A LITTLE TREATISE OF FORGIVENESS AND HUMAN NATURE

(Preprint version plus abstract of a paper published 2009 in *The Monist* **92**: 537-555.)

Abstract

The 'paradox of forgiveness' says: to forgive is to forgive a person for a particular culpable wrongdoing, but if the wrongdoing is culpable there is no reason to forgive him/her for it. The author's view is that an anthropology that states that man has only self-regarding desires can never resolve this paradox and make sense of prototypical forgiving. Positively, he argues that the paradox disappears if human nature is ascribed (as, for instance, by Hume) basic other-regarding desires of benevolence. Such benevolent desires have no meaning outside of themselves; they are their own meaning and value. The analysis implies that one cannot justify forgiving by means of reasons, only explain it by means of benevolence. This conclusion can also be put as follows: only nonjustifiable benevolence can justify forgiveness. Jacques Derrida's, Aurel Kolnai's, Charles S. Griswold's, and Leo Zaibert's analyses of forgiveness are explicitly discussed and criticized.

As has been made abundantly clear by now, the phenomenon of prototypical forgiving must not be identified with related phenomena such as excusing, pardoning, condoning, accepting an apology, and reconciling.[1] Neither should it be conflated with even more closely related phenomena such as self-forgiving, third-party forgiving (forgiving on behalf of another), and non-communicated forgiving (including even non-communicable forgiving such as forgiving a dead person).[2] The philosophical problem of understanding prototypical forgiving is the problem of coming to know the presuppositions and structure of the act in which a clearly culpable and living wrongdoer is told by the wronged person that he[3] is forgiven for his deed.[4] One thing can be said at once: as a memory refers back to an earlier perception, a forgiving refers back to an earlier resentment.[5]

Also, it should be noted, it is quite possible to forgive a person for one of his deeds but not for another. To forgive a person must not always be the same as to forgive a person *as a person*, even though such whole-person forgiving is necessary when forgiving is part of a reconciliation.

The actual forgiving action is a speech act, an utterance of the form 'I forgive you for your deed(s)'. I think the best general analysis and classification of speech acts made so far is made by Searle (1979: chapter 1), but, surprisingly, forgiving does not fit directly into any of his five taxa: assertives, directives, commissives, expressives, and declarations. A forgiving utterance has one feature in common with an expressive utterance such as 'I am just happy!',

and one in common with a commissive utterance such as 'I promise not to disturb you any more today'. A person who says 'I forgive you for your deed' does hereby publicly express two psychological states, both a feeling that his resentment has decreased or gone away, and an intention not in the future to blame the wrongdoer for his deed. (That is, there are two kinds of 'illocutionary acts' in play, both an expressive and a commissive act.) The intended ('perlocutionary') effect is that the listeners should understand that these psychological states are quite compatible with the fact that the speaker nonetheless finds the deed blameable, and holds the wrongdoer responsible for it.[6]

1. Forgiveness and Philosophical Anthropology

The so-called 'paradox of forgiveness' has been given a couple of different formulations (see the next two sections), but I think the essence can best be stated thus:

- to forgive is to forgive a person for a particular culpable wrongdoing, but if the wrongdoing is culpable there is no reason to forgive him for it.

The thesis of this paper is that the paradox – even with respect to *unconditional* forgiving – disappears as soon as one accepts and thinks through a philosophical anthropology that not only counts with desires for pleasure, aversions to pain, and other self-regarding desires, but ascribes human nature direct other-regarding desires or passions, too. Even though the intensity of the different desires may differ between individuals, human nature has three main kinds of desires wired-in:

1. self-regarding desires
2. benevolent other-regarding desires
3. malevolent other-regarding desires.[7]

Put negatively, my view is that an anthropology that states that man has only self-regarding desires can never make sense of forgiving.[8] Benevolent desires (not just benevolent feelings) are necessary for the kind of over-ruling of culpable wrongdoings that constitute forgiving, and malevolence is necessary for the retributive attitude that forgiving publicly proclaims to be gone. (As the term 'malevolent' is used here, to be malevolent towards

another person means *directly* to desire that some of the latter's self-regarding desires will become frustrated in such a way that it hurts; may this other-regarding desire have arisen out of pure sadism, out of a wish to give a villain a fair retributive punishment, or out of some mixture; on the other hand, to try to win a competition and, thereby, frustrate the competitors' desires to win, is normally not a desire to make the latter feel pain and so not a case of malevolence.)[9] One famous place where the mentioned anthropology can be found is David Hume's *A Treatise of Human Nature*. He writes:

> Beside good and evil, or in other words, pleasure and pain, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies [other-regarding malevolence], and of happiness to our friends [other-regarding benevolence]; hunger, lust, and a few other bodily appetites [self-regarding desires]. These passions, properly speaking, produce good and evil [pleasure and pain], and proceed not from them, like the other affections (2000: 281; *Treatise* 2.3.9).

What Hume here says about friends should be understood in the light of the following much more famous quotation from *Enquiry concerning the Principles of Morals*:

> We cannot without the greatest absurdity dispute that there is some benevolence, however small, infused into our bosom; some spark of friendship for human kind; some particle of the dove kneaded into our frame, along with the elements of the wolf and serpent (1975: 271).

Hume lived before Darwin, and what he regards as "perfectly unaccountable" we can today explain by means of gene-based group selection and kin selection theories in evolutionary psychology.[10] With respect to herd species, natural selection favors species that harbor individuals that can sacrifice themselves for their group or for their relatives. The sub-title of Hume's *Treatise* is *Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, and even though "the Experimental Method" to Hume merely meant making informally systematic everyday observations, I think it made him stumble upon an important truth. If one keeps a phenomenological first-person view of desires distinct from their underlying physiological and/or evolutionary causes, then it is impossible to reduce *all*

apparent acts of benevolent and malevolent behavior to behavior that aim only at satisfying self-regarding desires. Now, Hume did of course not only live before Darwin, he lived before Nietzsche, Freud and the focusing on human self-deception that has followed in their wake, too. However, even though many first-person apprehensions of benevolence may be self-deceptions, I am sure that they are not completely ubiquitous. In this paper, therefore, I will rest content with arguing that *if* Hume's anthropology is true, then the paradoxical aura around forgiveness disappears completely. Hume himself, it should be noted, does never discuss forgiveness.

What then is a desire from a first-person perspective? First, it is an intentional phenomenon with, to speak with Searle (1983), a world-to-mind direction of fit; one wants the world changed in such a way that the desire in the mind becomes satisfied. If you are hungry and want to eat the sandwich in front of you, then you want the world changed in such a way that your felt hunger becomes satisfied. Every so desired object has at least a subjective value, and is in this desire-dependent sense a reason for the desiring person to act in order to satisfy the desire.

Second, desires contain, as do all intentional phenomena, an act-object or state-object distinction, and both the act/state and the object can be classified as being of a certain kind. Hunger, sexual desiring, and longing for confirmation are different kinds of (self-regarding) *intentional states*; rice, potatoes, meat, fish, and vegetables are different kinds of *intentional objects* (for the state of hunger). The intentional object of a benevolent other-regarding desire is the satisfaction of another being's desire; the object of malevolence is a painful frustration (dissatisfaction) of another being's desire.[11]

Third, desires can either be basic, or they can be merely means towards the satisfaction of basic desires. What is to count as the one or the other may require some analysis, and even change from one situation to another. Should, for instance, hunger be regarded as a basic desire or merely as a means for a desire to continue to live? Seen from the first-person perspective I am talking about, hunger is normally a basic desire, i.e., the desiring is experienced only as a desire to get rid of the hunger. In extraordinary circumstances, however, it becomes a means for and fused with the desire to live. Something similar applies to sexual desire. Normally, it is from a first-person perspective simply a desire to have sex; whatever the evolutionary natural selection story looks like. There is usually no other reason to have sex than the existence of the desire to have sex. But, of course, sometimes for some persons, the sexual desire becomes fused with a desire to have a child or (in our modern 'stay fit'

culture) a desire to stay in a better psychological mood and physiological state by means of sex. My point about benevolence is structurally similar. Sometimes we are benevolent only as a means to get benevolence back, but sometimes we are benevolent for no other reason than that we have a desire to be benevolent; whatever the natural selection story for the emergence of benevolent behavior may look like. Such a benevolent desire has no meaning outside of itself; it is, like the normal desire for sex, its own meaning and subjective value.

Fourth, two desires can be conflicting in the sense that they pull one, like Buridan's ass, in two opposed directions. For instance, a fasting person may desire both not to eat and to eat. Also, there can be what Searle (2001) calls 'desire-independent reasons' for an action, and these may just like a counter-desire cancel a desire. If, following Searle, one has made a promise, then there is a desire-independent reason to keep the promise. A Catholic priest has both such a reason not to have sex and (I take it for granted) a desire to have sex. This means: a desire to be benevolent need not necessarily result in a benevolent action. It is, though, always a sufficient condition for such actions when no counter-desires or countering desire-independent reasons are present. On the other hand, two non-conflicting desires can sometimes be satisfied by one and the same action. For instance, if a friend who likes to cook invites one to dinner, then one can by attending the dinner satisfy both one's self-regarding desire for food and one's benevolent other-regarding desire to be nice to one's friend. In many situations, self-regarding and other-regarding desires work in tandem.

Fifth, most desires come and go and have a temporally intermittent structure; and they seem to be able to be triggered in two different ways, externally and internally. Quite obviously, hunger and sexual desire can come into being because of a sudden external appearance of a certain kind of food or a certain kind of person, but both of them can also first be felt as an intense internal desire that *either* as yet has found no determinate intentional object at all *or* has a determinate object that is absent. In these kinds of cases (counter-desires and countering desire-independent reasons disregarded), one has either to start to search for some objects or persons that might be able to satisfy one's desire, or try to come near to the determinate objects that one already desires. This distinction between external and internal triggering applies to benevolence and malevolence, too. Benevolence is often triggered by the mere appearance of a beloved one or a close friend, but there are many people who first have an indeterminate desire to be benevolent towards someone, and then try, for instance, to adopt a child or become members of some community where there are people they can be friendly towards. Similarly, malevolence can be triggered by the appearance of a person who in some

way or other has been nasty towards oneself, but it is also a well known fact that people can harbor a general resentment that is intensely seeking for an outlet. Their indeterminate malevolent other-regarding desire is looking for some determinate intentional object in a way similar to the way indeterminate hunger and sexual desire can be searching for some determinate intentional object.

My sixth and last point about desires is the following. Independently of whether the intentional object of one of my desires is my own self (a self-regarding desire) or another person (an other-regarding desire), the desire is always a desire of mine. That is, necessarily when one of my benevolent other-regarding desires is satisfied there are two satisfactions: one in me (be it very pleasurable or just a tiny feeling of satisfaction) and one in the other person. And when one of my malevolent other-regarding desires is satisfied, there is satisfaction in me but painful dissatisfaction in the other. The existence of this inevitable 'trivial egoism' does not turn other-regarding desires into self-regarding desires. A desire of a person is not just a property of this person such as height and weight, but an intentional property that has directedness towards something. Self-regarding desires are directed from oneself towards oneself, other-regarding from oneself towards someone else.

This being stated about the nature of desires in general, and of benevolence in particular, I can present my solution to the paradox of forgiveness. If the notion of 'benevolent desire' that I have sketched is found coherent and regarded as having possible referents, then the following becomes at once a reasonable claim:

- to forgive is to forgive a person for a particular culpable wrongdoing, and this is possible since a desire to be benevolent can overrule culpability; therefore, benevolence can give rise to forgiveness even where there is no repentance on part of the wrongdoer.

On the account that I have given, one cannot ask someone for an objective reason why he is benevolent. The desire to be benevolent appears as a basic desire in the situation at hand. Here to ask 'give me a reason why you are so benevolent that you simply forgive him' would be like asking persons questions such as 'give me a reason why you are hungry' and 'give me a reason why you would like to have sex'. In relation to basic desires there are no objective reasons, only purely subjective values and causal explanations. As one cannot justify, only explain, why one is hungry or desires sex, one cannot justify, only explain by means of

benevolence, why one is unconditionally forgiving. If I am right that objectively unjustifiable benevolence is the crucial thing in unconditional forgiving, then this fact also places such forgiving outside of contexts of justification; but not outside of human nature.

People who like each other can out of pure benevolence spontaneously give unmerited presents and gifts to each other. Similarly, pure or unconditional forgiving is a spontaneous unmerited present from a wronged person to the wrongdoer. As one author on forgiveness has it:

> Pure forgiveness is not an instrumental good, a prudent management technique or a damage limitation exercise; it is an intrinsic good, an end in itself, a pure gift offered with no motive in return (Holloway 2000: 78).

My analysis fits well with the fact that 'forgiving' linguistically presents itself as a special kind of giving, 'for-giving', but does it fit this common view: "forgiveness is the process of ceasing to feel resentment, indignation or anger against another person for a perceived offense, difference or mistake" (Wikipedia, June, 2008)? No, it does not fit immediately, but it can easily make sense of the Wikipedia view. Where there is benevolence towards a certain person, there can be no strong resentment, indignation, anger, or some other kind of strongly malevolent attitude towards this person.[12] Nonetheless, benevolence is not identical with the non-existence of strong resentment, indignation or anger; in-between benevolence and malevolence there is indifference. But out of indifference comes no forgiving; indifference cannot cancel culpability. This notwithstanding, there are things to be said about the interplay that can arise between benevolence and malevolence, but this topic belongs to Section 3, where *conditional forgiveness* (forgiveness that requires a change in the wrongdoer) is discussed; first, in order to make the core of my analysis even more clear, some more words about unconditional forgiveness.

2. Is Unconditional Forgiveness Unintelligible?

The most famous and most radical formulation of the paradox of forgiveness is a sentence of Jacques Derrida condensed into the dictum 'to forgive is to forgive the unforgivable'. He writes:

> 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 (2001: 32-33).

This does not mean that Derrida thinks that there can be no forgiveness. To the contrary, he believes in the existence of unconditional forgiveness. The quotation above should be placed within Derrida's overarching philosophical project (with which I violently disagree), which is to show that there really are no such things in the world as Reason and The Given; be this Given an epistemologically, semiologically, or ontologically positive entity.[13] Therefore, Derrida thinks he can be in favor of forgiveness despite the fact that he finds the notion of 'forgiveness' self-contradictory. He may even in part like it *because* he thinks the notion is unintelligible. What from traditional Western metaphysics looks mad and unintelligible can, according to Derrida, in some metaphorical sense of truth be the 'truth'. He says:

- Forgiveness is thus mad. It must plunge, but lucidly, into the night of the unintelligible (2001: 49).

According to the views I have put forward in the first section, Derrrida's message has to be exchanged for the following:

- Forgiveness is thus a contingent natural expression of human benevolence. This fact must move, lucidly, into the daylight of the intelligible.

So far, I have discussed only unconditional (absolute, pure) forgiving, and claimed that it contains no paradox. Several philosophers of forgiveness have taken another route. They agree with Derrida that unconditional forgiving is unintelligible, but, of course, disagree with his acceptance of this presumed unintelligible phenomenon. They reject unconditional forgiveness, and rest content with trying to make philosophical sense of forgiveness that is conditional upon repentance or some other change in the in the mind of the wrongdoer. Derrida, on the other hand, claims that the possibility of unconditional forgiveness is also a necessary presupposition for the possibility of conditional forgiveness (2001: 34). Taken in abstraction from Derrida, I agree with the last claim. More precisely, I will show that benevolence is as crucial to conditional forgiveness as it is to unconditional forgiveness.

3. Conditions on Forgiveness

From a purely logical point of view, conditions on forgiving can be divided into the four different kinds below. One can claim that forgiving requires:

(i) a change in the wronged person
(ii) a change in the wrongdoer (this is *conditional forgiving* as opposed to unconditional)
(iii) changes in both the wronged one and the wrongdoer
(iv) something that is external to both the wronged and the wrongdoer.

Let me comment on these requirements one by one.

(i)

By definition, all kinds of forgiving requires a change in the wronged person. At first he resents, but by the time of forgiving this resentment should be radically reduced; in the best of cases completely gone. What makes this definitional condition worth speaking of without discussing any details in the change is the fact that there can be insincere forgiving.[14] A person who says 'I forgive you for your deed' does hereby *publicly express* both a feeling that his resentment is not especially strong and an intention not to blame the wrongdoer anymore. But such a speech act can be performed even by a speaker who consciously hides strong resentment in his heart, and has no intention whatsoever to stop blaming the wrongdoer that he is pretending to forgive. Insincere forgiving is just as possible as faked emotions and false promises.

(ii)

I will take Aurel Kolnai (1973) as a good representative of the view that a necessary requirement for forgiving is that the wrongdoer has undergone some change of mind since he made his deed. Be it that the wrongdoer afterwards repents, asks for forgiveness, or does something else of a similar kind; Kolnai focuses on repentance. He regards unconditional forgiving as an illusion that should be rejected, and summarizes the presumed paradox of such forgiveness as follows: "Briefly, [such] forgiveness is either unjustified or pointless" (1973: 99). It is unjustified in case the deed "is still flourishing, the offence still subsisting: then by 'forgiving' you accept it and thus confirm it and make it worse" (1973: 98); and it is pointless in case "the wrongdoer has suitably annulled and eliminated his offence" (1973: 98-99).

Kolnai's way out of a complete denunciation of forgiving is to make forgiving conditional on repentance or '*metánoia* ("Change of Heart")' (1973: 99).

According to Kolnai's analysis, forgiveness without a preceding repentance by the wrongdoer is not a forgiving but a *condonation* (i.e., in spite of morally disapproving an action, one deliberately refrains from all retributive attitudes and responses). Kolnai must mean that if the wrongdoer has repented, then the offence is neither "flourishing" (first horn of the dilemma) nor "annulled" (second horn). But even so, I will argue, repentance cannot be a sufficient reason for forgiving; even conditional forgiving requires something more, namely benevolence.

Since my argument is quite general, let us in the abstract conceive of a wrongdoer that in a culpable way has offended a person, 'the wronged', who does not condone but reacts with indignation and a retributive attitude. In other words, the wronged person starts to harbor malevolent desires towards the wrongdoer. According to my assumed philosophical anthropology, it might happen that one day these malevolent desires simply have melted away in the face of the general human sympathy and benevolence that the wronged one felt for the wrongdoer before the deed; and that the wronged party, out of this re-gained benevolence, unconditionally forgives the wrongdoer. But let us now disregard this possibility, i.e., the wronged person clings to his retributive attitude. However, when he is told that the wrongdoer is sincerely repenting, he starts to reflect on his attitude, and after some time he forgives the wrongdoer. To other people the wronged person says: 'okay, now when he repents I can forgive him'. The question is whether this repentance by itself can make such a difference to the forgiver. Let us see.

Of course, a repentant wrongdoer is from a moral point of view a different kind of person than a non-repentant one, and this difference ought to have some consequences. Think of the difference we make between murder and manslaughter. The difference of mind, which we ascribe to the two kinds of killers, makes us require a less hard punishment for manslaughter than murder. Analogously, the difference between a repentant and a non-repentant wrongdoer has to be taken into account when estimating what degree of blame and punishment we should burden each of them with. In neither the murder/manslaughter case nor the non-repentant/repentant case, however, should the blame or retributive attitude be allowed to be *completely* withdrawn. But this is exactly what is allowed to happen in a forgiving. When saying 'I forgive you', the wronged person publicly promises not at all in the future to blame the wrongdoer for his deed; and if the forgiver still harbors some resentment in his heart he

should be happy if it vanishes. That is, repentance from the wrongdoer leads naturally to a less hard condemnation from the wronged person, but it cannot by itself lead to forgiving. In order to annul the retributive attitude completely, benevolence is needed. Conditional forgiveness is at bottom conditional benevolence. Benevolence can have many different causes and counter-forces, and it is by no means psychologically odd that repentance by the wrongdoer can make room for benevolence in the wronged person. Obviously, this psychological observation was one reason behind the truly great social innovation that saw the light at the end of the last century, the South African 'Truth and Reconciliation Commission'.

(iii)

In a recent book, Charles L. Griswold (2007) has, like Kolnai, claimed that unconditional forgiveness or forgiveness as a pure gift is an illusion. In the prologue of the book, he formulates the paradox of forgiveness thus: "It may seem at the outset that [...] forgiveness [...] aspire[s] to something impossible: knowingly to undo what has been done (2007: xv)." And at the end of the book he writes:

> I hope to have shown that forgiveness is fundamentally an interpersonal process whose success requires actions from both parties. Anything an individual can accomplish here on his or her own regarding forgiveness is less than fully adequate. Consequently, forgiveness should not be understood as a "gift" that may be bestowed at the discretion of the injured party (2007: 212).

Griswold argues that forgiveness has six conditions in relation to the forgiver, six in relation to the wrongdoer, and one that is external to both the forgiver and the wrongdoer, namely that the deed must in principle be forgivable (2007: 59). I would call the last condition 'logical', and see no reason to comment on it here. According to Griswold, at the moment of forgiving, the forgiver: 1) has to forswear revenge, 2) has moderated his resentment, 3) has to commit himself to let the rest of the resentment go, 4) has "reframed" his mind in relation to the wrongdoer, 5) feels no moral superiority in relation to the wrongdoer, and 6) talks directly to him (2007: 54, 58). The man to be forgiven, on the other hand, must have: 1) accepted responsibility for the deed, 2) repudiated the deed, 3) expressed regret, 4) expressed a will to become a person that does not inflict injury, 5) shown a first-person understanding of the injury done, and 6) offered a narrative account of how he came to do wrong (2007:49-51).

When all the thirteen necessary conditions are met, Griswold thinks there is a *sufficient reason* for forgiving (2007: 58-59). Let it be clear that he also (rightly) thinks that the wrongdoer never has a *right* to forgiveness, and that the forgiver never can be *compelled* to forgive. Forgiveness is always voluntary; the sufficient reason mentioned is a morally sufficient reason in the sense that it makes the corresponding (conditional) forgiving morally praiseworthy (2007: 67-69). I claim, to the contrary, but in conformance with my remark in relation to Kolnai, that the thirteen conditions can only be a morally sufficient reason to *reduce* blame, punishment, and a retributive attitude. Since the culpability is still there, something more – or other – than a moral reason is needed in order completely to overrule the culpability. And as far as I can see, the only thing available is benevolence.

Griswold regards his analysis of conditional forgiveness as also showing that such forgiving must be the result of a preceding "interpersonal process," and that, therefore, forgiveness is *always* "dyadic forgiveness" (2007: 47).[15] In my opinion, by not seeing the phenomenon of basic benevolence clearly, Griswold takes away from forgiving the asymmetry with which both unconditional and conditional forgiving have traditionally been connected. He is not explicating conditional forgiving, he is completely redefining it. This could in principle have been acceptable, but, in Griswold's case, it is not. The end result of his analysis contains a very odd feature. On the one hand, he claims that the paradigm case of forgiveness is dyadic, and on the other hand that forgiveness is the expression of a virtue, the virtue of forgivingness: "Forgiveness is what [the virtue of] forgivingness expresses" (2007: 17). These claims contradict each other. No virtue is dyadic; a virtue is a monadic property of a person. Of course, there is no inconsistency in saying that: a) '*willingness* to seek conditional forgiving' and '*willingness* to repent' are virtues, b) actual conditional forgiving is necessarily dyadic, and c) such a forgiving can only spring from the mentioned *two virtues in interaction*.[16] But this is not what Griswold says.

That Griswold overstates how much 'dyadicity' there is in forgiving is clear also from the following quotation: "the offender depends on the victim in order to be forgiven, and the victim depends on the offender in order to forgive (2007: 49)." Here, the logical truth that there can be no forgiving without both an offender and a victim, is mixed up with the intimated (but false) view that both the offender and the victim depends equally on each other in order to get something they want in life. In fact none of them, and especially not the victim, may think he has anything to gain by an act of forgiving.

Even though Griswold at some length discusses the phenomenon of a general human sympathy, which is intimately connected with benevolence, his analysis of forgiveness does in the end only take self-regarding desires and objective moral reasons into account – not benevolent other-regarding desires. This fact makes him unable to accept the existence of unconditional forgiving, and it makes him cancel the asymmetry between forgiver and forgivee that is there even in conditional forgiving.

(iv)

In the paper "The Paradox of Forgiveness," Leo Zaibert (2009) tries to rescue unconditional forgiveness. He discusses Derrida's and Kolnai's versions of the paradox, finds Derrida "too obscure to be by itself helpful," and argues (rightly to my mind) that Kolnai's appeal to repentance "faces more problems than it solves" (2009: abstract). Zaibert's way out of the reduction of forgiveness to something unintelligible or to conditional forgiveness is to claim: *if* (a) one focuses on the definition of forgiveness without bothering about how to justify particular acts of forgiving, and (b) looks at what forgiving amounts to as a mental phenomenon in the forgiver's head, *then* "forgiveness is not quite as paradoxical after all (2009: abstract)." As can be seen from this quotation, he is a bit hesitant to claim that he has made complete sense of traditional unconditional forgiving; and this hesitancy has a good reason, as I will show.

Zaibert stresses the fact that retributive punishment and forgiving are different and incompatible reactions to one and the same thing,[17] call it 'a culpable wrongdoing' or 'a blameworthy action'; Zaibert makes a distinction between the latter that can be disregarded here. He claims that when a person A blames another one B for an action X, then A has in his mind six different kinds of beliefs and one kind of feeling.[18] These seven mind states are consequently regarded as presuppositions common to a retributive and a forgiving mind; two versions of a belief (eighth mind state) create the difference between them (2009: 17):

> Punishment (8'): A does something to B, which A believes it is painful for B to endure, as a response to B's having Xed.
>
> Forgiving (8): A believes that the world would in fact be a worse place if A did something to B in response to B's wrongdoing, and thus A deliberately refuses to try to offset B's wrongdoing.

In summary: “to forgive is to deliberately refuse to punish (2009: 3).” This applies independently of whether there is repentance or not. However, as seen from (8), the forgiver is assumed to have a believed justificatory *reason* for his refusal. About this Zaibert is quite explicit: “Forgiveness, like punishment, is the sort of phenomenon which stands in need of justification (2009: 5).” The justificatory reason found is neither a belief that there is a change in the heart of the wrongdoer, nor a belief that the forgiver’s own resentment has disappeared, but a belief that the world would in fact be a worse place if the potential forgiver himself would attempt to enforce some retribution. Relating himself to Jeffrie G. Murphy (Murphy and Hampton 1988: 24), Zaibert says: “forgiveness is something we do ‘for a moral reason’ (2009: 5).”

I have no qualms in letting the phenomenon that Zaibert describes be called a kind of forgiving. Surely, it should not be called ‘*conditional* forgiving’, since (in contradistinction to Murphy) he puts no moral conditions on the wrongdoer. Therefore, in a sense, he has managed to be true to his intuition that there is more to forgiving than what can be found in conditional forgiving. However, the phenomenon he delineates is not *unconditional* forgiving *in the traditional sense* either. In such forgiving, the wrongdoer is forgiven *for his sake*, but in Zaibert’s kind of unconditional forgiving the wrongdoer is forgiven *for the sake of morality*. There are many kinds of actions that differ from both conditional and traditional unconditional forgiving, but go under the name of forgiving, e.g., excusing, self-forgiving, and third-party forgiving, and I think that the kind of forgiving that Zaibert highlights had better be called ‘consequentialist moral forgiving’ (which does not imply that Zaibert is in general a consequentialist; he is not).

What leads Zaibert away from traditional unconditional forgiveness, and necessarily so, is his stated requirement that unconditional forgiveness requires a justificatory reason. But this is wrong; unconditional forgiveness is exactly the kind of phenomenon that, unlike punishment, cannot be given such a reason; desire-*de*pendent reasons are not justificatory only explanatory. It is enough that the forgiver tells interrogators: ‘I forgive him because I desire to forgive him’. But a punisher cannot similarly say: ‘I punish him because I desire to punish him’.

My analysis implies that one can make a distinction between *deserved* and *justified* forgiving. According to my analysis, *no* forgiving can be completely justified; non-justifiable benevolence has always some role to play. But one may well call conditional forgiving deserved forgiving and unconditional forgiving undeserved forgiving. Note, though, that it is

possible unconditionally to forgive a repentant wrongdoer; this is the case when the forgiver does not care about the repentance in question.

Zaibert has not found traditional unconditional forgiveness, even though he has found a kind of forgiveness that differs from conditional forgiveness. His analysis does not make forgiveness depend on changes in the wrongdoer, but it makes it depend on what the future world may look like. Since prudent persons take the future into account, some words about forgiving and prudence are necessary, too.

4. Words of Conclusion: Stupid Forgiving is Still Forgiving

At the beginning of this paper, I introduced David Hume's philosophical anthropology; now, being close to the end, I will mention a thinker Hume held in high esteem, Joseph Butler. Griswold, like many other contemporary philosophers of forgiveness, thinks very highly of Butler, too, and regards him as being a very seminal thinker in the project of obtaining a deeper understanding of forgiveness. However, Griswold does so without stressing Butler's view that man contains desires to be benevolent. Let me now quote a paragraph from C. D. Broad's *Five Types of Ethical Theory* (first published 1930) that mentions Butler:

> It was fashionable in Butler's time [the early 18th century] to deny the possibility of disinterested [not self-regarding] action. This doctrine, which was a speculative principle with Hobbes, has always had a certain vogue. It is not without a certain superficial plausibility, and it has naturally been popular both with vicious persons who wanted a philosophical excuse for their own selfishness and with decent people who felt slightly ashamed of their own virtues and wished to be taken for men of the world. One of Butler's great merits is to have pointed out clearly and conclusively the ambiguities of language which make it plausible. As a psychological theory it was killed by Butler; but it still flourishes, I believe, among bookmakers and smart young businessman whose claim to know the world is based on an intimate acquaintance with the shadier side of it. In Butler's day the theory moved in higher social and intellectual circles, and it had to be treated more seriously than any philosopher would trouble to treat it now. This change is very largely the result of Butler's work; he killed the theory so thoroughly that he sometimes seems to the modern reader to be flogging dead horses. Still, all good fallacies

> go to America when they die, and rise again as the latest discoveries of the local professors. So it will always be useful to have Butler's refutation at hand (1979: 54-55).

(Let me leave Broad's unfair comment on America aside; from my perspective, Europe does not fare any better.) What Broad regards as "fashionable in Butler's time" seems to me to be just as fashionable in ours. Therefore, I think that the arguments for benevolence put forward by thinkers such as Butler, Hume, and Broad are necessary to rehearse even today. In this paper, however, as I said at the beginning, I have only wanted to show that *if* there is benevolence then there is no paradox of forgiveness. On the other hand, if there is no benevolence, then, I would say, there is a paradox. In a world consisting only of complete egoists, it is not only the case that there would be no true acts of forgiving, philosophers in such a world would be able to show that the very notion of 'forgiveness' is unintelligible.

Intelligible actions need not be prudent. My defense of the possibility of unconditional forgiving does by no means rule out the fact that a particular such forgiving may be stupid, i.e., a *prudent* benevolent man should in the situation at hand not have forgiven the wrongdoer. Why? Because a forgiver cannot be sure that a non-repentant wrongdoer understands that his deed despite the forgiving is morally wrong and culpable. Perhaps such a wrongdoer thinks that the forgiving takes the culpability away. Therefore, accepting that stupid unconditional forgiving is still forgiving, we should ask: *can it not truly be said that at least prudent persons should always reject unconditional forgiving?* No, it cannot; this position is open to an obvious objection. How interpersonal relations will develop is extremely hard to predict; even for the wisest of wise persons. As noted by several thinkers on forgiving (e.g., Holloway 2002: 82), a seemingly stupid unconditional forgiving may *start* exactly the kind of repentance process that Kolnai and Griswold thinks have to be unfolded *before* any real act of forgiving can take place. Their mistaken philosophical reduction of unconditional forgiving to conditional forgiving is not without its non-philosophical consequences.

Ingvar Johansson

Professor emeritus in theoretical philosophy
Umeå University, Sweden

ACKNOWLEDGMENT

I would like to thank Per Bauhn, Kevin Mulligan, and Leo Zaibert for comments on an earlier version of this paper.

NOTES

[1] Even a philosopher such as Charles L. Griswold, who puts very strong emphasis on reconciliation, claims that forgiveness is neither a sufficient nor a necessary condition of reconciliation (2007: 93).

[2] 'Non-communicated forgiving' could equally well have been called 'unilateral forgiving', had not Griswold (2007) under this label lumped 'forgiving the dead' together with 'forgiving without repentance'.

[3] Here, and in what follows, I will use the personal pronoun 'he' to cover individuals of all sexes. I think that as long as there is no good gender neutral personal pronoun around, biologically male authors could be allowed to write 'he' and biologically female authors 'she'.

[4] Perhaps this choice of what is prototypical displays my secular outlook. Some Christian theologians make the inner act of the forgiver prototypical, be it communicated or not, since they put more stress on the 'change of heart' of the forgiver than the social act that involves both the forgiver and the forgive; see, e.g., D. von Hildebrand (1980: 315, 335).

[5] I am in the whole paper using the notion of 'resentment' as it is delineated by Griswold (2007).

[6] Compare Griswold: "But the aim of forgiveness is something quite different: to understand, to relinquish revenge and resentment, all the while holding the offender responsible (2007: 47)."

[7] Even self-regarding desires can be divided into benevolent and malevolent desires. Truly self-destructive persons have self-regarding malevolent desires, but for simplicity's sake I have left this complexity out of account.

[8] I will not discuss the Christian view that there is for Christian people a moral obligation to forgive; only state that normally even this view posits benevolence, if only implicitly. For instance, von Hildebrand distinguishes between two kinds of human forgiving ("verzeihen"): Christian and natural forgiving. The Christian form necessarily involves charity/brotherly-love ("Nächstenliebe"), and the non-Christian form springs from generosity/magnanimity ("Grossmut") (1980: 331, 341 n28). Now, charity and brotherly love are very strong forms of benevolence, and I think some weak form of benevolence is a necessary presupposition for the existence of both generosity and magnanimity. It might be added that von Hildebrand distinguishes human forgiving from God's kind of forgiving ("vergeben"), which is the only one that von Hildebrand thinks can cancel objective moral guilt (1980: 254 n1, 324).

[9] Punishment that (without any self-deceit) is based *only* on utilitarian considerations about the future happiness of mankind does by definition contain no malevolent desires. It is as logically impossible to punish retributively by means of utilitarian reasons as it is to love for such reasons.

[10] See, e.g., Wright (1996: chapter 7).

[11] The term 'being' is meant to include animals, too. We are not always emotionally indifferent in relation to animals. Benevolence towards pet animals and malevolence towards animals that appear threatening are quite common phenomena.

[12] I am writing 'no *strong* resentment' instead of simply 'no resentment' since I am of the opinion that benevolence can co-exist with some malevolence just as love can co-exist with some hatred. Using the concept of 'organic unity', my view is that benevolence towards a person can be an organic unity that allows spots of malevolence as parts. Of course, the fact that a person A has a malevolent *attitude* towards one specific person B does not imply that A is a malevolent *person.*

[13] See, e.g., Carlshamre (1986: chapter 1).

[14] For discussions of the "details" hinted at, see (Murphy and Hampton 1988) and (Griswold 2007).

[15] Griswold seems to me to think that he differs radically from Kolnai, but, in my opinion, it is only the requirements four and five on the forgiver that can be called 'non-Kolnaian'.

[16] Note that I am talking only about 'willingness to seek *conditional* forgiving' as a virtue. For reasons of self-respect one should not always be willing to forgive. This fact has been made quite clear by Murphy (1988, 2003).

[17] This incompatibility is an incompatibility only for a single person at a particular moment. Of course, a person can first punish a wrongdoer and later forgive him (but not vice versa). Also, when the punishment is in the hands of society and not the forgiver, the forgiver may well forgive the wrongdoer but nonetheless accept a punishment from the society.

[18] According to Zaibert: "A [adequately] 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 (2009: 17)."

REFERENCES

Broad, C. D. 1979 *Five Types of Ethical Theory*, London: Routledge & Kegan Paul

Carlshamre, Staffan 1986 *Language and Time. An Attempt to Arrest the Thought of Jacques Derrida*, Gothenburg: Acta Universitatis Gothoburgensis.

Derrida, Jacques 2001 *On Cosmopolitanism and Forgiveness*, London: Routledge.

Griswold, Charles L. 2007 *Forgiveness. A Philosophical Exploration*, Cambridge: Cambridge University Press.

Holloway, Richard 2002 *On Forgiveness*, Edinburgh: Canongate Books.

Hume, David 2000 *A Treatise of Human Nature* (ed. the Nortons), Oxford: Oxford University Press.

Hume, David *Enquiries concerning Human Understanding and concerning the Principles of Morals* (ed. Selby-Bigge), Oxford: Clarendon Press.

Kolnai, Aurel 1973 "Forgiveness," *Proceedings of the Aristotelian Society*, 74, 91-106.

Murphy, Jeffrie G., and Jean Hampton 1988 *Forgiveness and Mercy*, Cambridge: Cambridge University Press.

Murphy, Jeffrie G. 2003 *Getting Even. Forgiveness and Its Limits*, Oxford: Oxford University Press.

Searle, John 1979 *Expression and Meaning*, Cambridge: Cambridge University Press.

Searle, John 1983 *Intentionality*, Cambridge: Cambridge University Press.

Searle, John 2001 *Rationality in Action*, Cambridge MA: MIT Press.

von Hildebrand, Dietrich 1980 *Gesammelte Werke IX. Moralia*, Regensburg: Verlag Josef Habbel.

Wright, Robert 1996 *The Moral Animal*, London: Abacus.

Zaibert, Leo 2009 "The Paradox of Forgiveness," *Journal of Moral Philosophy*, 6, ??-??.