



The Paradox of Forgiveness

Leo Zaibert

Amherst College, 74 College Street, Amherst, MA 01002, USA
zaibert@gmail.com

Abstract

Philosophers often claim that forgiveness is a paradoxical phenomenon. I here examine two of the most widespread ways of dealing with the paradoxical nature of forgiveness. One of these ways, emblemized by Aurel Kolnai, seeks to resolve the paradox by appealing to the idea of repentance. Somehow, if a wrongdoer repents, then forgiving her is no longer paradoxical. I argue that this influential position faces more problems than it solves. The other way to approach the paradox, exemplified here by the work of Jacques Derrida, is just too obscure to be by itself helpful. Yet, I argue that what I take to be its spirit is on the right track. I recommend distinguishing between (1) the definition and the justification of forgiveness, and also between (2) forgiveness understood as (a) a mental phenomenon and (b) an overt, communicative act. These distinctions are not given their due in the specialized literature, and I expose the nefarious consequences of this neglect. By focusing on forgiveness as a mental phenomenon I seek to analyze the root of the talk of paradoxes which surrounds the discussion of forgiveness. Finally, I present an analysis of forgiveness as a pure mental phenomenon, and argue that this analysis is the most important step in understanding forgiveness in any other sense. While my analysis reveals interesting aspects of forgiveness, it reveals, too, that forgiveness is not quite as paradoxical after all.

Keywords

blame, condonation, definitional stop, forgiveness, justification, punishment

Recent contributions to the literature on forgiveness often begin with a lamentation to the effect that philosophers have not given the phenomenon of forgiveness enough attention. Typically the philosophers' lack of attention is all the more poignant in 'comparison to the attention they give the related concept of punishment'.¹ While this lamentation is, precisely in light of the

¹ Berel Lang, 'Forgiveness', *American Philosophical Quarterly* 31 (1994): 105-17, at 105. Similar laments are found at the outset of Claudia Card, 'On Mercy', *The Philosophical Review* 81 (1972): 182-207; in R.S. Downie, 'Forgiveness', *The Philosophical Quarterly* 15 (1965): 128-34; P.F. Strawson *Freedom and Resentment and Other Essays* (Oxford: Methuen, 1974), *passim*; Joram Graf Haber, *Forgiveness: A Philosophical Study* (Boston: Rowman & Littlefield, 1993), *passim*, and in many other influential works.

wealth of recent discussions of the phenomenon, somewhat passé, much remains to be clarified regarding the analytical contours of forgiveness. What is not passé is to see forgiveness as an extraordinarily complex, indeed a philosophically paradoxical phenomenon, in spite of the fabulous success of (self-help) books which simply-mindedly hype forgiveness as the solution to all sorts of vicissitudes.² Philosophers—in contrast, it seems, to the general public—agonize over the nature of forgiveness, and they allude to a veritable ‘paradox of forgiveness’. By way of introducing the problem that shall occupy my attention, I would like to address, in turn, two different ways of formulating the paradox of forgiveness.

The first way is Aurel Kolnai’s. Kolnai died before presenting his ‘Forgiveness’ to the Aristotelian Society, but the version of this piece included in the *Proceedings* has proven extremely influential. In this article, which Kolnai considers to be ‘chiefly *logical*’,³ he presents the following paradox:

[either] forgiveness is objectionable and ungentle inasmuch as *there is no reason to forgive*, the offender having undergone no *metánoia* (‘Change of Heart’), but persisting in his plain identity *qua* offender⁴ ... [or] at the other end of its spectrum, forgiveness seem to collapse into mere *redundancy*, or the mere *registering* of moral value in the place of moral disvalue.⁵

Kolnai’s formulation of the paradox, without further clarifications, is not immediately enlightening. In fact, some have concluded that ‘the [Kolnaian] paradox might move some to conclude that the concept of forgiveness is internally incoherent’.⁶ Schematically, at first, Kolnai’s paradox can be brought into sharper focus by considering the famous request that Jesus made in the cross: *Father, forgive them; for they know not what they do*. The paradox is then the following: if, on the one hand, they really did not know what they were doing, and assuming that this ignorance was not itself culpable, then surely they should be *excused*, not forgiven. Forgiveness, as a matter of sheer logic, presupposes (perceived) culpable wrongdoing. On the other hand, if they did know that they were doing something wrong, then presumably they should have been *punished*, and, again, not forgiven. (This presumption’s strength is tied to

² Consider a small sample of best-selling books on forgiveness: *Forgiveness Is a Choice: A Step-By-Step Process for Resolving Anger and Restoring Hope*; *Forgiveness: How to Make Peace With Your Past and Get on With Your Life*; *Forgiveness: The Greatest Healer of All*; *The Gift of Forgiveness*; *The Art of Forgiveness, Lovingkindness, and Peace*; *Forgiveness: A Bold Choice for a Peaceful Heart*.

³ Aurel Kolnai, ‘Forgiveness’, *Proceedings of the Aristotelian Society* 74 (1973): 91–106, at 91.

⁴ Kolnai, ‘Forgiveness’, p. 97.

⁵ Kolnai, ‘Forgiveness’, p. 98.

⁶ Jean Hampton, ‘Forgiveness, Resentment and Hatred’, in Jeffrie G. Murphy and Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), p. 42.

the strength of retributive intuitions). In other words, putative cases of forgiveness are either not cases of forgiveness at all, or else they seem to be cases where forgiveness is presumably unjustified.⁷

Thus, the Kolnai paradox has two parts. The first part is indeed ‘chiefly logical’: putative instances of forgiveness are in fact spurious, since they are instances of other sorts of phenomena: often what looks like forgiving is something else: excusing, exonerating, forgetting, exculpating, or, crucially for Kolnai, condoning. The second part of the paradox, however, is not ‘chiefly logical’, since it relates to the robust normative discussion concerning the justification of forgiveness. The tension between normativity and description shall occupy my attention throughout, for, as we shall see, the conflation of the merely definitional (descriptive) and the justificatory (normative), complicates the already difficult discussion of forgiveness.

The second way of formulating the paradox is Jacques Derrida’s, whom I would like to pit against Kolnai. In his characteristic style, Derrida states that in order to understand

the very concept of forgiveness, logic and common sense agree for once with the paradox: it is necessary ... to begin from the fact that, yes, there is the unforgivable. Is this not, in truth, the only thing to forgive? The only thing that calls for forgiveness? ... One cannot, or should not, forgive; there is only forgiveness, if there is any, where there is the unforgivable. That is to say that forgiveness must announce itself as impossibility itself. It can only be possible in doing the impossible.⁸

Not surprisingly, perhaps, Derrida is led to the sort of grandiloquent statement for which he is (in)famous: ‘Forgiveness is thus mad. It must plunge, but lucidly, into the night of the unintelligible’.⁹

The main reason why I pit Kolnai and Derrida against each other is that that the former is primarily interested in what is widely known as ‘conditional forgiving’, whereas Derrida can be seen as primarily interested in ‘unconditional forgiveness’, and the opposition between these two views of forgiveness shall prove important for my purposes. As we shall see, Kolnai ‘resolves’ the paradox of forgiveness by appealing to the idea of repentance (*metánoia*, in the preceding quotation): what renders forgiveness ‘genuine and unobjectionable’ is that the wrongdoer has (in the forgiver’s eyes, at least) repented—as we shall

⁷ The reference to the biblical passage merely seeks to illustrate a philosophical problem; this article approaches forgiveness in a wholly secular way.

⁸ Jacques Derrida, *On Cosmopolitanism and Forgiveness* (London: Routledge, 2001), pp. 32–33.

⁹ *Ibid.*, p. 49.

also see, this is a rather influential view. In what follows, I shall criticize the Kolnai move, arguing that the appeal to repentance does not help us overcome the paradoxical nature of forgiveness.

In contrast, by linking forgiveness to the unforgivable, Derrida simply refuses to ‘resolve’ the paradox of forgiveness; and he, rightly in my opinion, suggests that the phenomenon of forgiveness chiefly worth our attention is independent of whether the wrongdoer repents or apologizes. I will argue that forgiveness is also independent of many things with which it is usually considered (even by Derrida himself) to be essentially linked. Unfortunately, however, Derrida’s formulation of the paradox is yet more unhelpful than Kolnai’s, as Derrida fails to adumbrate, even in rough outline, what the analysis of this forgiveness-without-repentance which he describes could turn out to be, and thus I cannot defend his position on the matter. I think, however, that, if I understand him correctly, I will here defend the *spirit* of the Derridian position (or at least part of it). Going beyond Derrida, I shall present an *analysis* of the sort of forgiveness I think he has in mind. Such analysis comes in the last section of this article—much needs to be clarified before getting there, and in particular, much needs to be said about the infelicities of conditional forgiveness. But, merely to fix ideas, let me sketch the position that I will defend later in this article: to forgive is to deliberately refuse to punish. This account sounds, perhaps, too broad, and while it may still sound problematically broad after I defend it below, I hope to be able to dispel doubts about the potential infelicities concomitant to this breadth.

1. The Unforgivable and other Methodological Preliminaries

Derrida’s style may be unique, but in the passage quoted he commits a rather ordinary mistake. The error has to do with the ambiguous use of the term ‘unforgivable’. Although there is great obscurity as to what exactly this locution means, it typically seeks to suggest that some wrongs are of such immense gravity that they *should not* be forgiven. Nazi atrocities, say, or child molestation, are customarily held to be in this sense unforgivable; arguably the point being made is that these very serious wrongs *should* be punished, and *should not* be forgiven.

To be sure, it is only by assuming that Derrida uses ‘unforgivable’ in the sense of ‘ought not to be forgiven’ (and not in the sense of ‘cannot (logically) be forgiven’) that his quoted views make sense. Forgiveness presupposes, on my interpretation of Derrida, that the act to be forgiven is, *prima facie* at least, something that we *should not* forgive. For if we *ought* to forgive a given act, then, and indeed paradoxically, this act would no longer be as interesting a

case for being forgiven (it will often just be a case in which the alleged wrongdoer is excused, justified, not really a wrongdoer, etc.). What Derrida means by ‘there is only forgiveness ... where there is the unforgivable’, is, then, that forgiveness can only occur as a response to something that should, in principle, in some sense, not be forgiven.

As a matter of sheer definition, forgiveness is a way of reacting to ‘the guilty as guilty’, to use R.G. Collingwood’s apt phrase (which Derrida uses elsewhere as well).¹⁰ This phrase captures an important point about forgiveness, and to which I shall return in the next section. For now, I will just offer a one-liner of my own, which I hope helps to unpack the importance of Collingwood’s: to ‘forgive’ the no longer guilty is no longer to forgive.

There is, then, a purely analytic connection between punishment and forgiveness worth our attention: only what is punishable is forgivable, and only what is forgivable is punishable. Not merely wrongdoing, however, but *culpable* wrongdoing, is a logical precondition of both punishment and forgiveness.¹¹ As Hannah Arendt, amongst others, would have it: we ‘are unable to forgive what [we] cannot punish’.¹² Now, this thesis entails neither that (1) whenever we blame someone for her wrongdoing, we must either punish or forgive her, for punishment and forgiveness are not jointly exhaustive—not only can we do other things too, but we can refuse to do anything at all, as we shall see in due course, nor that (2) if we do punish the culpable wrongdoer, then we cannot forgive her (or vice versa). While I will eventually suggest that punishment and forgiveness are mutually exclusive (synchronically), the analytic connection just uncovered does not by itself support the mutual exclusivity thesis.¹³ My suggestion thus far is strictly this: (perceived) culpable wrongdoing is a necessary condition for the very possibility of forgiving X, just as it is for punishing X. Strictly speaking, then, whatever *on these grounds* cannot be forgiven cannot be punished either (and vice versa)—of

¹⁰ R.G. Collingwood, *Essays in Political Philosophy*, ed. David Boucher (Oxford: Clarendon Press, 1995), p. 127. The theme of the ‘guilty as guilty’ runs throughout Derrida’s *Cosmopolitanism and Forgiveness*, but he uses the very expression on p. 34.

¹¹ My account of punishment differs in significant ways from the standard account found in the literature. For my purposes here, the main noteworthy peculiarity of my account of punishment is that it sees punishment as a manifold phenomenon. The punisher, on my account, need not have the authority to punish, need not wish to communicate anything to the punishee, and so on. For more on my account, see my *Punishment and Retribution* (Aldershot: Ashgate, 2006).

¹² Hannah Arendt, *The Human Condition* (Chicago: University of Chicago Press, 1958), p. 241.

¹³ See Christopher Bennett, ‘Personal and Redemptive Forgiveness’, *European Journal of Philosophy* 11 (2003): 127–44, for a defense of the possibility of punishment and (one peculiar form of) forgiveness coexisting simultaneously.

course, something may, *on some other grounds*, be unpunishable but forgivable (and vice versa). But whether something *should* be punished or forgiven is a different discussion. The paradox, again, is that sometimes an act which presumably ought to be punished (and which, therefore, is simultaneously punishable and forgivable), somehow ought to be forgiven as well.

The term ‘unforgivable’ is thus ambiguous: it refers (in its strictest sense) to the impossibility of forgiving, and also (in a looser but much more widespread sense) to the inappropriateness of forgiving. To fail to distinguish the two senses is to be guilty of committing the same very famous mistake that John Stuart Mill committed as he equivocated between two different senses of ‘desirable’.¹⁴ The error, in effect, is to confuse descriptive and normative enterprises, which I noted at the outset. In dealing with the paradox of forgiveness we must be particularly careful, for the paradox of forgiveness reveals the difficulties facing, on the one hand, an *account* (or definition, or description, or theory) of the phenomenon of forgiveness itself, and the difficulties facing a *justification* of the phenomenon, on the other.

Both Kolnai and Derrida, however, like most authors, lump together these issues, as they in the same breath refer to ‘objectionable and ungentle’ forgiveness (Kolnai), or to what ‘we can or should do’ in connection to forgiveness (Derrida). This is terse to a fault. We can bring to bear a lesson from the philosophy of punishment to the discussion of forgiveness: our understanding of punishment, qua phenomenon, has not been advanced by the recalcitrant tendency to confuse the problem of its definition with the problem of its justification. Indeed many justifications of punishment are misleadingly advertised as if they were theories of punishment.¹⁵ This error in the philosophy of punishment, which goes back at least until Plato,¹⁶ has, however, recently been offered as the beginning of wisdom by one of today’s most insightful forgiveness and punishment theorists. This is how Jeffrie G. Murphy puts it:

¹⁴ See John Stuart Mill, *Utilitarianism* (London: Longmans, Green & Co., 1901), pp. 52ff.

¹⁵ See, e.g., the (largely unheeded) warning in Antony Flew, ‘The Justification of Punishment’, *Philosophy* 29 (1954): 291–307, especially at 297. This is not to deny that sometimes, particularly regarding value-terms, the relationship between definition and justification is very close. (See, on this, Gerald Dworkin, *The Theory and Practice of Autonomy* [Cambridge: Cambridge University Press, 1988], pp. 1–33, and Hilary Putnam, *Reason, Truth and History* [Cambridge: Cambridge University Press, 1981], pp. 201–216.) My point is merely that there is a way of confusing definitional and justificatory enterprises which is famous in the case of punishment, and which, although not at all famous, is also present in the case of forgiveness.

¹⁶ See Plato, *Protagoras* 324a–b, in W.K.C. Guthrie (trans.), Edith Hamilton and Huntington Cairns (eds.), *The Collected Dialogues of Plato* (Princeton: Princeton University Press, 1989), pp. 308–52.

The Question ‘What is Forgiveness?’ cannot after all be sharply distinguished from the question ‘How is forgiveness justified?’ ... We cannot define forgiveness and *then* ask what moral reasons make it appropriate.¹⁷

Later I will have more to say about Murphy’s illuminating views on forgiveness, but for now I just wish to focus on the fact that this amalgam of definition and justification is inconvenient in the case of forgiveness (just as it is famously inconvenient in the case of punishment). Murphy, sensibly, disagrees with a certain widespread view of forgiveness according to which forgiveness is the simple overcoming of resentment. Thus Murphy suggests that forgiveness is ‘foreswearing resentment on moral grounds’, and that forgiveness is something we do ‘for a moral reason’.¹⁸ I think that these remarks are fundamentally correct. Yet, they do not, I think, entail quite what Murphy thinks: namely, that we cannot separate the definition and the justification of forgiveness. Consider an analogous example, that of ‘white lies’: these are lies we tell for a moral reason, say, kindness, or benevolence (that is, in fact, part of what distinguishes them from ‘normal’ lies), but this is by no means to *justify* white lies. There is a difference between asking ‘Why do you do X?’ and asking ‘Why should you do X?’

In spite of the fact that a given white lie is *motivated* by benevolent reasons, the question as to whether or not one is *justified* in lying (even whitely) remains pertinent. The fact that one always forgives for a moral reason—as Murphy correctly has it—does not entail that this moral reason is a good, or definitive one. Furthermore, even if it were true that whenever someone tells a white lie, she herself, at least, would have to believe that she is justified, this would not entail that she in fact is justified. In other words, it is one thing to define a white lie, and another to justify it—even if in defining we refer to some (moral) reasons that the white liar has for lying. Similarly, one thing is to describe what it is that we do when we forgive and another thing is to justify our forgiving—even if in the former enterprise we refer to the (moral) reasons that the forgiver has for forgiving. Forgiveness, like punishment, is the sort of phenomenon that stands in need of justification, and independently of issues of motivation which may be built into the very definition of the phenomenon.

¹⁷ Jeffrie G. Murphy, ‘Forgiveness and Resentment’, in Jeffrie G. Murphy and Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), p. 23.

¹⁸ Murphy, ‘Forgiveness and Resentment’, p. 24. For some interesting modifications (which do not affect the substance of my concerns in this article) of Murphy’s views see his ‘Remorse, Apology, and Mercy’, *Ohio State Journal of Criminal Law* 4.2 (2007): 423–53, and his American Philosophical Association presidential address, as published in his ‘Legal Moralism and Retribution Revisited’, *Criminal Law and Philosophy* 1.1 (2007): 5–20.

Confusing the definitional and the justificatory enterprises is of no help in answering the justificatory question. In particular, we should resist the following move regarding forgiveness (itself a parallel to a very famous move regarding punishment): forgiveness which is not (believed by the forgiver) to be deserved is not forgiveness. That is, just as some have built retributivism into the very definition of punishment, some may wish to build (something akin to) retributivism into the very definition of forgiveness. Elsewhere I have argued against this move in the case of punishment—a move which in effect caricaturizes the justification which is incorporated into the definition of punishment.¹⁹

While it remains to be seen whether forgiveness can ever be *deserved* (maybe granting forgiveness is always supererogatory),²⁰ the arguments against building the justification of punishment into its definition do apply to forgiveness as well. For the problem here is not with desert itself (it just happen to be the case that the only justification which is customarily built into the definition of punishment is desert-based retributivism), but with the general smuggling of justifications into definitions. The objection, broadly, is against the thesis that asserts that a punishment which is not believed by the punisher to be the right thing to do is not punishment. And, similarly, we can object to the thesis that asserts that forgiveness which is not believed by the forgiver to be the right thing to do is not forgiveness.

The most decisive objection to this sort of move is captured by H.L.A. Hart's famous discussion of what he dubbed 'the definitional stop'.²¹ One cannot simply evaporate deep moral questions *deus ex machina*, by turning them into merely conceptual questions. Turning retributivism into a merely logical thesis, à la Antony Quinton (et al.),²² and then saying that punishment which is not deserved is simply not punishment has not convinced any retributivists, and it has not advanced the normative debate between retributivists and consequentialists. Similarly, claiming that simply as a matter of definition, forgiveness which is not motivated by X (X being whatever reason, moral or otherwise) is not forgiveness is not likely to solve any interesting question as to whether or not we *should* forgive.

¹⁹ In addition to Zaibert, *Punishment and Retribution*, see Leo Zaibert, 'The Fitting, the Deserving, and the Beautiful', *Journal of Moral Philosophy* 3.3 (2006): 331-50.

²⁰ On the relation between retributivism and forgiveness see Kathleen Dean Moore, *Pardons: Justice, Mercy, and the Public Interest* (Oxford: Oxford University Press, 1989), pp. 89-165.

²¹ H.L.A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford: Clarendon Press, 1968), pp. 5ff.

²² Antony Quinton, 'On Punishment', *Analysis* 14 (1954): 133-42.

In the next section I shall elaborate on how the amalgam of the descriptive and the normative plays out within the specific context of requiring repentance in order for forgiveness to be possible (or to be morally appropriate). In closing this section, I would like to refer to one more general unfortunate side-effect of this amalgamation of definition and justification, namely a view whereby punishment and related phenomena are (mostly or usually) bad, whereas forgiveness and related phenomena are (mostly or usually) good—or, at the very least, that the two families of phenomena are radically opposed.²³ Hastings Rashdall, for example, openly opposes the ‘ethics of forgiveness’ to the ‘irrational’, ‘immoral’ (and ‘wholly unchristian’) ethics of (retributive) punishment.²⁴ This sort of assumption is visible even in some of the most valuable recent contributions to this discussion found in the specialized literature. Claudia Card, for example, begins her influential ‘On Mercy’ claiming that: ‘Mercy [and with it forgiveness] and retribution have been thought to presuppose fundamentally different ethical orientations’,²⁵ and a corollary to these views which she diagnoses (a corollary which she does not endorse) is that one of these orientations is good whereas the other one is bad. Similarly, Martha Nussbaum opposes the mildness and gentleness of *epieikeia* (equity, leniency, or mercy—all related to forgiveness), with the harshness and rigidity of *dike*, justice or retribution, or retributive punishment.²⁶ While few authors would deny that some instances of punishment are good and that some instances of forgiveness are bad, there nonetheless exists a certain pattern of approaching these issues whereby punishment and related phenomena are to be shunned and forgiveness and related phenomena are to be celebrated.

One noteworthy exception to this simplistic view, particularly useful for my purposes, is Collingwood, who eloquently disagrees with views asserting the moral superiority of punishment over forgiveness, and vice versa. Oddly, however, Collingwood concludes that ‘punishment and forgiveness are thus not only compatible but identical; each is a name for the one and only right attitude of a good will towards a man of evil will’.²⁷ But, for the reasons just

²³ Consider the title of Martha Minow’s famous book: *Between Vengeance and Forgiveness* (Boston, MA: Beacon Press, 1998); or, similarly, the title of Trudy Govier’s *Forgiveness and Revenge* (London: Routledge, 2002).

²⁴ See, e.g., Hastings Rashdall, ‘The Ethics of Forgiveness’, *International Journal of Ethics* 10.2 (1900): 193.

²⁵ Card, ‘On Mercy’, p. 182.

²⁶ Martha Nussbaum, ‘Equity and Mercy’, *Philosophy and Public Affairs* 22 (1993): 83–125, 85ff., and *passim*.

²⁷ Collingwood, *Essays in Political Philosophy*, p. 132.

canvassed, Collingwood's thesis is also to be rejected. His is, after all, but another way of conflating the justificatory and the descriptive enterprises. Collingwood is right, of course, in that the *motivation* behind forgiving and behind punishing could be identical, but this does not render forgiveness and punishment themselves identical phenomena.

2. Repentance and Forgiveness

The widespread claim that forgiveness requires repentance is often hard to assess because it is not clear whether what is meant is that, as a matter of logic, it is impossible to forgive the unrepentant, or, rather, that as a normative matter, forgiving the unrepentant is objectionable. This is not only the result of the looseness with which many authors shift between descriptive and normative enterprises which we have just discussed. For, as we have also seen, the very paradox of forgiveness which occupies our attention has a part which is purely descriptive and a part which is to an extent normative. I shall now try to show how the appeal to repentance is unsuccessful in resolving the paradox.

Although there are numerous defenses of the thesis that forgiveness requires repentance, the best starting point for this discussion is Kolnai's seminal article, to which I referred at the outset. In this article Kolnai forcefully argues that if so-called forgiveness were to be granted to the unrepentant, then it would not really be forgiveness, but, at best, mere *condonation*. Kolnai's distinction between forgiveness and condonation occurs in the context of distinguishing forgiving from all sorts of related phenomena, such as pardoning, absolving, forgetting, and the like. Surely this is a salutary enterprise, which by now has become rather common.²⁸ In fact, Kolnai's distinction is particularly valuable in that it also contains a rather enlightening account of the terms used to refer to forgiveness and related phenomena in different languages, thus avoiding the insularity of some 'ordinary language' analyses.²⁹

The project of distinguishing forgiveness from related phenomena is not, however, without risks, and I think that one of these risks materializes in Kolnai's distinction between forgiveness and condonation. The main risk is the abuse of stipulation: to draw sharp analytic distinctions between allegedly

²⁸ See Haber, *Forgiveness*, pp. 11ff., 59ff., and *passim*; Lang, 'Forgiveness', pp. 113–15.

²⁹ The linguistic distinction between 'pardon' and 'forgiveness', in English, for example, does not have currency in some other languages, like French, as Kolnai points out—and indeed not, either, in romance languages in general. The general point is that the peculiarities of this or that natural language *need not* be philosophically meaningful.

diverse phenomena by mere fiat. Again, the philosophy of punishment offers illustrative examples: the infamous use of the ‘definitional stop’, or the common attempt to distinguish punishment from revenge (or from monstrosities, or from barbaric cruelty), are illustrative cases.³⁰ Of course, to define any entity, is, ipso facto, to admit the existence entities other than the one being defined. Not all ‘definitional stops’ are vicious or to be avoided (if they were, then all definitions would be vicious). I will argue that Kolnai is guilty of employing the definitional stop in a vicious way—as are, too, many of those who require repentance for the definition (or the justification) of forgiveness.

Kolnai defines condonation as a state of affairs whereby someone ‘deliberate refrains from any retributive response’ to the wrong, even though he sees it as a wrong, and does consequently disapprove of it.³¹ Now, this refraining can be motivated by a variety of reasons, although they are not as varied so as to render condoning indistinguishable from ‘indifference’, ‘laziness’ ‘light-mindedness’, ‘exculpation’, ‘absolution’ and other phenomena. In Kolnai’s view, the way in which the condoner ‘acquiesces in the offence’ differs from the way in which the indifferent, the light-minded, the lazy, the exculpator, the absolver, etc., deal with the offence. Unlike all these other forms of refraining from inflicting punishment, condonation constitutes for Kolnai ‘a more consciously decisional act [more than the other ones just mentioned] and so far closer to a *simulacrum* of forgiveness proper’.³² Interestingly, then, while Kolnai’s central thesis is that condonation and forgiveness are dramatically different, condonation is still closer to forgiveness than all these other related phenomena—indeed condonation is, in an unexplained sense, a ‘*simulacrum* of forgiveness’.³³ Yet, Kolnai and followers seem to forget this very point: that condonation is in fact rather similar to forgiveness³⁴—a point to which I will return in the last section of the article.

In Kolnai’s opinion, condonation differs ‘sharply’ from forgiveness, however, in that ‘it does not presuppose and nullify the original retributive

³⁰ See Zaibert, ‘Punishment and Revenge’, *Law and Philosophy* 25 (2006): 81–118.

³¹ Kolnai, ‘Forgiveness’, p. 95.

³² Kolnai, ‘Forgiveness’, p. 96.

³³ It could be argued that the similarity between forgiveness and condonation that I am emphasizing is merely structural, and thus not too interesting. So, my argument resembles the claim that a lie is more similar to a truth than are groans or exclamations. But this argument is not persuasive, since the other phenomena which Kolnai analyzes share that similar structure too, and since the similarities Kolnai admits between forgiveness and condonation are, in any event, not merely formal.

³⁴ As evidenced, for example, by the fact that more than one of the quotations which the *Oxford English Dictionary* presents as illustrating the meaning of ‘condonation’ equate it with forgiveness.

position'.³⁵ Sadly, however, Kolnai says precious little as to what this retributive position is, or about how it is that forgiveness *alone* nullifies it. Surely it is part of the logical structure of forgiveness that it is a response to (perceived) *culpable* wrongdoing—and with this Kolnai agrees, although he has not shown that this need be different regarding condonation. In fact, Kolnai hardly discusses the reasons why the condoner condones. Abruptly, and somewhat gratuitously, Kolnai tells us that: 'condonation is ... virtually "conniving" and immoralistic; in its gravest forms, it is not only undignified and self-soiling but unfair'.³⁶

Not only is this dismal opinion of condonation odd in light of the fact that by his very own admission, condonation is more like forgiveness than are the host of other related phenomena which he discusses, but also in light of the fact that Kolnai asserts that:

to condemn *all* condonation might, however, amount to over-severity; for it seems plausible that without condoning *some* faults we could not possibly live together, nor, for that matter, with ourself.³⁷

Maybe it is true that we cannot possibly live together or even in peace with our very selves if we did not do some undignified, self-soiling, or unfair things. But if condonation in some cases allows us to live with others and with ourselves, and these things are valuable, then we should pause before we flatly condemn condonation. And yet, condemning condonation flatly is precisely what Kolnai does. He further distinguishes condonation from forgiveness in that the latter is 'supposed to contribute to the eradication of wrongdoing' whereas the former somehow contributes 'to the fostering of it'.³⁸ This familiar amalgam of the justification and the definition of forgiveness (and of condonation) should by now be admitted to be clearly problematic. It is not clear that forgiveness contributes to the eradication of wrongdoing, or that condonation does not.

The standard view is that punishment (and not the refusal to punish—of which *both* forgiveness and condonation are instances) contributes to the eradication of wrongdoing. Whether or not punishment indeed contributes to the eradication of wrongdoing, however, says little about the logical structure of the phenomenon of punishment itself; similarly, whether or not condonation and forgiveness would differ as to their effects vis-à-vis the eradication of wrongdoing is not to pick out a difference in the structure of the two phenomena. Kolnai, however, summarily and facilely decrees that condonation and forgiveness are 'sharply' different, and he confidently moves

³⁵ Kolnai, 'Forgiveness', p. 96.

³⁶ Kolnai, 'Forgiveness', p. 96.

³⁷ Kolnai, 'Forgiveness', p. 96.

³⁸ Kolnai, 'Forgiveness', pp. 97–98.

along, focusing on the ‘unobjectionable and genuine’ forgiveness which occurs only when we forgive the sincerely repentant.

A one-line summary of Kolnai’s position is that (the disposition) to forgive (the repentant) is a good thing, whereas the mere condoning of the unrepentant is not—which is problematically reminiscent of the sort of simplistic moves (discussed above) regarding the goodness of forgiveness and the badness of punishment, except that now the culprit is not punishment but condonation. Kolnai’s move is, I submit, just ‘old wine in new skins’: it is a vicious appeal to the definitional stop. If a certain behavior is a ‘consciously decisional’ response to wrongdoing, independently of whether or not the wrongdoer repents, it is condonation *and a bad thing*; if the consciously decisional response is the result of believing that the wrongdoer repents, then it is forgiveness *and a good thing*. Thus, it problematically turns out that acts of forgiveness are, as a matter of definition, good—provided that we are justified in believing that the wrongdoer really repents, we cannot be mistaken in forgiving her.

Nowhere does Kolnai tell us how it is that something—condonation—which, by his own admission, resembles forgiveness so closely, can nonetheless be so bad, in spite of the fact that forgiveness is a good thing. This reeks of gratuitous stipulation, and, in the last section of the article I shall suggest that the difference between condonation and forgiveness is much more tenuous and much more complicated than Kolnai (and others) take it to be. For now, I wish to discuss other problems facing the thesis that forgiveness requires repentance.

The essential reason why repentance is taken by Kolnai and others to be so important is that, by repenting, the wrongdoer somehow distances herself from her act. Thus, Kolnai and others think, repentance facilitates forgiveness, for the forgiver can simultaneously condemn the wrong and forgive the wrongdoer. That is, they urge, there is a difference between what we do *to* the wrongdoer, and what we do *about* the wrongdoing. As a matter of fact, the repentant wrongdoer can join forces with the forgiver in condemning the wrongdoing—to repent means, in part, to condemn the wrong one committed. This idea includes the ‘hate [and blame, and punish] the sin but love [and forgive] the sinner’ line, whose lineage goes back to Saint Augustine—a position still much en vogue nowadays, for example, in the Catholic church’s attitude regarding homosexuality.

Kolnai sensibly disagrees with this suspicious and overly ‘neat separability between the sin and the sinner’, which he in fact dubs ‘fictitious’.³⁹ Yet, Kolnai does endow repentance with this alchemistic property of rendering the

³⁹ Kolnai, ‘Forgiveness’, p. 97.

repenting wrongdoer significantly different from the unrepenting wrongdoer, and from herself before she repented—and thus somehow worthy of forgiveness. Similarly, Murphy admits that, in principle, ‘it is, of course, impossible to hate the sin and not the sinner’, but he admits that there is at least one exception to this ‘impossibility’: the case in which the sinner is no longer ‘intimately identified with his sin’.⁴⁰

Thus, for Kolnai, Murphy, and others, the most important way in which a wrongdoer can separate herself from her wrongdoing is through repentance. But this suggestion presupposes a problematically communicative account of wrongdoing and of forgiving. For example, Murphy assumes that one important reason why we resent instances of wrongdoing (in fact, one important reason why these acts are wrong) is that they are ‘*messages*—symbolic communications. They are ways a wrongdoer has of saying to us ‘I count but you do not’, ‘I can use you for my purposes’, or ‘I am here up high and you are down there below’.⁴¹ These messages are disrespectful, and it is mainly this disrespect that we resent.

While sometimes wrongdoers may wish to send those messages, or sometimes certain actions *mean* (independently of the wrongdoer’s intentions) sending this or that message, sometimes things are otherwise. In fact, while Murphy comes close to tacitly admitting that his model really applies only to intended wrongdoing,⁴² he seems to fail to see the implications of this admission. For surely unintended culpable wrongdoing exists, and as such it is both punishable and forgivable, even though, presumably, these instances of wrongdoing do not constitute or entail sending the messages that Murphy describes as one important reason why we resent these wrongs. Moreover, as Garrard and McNaughton have perspicuously suggested, the communicational view of wrongdoing fails even in cases of fully intended wrongdoing:

if my car is stolen by someone wholly unknown to me, it would be odd to say, without a special context, that I resent it, and this is because it would be odd for me to regard myself as being personally slighted [and wrong for me to think that the wrongdoer was communicating anything whatsoever—to me or to anybody else]. The unknown car thief does not pick on *me* [or on anyone else] (any car would do, mine just happens to be there).⁴³

⁴⁰ Murphy, ‘Forgiveness and Resentment’, p. 24.

⁴¹ Murphy, ‘Forgiveness and Resentment’, p. 25.

⁴² See, e.g., Murphy, ‘Forgiveness and Resentment’, p. 25: ‘Intentional wrongdoing *insults* us and attempts (sometimes successfully) to *degrade* us’. Some forms of unintentional wrongdoing, say borderline cases of extreme recklessness, can also send messages, but they tend to be messages of indifference rather than of a directly *insulting* or *degrading* nature.

⁴³ Eve Garrard and David McNaughton, ‘In Defence of Unconditional Forgiveness’, *Proceedings of the Aristotelian Society* (2003): 39–60, at 42–43.

Still, for the defenders of the forgiveness-requires-repentance thesis, the wrongdoer who had initially sent messages along the lines Murphy sketches, as she repents, she then sends different messages, to the tune of ‘I no longer stand behind the wrongdoing, and I want to be separated from it. I stand with you in condemning it’.⁴⁴ It is in virtue of this message of repentance, then, that we forgive her. In light of the limitations arising from the fact that this sort of view is a non-starter in cases of unintended wrongdoing, and that it does not work either in the many cases resembling Garrard and McNaughton’s example, then, even if Murphy and Kolnai were correct, it would still be true that their account/justification (their amalgam) of forgiveness would only work in a rather limited number of cases. And yet it is a perfectly common occurrence to forgive unintended wrongdoing, and wrongdoing of the sorts described by Garrard and McNaughton.

But there is more which is problematic in the communicative view of wrongdoing and forgiveness presupposed by the forgiveness-requires-repentance view. If wrongdoing need not involve communicating anything (as I have glossed over Garrard and McNaughton’s example in order to emphasize), why should repentance perforce involve communication? Why, moreover, should what the repentant wrongdoer experiences, even if she wishes to communicate it, have the force to perform the alchemist trick of turning what would otherwise have been merely bad condonation into good forgiveness (or, indeed, the force to make punishment somehow inappropriate)? Lurking in the background of these views is the additional view whereby forgiveness itself is communicational: by forgiving we send a message to the forgivee to the effect that she is now welcome back into our community, or that we no longer hold any grudge against her, etc. The same questions regarding the communicational analysis of wrongdoing and of repentance are pertinent in the case of the communicational account of forgiveness: Why should forgiveness communicate anything at all? Could we not forgive privately, even in secret?

The mystery surrounding the quasi-magical powers of repentance is not solved by looking at the important role that repentance has in many different religions. As is well known, the idea that true repentance on the part of the wrongdoer requires that the aggrieved party forgives her is ‘fundamental to Judaism [and to other religions as well]’.⁴⁵ As Haber points out:

⁴⁴ Murphy, ‘Forgiveness and Resentment’, p. 26. Murphy admits that there exist legitimate moral reasons for forgiving other than repentance, but he is emphatic in that repentance is ‘the clearest way in which a wrongdoer can sever himself from his past wrong’. *Ibid.*

⁴⁵ Aaron Lazare, *On Apology* (Oxford: Oxford University Press, 2004), p. 241.

Under Rabbinic law, on the eve of Yom Kippur, wrongdoers are obligated to ask forgiveness from those they have wronged. Should their request be denied, they are obligated to ask two more times, at which point, should their request be denied, it is the victim—not the wrongdoer—who is morally to blame.⁴⁶

It would be interesting to see how many rabbis would truly forgive, say, Hitler, after he sincerely apologized three (or three thousand) times, and why, if some would indeed forgive him, this would be the right thing to do. Some people would compellingly refuse to forgive certain acts, again, say Hitler's acts, no matter how sincerely (or often) he repented—Hitler's acts could be seen as unforgivable, in the sense that they should never be forgiven. Yet, the question as to what exactly this awesome power of repentance is remains unanswered. Unfortunately, the defenders of the forgiveness-requires-repentance thesis say precious little of help in answering this question. Merely to assert that repentance is communicative along the lines that these defenders sketch is not to explain why this communication has the effects that they claim it has.

There are, of course, immense epistemological limitations regarding our ability to know when someone else's (or even our own) repentance is full and sincere. But, putting these limitations aside, I suggest that there are two options (one of which is further subdivided) as to what repentance may entail, none of which is useful for Kolnai, Murphy, and other endorsers of the forgiveness-requires-repentance thesis.

On the one hand, if the repenting wrongdoer would merely, and in more or less jejune ways, 'feel bad' about what she has done, then it seems to be more or less inadequate to grant forgiveness—to forgive in this case would be an instance of what Dietrich Bonhoeffer calls 'cheap grace'.⁴⁷ If merely feeling bad about my wrongdoing were sufficient for being forgiven, it is hard to imagine how forgiveness could possibly be conducive, as Kolnai believes it is, to the eradication of wrongdoing. On the other hand, however, even if the repenting wrongdoer displays a robust, sobering owning up to her own wrongdoing, in such a way that she *deeply* suffers for having behaved as she did, this will still be problematic. For now two further alternatives suggest themselves. First, if someone really repented this maturely, arguably she would punish herself, and, at any rate, she may not accept to be forgiven, for otherwise she would rightly see this as the cheap grace of the previous case. Second, it is possible that this suffering is by itself constitutive of punishment. But if this is so, then so-called forgiveness turns out to be, again, spurious: we are not quite forgiving this wrongdoer, we are merely refraining from

⁴⁶ Haber, *Forgiveness*, p. 103 and *passim*.

⁴⁷ Quoted in Lazare, *On Apology*, p. 241.

victimizing her, given that she has already been punished (even if she punished herself) for what she did.

As it turns out, invoking repentance does not resolve the paradoxical nature of forgiveness, for, after much fancy footwork, we remain in roughly the same place we started. Either putative instances of forgiveness are spurious (in the case of the truly repentant wrongdoer who has already punished herself), or, when the putative instances of forgiveness are not spurious, then they seem more or less like ‘cheap grace’, more or less unjustified. Thus, resolving the paradox of forgiveness via the appeal to repentance seems doomed. I thus turn to the discussion of the type of forgiveness worth our attention: pure forgiveness.

3. Pure Forgiveness and Treating the Guilty as Guilty

Some (unbridled retributivists) have asserted that culpable wrongdoing is *sufficient* (not merely necessary) for the justified punishing of X. These influential authors (say, Immanuel Kant or Michael Moore)⁴⁸ would then have difficulty admitting that forgiveness is ever possible (or ever justified). Most contemporary authors seek to justify punishment combining retributivist and consequentialist rationales. Independently of the success of these mixed justifications of punishment (about which I am rather skeptical),⁴⁹ it is clear that the force of retributivism is irresistible: a wholly non-retributive justification of punishment is not very attractive. And yet retributivism is, in some sense, at odds with forgiveness. Thus, while forgiveness poses a much more serious problem to unbridled retributivist positions, it nonetheless poses problems to any justification of punishment that contains a retributivist element.

I am unaware of philosophers who flatly assert that culpable wrongdoing is sufficient for the justified forgiving of X, but if there are, then there surely would arise the ‘paradox of punishment’ along the lines of the paradox of forgiveness.⁵⁰ What seems downright inconsistent is for someone to assert that

⁴⁸ See, e.g., Immanuel Kant, *The Metaphysical Elements of Justice*, trans. John Ladd (New York: Macmillan, 1965), *passim*; Michael Moore, *Placing Blame* (Oxford: Clarendon Press, 1997), *passim*. Moore is more nuanced than Kant in his discussion of conflicts of duties, but if these nuances allowed him to tolerate forgiveness, then they would create other problems for his views. See my *Punishment and Retribution*, pp. 173ff., where I discuss unbridled retributivism.

⁴⁹ See, e.g., Zaibert, *Punishment and Retribution*, ch. 1.

⁵⁰ The famous biblical passage ‘If someone strikes you on the right cheek, turn to him the other also’ (Matthew 5) may be seen as suggesting that wrongdoing is indeed sufficient for the justified forgiving of X. Assuming, somewhat plausibly but by no means certainly, that the turning of the other cheek involves forgiving, will not quite turn this into an objection to my claim that there is no paradox of punishment. For these sorts of claims need to be understood within

culpable wrongdoing is sufficient for *simultaneous* justified punishment and justified forgiveness. In any event, even if one does not endorse extreme positions, in some cases at least we think that punishment is the right thing to do, and then it seems that in those cases forgiveness would perforce not be the right thing to do. In some cases at least, ‘either punishment is right and forgiveness wrong, or forgiveness is right and punishment wrong’, as Collingwood would have it.⁵¹ And yet, it seems that it is precisely in cases where punishment is or seems right that forgiveness would be interesting.

In fact, I think this is part of what describes the sort of case that Derrida has in mind when he talks about ‘pure forgiveness’ or ‘unconditional forgiveness’. Derrida distinguishes between an ‘economic’ type of forgiveness and another ‘aneconomic one’; only the latter is ‘pure’ in his sense. Derrida opposes, and in my opinion rightly, the view that forgiveness is conditioned by certain actions by the wrongdoer, in such a way that if the wrongdoer does not engage in those actions forgiveness would be logically impossible (or, in other formulations, that it would be unjustified). As Derrida puts it:

I would be tempted to contest this *conditional* logic of the *exchange*, this presupposition, so widespread, according to which forgiveness can only be considered *on the condition* that it be asked, in the course of a scene of repentance attesting at once to the consciousness of the fault, the transformation of the guilty, and the at least implicit obligation to do everything to avoid the return of evil.⁵²

In contrast to this ‘economic’ sort of forgiveness (the term ‘economic’ suggests the idea of ‘transaction’ or ‘exchange’ in the case of conditional forgiveness; the term ‘aneconomic’ describes unconditional forgiveness, in which this transaction or exchange is absent), Derrida describes pure forgiveness as:

the *unconditional*, gracious, infinite, aneconomic forgiveness granted *to the guilty as guilty*, without counterpart, even to those who do not repent or ask forgiveness.⁵³

For analytic purposes, I would like to steer clear of the discussion of grace or of infinitude, and focus exclusively on the idea of forgiveness without repentance (or apologies). This strategy, moreover, allows me to remain in close contact with the *secular* literature on forgiveness.

the context of the view that the reason why we are to turn the other cheek is because it is God himself—not us—who would do the punishing (see, e.g., Romans 12). So, punishment is not really eliminated in favor of forgiveness; rather we are asked to engage in inaction (it is an interesting question whether this inaction is indeed to forgive), insofar as God will act (on our behalf).

⁵¹ Collingwood, *Essays in Political Philosophy*, p. 124.

⁵² Derrida, *On Cosmopolitanism and Forgiveness*, p. 34.

⁵³ Derrida, *On Cosmopolitanism and Forgiveness*, p. 35.

I think that turning towards something akin to what Derrida calls pure forgiveness is crucially important if we wish to better understand the paradox of forgiveness. Derrida is emphatic about the primacy of his version of pure forgiveness:

despite all the confusions which reduce forgiveness to amnesty or to amnesia, to acquittal or prescription, to the work of mourning or some political therapy of reconciliation, in short to some historical ecology, it must never be forgotten, nevertheless, that all of that refers to a certain idea of pure and unconditional forgiveness, without which this discourse would not have the least meaning.⁵⁴

And yet, I think that Derrida does not go far enough in isolating pure forgiveness, and in separating it away from these other reductive manifestations. For example, in the second to last quotation he talks about pure forgiveness being *granted*, and this seems to me to be unnecessarily close to precisely those sorts of economic transactions which he wishes to avoid. Moreover, as he discusses the right of grace (and executive pardons),⁵⁵ he refers to a ‘personal head-to-head or face-to-face [encounter], which one could think is required by the very essence of forgiveness’.⁵⁶ And Derrida does not distance himself from the necessity of a ‘face-to-face’ encounter between the wrongdoer and the forger—in fact, Derrida problematizes cases in which language may render this encounter moot, namely, when the wrongdoer and the forger do not speak the same language, or even when they do not share the same values. Given Derrida’s interest in pure forgiveness, it is puzzling to see how in his book, *Cosmopolitanism and Forgiveness*, he discusses forgiveness in connection to political phenomena which, though surely important, are not ‘pure’ in his sense.

Thus Derrida, in my opinion, does not sufficiently distance himself from conditional forgiveness. Granted, he distances himself from one of those conditions, in fact from the commonest of these conditions (in terms of its treatment in the specialized literature): repentance. But he does not get around to analyzing this pure forgiveness which he thinks is so important. And while it is true that the distinction between conditional and unconditional forgiveness is drawn up in terms of whether or not one requires that the wrongdoer repents, this is, logically speaking, obviously not the only condition that may be imposed on forgiveness.

⁵⁴ Derrida, *On Cosmopolitanism and Forgiveness*, p. 45.

⁵⁵ See, for more on Derrida’s treatment of forgiveness in connection to public policy, Austin Sarat, *Mercy on Trial: What it Means to Stop an Execution* (Princeton: Princeton University Press, 2005), pp. 71ff.

⁵⁶ Derrida, *On Cosmopolitanism and Forgiveness*, p. 48.

Some of the most important defenses of unconditional forgiveness are in fact defenses of the morality of forgiveness without repentance, and only tangentially touch upon those purely conceptual or analytic aspects which I discuss here.⁵⁷ That is, they limit themselves to claiming that forgiveness need not be conditional on the wrongdoer's repentance (or on her apologies) in order for it to be 'morally permissible'.⁵⁸ Moreover, defenders of unconditional forgiveness typically do not address the issue of whether or not even what they call unconditional forgiveness needs to be *granted*, or *communicated*, to the wrongdoer. In contrast to these standard moves, I wish to emphasize two aspects of my account. First, I am not primarily interested in the morality of forgiveness (except for those moral aspects which may be necessary to resolve the paradox); my 'defense' of unconditional or pure forgiveness is simply an attempt to show how it can exist. Second, my account of pure forgiveness is non-transactional through and through; it is wholly aneconomic, to deploy Derrida's term—it need not be granted or communicated at all, and thus it is irrelevant for my purposes whether or not the forgiver and the wrongdoer speak the same language, share the same values, and so on.

My pure forgiveness is thus *purier* than even Derrida's pure forgiveness. Furthermore, my account of pure forgiveness is not quite identical to what authors in the literature mean by unconditional forgiveness. For example, as they discuss Gordon Wilson's reaction to the death of his daughter as a result of the IRA's infamous Remembrance Day Bombing on 8 November 1987—he forgave the assassins—commentators typically assert that this is an instance of unconditional forgiveness, simply because it was not conditioned by the requirement that the assassins repented.⁵⁹ And yet, quite obviously, this case could have been conditioned by all sorts of other considerations, such as the belief, on Wilson's part, that such an explicit and sensational case of forgiveness may have contributed to the end of violence, the attainment of peace, and so on. That is, it is often summarily assumed that forgiveness given without requiring repentance is thereby unconditional—although it is obvious that, in addition to repentance, there could be all sorts of conditions placed upon forgiveness.

⁵⁷ The best defense of unconditional forgiveness in the recent literature is, in my opinion, Garrard and McNaughton, 'In Defence of Unconditional Forgiveness', pp. 39–60.

⁵⁸ Garrard and McNaughton, 'In Defence of Unconditional Forgiveness', p. 39.

⁵⁹ See Garrard and McNaughton, 'In Defence of Unconditional Forgiveness', p. 39; see also Bennett, 'Personal and Redemptive Forgiveness'. Similarly, both Garrard and McNaughton and Bennett praise Wilson's forgiveness, and in fact the former authors explicitly assert that 'it is always admirable to forgive' ('In Defence of Unconditional Forgiveness', p. 40). For reasons which by now should be clear, that I am not discussing the morality of what Wilson did.

My account of pure forgiveness is related to my own work on punishment theory, where I have urged that, in order to understand the phenomenon of punishment itself, it is advisable to focus on cases purer than the one that tends to monopolize the attention of specialists (criminal punishment carried out by the State). This suggestion is not at all motivated by the view that criminal punishment carried out by the State is not a very important topic (undoubtedly it is), but by the view that the discussion of such a type of punishment is contaminated by all sorts of (again, terribly important) discussions pertaining to political philosophy, thus obscuring the phenomenon of punishment itself, *qua* phenomenon. Thus, with this eidetic reduction of sorts, I have tried to analyze what happens when a person punishes, independently of whether this person has the authority to do so, wishes to communicate anything, is doing the right thing, and so on. Here I wish to apply this method to the analysis of forgiveness.

When analyzing punishment, I have suggested a much tighter connection between punishment and blame than is typically acknowledged. To punish is to do something about what we find blameworthy. And then the obvious question suggests itself: blame is to punishment is like _____ is to forgiveness? Notice that blame, punishment, and forgiveness are all responses to wrongdoing. Notice, too, that if the forgiveness-requires-repentance thesis were correct, forgiveness would not be a response to wrongdoing simpliciter, but rather to wrongdoing-which-is-repented; presumably, adherents to this thesis would also say that punishment is not a response to wrongdoing simpliciter, but rather to wrongdoing-which-is-not-repented. But insofar as punishing is to go beyond mere blaming; the question remains: what is that more basic phenomenon regarding which going beyond entails that one is forgiving? Or is the question rather: blame is to punishment like forgiveness is to _____? The fact that it is not clear whether the blank in these questions should be to the left or to the right of ‘forgiveness’ suggests the answer: the blank, in either case, should be filled by repeating the very term ‘forgiveness’: blame is to punishment like forgiveness is to forgiveness.

In other words, the term ‘forgiveness’ is ambiguous—and this ambiguity is not customarily noticed. The ambiguity is between forgiveness as a pure mental phenomenon, and forgiveness as the communication of this mental phenomenon. To be sure, that there exists something like private forgiveness and something like communicative forgiveness is not an entirely novel thesis. Solomon Schimmel, for example, presupposes a distinction between ‘private’ and ‘interpersonal’ forgiveness, which in some ways resembles my distinction.⁶⁰

⁶⁰ Solomon Schimmel, *Wounds not Healed by Time: The Power of Repentance and Forgiveness* (Oxford: Oxford University Press, 2002), pp. 42ff.

Christopher Bennett has recently interestingly distinguished between ‘redemptive’ and ‘personal’ forgiveness; a distinction which also bears resemblance to Schimmel’s distinction and to mine. Some of the main differences between my distinction and others (such as Schimmel’s or Bennett’s) shall become apparent below, as soon as I present my analysis of each of my two types of forgiveness. But some important differences between my views and similar theses stem from their emphasis on the healing power of forgiveness, its power to restore relationships, and with the sorts of considerations regarding the loss of status that we suffer when we are wronged, and which I discussed above, in connection to Murphy’s communicative views.

For example, Bennett believes that to ‘grant’ personal forgiveness (while at the same time withholding redemptive forgiveness) to the unrepentant wrongdoers is rather admirable. I do not make such pronouncements; and I suspect that this alleged admirableness would have to be determined on a case-by-case basis. Moreover, these other forms of ‘personal’ or ‘private’ forgiveness are not the sort of purely mental phenomena with which I am interested, insofar as they tend to require that through ‘personal’ or ‘private’ forgiveness, the forgiver may come to ‘treat the wrongdoer as an equal’. In fact, as Bennett examines the possibility that what he calls personal forgiveness be not forgiveness at all, he concludes that it must be forgiveness insofar as (1) ‘it involves the overcoming of hostile feelings towards the wrongdoer occasioned by the wrongdoing’ and (2) ‘it involves the resumption of relations between two individuals’.⁶¹ While I agree with Bennett’s thesis that personal forgiveness is indeed forgiveness, I do not find these reasons compelling (as will become clear immediately—I believe we can forgive without overcoming hostile feelings, and without necessarily wishing to resume any relationships).

Thus, my distinction between pure and communicative forgiveness is different from similar extant distinctions. Moreover, to focus on the dangers of the ambiguity inherent to these two types of forgiveness which nonetheless go by the very same name is not terribly common. That is, those authors who admit that there are different types of forgiveness are not primarily interested in explaining the ambiguity that I am exposing here. Less common yet is the attempt to see the connection between this ambiguity and the paradox of forgiveness.

Hopefully, it is immediately clear that regarding the pure mental phenomenon of forgiving, few limitations obtain (fewer, in any case, than regarding the phenomenon of its communication). For example, I may forgive, as a pure mental phenomenon, wrong acts which have not harmed me in any way, and

⁶¹ Bennett, ‘Personal and Redemptive Forgiveness’.

in this sense I can forgive dead people, myself, and so on; whereas, for obvious reasons, it is not clear whether this is possible regarding forgiveness as a communicative phenomenon. The analyses of each of these phenomena should make these differences explicit. These analyses are extensions of my account of blame (which I have defended elsewhere).⁶² A blames B for Xing, when:

- (1) A believes that X is wrong,
- (2) A believes that X is an action of B,
- (3) A believes that B is a moral agent,
- (4) A believes that there are no excuses, justifications or other circumstances which would preclude blame,
- (5) A believes that the world would have been a better place had B not done X.
- (6) A believes that the world would be a better place if something would happen to B, something which would somehow offset B's Xing.
- (7) B's having Xed tends to make A *feel* something negative, i.e., a reactive emotion, like outrage, indignation or resentment.

A forgives B (as a pure mental phenomenon) when, in addition:

- (8) A believes that the world would in fact be a worse place if *she* did something to B in response to her wrongdoing, and thus she deliberately refuses to try to offset B's wrongdoing.⁶³

A forgives B (in the communicative sense) when, finally:

- (9) A communicates to B, or to someone else, that she has forgiven (in the sense of a pure mental phenomenon) B.

If instead of (8) and (9), I would simply write (8'): 'A does something to B, which A believes it is painful for B to endure, as a response to B's having Xed', and leave it at that, then we would have an analysis of punishment.⁶⁴ Since it is obvious that one cannot consistently satisfy (8) and (8') simultaneously, then it is clear that, on my analysis, to punish and to forgive are indeed mutually synchronically exclusive. Of course, one may at one time satisfy (8) and at another time satisfy (8')—one can, indeed, oscillate between these two states.

⁶² Leo Zaibert, *Five Ways Patricia Can Kill her Husband: A Theory of Intentionality and Blame* (Chicago: Open Court, 2005).

⁶³ The appeal to the world being a 'better (or worse) place' is very general; for more on what this appeal entails see my *Punishment and Retribution, passim*.

⁶⁴ Again, for a defense of this admittedly broad account of punishment, see my *Punishment and Retribution, passim*.

What one cannot do, on pain of irrationality, is to satisfy both (8) and (8') simultaneously. To concentrate on forgiveness itself, the first thing that my analysis brings out is that while the two types of forgiveness have a lot in common, the purest form of forgiveness is mental forgiveness; for whether or not we should communicate our forgiveness is to add a new dimension to the analysis. I will thus concentrate only on what goes on in someone's head when she forgives, though much of what I say may apply as well to communicative forgiveness.

There are several peculiarities of my analysis of pure forgiveness, beyond the by now obvious point that it is not conditioned in any of the usual senses, and certainly not conditioned by the wrongdoer's repentance, her apologies, or indeed by any behavior on her part. Of course, the account may be seen as 'conditioned' by the fact that the wrongdoer acted, that her action is wrong, and that she is not excused, justified, etc.—but these are merely innocuous matters of logic, not smuggled normative issues.

My account of forgiveness makes explicit that merely to claim that forgiveness, like punishment, as a matter of sheer logic requires wrongdoing, or even culpable wrongdoing, is not to go far enough. We can only forgive, just like we can only punish, that which we find *blameworthy*. This realization has important implications. In contrast to the overwhelming majority of views, my account shows that not only is it possible to forgive and continue to blame, but that this is the most natural form of forgiveness, at least in the sense that it is the one that happens first. To come to believe (8) in no ways entails ceasing to believe any of the elements contained in (1) through (7). Moreover, and since my account of blame includes an emotional component (feelings of resentment, outrage, or indignation), it follows, too, that we can forgive and continue to resent, continue to hold (emotional) grudges against the forgivee. In addition to the mental phenomena described in (1) through (8) above, a forgiver may also wish to get rid of the emotional component described in (7), or may come to cease to believe some of the elements contained in the analysis, but this is by no means necessary in order for forgiveness to exist. All that is necessary, given that A blames B, is that A would also comply with (8), and this compliance does not entail (nor preclude) the abandonment of any of the other elements in the analysis.

The overwhelmingly standard view, in contrast, has it differently: to forgive is, explicitly, to overcome resentment (even if this overcoming is done for a moral reason). That is, on this standard understanding of forgiveness, as you forgive, you no longer hold on to anything resembling my condition (7). This is not necessarily to say, however, that you must believe that as you forgive, insofar as you cease to resent, you also cease to blame, since it is not necessarily

the case that your account of blame involves this emotional component, as mine does. (Bennett, for example, believes that as you grant personal forgiveness, you cease to resent, though he believes that this does not entail that thereby you also cease to blame.)⁶⁵ Still, the standard line is that as you forgive you cease to resent and to blame, and indeed you cease to believe anything resembling my condition (6) as well.

The interplay between conditions (6) and (8) in my analysis deserves close attention. On first approximation, it may look as if for A to believe (6) and (8) simultaneously would be inconsistent. While I do not think it is inconsistent at all (as I will show immediately), it is to an extent understandable that it may appear odd, since this is, I submit, the root phenomenon giving rise to all the versions of the paradox of forgiveness: the forgiver believes that if a certain bad thing would befall the wrongdoer, this would be an acceptable state of affairs, and yet she refuses to bring about this state of affairs herself.

Imagine that a friend of yours has forgotten that today is your birthday, and that you wish she had not; you may find it acceptable if a common friend reminds your forgetful friend of your birthday, while at the same time refusing to remind your friend of it yourself. I am neither defending your position in this example nor tackling the problem of what the reasons justifying your position may turn out to be: I am merely arguing that this peculiar relation between (6) and (8) is not an uncommon phenomenon. We may want (and expect) people to treat us kindly, even if we do not think it appropriate for us to be telling others what kindness is; Socrates may have thought that the world would have been a better place had he not been convicted, but not if he would personally engage in certain actions; A thinks that it would be a better place if B pays his debt, and yet she refuses to remind B about the existence of the debt, and so on. Consider some of the peculiarities of forgiveness.

Imagine that A experiences the series of mental phenomena described in (1) through (8); imagine further that although A refrains from intervening in making sure that B's wrongdoing is offset, somehow the exactly offsetting event occurs (that is, an event that makes B suffer the exact amount which would, in A's estimation, offset her wrongdoing). Would A be sad or disappointed as a result of what happened to B? It seems to me that unlike the standard accounts of forgiveness, my account allows us to answer this question negatively. Standard accounts, after all, assume that A no longer blames, or resents, or believes that the world would be a better place if something offsetting B's wrong would happen to B, and thus, that whatever suffering may

⁶⁵ Bennett, 'Personal and Redemptive Forgiveness'.

afflict B, A would regret. But on my account the only inappropriate thing would be for A, *herself*, to inflict the suffering upon B—and not simply for B to suffer. Imagine A has come to conclude that B deserves to undergo an immensely painful experience, as the only way to offset her similarly extraordinary wrongdoing. A may nonetheless refuse to engage herself in actually making sure that B suffers so dramatically, though she may welcome, or perhaps be indifferent, if this suffering befalls B through other causes.

Many views on forgiveness presuppose that to forgive someone is to behave as if the wrongdoing would in a sense not have happened—the images of wiping the slate clean, or of turning the page, etc., all seem to capture this idea. But on my account this is not a plausible thesis. Imagine two people, Charles and Duncan; imagine that Charles has done you wrong, though you have come around to forgive him, whereas Duncan has, in your estimation, never done wrong to you or anyone else. Further imagine that each of them suffers a perfectly quantifiable amount of pain, and that this amount of pain was exactly what you thought would offset Charles's wrongdoing. I think that you would find Duncan's suffering more tragic, more upsetting, than you would Charles's—even though you have forgiven Charles. I am not suggesting that you will necessarily celebrate Charles's suffering; rather, I am suggesting that your reaction to it must *differ* from your reaction to Duncan's suffering, though it would *differ*, too, from your reaction to Charles's suffering had you not forgiven him. And this difference is connected with the requirement (8): *you* will not punish, but if someone else punishes you may not be upset.

Yet, it may seem that my account of forgiveness is still too broad (indeed too similar to what Kolnai calls condonation), in that the term forgiveness should be restricted to those cases in which refraining from punishing is commendable, when the refraining from punishing is caused by a moral reason, along Murphy's lines. However, I think that rather than constituting a problem, and independently of the problems facing Kolnai's distinction between forgiveness and condonation, the breadth of my account is one of its strengths.

After all, the more limited scope of standard accounts of forgiveness is based on the appeal to moral reasons, which, though plausible, is not entirely convincing. Consider the sort of case that Card discussed: refusing to punish a powerful gangster due to fear that his associates would hurt innocent people seems not to be truly to forgive the gangster.⁶⁶ But is this not a moral reason?

⁶⁶ See Card, 'On Mercy', p. 187.

Why is this not to forgive the gangster? If this is indeed a moral reason, then standard accounts may have difficulty showing that this is *not* to forgive the gangster. Much work remains to be done in further specifying the *type* of moral reason which is peculiar to forgiveness, given that not just *any* moral reason seems to suffice. I think that my account allows me to argue that this is not to forgive the gangster, in that if I refuse to punish the gangster in order to avoid harming innocents, I would want others to refrain from harming innocents too. In other words, it is not that the world will be a worse place if I punish the gangster: insofar as punishing him will harm innocent people, this is not a matter of a difference between myself and others—no one should punish him.

Moreover, it is simply not the case that any instance of not punishing a wrongdoer will on my account count as forgiveness. For not only do I require that this refusal to punish be *deliberate* (thus limiting considerably what the candidates to forgiveness could possibly be), but this deliberation is linked to the admittedly complex realization that while the world would be better if the wrongdoing were somehow offset, this offsetting should not come from oneself. Thus, refusing to punish a wrongdoer who deserves to be punished because one has a plane to catch, or because one is lazy, or because one loves (or fears) the wrongdoer, or because by punishing her we would harm innocent people, etc., are not, on my account, necessarily instances of forgiveness. In cases like these, either the belief contained in (8) or the deliberate refusal tied to this belief also contained in (8) (or both) are absent.

But, via this deliberation, I may be lapsing into Murphy's appeal to moral reasons after all. This is not a serious problem, insofar as I said at the outset, on *this* point, i.e., on the importance of specifying the type of reason which is relevant for forgiveness, I agree with Murphy. My account differs from his in that mine is indeed liberally broad as to the reasons which it would allow as leading a person who satisfies (1) through (7) to also endorse (8), whereas he focuses more than I do on repentance. (An additional difference I mentioned above is that the concern with reasons, which I share with Murphy, does not lead me to suggest that the definition and the justification of forgiveness are not sharply distinguishable.) That reasons to forgive are variegated, and that some are good and some are bad, distinguishes my account from the sort of forgiveness-requires-repentance accounts that I have discussed above, of which Murphy's is an example. After all, the forgiveness-requires-repentance thesis is usually accompanied by the repentance-justifies-forgiveness, or the repentance-precludes-punishment corollaries. And I have argued that these sorts of corollaries are problematic in that they are vicious instances of definitional stops.

In sum, my account of pure forgiveness does limit what counts as forgiveness: it is not the case that any and all instances in which we fail to punish someone whom we deem blameworthy are, willy-nilly, instances of forgiveness. But, admittedly, my account does allow a variety of reasons to enter someone's deliberation as she is about to forgive. I do this, however, precisely in order to avoid the sort of definitional stop visible in the forgiveness-requires-repentance views whereby all instances of forgiveness are, *eo ipso*, instances of justified, or otherwise appropriate, forgiveness. The comparatively large scope of my account of forgiveness allows us to judge that some instances of forgiveness are morally objectionable while others are admirable. This is, I think, an advantage of my account over the forgiveness-requires-repentance/repentance-justifies-forgiveness approach.

Can, on my account, one of the reasons for forgiving be that the forgiver believes that the wrongdoer has repented? Perhaps. The problem with presenting a straightforward affirmative answer to this question is the following. If repentance is endowed with the sorts of wholly expiatory powers with which it is typically endowed, it is hard to see how the alleged forgiver could possibly meet requirement (6). That is, if repentance is understood in this way, then the repentant person is no longer blameworthy, and thus, as a matter of sheer logic, no longer forgivable. If repentance is understood differently, then it may be a reason for forgiveness in my account—though it importantly remains an open question whether it would be a good reason.

In contrast, then, to the 'forgiving' of the initially-guilty-but-through-repentance-no-longer-guilty (or at least no-longer-as-guilty) with which Kolnai, Murphy, and others are concerned, mine is an account of the forgiving of the guilty truly as guilty. Admittedly, it sounds odd to say things to the tenor of 'I forgive you, but I still resent you and blame you for what you have done', but that something sounds odd is not a reason to reject it. There are, moreover, problems with the opposing, widespread view that to forgive means, or entails, that one *eo ipso* no longer resents or blames. I have not discussed how common cases in which we forgive while we continue to resent and/or blame are; my claim is only that these cases are possible—sometimes the forswearing of resentment sets in gradually, and it takes place *after* one has forgiven. In recognizing this possibility, my account of forgiveness avoids the pitfalls of other accounts. Why we may sometimes be justified in forgiving the guilty as guilty is an issue that I have not discussed either; and it surely involves discussions of moral positioning and distance, of aesthetic valuation, and of fittingness, which are beyond the scope of this article. But I hope to have succeeded in exposing the shortcomings of those 'solutions' to the paradox of forgiveness which are linked to repentance, and that thus render forgiveness *definitionally* good.

I hope to have succeeded too in showing that the paradox of forgiveness emanates from a purely mental phenomenon: that we think that it would be good if a certain wrongdoer would suffer (by way of offsetting her wrongdoing), and yet we think that it would be wrong for us to inflict this suffering ourselves. While surely an interesting and under-researched phenomenon, this is not at all an uncommon situation in which to find oneself; it is not 'mad', and it is not quite to 'plunge, but lucidly, into the night of the unintelligible'.⁶⁷

⁶⁷ With thanks to the Alexander von Humboldt Foundation for supporting my research, and to the participants of my seminar on forgiveness in the Spring of 2007. Thanks are due, too, to Christopher Bennett, Thom Brooks, Antony Duff, Pierre Grenon, Ingvar Johansson, Jeffrie Murphy, Gerhard Øverland, Anna Schur, Jonathan Simon, Barry Smith, Aaron Snyder, and the two anonymous referees for this journal, for comments on earlier versions of this article.