanni, introduced by R. M. Adams, (Cambridge: Cambridge University Press, 1998).

ⁱⁱ My analysis focuses mainly on the passage of the *Metaphysics of Morals*, *TL*, 6:460-1. I will refer to this passage as the *TL* passage.

ⁱⁱⁱ D. Sussman, 'Kantian Forgiveness', *Kant-Studien*, 96/1 (2005), 85-107.

^{iv} Sussman, 'Kantian Forgiveness', 90.

^v Sussman, 'Kantian Forgiveness', 88.

^{vi} Sussman, 'Kantian Forgiveness', 86. Sussman appeals to Kant's views on God's grace in order to develop a 'recognizable Kantian account of forgiveness' (85). In contrast, I argue that Kantian ethics already has the necessary resources to provide a theory of forgiveness without the need to appeal to the doctrine of divine grace.

^{vii} See P. Satne, 'Forgiveness and Moral Development', *Philosophia* (2016), DOI 10.1007/s11406-016-9727-6.

^{viii} This interpretation has been developed in more detail in Satne, 'Forgiveness and Moral Development'.

^{ix} Interpretations along these lines have been presented by both J. Murphy, 'Hatred: A Qualified Defense' in J. Murphy and J. Hampton (eds), *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1990), pp. 96-103 and E. Garrard and D. McNaughton, 'In Defence of Unconditional Forgiveness', *Proceedings of the Aristotelian Society* 103/1 (2003),

39-60. Both interpretations are discussed and rejected in Satne, 'Forgiveness and Moral Development'; see especially Section III.

^x See T. Brewer, 'Maxims and Virtues', *Philosophical Review* 111 (2002), 539-572 for a good discussion of the main features of maxims.

^{xi} See Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).

^{xii} M. Caswell, 'Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil', *Kant-Studien* 97/2, 184-209.

^{xiii} Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 136-145; M. Caswell, 'Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil', 191-196.

^{xiv} For Kant's argument in support of rigorism, see P. Satne, 'Reliability of Motivation and the Moral Value of Actions', *Studia Kantiana* 14 (2012), 5-33.

^{xv} M. Caswell, 'Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil', 199 and H. Allison, *Kant's Theory of Freedom*, p. 153.

^{xvi} See S. Morgan, 'The Missing Formal Proof of Humanity's Radical Evil in Kant's Religion', *Philosophical Review* 114/1 (2005), 63-114, for an overview of the difficulties associated with Kant's theory of radical evil.

^{xvii} Ch. Drogalis, 'Kant's Change of Heart: Radical Evil and Moral Transformation', *Dissertations*. Paper 512. http://ecommons.luc.edu/luc_diss/512, pp. 3-4; Mary Gregor, *The Metaphysics of Morals*, (Cambridge: Cambridge University Press, 1991), Introduction, p. 18.

^{xviii} In Satne, 'Reliability of Motivation and the Moral Value of Actions', I argued that a good *Gesinnung* is a necessary condition for the possibility of developing a virtuous character, but not for the possibility of acting from duty and actions possessing moral worth.

^{xix} See S. Loncar, ‘Converting the Kantian Self: Radical Evil, Agency, and Conversion in Kant’s Religion within the Boundaries of Mere Reason’, *Kant-Studien* 104/ 3 (2013), 346-366 and L. Pasternack, ‘Kant on the Debt of Sin’, *Faith and Philosophy*, 29/1 (2012), 30-52.

^{xx} Kant uses temporal language to refer to the revolution (*Rel*, 6:47), but he also claims that the choice of evil *Gesinnung* is an ‘intelligible deed’ that does not occur in time (*Rel*, 6:31). There are competing accounts of how to explain the relationship between the timeless revolution and the gradual process of moral- self-improvement.

^{xxi} See also *Rel*, 6: 66-67.

^{xxii} Kant does not attempt this derivation, so here I am completing his argument.

^{xxiii} See T.E. Hill, *Dignity and Practical Reason in Kant’s Moral Theory* (Ithaca: Cornell University Press, 1992), pp. 40-1; B. Herman, *The Practice of Moral Judgement* (Massachusetts and London: Harvard University Press, 1993), p. 238 and O. O’Neill, *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), pp. 137-8.

^{xxiv} M. Holmgren defends an interpretation along these lines in ‘Forgiveness and the Intrinsic Value of Persons’, *American Philosophical Quarterly* 30/4 (1993), 341-52. I discuss and reject her view in Section V of ‘Forgiveness and Moral Development’.

^{xxv} See A. W. Wood, ‘Duties to oneself, duties of respect to others’, in T.E. Hill Jr. (ed.), *The Blackwell Guide to Kant’s Ethics* (Oxford: Blackwell Publishing, 2009), p. 234.

^{xxvi} I read the last part of the *TL* passage as establishing a perfect duty to oneself not to forgive unrepentant wrongdoers, but I do not develop this aspect of Kant’s view in this chapter.

^{xxvii} It is important to stress that this is a duty to be forgiving (i.e. a duty to develop a forgiving character), rather than a duty to forgive. As such it has some affinities with what Robert Roberts has called ‘forgivingness’. See his ‘Forgivingness’, *American Philosophical Quarterly* 34/2, 289-306.

^{xxviii} Further considerations in support of the conditional nature of this duty are provided in Satne, 'Forgiveness and Moral Development', pp. 21-22.

^{xxix} Duties of virtue are not any less obligatory than perfect duties; see Wood, 'Duties to oneself, duties of respect to others', p. 229.

^{xxx} See P. Stratton-Lake, 'Being virtuous and the virtues: two aspects of Kant's doctrine of virtue' in M. Betzler (ed.), *Kant's Ethics of Virtue* (Berlin: Walter de Gruyter, 2008), pp. 101-122, especially pp. 104-106.

^{xxxi} See also *R*, §49E, 6:331 and *Rel*, 6:73.

^{xxxii} Cf. *TL*, 6:383-4.

^{xxxiii} D. Scheid, 'Kant's Retributivism', *Ethics* 93 (1983), 262-82, pp. 262-5.

^{xxxiv} S. Byrd, 'Kant's Theory of Punishment: Deterrence in its Threat, Retribution in its Execution', *Law and Philosophy* 8 (1980), 151-220, pp. 184-98.

^{xxxv} T.E. Hill Jr., "Punishment, conscience, and moral worth", in M. Timmons (ed.), *Kant's Metaphysics of Morals. Interpretative Essays* (Oxford and New York: Oxford University Press, 2002).

^{xxxvi} This passage seems to undermine the interpretation of Kant as a full-blown legal retributivist.

^{xxxvii} T. Kotkas, 'Kant on the Right of Pardon: A Necessity and the Ruler's Personal Forgiveness', *Kant-Studien*, 102/4 (2011), 413-421, p. 418.

^{xxxviii} See his 'Kant on the Right of Pardon'.

^{xxxix} Kotkas 'Kant on the Right of Pardon', p. 419.

^{xl} Pasternack, 'Kant on the Debt of Sin', p. 33.

^{xli} Pasternack, 'Kant on the Debt of Sin', p. 41.

^{xlii} Pasternack, 'Kant on the Debt of Sin' p. 41.

^{xliii} *TL*, 6:460. Cf. *Rel*, 6:73.

^{xliv} This point is made by Pasternack, 'Kant on the Debt of Sin', p. 42.

^{xlv} Whether it is ever possible for human beings to be freed from this debt is an open question but is a more pressing problem from a religious (rather than strictly moral) point of view.

^{xlvi} But note that for Kant, we do not completely overcome emotions but we *incorporate* (or *refuse to incorporate*) them into maxims (*Rel*, 6:24) by critically evaluating them and refusing to endorse those that on reflection we deem morally inappropriate.

^{xlvii} P. Hieronymi, 'Articulating an Uncompromising Forgiveness', *Philosophy and Phenomenological Research* 63/3 (2001), 530.

^{xlviii} This would put Kant's account in some respects in line with Bishop Butler's definition, which is widely accepted by contemporary theorists. See Bishop Butler, *Sermons* (Boston: Hilliard and Brown, 1827).

^{xlix} See G. Scarre, *After Evil: Responding to Wrongdoing* (Aldershot: Ashgate Publishing Limited, 2004), pp. 25-31.

¹ Scarre, *After Evil*, p. 31.