

³⁴ Interestingly, in offering a specifically Christian model of forgiveness Marilyn Adams (1991) argues that forgiveness essentially depends on a different kind of perspective-taking: namely, that taking the perspective, to the extent possible, of God. Seeing the situation from God's perspective allows one to see the offender in a different light. On the debt-release model, which is compatible with a variety of reasons for forgiving, this could be one among the possible reasons for forgiveness.

³⁵ See Garrard and McNaughton 2010, p. 23 and p. 91.

³⁶ Outrage might be interestingly different, and though I set it aside here, note that it deserves its own discussion.

³⁷ Because Garrard and McNaughton see forgiveness primarily in terms of overcoming ill will, they conclude that there is always good reason to forgive if one does so for the right reasons. But once we shift to a debt release model, and see the overcoming of other aspects of resentment as the more immediate consequences, we do not have the same reasons for a universal endorsement of forgiveness. Similarly, but for somewhat different reasons from Garrard and McNaughton, Holmgren (2012) argues that unconditional forgiveness is always called for (consistent with the forgiver's having addressed the moral wrong). But here it may be that the disagreement is not straightforward, since on her view, forgiveness is always "appropriate and desirable from a moral point of view" (p. 10), and this is not the same as something one is morally obligated to do.

³⁸ A version of this paper was presented at the Bowdoin Workshop on Free Will in October 2011, and I am very grateful for the excellent commentaries by Laura Ekstrom and Daniel Speak, and thank all of the participants for their very valuable feedback, including Mark Balauger, Randolph Clark, Peter Graham, Eddy Nahmias, Scott Schon, Neal Tognazzini, and Manuel Vargas. I also benefited greatly from input from my colleagues in the UCSD Chancellor's Interdisciplinary Collaboratory on the Nature of Forgiveness and Its Role in Cardiovascular Health (2010-2012)—Nicholas Christenfeld, Ryan Darby, Christine Harris, Britta Larsen, and Per Milam. Finally, I am indebted to Eve Garrard, Margaret Holmgren, David McNaughton, Derk Pereboom and Sam Rickless for their very helpful and insightful comments on previous drafts.

CHAPTER NINE

MORAL RESPONSIBILITY, FORGIVENESS AND CONVERSATION

BRANDON WARMKE
AND MICHAEL MCKENNA

Introduction

All but the hardest of us have forgiven someone for something. Many of us, probably most, have not forgiven someone we should have, and have forgiven someone we shouldn't have. Based on this very simple observation about our everyday lives, we note two questions one might press, one analytical, another normative: First, what is forgiveness? Second, what norms govern the propriety of it?

In this essay we shall attend primarily to the first of these questions, but our hope is that what we say about it will also be suggestive of how one might take up the second. As for the first, we shall answer it within the context of a theory of moral responsibility. Naturally, for this reason, we shall restrict our attention to what might awkwardly be called *moral forgiveness*. Just as there might well be other kinds of responsibility other than moral, such as legal, practical, aesthetic, or even etiquettical, so too on our account there might be other kinds of forgiveness.¹ Those who disagree on this score—that is, those who think all forgiveness is moral—can handily set aside their dispute with us on this point, since, if all forgiveness is moral, then we can be taken to be offering an unrestricted rather than a restricted treatment, in which case, we'll happily drop the modifier 'moral' as it would be an obvious redundancy. We'll begin by sketching a conversational theory of moral responsibility. Then, drawing upon this theory, we shall develop an account of forgiveness, one that, admittedly, will not be exhaustive (nothing short of a book-length treatment could be). We'll close by discussing benefits of this approach and suggesting how one might begin to settle the normative question.

1. A Conversational Theory of Moral Responsibility

For the purposes of this paper, we shall assume a broadly Strawsonian theory of moral responsibility², at least as regards three central tenets of Strawson's view. Although controversial, these tenets are shared by many theorists about moral responsibility. One is that being morally responsible—in both the sense of being a morally responsible agent, and being morally responsible for something (for an act, an omission, or the consequences of an act or omission)—is conceptually connected to holding morally responsible and so to the liability of praise and blame. One *cannot* provide an adequate understanding of the former without attending to the nature of the latter. A second tenet is that the reactive attitudes, and especially the morally reactive attitudes, such as resentment, moral indignation, and guilt, are crucially implicated in the nature of holding morally responsible so that, by attending carefully to them, and in light of the first Strawsonian tenet, we can gain much understanding about the nature of being morally responsible.³ A third tenet is that what matters most fundamentally as regards an agent's praiseworthiness and blameworthiness is the quality of will with which she acts (or fails to act)—that is, her regard or lack of it for others and for the morally salient considerations bearing on the context of her action.⁴ Of course, the clearest indicator of the moral quality of an agent's will is whether, for example, she freely does what she knows to be morally wrong, or instead freely fails to do what she knows to be morally right.⁵ But these considerations will not exhaust what is involved in the quality of her will. Though knowingly and freely doing morally wrong, an agent might have acted with great regret, or with trepidation, or perhaps with ambivalence; or instead, she might have done so with glee over the very thought of her maliciousness. These variables will also contribute to the quality of her will, and so, for instance, to the nature and degree of her blameworthiness. Naturally, with variations in nature and degree of blameworthiness come variations in the nature and degree of (appropriate) blame. Similar remarks apply to praiseworthiness, though for transparent reasons we will for the most part restrict our attention to blameworthy conduct and the response of (overt) blame.

On a plausible version of a Strawsonian thesis, to hold another person morally responsible for a blameworthy act is, at the very least, to be susceptible to a negative morally reactive emotion, such as resentment or indignation, or to believe that it would be appropriate to experience and respond to the person via such a reactive attitude (e.g., see Wallace, 1994). But this common characterization of Strawson's view, though correct,

under-describes the phenomena of holding morally responsible insofar as it leaves out an important element, namely our interpersonal practices whereby the relevant reactive emotions are revealed. As various commentators have observed (e.g., Watson, 1987), on Strawson's view, holding another morally responsible for something (a morally wrong act, for instance) means something in practice. Normal modes or patterns of conduct are withdrawn or altered and other patterns take their place. Sometimes courtesies are withheld, and relationships can be severed or deeply impaired. In minor cases, where the offense is no more than a peccadillo, perhaps the altered behavior comes to no more than a disapproving grimace or a scolding nod. When the infraction is substantial, the patterns can take on various forms, depending on the bonds that held between wrongdoer and blamer prior to the offending act. A marriage might be ended or at least suspended. A tongue-lashing or an outright denunciation of the guilty party and her conduct might be fitting. Banishment or at least temporary withdrawal from a circle of coworkers, friends, or intimates is another possibility. In the case of relations between complete strangers, the relationship prior to the offense is merely one of the basic respect that each of us owes others simply as persons. Modifications to the practices whereby that respect is shown can then vary accordingly.

The salient point here is that the morally reactive attitudes provide a motivational base for the pertinent alterations to our practices—to the very alterations that are constitutive of blaming practices. These alterations are taken to be natural manifestations or expressions of the reactive emotions, and they are subject to subtle norms of fittingness or propriety (one can overreact or under-react to another's wrongdoing, for instance).

In his highly suggestive paper "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme" (1987), Gary Watson proposed a way of further extending and elucidating a Strawsonian theory of moral responsibility. Watson advised that we understand our morally reactive attitudes and their attendant practices as expressive of our moral demands and expectations towards those whom we hold responsible. They are, as Watson put it, "incipiently forms of communication" (p. 264). We can then understand competent morally responsible agency in terms of the facilities required to understand the moral demands, assessments, and expectations others express by way of their modes of holding us morally responsible through, for instance, their blaming practices. Those exempted from that category of morally responsible agents, such as young children and the severely mentally retarded, are exempted precisely because they are incapable of understanding the forms of communication

others in the moral community might put to them via pertinent alterations in the complicated web of interpersonal relations and practices constituting such phenomena as blaming (and, we might note here, forgiving). They are, thus, not capable candidates for the full spectrum of interpersonal transactions and expectations involved in moral life.

Although Watson proposed an expressive and communication-based theory of moral responsibility, he did not develop it. One of us, however, has done so by modeling our moral responsibility practices along the lines of a conversation.⁶ The model develops an analogy: Our moral responsibility practices are not *literally* conversations, but they are like them in ways that are theoretically illuminating. A competent speaker of a natural language must have the facility to appreciate the conversational maxims of felicitous speech, just as she must be able grasp various Gricean implicatures in order to appreciate others' intended meanings and also convey her own. So too, a morally responsible agent must have the facilities needed to appreciate the demands that others put to her when they hold her morally responsible, for instance, by blaming her or praising her, or instead by standing ready to praise or blame her depending upon how she acts. She must, furthermore, be able to do likewise—that is, to hold others morally responsible and praise and blame them. Her appreciation for the conventions and practices constituting these forms of expression and communication will color her own conduct. Her quality of will in acting as she does, if she is, so to speak, sufficiently articulate, is liable to be sensitive to the means of expressing or communicating moral demands and expectations by members of her moral community. Because of this, her actions themselves are candidate bearers of meaning, meaning that can be understood in terms of the quality of her will. (We shall return to this point momentarily.)

The basic model works on analogy with a linguistic transaction in which two (or more) competent speakers of a shared language converse. Suppose an agent performs what at least appears to be a morally blameworthy action; her conduct seems to reveal moral ill will for some innocent party, and she, let us suppose, does something that is morally wrong. Perhaps she lies to a friend for the purpose of personal profit at her friend's expense. Her action can be understood as bearing a certain meaning regarding the quality of her will and thus as a basis for initiating a conversation that makes certain responses fitting or appropriate (meaningful or felicitous). When another blames her, perhaps by withdrawal of goodwill or typical means of conveying friendship, her doing so is itself expressive of her resentment or moral indignation. This leaves open the possibility of further responses from the putative wrongdoer. She might

offer a justification (she had good reason to lie), or an excuse (such as duress). She might deny the very assumption that she lied or the apparent motive of self-interest. Or she might defiantly acknowledge the wrong done and refuse to display any regret for the wrong. These three stages can be labeled Moral Contribution, Moral Address, and Moral Account.

An especially useful point about understanding a moral responsibility exchange as sketched above has to do with the way that excuses and justifications work. Suppose that in the case above, at the stage Moral Contribution, the would-be liar did not lie to a friend and was not motivated out of personal gain. Nevertheless, her action did appear as if it was indicative of this sort of moral ill will. When at the stage Moral Address another party blames in whatever pattern of activity her resentment or indignation manifests itself, her blaming and its communicative role is premised upon a false interpretation of the meaning of the agent's action—that action did not have the meaning originally assigned to it. It did not manifest the morally ill will it was thought to manifest. Imagine that, in offering an excuse, our would-be liar convincingly explains that she non-culpably came to have some poor information. She could see how one might think that it could be purposefully used to deceive a friend for one's own advantage. But as it transpired, she was not intentionally deceiving her friend. She simply and innocently had a false belief. Here, at the stage Moral Account, the blaming party is given reason by the blamed party to reinterpret the original quality of will with which the blamed party originally acted. As this (analog to a) conversation evolves, or comes to completion, we have a transaction that has left both parties with a relationship that has unfolded in certain ways, ways that reveal our moral responsibility practices to be dynamic rather than static and that help to sustain and constrain our interpersonal relationships.

One further point before turning to the nature of forgiveness: we had previously claimed that, on the proposed conversational theory, a morally responsible agent must be able to master the interpretive resource needed to understand the forms of communication put to her by those who hold her morally responsible. We pause here to explain why this proposed necessary condition for morally responsible agency is built into the conversational theory. Does it not commit to too much? Could there not be agents who freely and knowingly do moral wrong and act from a morally objectionable quality of will but yet lack the interpretive resources and abilities to appreciate the demands and expectations expressed by those who hold them morally responsible? Perhaps. If so, the theory should be weakened. But here is something such agents would be

incapable of doing. Not being able to appreciate the demands put to them by others via the social practices of adult interpersonal life (because they lack the interpretive resources), they would not be able to consider ways of adjusting their behavior so as to avoid the misimpression that they are flouting pertinent demands or ignoring reasonable expectations. So the meaning or significance of their actions would be limited to the extent that the qualities of their wills *could not* be sensitive to salient interpersonal consideration. This sort of sensitivity, while complicated, is so pervasive in our moral lives that it is easy to go unnoticed. The nurse who is busy in triage aiding two patients becomes aware that his attention to them might lead a third patient to think he has no concern for her. Maybe, she thinks, he is a racist, or instead he favors these two patients because they have money, and she is poor and homeless. Worried that his conduct might be misconstrued, while caring for the other two patients, he pauses to offer a comforting glance to this third patient. Maybe he behaves in ways to make perspicuous to the third patient that the other two patients really are in more urgent need of care. Our moral lives are chock-full of these sorts of micro-adjustments to our conduct, all explained by our sensitivity to how others might naturally construe (or misconstrue) the quality of our will when it is open to potential misinterpretation given how we are acting.

We now turn to the topic of forgiveness. We'll first set out a problem for understanding it, and then proceed to nest it within the conversational theory of moral responsibility.

2. Accounting for Forgiveness: A Problem and Diagnosis

The phenomena counting as forgiveness are diverse. Sometimes we forgive "in our hearts" those who have wronged us: we might stop being resentful, for example. Other times, we express our forgiveness in word or deed: we might say "I forgive you" or give a close friend a knowing look that they understand to be an expression of forgiveness. Sometimes, forgiveness seems to happen in a moment—we are done a minor wrong and when faced with the option of holding it against the wrongdoer or forgiving, we simply decide to forgive and move on with our lives.⁷ Still other times, the process of forgiving seems to take years, perhaps even a lifetime. The many flowers of forgiveness bloom.

Upon reflection, then, it should come as no surprise that the seemingly innocuous question, "What is forgiveness?" has solicited a diversity of answers from philosophers.⁸ For example, one currently popular account of forgiveness claims that to forgive someone who has wronged you is (roughly) to overcome the resentment you have towards them because of

what they did to you (Murphy 2003, Holmgren 1993, Hughes 1993).⁹ Others have connected forgiveness to punishment and therefore see forgiveness as involving the forswearing of punishment (e.g. Hobbes 1969, Zaibert 2009). Bishop Butler famously contended that forgiveness is the virtue that leads us to forswear revenge on those who wrong us (1846). Jean Hampton has argued that forgiveness requires a decision to see the wrongdoer differently—to revise one's judgments about her (Murphy and Hampton 1989, pp. 84-85). Still others have focused on the illocutionary force of our forgiveness utterances (Haber 1991, Pettigrove 2004).

Given this diversity in philosophical accounts of forgiveness, which view is preferable? Even more fundamentally, what philosophical strategies should be employed for determining which view is preferable? This is a problem. To solve it, we introduce a methodology for theorizing about forgiveness. But before we do, it will prove instructive to see what goes wrong in much theorizing about forgiveness. Our foil here is the Resentment View, by far the most commonly held account of forgiveness.¹⁰ According to this view, forgiveness is the overcoming of resentment. Actually it might be best to think of the Resentment View as a *family* of related views that regard forgiveness as "primarily concerned with how one feels instead of how one acts."¹¹

Most defenders of the Resentment View have been keen to clarify that it is not just *any* kind of overcoming of resentment that may qualify as forgiveness. Were you purposely to hit your head on a rock, thereby causing your resentment to be abated and removed, you would not have forgiven. Defenders of the Resentment View therefore usually require that the forgiver's resentment be overcome for the *right reasons*.¹² For example, Jeffrie Murphy, one of the more well-known proponents of the Resentment View writes that, "Forgiveness may be viewed as the principled overcoming of feelings of resentment that are naturally (and perhaps properly) directed toward a person who has done one a moral injury" (2003 p. 561). Charles Griswold has in mind "the letting go of resentment for moral reasons, as well as revenge, without forgetting the wrong that was done, and even in some cases (re)accepting the offender as a friend. This is what we are calling forgiveness" (2007, p. 40). Stephen Darwall has also advocated a Resentment View of forgiveness: "To forgive is, roughly, to forbear or withdraw resentment" (2006, p. 72). P.F. Strawson connects forgiveness to resentment as well:

[T]o ask to be forgiven is in part to acknowledge that the attitude displayed in our actions was such as might properly be resented and in part to repudiate that attitude for the future (or at least for the immediate

future); and to forgive is to accept the repudiation and to forswear the resentment. (1962 [2003], p. 76)

One way of understanding the Resentment View is as involving the claim that it is a necessary condition for forgiveness that the forgiver overcomes resentment. We could gloss it as:

RV-N: If agent S forgives agent P (for P's conduct), then S overcomes S's resentment towards P (for P's conduct).

RV-N entails that one cannot forgive unless one has felt resentment.¹³ Is this true? As we see things, RV-N is not obvious and indeed, we think there is good reason to deny it. Take for example, the view that resentment involves feelings of ill-will. R. Jay Wallace gives an example (1994, pp. 76-77) in which an especially charming colleague has wronged you by cheating and lying to you. However, because of her sweet and charming disposition, it's hard for you to work up any feelings of ill-will towards her. And yet, in light of what she has done to you, you might very well hold her morally responsible by altering the terms of the relationship, by withholding friendly relations, expressing disapproval for what she has done, or making moral demands on her. You see that it is appropriate to hold her responsible and blameworthy for what she has done, and yet you feel no resentment or hostility. Were your charming colleague to apologize and ask for forgiveness, we see no good reason to think that you would be barred from forgiving her simply because you felt no resentment towards her in the first place. You might stop avoiding her, start meeting her for lunch again, and then treat her as if she had not harmed you. Why not admit that these sorts of actions would reveal that you had forgiven her after all? These kinds of cases suggest that in order to forgive it is not necessary to overcome or forswear resentment.¹⁴

One may try to avoid this consequence by conceding that it is not essential to forgiveness that one overcome resentment, and go on to modify the Resentment View to claim that *if* there is resentment, *then* it must be overcome. This makes the requirement to overcome resentment conditional:

RV-CN: If agent S forgives agent P (for P's conduct), then if S felt resentment towards P (for P's conduct), then S overcame S's resentment towards P (for P's conduct).

But consider the set of cases in which one forgives without overcoming resentment. We can then ask: in virtue of what has

forgiveness taken place? If it is not necessary to overcome resentment in order to forgive, then what does forgiveness amount to in those cases where there is no resentment to overcome? There must be some underlying phenomenon—some *further* fact—that explains why forgiveness can take place even when there is no resentment to overcome. But here the Resentment View is silent. We conclude then that one must neither feel resentment nor overcome it in order to forgive.¹⁵

Of course another way to understand the Resentment View is as claiming that overcoming resentment is sufficient for forgiveness:

RV-S: If agent S overcomes resentment towards agent P (for P's conduct), then S forgives P (for P's conduct).

But this will not do either.¹⁶ For suppose Michael overcomes the faint resentment he has towards Brandon for what Brandon did (we won't tell you), and yet Michael avoids Brandon and withdraws what were previously friendly relations. He expresses his disapproval of what Brandon has done, and requests an apology. He sees these as necessary steps to take in order to express his disapproval and to make demands, given the harm Brandon has caused. Were Michael deliberately to engage in all of these overt actions and *at the same time* insist that he has forgiven Brandon for what he has done, Brandon would rightly be dumbfounded. "If you have forgiven me, how is it that you can overtly blame me and do so with such deliberateness, intending to express your disapproval and make your demands" (not that Brandon would actually use those words)? We conclude then, that overcoming resentment (even when done for the "right reasons," whatever those are) is neither necessary nor sufficient for forgiveness.

Return now to the question we posed above: in light of the apparent diversity in the phenomena of forgiveness, how should we proceed to theorize about it? The point we have tried to bring into relief here is that it is best not to theorize about forgiveness by giving necessary and sufficient conditions. This is the problem with the Resentment View, and we think this will be a problem for any account of forgiveness that seeks to give necessary or sufficient conditions. The phenomena pertaining to forgiveness are too diverse and too diffuse to admit of this kind of analysis.¹⁷ We cannot, of course, give an argument that no (illuminating, non-trivial) analysis of forgiveness would succeed, but we do offer two things in return. First, we offer the fact that no analysis of forgiveness we know of has so far succeeded in capturing all cases of forgiveness—indeed, none has even come close to doing so. Second, we offer a

replacement methodology for theorizing about forgiveness. Given that the many flowers of forgiveness bloom, we suggest theorizing about forgiveness by first explaining an exemplar sort of forgiveness, and then considering other cases by reference to the exemplar ones. The explanation proffered for the exemplar cases should not be construed in terms of necessary and sufficient conditions that other cases should also satisfy so as to count as forgiveness at all. Rather, departures from the exemplar cases can be understood as gaining explanatory insight by reference to their similarity with the pertinent exemplars.

3. In Defense of a Methodological Proposal

What, then, are the exemplar cases of forgiveness? To answer this question we take a brief detour to discuss moral blame. Blame, like forgiveness, is a diverse phenomenon. In other work, McKenna has distinguished between *private blame*, *overt blame*, and *directed blame* (2013). Private blame involves adopting a blaming attitude toward someone but concealing (perhaps intentionally, perhaps not) the outward behavioral manifestations. We might, as it were, keep our blame to ourselves. Overt blame involves adopting such an attitude *and* making it manifest in one's conduct. This can be done in the absence of the blamed. One might go out of one's way to avoid the blamed, or one might verbally blame the dead. Directed blame is a form of overt blame aimed at addressing the blamed party. Indeed, it is outwardly directed *at* the blamed party. One might reprimand, or make moral demands, or express disapproval *directly* to the blamed.

A similar set of distinctions can be made with respect to forgiveness. *Private forgiveness* involves adopting a forgiving attitude towards the forgiven. One might moderate one's resentment and adopt an attitude of good-will (although even here, we do not think that moderating resentment is a necessary condition). *Overt forgiveness* involves adopting such a forgiving attitude *and* making that attitude manifest in one's conduct. And just as one can overtly blame in the absence of the blamed, one can overtly forgive in the absence of the forgiven. One might express forgiveness of the deceased in a eulogy, for example. *Directed forgiveness* is a form of overt forgiveness that is aimed at communicating with the forgiven. One might say to the forgiven, "I forgive you," or give the forgiven a knowing look, which the forgiven understands as an expression of forgiveness.

One might get the impression from the sheer amount of attention that is devoted to them that the exemplar cases of forgiveness are best construed as the private ones, and in particular, the cases in which one overcomes

resentment. Even if it is true that the private cases of forgiveness outnumber other kinds of cases, we do not think that it is most fruitful to treat them as the prototypical examples of forgiveness. Rather, it seems to us that the exemplar cases of forgiveness are better understood as the cases of directed forgiveness. Because this is contrary to (what we take to be) the widespread assumption that forgiveness is paradigmatically private (recall Murphy's claim that forgiveness is "primarily concerned with how one feels instead of how one acts"¹⁸), we pause here to defend this move.

Why ought cases of directed forgiveness be regarded as the exemplar ones? Here, we offer five reasons. Space will not permit a full defense of any of these considerations, and so we ask the reader to consider them collectively as offering an overall picture that speaks to the plausibility of our contention.

First, we have noted that on the Strawsonian theory, holding morally responsible means something in practice. And this applies, of course, to blame as a particular mode of holding morally responsible. In the case of blame, pertinent attitudes, such as resentment or indignation, provide the motivational base for fitting alterations to practices, the sort of alterations that modify relationships in ways characteristic of blaming. Hence, there is reason to take cases of directed blame to be the exemplars. So too, we contend, the forgiving attitude, to the extent that it involves characteristic emotions, or instead, the relinquishing of certain emotions like resentment, also provides the motivational base for fitting alterations to practices of a sort that involves modifying relationships.

Second, reflect on cases of private blame, and on the emotion of, say, resentment. What makes that emotion intelligible to the blaming party as an emotion of resentment rather than something like mere contempt, or raw hatred, or jealousy, or some other emotion? It is, we propose, certain criterial indicators of what would count as an expression of *that* emotion. We come to understand our own private emotional episodes for what they are precisely by understanding how one might intelligibly manifest them in word and deed. The same applies to forgiveness. One who privately forgives by, for instance, overcoming resentment is subject to a constellation of emotional dispositions which is such that it would lead one to respond and outwardly regard the blamed party in ways characteristic of forgiving.

Third, it is part of what blaming is—that is, part of its nature—that it is the sort of thing that can be subject to norms governing its propriety. So too, we say, for forgiving. Now, *some* of the norms bearing on blaming and also some of the norms bearing on forgiving are norms that concern how we ought to treat the (allegedly) blameworthy party. There is, in

these norms, an implicit presupposition that whether one ought to blame or instead forgive is within the ambit of the blamer's or forgiver's voluntary control. But notice that the sort of control presupposed here is ill-suited for treating paradigm cases of blaming or forgiving as the private ones. The private ones typically involve episodes of emotional responses that are not easily thought of as within a person's voluntary control. It's not simply a matter of deciding or choosing, for instance, to be angry or hurt or fearful. Emotions, including the dissipation of them, are not usually that malleable in response to our practical judgments and decisions. But our outward behavioral manifestations of them are—or at least can be. We do have some degree of voluntary control over whether we manifest certain emotions in our dealings with others. Hence, to the extent that we are open to judgments about whether we ought to forgive someone for a past wrong, and to the extent that there is a presupposition that these ought-claims involve things that we are able to bring about voluntarily, there is good reason to think that the cases at issue are best thought of as the overt, directed cases. These are cases where we have a greater degree of voluntary control.

Fourth, forgiveness is often a response to an apology, a request for forgiveness, an act of restitution, and the like. These things are all essentially overt behaviors directed at those positioned to hold responsible. Seen in this context, forgiveness is paradigmatically a *response* to these practices. Of course, we wish to emphasize here, as with the three reasons adumbrated above, that our contention is not that it is either necessary or sufficient that all instances of forgiveness be understood in this manner. But cases that are so featured cast light on the sort of interpersonal role directed forgiveness is able to play—one understood as a response. And, we maintain, other cases of forgiveness can be usefully thought of by reference to these sorts of paradigmatic cases.

Fifth, we also note a straightforward normative observation about forgiveness at its best, so to speak.¹⁹ As a moral ideal, forgiveness involves its expression to the forgiven. There is something incomplete or sub-optimal about an agent who privately forgives, but, given the chance, does not also express forgiveness to the forgiven. We see this as evidence that exemplary forms of forgiveness involve the communication of forgiveness to the forgiven. For those who are skeptical of our diagnosis, it is worth attending to what is involved in forgiveness's withdrawal of a disposition to directed blame. Directed blame can plausibly be regarded as constituting a potential harm to the one blamed. It has a characteristic sting, often understood in retributive terms. Blame hurts, or at least it can. For one to forgive but to refrain from directing her forgiving attitude to the

wrongdoer by word or deed leaves the wrongdoer exposed to the potential persisting harm of being, or minimally feeling, blamed. By not directing our forgiveness when we have the chance to do so, we also permit the forgiven to labor under the assumption that she is still to be the object of our blame. Worried that the forgiver might still withhold friendly relations, request further restitution, reprimand, or express demands, the forgiven may continue to alter *her* behavior accordingly—to behave not as one who has been forgiven, but as one who is to be blamed. Absent being forgiven directly, the blamed may simply be ignorant of her new moral standing in relation to the forgiver, and this could needlessly harm her and the relationship. If nothing else, a spirit of generosity invites the thought that, in the absence of further normative considerations, withholding directed forgiveness from one who is (privately) forgiven is stingy, mean-spirited, ungenerous; it involves a defect of character.

4. Nesting the Nature of Forgiveness within the Conversational Theory

In the previous section, we have suggested a methodological proposal departing from the familiar philosophical strategy of classical conceptual analysis. We have proposed that one theorize about forgiveness, not by offering necessary and sufficient conditions adequately capturing its complete extension, but by offering illuminating explanations of exemplar or paradigm cases of forgiveness. Here, we have argued that the cases theorists ought to focus upon are not the private cases of forgiveness, which might be more poetically described as forgiving in one's heart. Rather, one ought to attend to the publically accessible cases of overt forgiveness in the contexts of directing one's forgiveness to the one forgiven. Perhaps this will be regarded as a predictable "Strawsonian" move insofar as Strawsonians emphasize that moral responsibility means something in practice. We seek to explain forgiveness within the family of moral responsibility concepts, and so we grant that it is not surprising that we incline toward this strategy. Nevertheless, we believe we have enumerated several compelling reasons why one ought to tend first to directed blame. Let's consider now how one might fit forgiveness within the framework of the conversational model.

In setting out the conversational theory of moral responsibility (sec.1), we distinguished between three stages in a moral responsibility exchange: Moral Contribution, Moral Address, and Moral Account. Recall that the first stage involves a candidate morally significant contribution from a morally responsible agent, typically some act or omission, indicative of the

moral quality of an agent's will. Another holding that agent morally responsible, at the stage Moral Address, responds with a pertinently charged morally reactive attitude, and she does so directly by altering practices in ways that are expressive of her resentment or indignation. This can be understood on analogy with a conversational response to the significance or the meaning of a speaker's initiating conversational contribution. Then, at the third stage, Moral Account, a morally responsible agent is situated to respond to those holding her morally responsible by apologizing, defying, or perhaps merely acknowledging wrong done. Attempts at excuses and justifications are open at this stage too, but the cases we are focused upon here are not ones in which any excuse or justification applies. In the cases at issue, the agent is in fact blameworthy and is rightly blamed at the state Moral Address.

Given this picture, we now propose that directed forgiveness is best understood in exemplar cases as responses at a further (perhaps fourth) stage by the one blaming, and typically it is in response to the blamed party's fitting contribution at the stage Moral Account. Forgiveness, understood in this way, is a *further* conversational move in this sort of dynamic transaction between one held morally responsible and those so holding her. Though somewhat awkward, we shall call this further stage Moral Reconsideration.

By understanding forgiveness in terms of the conversational model and by reference to the further stage Moral Reconsideration, we can explain forgiveness as one amongst a range of potential further meaningful or intelligible responses to the blameworthy agent, responses naturally or typically (but not always) following upon a blameworthy agent's accounting for her objectionable conduct at the stage Moral Account. A distinct response might instead be to punish rather than forgive, or simply to persist in overtly blaming. The most familiar sort of dynamic is one in which forgiveness is a conversationally meaningful or intelligible response to an apology, an act of contrition, or some sort of effort at restitution. This is in contrast to a case in which an agent, at the stage Moral Account, responds to blame with contempt, defiance, or disregard. But note that there is nothing on our account as far as blame's nature goes that rules out responding to defiance with forgiveness. Not all conversations follow a typical script, and some go in ways that remain intelligible even while violating pertinent norms bearing upon what a blameworthy person deserves, or what constitutes a victim's respect for her own dignity.

As we envision our proposal, the exemplar or paradigm cases of forgiveness are cases of directed forgiveness in which the overt manifestations of forgiving have certain characteristic behavioral manifestations. Most

particularly, the relevant form of behavior concerns patterns of conduct that are characteristic of the common criterial indicators of relinquishing resentment. That is to say, just as manifesting resentment involves certain behavioral patterns that are familiar criterial indicators of expression of the emotion, so too does manifesting a retreat from or a relinquishing of resentment. Directed forgiveness, then, involves intentional alterations away from modes of conduct that typically indicate a disposition to manifest resentment toward a wrongdoer.

We stress here that as we conceive of the relevant range of exemplar cases of directed forgiveness, such instances do not *simply* involve intentionally engaging in overt behaviors, perhaps in the absence of any alterations to one's internal life. As we noted above, directed forgiveness (as a form of overt forgiveness) characteristically also involves the taking up of a forgiving attitude. In some cases this may involve a relinquishment of resentment, but as we have argued above it need not. For even if one has no resentment to relinquish, one can still alter one's conduct in ways that communicate to the forgiven that that one will no longer treat her in ways typical of manifesting resentment. One can, that is, engage in behavior that communicates to the forgiven that one will no longer hold the forgiven's blameworthy act against her. And this can be done without the forgiver experiencing antecedent resentment or moderating or eliminating it. As we understand it, then, the cases of directed forgiveness we wish to treat as exemplars do involve *some* kind of private forgiveness, just not any particular manifestation of it. Furthermore, our proposal is *not* that directed forgiveness is somehow metaphysically prior or more basic than private forgiveness. Directed forgiveness assumes, or naturally invites as an implicature, as it were, private forgiveness. Our claim is that these cases of directed forgiveness should enjoy explanatory privilege: the key to understanding the flora and fauna of forgiveness lies with first understanding these cases of directed forgiveness.

Note that our proposal helps show where the Resentment View gets something right, but also where it goes wrong. What it gets right is that there *is* an important conceptual tie to the role of resentment, or better put, to the role of diffusing it. But where it goes wrong is in featuring the place of actual private episodes of resentment in the etiology of forgiveness as either necessary or sufficient. It's possible to forgive, on our proposal, when one who holds responsible as a matter of fact did not experience any resentment (like in Wallace's case of the charming colleague). One does so by engaging in patterns of conduct that present themselves as ways one could express her relinquishing of resentment, were she in fact to resent. It's also possible, on our proposal, to forgive when one persists in

experiencing pangs of resentment. One engages in patterns of conduct typically indicative of relinquishing of resentment, and one adopts a policy or attitude regarding her residual resentment of doing her best to suppress or correct any outward expressions of it. (As most of us imperfect souls are often conflicted in our efforts to forgive, isn't this really how it usually goes?)

Because we think that there are no particular private episodes that are either necessary or sufficient for forgiveness, this might invite the objection that we have allowed for situations in which one "forgives" all the while harboring intense hatred toward the object of putative forgiveness. Does this mean that anything goes when it comes to forgiveness? To answer this objection we remind the reader of our methodological suggestion: to identify and account for the paradigm cases of forgiveness and explain other cases by reference to the paradigmatic ones. Some cases, then, will clearly count as instances of forgiveness, others less so. Cases of directed forgiveness, on our view, are the exemplar ones. But what about self-forgiveness, or forgiving those who are dead, or political forgiveness between groups or states? These are, we think, further from the paradigmatic cases. At the outer edges, however, it just may simply not be clear whether a certain activity should count as forgiveness. There may be considerable gray area given that the contours of forgiveness are not clearly circumscribed. Falling within this gray area are cases in which one exhibits forgiveness-like behavior but retains certain attitudes atypical of forgiving (we can call such cases *hollow forgiveness*, to borrow a term from Baumeister, et al. 1998). In some cases, one might engage in overt forgiving behaviors while still retaining moderate feelings of resentment (perhaps it is indulged in, or perhaps it is mostly unbidden). In other cases one might engage in overt forgiving behaviors all the while hoping that the blameworthy agent experiences unimaginable suffering. Are we to call all these cases of forgiveness? In some cases, probably so. In others it may not be clear at all. In still other cases, probably not. The main point here, though, is that the further we get from the paradigm cases, the less clear it is that forgiveness has taken place.

5. Concluding Remarks

As noted previously, our goal in this paper is most immediately to offer an account of the nature of forgiveness, not the norms governing it. However, we conclude with a few suggestions about how one might approach the normative question.

First, note that directed forgiveness, as a stage in a moral responsibility exchange, carries certain (analogous to) Gricean implicatures and felicity conditions. Directed forgiveness presumes that the person forgiven is indeed blameworthy, and that the forgiver had blamed the person previously, or at least held her morally responsible and blameworthy. It also presumes that the person forgiving is in a position to blame the guilty party and then, of course, is in a position to forgive her. Any of these presuppositions might be violated, and in such cases, a relevant norm would be at issue. Think here of the analogous cases of violating certain conversational felicity conditions, such as responding to someone under the assumption that she asserted something that, after all, she did not assert, or of "butting into" a conversation when it is not one's place to do so.

Here is a different set of considerations worth mining: We often evaluate conversational contributions in terms of *degrees* of meaningfulness, intelligibility, or aptness. Oftentimes, a person knows "just what to say," and we can evaluate her, indeed, congratulate her, on putting a point "just right," and so following through on a conversation in an especially insightful or meaningful way. And typically, these evaluations are based on how alive the conversationalist was to the particular details of the history of that conversational exchange. She spoke directly to the meaningfulness of her interlocutor's prior conversational contributions. Note also that, in such cases, there is no presumption, despite expressions like "said just the right thing", that there would not be other equally rationally defensible means of carrying on the conversation. There is a permissiveness, a sort of rational pluralism, as to how a good conversation might proceed. Here, norms of conversational aptness are at play, but they are not rigid or exclusionary. So too, we contend, for normatively warranted means of directed forgiveness. There can be more or less meaningful or intelligible ways to forgive, ways that are better or worse means of expressing to a blameworthy party one's altered stance toward her in light of the meaningfulness of the wrong done, the moral account offered (such as an apology), and so on. And just as is the case with conversational aptness, the fact that there is not just one fitting or apt way to forgive does not mean that there are ways of doing so that are not intelligible, not apt, not fitting. These observations do not fix upon normative considerations that permit of elegant, lean principles that can be easily stated. They will arise, as they do in the context of conversational norms, as a function of pragmatic considerations that make context crucial.

This point is important to keep in mind when we consider the question of what norms govern the vast flora and fauna of our forgiveness practices.

Is there a duty to forgive? If so, does this entail that the blameworthy party therefore has a corresponding right to be forgiven? Or is forgiveness not a duty, but something good and desirable nonetheless? Perhaps forgiveness is like gift-giving in that it is benevolent but that withholding it is not (at least generally) morally required, even if it may evince vice. Or is forgiveness supererogatory? Perhaps no one is ever blameworthy for not forgiving, but it is worthy of our moral aspirations nonetheless. As one might anticipate by now, we think it best not to expect that all putative cases of forgiveness will answer to the same set of norms. Consider, for example, the parents, partners, and children of the 77 persons who were killed during Anders Behring Breivik's 2011 bombing of Norwegian government buildings and shooting rampage through a summer youth camp. We find it hard to fathom that upon hearing of their deaths, dozens of parents had a moral duty to forgive Breivik for taking their children away from them, such that they were morally blameworthy for not immediately forgiving. On the other hand, suppose Fred tells his close friend Barney a white lie, which Barney holds against Fred for years. Barney simply refuses to forgive Fred, even though Fred has spent years apologizing, offering restitution, and showing that he has had a change of heart. Here, it strikes us that Barney is blameworthy for not forgiving, and may even have an obligation to forgive. (At the very least, Barney is being a big jerk.) To assume that all instances of forgiveness must answer to the same norms therefore seems to us to be an unnecessary theoretical straitjacket.

The Conversational Model allows for what we see as a desirable pluralism about the norms of forgiveness. Just like the appropriate or fitting responses in a conversation cannot be rigidly scripted, the appropriateness of forgiveness does not answer to rigid norms. Due to many contextual features (e.g. the seriousness of the wrong done, the nature of the relationship between the victim and the wrongdoer, the behavior and attitudes that the wrongdoer displays at the Moral Account stage, etc.), some putative instances of forgiveness may be supererogatory, while others may be obligatory. The relevant norms might also depend on how long the "conversation" has been going on. Forgiveness may be subject to different norms earlier in the conversation, as it were, compared to later in the conversation. This is because aspects of the unfolding conversation themselves can color the appropriateness of further future interactions. There is a kind of fluidity in evolving relationships that might make forgiveness more or less appropriate as the conversation unfolds. Barney may not have had an obligation to forgive Fred immediately after Fred's lie, and may not have been blameworthy for not doing so. But

given an unfolding conversation over the course of a decade, however, the norms may change.

Finally, we close by addressing a common theme in the philosophical literature on forgiveness, namely that forgiveness appears to admit of a kind of paradox.²⁰ Above, we noted a kind of pluralism about the norms of forgiveness. Not all putative instances of forgiveness answer to the same norms, and the Conversational Model has the resources to explain why this is the case. But there is a deeper issue—one not about how the norms governing forgiveness could vary from case to case—but about how forgiveness could ever be appropriate *in the first place*.

Here, in brief, is the puzzle: Consider the fact that putative cases of forgiveness involve forgiving a blameworthy agent. If an agent is deserving of blame, how could forgiving her be justified or appropriate? Isn't blaming her something that is required, is in fact a duty? "Forgiving" her appears to collapse into condonation—by undertaking a policy of not blaming a blameworthy agent, we take a bad situation and make it worse. And by not responding with appropriate blame we do not just condone the blameworthy conduct, we evince a lack of self-respect by not standing up for ourselves in holding the wrongdoer to blame. Forgiving wrongdoers as they stand in the full light of their blameworthiness therefore raises the concern that forgiveness in such cases is not appropriate—it reduces to an act of condonation and servility. When then might forgiveness be appropriate, given the worries expressed above? One might think that forgiveness is appropriate when the wrongdoer has had a change of heart, apologized, requested forgiveness, and the like. Perhaps something like this is correct, but if so, some sort of explanation is called for. Engaging in these activities does not achieve exculpation—one can still be blameworthy even when one has had a change of heart, apologized and asked for forgiveness. Breivik would remain morally blameworthy for his killings even had he immediately apologized and had a change of heart. So the puzzle remains: a wrongdoer remains blameworthy even after a change of heart, apology, etc. Hence, we return to the above worry—that to "forgive" agents who deserve blame reduces to condonation and an act of servility. Should we then jump to the conclusion that forgiveness is only appropriate if an agent is *not* blameworthy? If that were the case, what would we be "forgiving" the agent *for*? And what would it accomplish—if an agent is not worthy of blame, how could forgiveness constitute a meaningful and appropriate response to an agent? Additionally, to "forgive" an agent who is not blameworthy can itself constitute a wrong. This is why we are offended when someone tells us we are forgiven for something for which we know we are not blameworthy. The trick then is

to explain how forgiveness can be an appropriate and meaningful response to a blameworthy agent—to show that forgiveness can be justified *and* have a point.

Here, we briefly suggest a way to address this seeming paradox. We begin with an observation about blameworthiness itself. In other work, McKenna has argued that deserved blame can be helpfully characterized in terms of it being permissible to cause harm of a certain sort (2012).²¹ Perhaps some cases of deserved blame involve a duty or obligation to cause a pertinent harm, but it needn't be thought of as part of what deserved blame is that it unilaterally involves a requirement to harm the blameworthy party. The fact that a wrongdoer is blameworthy, therefore, does not entail that the victim of the wrong has a duty to blame—blameworthiness thus provides relatively flexible *pro tanto* reasons to blame.²² Indeed, there is an absurdity in thinking that blameworthiness for conduct as such creates duties to blame: given that we all are blameworthy for lots of things that we have done to others, we would all constantly be under an obligation to blame those who have wronged us. Morality may be demanding, but not in that way. If this is correct, then there is nothing inherently inappropriate about forgiving a blameworthy agent instead of (continually) blaming her. Nor is one guilty of condonation or servility in virtue of forgiving a blameworthy agent. For when one undertakes a policy of directed forgiving, one may still retain certain pertinent beliefs, such as the belief that the wrongdoer did do wrong, and is blameworthy for having done so. She may believe that the wrongdoer deserves certain kinds of punishment by a civil or administrative authority (and may even turn the wrongdoer over to the authorities), and she may believe that the wrongdoer has various defects of character and try to help (in appropriate ways) the wrongdoer to rid herself of them. In some cases, forgiveness may accompany only limited reconciliation. A battered wife may forgive her spouse without deciding to remain in the relationship. These are not the beliefs and behaviors of condonation or servility, and yet they seem perfectly compatible with directly forgiving those who are blameworthy for wronging us. Finally, by showing that forgiving blameworthy agents can be appropriate, we are able to show that forgiveness is not pointless. In forgiving we undertake new policies for how we will treat and regard those who are blameworthy for having wronged us. Note also, regarding this final point, the clear presupposition of undertaking a *new* policy, one that involves a transition from a prior policy. In particular, it involves a transition from a prior orientation, which typically will be to blame. When we understand forgiving as a move in an unfolding conversation, as the conversational theory of moral responsibility would have it, there is

nothing puzzling about the relationship between being blameworthy, blaming, and then forgiving. There is, therefore, no paradox in the idea that we might (now) not blame the blameworthy. In a typical sort of case, the blameworthy person *was* addressed in a certain way in blaming her; the forgiving, as a further response, does not “wash away” the fact that the prior exchange involved blaming. It just treats the antecedent blame as part of a process that is dynamic, not static.

Our primary goal in this paper has been to show that an illuminating account of forgiveness can be nested within a conversational theory of moral responsibility. At a stage we called Moral Reconsideration, victims of wrongdoing may respond to wrongdoers by way of forgiveness. And we have argued that instead of trying to account for the diverse phenomena of forgiveness in terms of necessary or sufficient conditions, a better methodology involves identifying and explaining the exemplar cases and accounting for other instances of forgiveness by reference to the exemplar ones. We then suggested that the exemplar cases of forgiveness can helpfully be understood as involving conduct typically indicative of the relinquishment of resentment. Such “directed forgiving” therefore involves alterations away from modes of conduct that indicate a disposition to manifest resentment toward a wrongdoer. We then turned to a brief discussion of the norms of forgiveness given a conversational theory of responsibility. There, we suggested not only that there are more or less appropriate ways of forgiving, but that the norms of forgiveness themselves may differ slightly from case to case. For just like the felicitous moves in a literal conversation cannot be regimented by easily identifiable norms, neither can the norms of forgiveness be universally regimented across all responsibility “conversations.” We see this as a desirable feature and not a bug of this way of thinking about forgiveness. Finally, we addressed the worry that forgiveness admits of a kind of paradox—that it is either always unjustified or that it is simply pointless. We argued that this objection falls away if we understand blameworthiness as not always morally requiring that victims blame. Therefore forgivers are (at least sometimes) justified in adopting a policy of forgiving instead of blaming—an alteration of conduct that is manifestly not pointless.²³

Notes

¹ For a philosopher who convincingly distinguishes between kinds of responsibility, see Haji (1998).

² One of us, Michael McKenna, is fully committed to the Strawsonian enterprise. The other, Brandon Warmke, remains undecided, but will entertain the view for the purpose of this paper.

³ How they are crucially implicated is a matter we will leave unsettled. Most Strawsonians seem to assume that the relation is essential, so that the morally reactive attitudes are necessarily implicated in holding morally responsible. But one of us (McKenna, 2012) has argued that the relationship could be contingent and nevertheless deep and pervasive, so that it would remain theoretically profitable to explore the nature of holding and being morally responsible by attending to the reactive attitudes, even if there are possible worlds (not our own) in which agents are morally responsible and others hold them morally responsible, but yet no one is susceptible to the pertinent moral emotions.

⁴ We reject a fourth Strawsonian tenant (assuming Strawson even meant to commit to it). This is that holding morally responsible is conceptually and explanatorily more basic or primitive than being morally responsible, so that one can give a full account of the latter in terms of the former.

⁵ Objective moral wrongness might not even be necessary for an agent's acting from a blameworthy quality of will. Perhaps all that is required is that an agent believe herself to have done what is objectively morally wrong. (For a defense of this subjectivist view, see Haji 1998; and Zimmerman 1988.) We remain neutral on this issue.

⁶ See McKenna 2012.

⁷ For ease of expression we will use the term 'wrongdoer' to refer to an agent who is a putative candidate for being forgiven, and the terms 'wronging' and 'wrongdoing' to refer to the conduct for which a wrongdoer is forgiven. We will also use the term 'victim' to refer to an agent who is putatively in a position to forgive a wrongdoer. We intend for nothing more to be inferred from the use of these terms.

⁸ Psychologists working on forgiveness have also discovered this diversity in trying to operationalize forgiveness. Everett Worthington writes, "It is not surprising that the major issue characterizing this new science of forgiveness has been how forgiveness ought to be defined" (2006, p. 3).

⁹ But even here, there is considerable disagreement about how this view should be understood. For example, there is disagreement about what resentment is. Some think of it as a "hostile feeling" which aims at inflicting harm on the offender (Garrard and McNaughton 2002), while others think of it as a feeling of "moral protest" (Hieronymi 2001), while some others think of resentment as a kind of "moral anger" (Hughes 1995). Some have understood the view to mean that resentment must be permanently eliminated (Haber 1991, p. 7), while others argue that resentment must only be moderated (Holmgren 1993, pp. 341-42). There is also disagreement about whether it is resentment alone that must be overcome, or

more generic "moral anger" (Hughes 1993), or, to widen the position further, whether any kind of negative reactive emotion at all, including disappointment and sadness, must be overcome (Richards 1988).

¹⁰ It is a testament to this view's popularity among philosophers that John Kekes (2010, p. 490) and Leo Zaibert (2009, p. 388) have both recently labeled it as the "standard view."

¹¹ Cf. Murphy in Murphy and Hampton (1988, p. 21): "Forgiveness is primarily a matter of how I *feel* about you (not how I treat you), and thus I may forgive you in my heart of hearts or even after you are dead." Similarly, Paul Hughes claims that, "forgiveness is essentially a matter of how one *feels* about another" (1995, p. 108, emphasis is original).

¹² Hence Hughes: "forgiveness seems to require overcoming personal moral anger for a reason that is compatible with morality" (1995, p. 107).

¹³ Griswold appears to endorse this view explicitly: "If one felt *no* resentment in response to someone's injurious action against oneself, it would make no sense to forgive them for their deed" (2007, p. 40).

¹⁴ Admittedly, we have only targeted a certain way of understanding resentment here, that which understands resentment as involving feelings of ill-will. But what of other accounts of resentment, such as one that understands it as "moral protest?" But here, too, we think that it is not necessary that one give up one's protest against a wrongdoing in order to forgive. Dana Nelkin makes the same point in her excellent paper in this volume. Indeed, we strongly suspect that for any extant account of resentment, there will be cases of forgiveness that do not involve overcoming or forswearing it.

¹⁵ Swinburne (1989, p. 87, fn. 8) and Digeser (2001, p. 26) also hold this view. Digeser gives the following example: "Jack may have promised to send an article to Jill but then forgot all about it. Jill need not resent Jack for the breaking of the promise to be a wrong or to forgive Jack. When Jack remembers the broken promise, he may feel bad, even if Jill doesn't, and he may consequently apologize. It is possible to forgive without resentment entering the scene" (p. 26).

¹⁶ Among those who hold that overcoming resentment is necessary for forgiveness, some explicitly deny that it is also sufficient (e.g. Griswold 2007, p. 16; Garrard and McNaughton 2003, p. 44; Govier 2002, p. 59).

¹⁷ Although far from exhaustive, one might, for example, consult Haber's (1991) criticism of Hampton's account, Warmke's criticisms (2011, forthcoming) of the punishment-forbearance account, Griswold's criticism (2007) of Butler's account, and Nelkin's criticisms (this volume) of Hieronymi's, Garrard and McNaughton's, and Murphy's accounts.

¹⁸ Cf. Hughes: "[S]ince resentment necessarily involves the belief that you have been wrongfully harmed and is thus a central case of personal moral anger, forgiveness will paradigmatically involve overcoming resentment" (1995, p. 108).

¹⁹ Some may balk at a willingness to incorporate the moral ideal of forgiveness into a discussion of the exemplar cases of forgiveness. (We thank Steve Wall for raising this concern.) One reason for concern here is that there is substantive disagreement about just how moralized an account of the nature forgiveness should

be and that our proposal should not settle this dispute from the outset. For suppose one thought that forgiveness is a moral virtue. If so, then whenever forgiveness is offered, something morally good has been accomplished. Or suppose that one thought that in order to qualify as forgiving, one must forgive on the basis of certain right moral reasons. To “forgive” on the basis of other reasons would not be to forgive at all. These would count as highly moralized accounts of the nature of forgiveness. On the other hand, however, if one thought that one could engage in full-throated forgiveness that was somehow inappropriate, vicious, or morally wrong, then it is less clear that we could identify the paradigm cases of forgiveness by focusing on the morally ideal ones. We do not wish to enter into this debate here, and so we offer this fifth argument in favor of directed forgiveness with the above caveat. Briefly, though, here is a possible reason in support of the viability of this moral ideal strategy. Suppose one could “line up” a diversity of instances of forgiveness and have agents familiar with the practice choose which of the instances were the clearest cases of forgiveness. Although we have no evidence, we suspect that agents would be very likely to rank the morally ideal cases of forgiveness among the clearest of cases.

²⁰ Actually a number of different concerns about forgiveness have been raised under the banner of “paradox.” Aurel Kolnai, for example, put the worry this way: “Either the wrong is still flourishing, the offence still subsisting: then by “forgiving” you accept it and thus confirm it and make it worse; or the wrongdoer has suitably annulled and eliminated his offence, and then by harping on it further, you would set up a new evil and by “forgiving” you would only acknowledge the fact that you are no longer its victim. Briefly, forgiveness is either unjustified or pointless” (1973, pp. 98-99). We will formulate the objection slightly differently.

²¹ We should point out, however, that it is not our view that blaming requires that one intend to cause harm. Rather, it is just that blameworthiness makes the causing of certain harms morally permissible.

²² We recognize that there may be cases in which the wrong done is so serious or harmful that one might have an obligation to blame, and that any reasons to forgive fall short of providing an obligation to forgive, especially if one has not already engaged in directed blame (and perhaps punishment).

²³ We wish to thank Hannah Tierney for her helpful comments on a previous draft of this paper as well as the University of Arizona Center for the Philosophy of Freedom, which invited us to share some of these ideas in a talk we gave in October 2012. We are grateful for the feedback we received from a generous and insightful audience.

CHAPTER TEN

MORAL RESPONSIBILITY WITHOUT DESERT

DERK PEREBOOM

Our practice of holding people morally responsible features a number of different senses of ‘moral responsibility’. Essential to one especially prominent sense is the notion of desert: someone who has done wrong deserves to be blamed and perhaps punished just because he has done wrong, and someone who has performed a morally exemplary action deserves credit, praise, and perhaps rewards just because she has performed the morally exemplary action (Feinberg 1970). Another related sense invokes the reactive attitudes of indignation, moral resentment, and guilt (Strawson 1962). On one version, to hold someone blameworthy essentially involves regarding her as the appropriate target of such a reactive attitude. The appropriateness here arguably also involves the notion of desert. At the same time, a number of senses of ‘moral responsibility’ clearly do not involve the notion of desert. For example, in George Sher’s conception, to blame someone, and, by extension, to hold someone blameworthy for a bad action is to believe that it was bad and desire that the agent not have performed it (Sher 2006). Some senses are distinctively forward-looking, and thus do not involve the essentially backward-looking desert component. In Moritz Schlick’s (1936) view, the point of blame and holding blameworthy is to reduce the incidence of bad action.

Sometimes one such sense is proposed as a comprehensive account of moral responsibility or of its negative component, blameworthiness. As a general strategy I think this is implausible. Consider the desert-involving senses. It’s often the case that when parents blame children for bad behavior, only forward-looking senses are at play. And when people blame foreign dictators for rampant torture and murder, desert is invoked absent any forward-looking notion. At times when we blame people who have been dead for centuries, we’re doing little more than noting that they’ve acted badly and expressing disapproval. I think it’s fruitless to claim that