Philosophers' Imprint

volume 21, no. 6
may 2021

Can We Un-forgive?

Monique Wonderly

University of California, San Diego

In sara greenslit's award-winning novel *As If a Bird Flew by Me*, we find the following passage: "'You can't say I ever forgave them,' she'll hear herself say, not knowing anymore what it means to forgive. Can you take back forgiveness, nullify it? Was it ever forgiveness then, or what was it called instead?" (2011: 83) The unnamed narrator neither provides a context for the passage nor makes any attempt to answer its queries, but simply leaves the matter to drop.[1] Yet, as dropped matters go, this is one worth recovering. The questions at its core represent a rich moral psychological puzzle — a puzzle that, though relatively neglected in the philosophical literature, has implications for how we understand the nature of forgiveness.

In this paper, I address the puzzle posed by Greenslit's curious passage. I consider whether and in what sense forgiveness is rescindable, retractable, or otherwise reversible. In other words, I consider what it might mean to say that a victim who forgave her offender for a particular act of wrongdoing later *un-forgave* that individual for the very same act.

The paper proceeds as follows. In §1, I clarify and motivate the paper's animating question. In §2, I employ extant literature on the nature of forgiveness to introduce and elaborate two broad, overlapping conceptions of what we do when we forgive. In §3, using the preceding conceptions of forgiveness, I formulate two parallel conceptions of un-forgiving, and I examine some factors that bear on their plausibility. In §4, I offer what I take to be a case of un-forgiveness, and I consider some objections against this interpretation. In §5, I briefly explore some potential applications of this analysis for assessing theories of forgiveness. Finally, in §6, I offer concluding remarks.

1. The Puzzle

The question "Can we un-forgive?" admits of multiple interpretations. The questioner, for example, might mean to ask whether we can

1. I might just as well have described the matter as left to hang curiously between two (seemingly) unrelated passages. The author employs an unconventional writing style that often incorporates elements that are more reminiscent of poetry than standard fiction prose.

un-forgive permissibly — i.e., without violating some moral or rational requirement. My target, however, is a prior question, one to which inquiries about the permissibility of un-forgiveness presuppose an answer. I am concerned with whether and how we can make sense of forgiveness being reversed or undone by the forgiver.

Relatively few theorists have addressed this question directly, and we find little consensus among those who have done so. Though he takes no definitive stance on the issue, in his defense of "economic models" of forgiveness, Brandon Warmke suggests that it may be possible, and indeed justifiable, to retract forgiveness in certain circumstances (2014: 584–585). Geoffrey Scarre — who, to my knowledge, offers the only sustained philosophical treatment of withdrawing forgiveness — argues that typically, "forgiveness, once truly granted, cannot be taken back" (2016: 933).[2] According to H.J.N. Horsbrugh, while the process of forgiveness is subject to reversal, complete forgiveness is necessarily permanent (1974: 279). Finally, David Owens, who characterizes forgiveness as "irrevocable", posits that we cannot "unforgive someone" for what they've done (2012: 53).[3]

There does seem something strange about the idea that we can "unforgive". We find one of the earliest mentions of the term in a work by grammarian Evan Daniel, where in a chapter on parsing, he identifies "unforgive" as a verb construction that the English language does not accommodate (1891: 73). To be sure, not every act of ϕ-ing admits of a plausible notion of un-ϕ-ing. While some words should remain unspoken, we cannot unspeak them. I can unlock a door, unwrap a gift, or uninstall software from my computer. But I cannot un-dream a frightening dream or un-attend a vapid lecture. And importantly, this is not just a matter of linguistic style or etiquette, but one of intelligibility. When I report to a friend that I un-walked my dog, I am not merely being gauche, but incoherent. Doubtless, people have been un-friended on social media sites for lesser offenses.

On the other hand, while "to un-forgive" is something of a tortured term, we might acknowledge that — unlike, say, "to un-swim" — it is not obviously a meaningless one. Though relatively rare, this locution is not entirely absent from our vernacular. Consider the lyrics of "Unforgiven", a 2001 single by American rock band The Go-Go's:

> Once I forgave you, but I did not forget. Now I am taking back everything I said. You're unforgiven so go on living, knowing that I've unforgiven you. And my Thanksgiving came the day I saw it was okay to unforgive you. (The Go Go's 2001)

The use of the term here does not strike us as mere gibberish but seems to express a recognizable idea. And it is presumably the same idea that psychotherapist Susan Forward means to convey when she describes counseling a patient to "unforgive" her parents in order "to get in touch with her anger" (1989: 180). These observations suggest that we at least have some concept of what it might mean to withdraw or to reverse forgiveness.[4]

Where does this leave us? Neither the extant philosophical literature on forgiveness nor observations about our ordinary language use affords us an easy answer to this paper's animating question. Crucially, whether and how we can make sense of "un-forgiving" will depend on whether forgiveness is *the sort of thing* that admits of retraction or reversal and if so, on what is involved in enacting the relevant change. To this end, a brief excursion into the philosophical literature on the nature of forgiveness will be instructive.

2. I say "typically" because though Scarre sets out to defend the thesis that "not taking back one's forgiveness is a necessary condition for forgiving" (2016: 932–933), he later states, "…that forgiveness *can* be withdrawn in (and, I believe, only in) the special case where it has been granted in error" (ibid: 938, his emphasis).

3. In a forthcoming work on forgiveness and commitment, I assume but do not argue for the possibility of un-forgiveness (Wonderly forthcoming).

4. Notably, while "un-forgive" remains absent from most formal dictionaries, it does have its own entry in the English-language Wiktionary, where it is listed as a transitive verb meaning: "To revoke or rescind forgiveness of" (unforgive. *Wiktionary*).

2. To Forgive

Un-forgiving can only make sense where there is some earlier act of forgiveness to "undo". We can express the generic description of the earlier act as follows: Victim V forgave offender O for wrongdoing w. To understand how forgiveness can be undone, we must first consider what was done — i.e., what occurred between victim V, offender O, and wrongdoing w — when the victim first forgave. Though theorists often diverge on this issue, we can make progress by focusing on two broad formulations of the relevant change(s).

> (F1) V had a shift in affective attitude such that her resentment (and/or other negative emotions) toward O for w has been eliminated or substantially reduced.
>
> (F2) V altered O's relationship status or normative position — e.g., by reaccepting O into a moral relationship or by releasing O from (certain) obligations to V on account of w.

In their current state, these formulations are vague and incomplete, but we can develop them by attending to the philosophical literature on forgiveness.

Let's start with (F1), according to which forgiveness centrally involves a shift in affective attitude. Contemporary forgiveness theorists often take as their starting point a view that is typically attributed to Joseph Butler: namely, that to forgive is to overcome or to forswear resentment (Butler 1897).[5] On this view, the victim's affective change is not one that merely happens to her as in the cases of forgetting or of one's resentment naturally abating over time. The victim "overcomes" or "forswears" resentment, thus playing some active agential role in bringing about the relevant shift. Theorists sometimes try to capture the victim's agential role by emphasizing that she must enact the shift "voluntarily" and/or for a particular type of reason.[6]

Notice that insofar as forgiveness involves a shift in affective attitude, it is not voluntary in the sense that the victim can simply enact the relevant change at will. When confronted by an apologetic victim who asks for my forgiveness, I can hand her a tissue with which to dry her guilty tears, but I cannot eliminate my resentment on request.[7] More controversially, we might imagine a case in which forgiveness, while agential in the sense of being reasons-responsive, occurs *despite* the victim's efforts to prevent it. Consider, for example, an agent who feels empowered by her angry blame and wishes to retain it but is moved by her offender's acts of contrition to "let go" of her resentment despite her clenched fists. One might think that her resistance and subsequent regret notwithstanding, she genuinely forgave her offender.

Just as attending to the victim's agential role in forgiveness adds a layer of complexity to (F1), so, too, does attending to the nature of her affective change. Theorists tend to follow Butler (1897) in identifying the emotion that the forgiver overcomes as resentment, the reactive attitude paradigmatically felt by an agent in response to having been wronged. Some, however, posit that the forgiver must also overcome other negative attitudes, such as contempt or grief (Richards 1988; Bell 2008; Garrard & McNaughton 2011; Blustein 2014). And some theorists maintain that the negative emotions must not be merely overcome, but replaced with positive attitudes such as benevolence or good will (Roberts 1995; Pettigrove 2004; Martin 2010; Garrard & McNaughton 2011; Blustein 2014). This approach construes forgiveness

5. As Ernesto Garcia (2011) argues, this is a misattribution since Butler's actual claim is that forgiveness involves overcoming excessive resentment.

6. For example, Horsbrugh (1974), Haber (1991), Bell (2008), Owens (2012), Nelkin (2013), and Blustein (2014) all hold that forgiveness must be in some sense voluntary. For those who argue that forgiveness is done, or at least sometimes can be done, for a particular type of reason, see Murphy & Hampton (1988), Hieronymi (2001), Griswold (2007), and Milam (2019).

7. Of course, I can agree at will to (try to) give up or to repudiate my returned resentment, and some may regard this act as sufficient for forgiveness.

not merely as an absence of some emotion, but as having a more substantive character of its own.

Regardless of the particular emotions that the victim eliminates or acquires via her forgiveness, theorists often emphasize that forgiveness is not just a matter of how one feels. Proponents of (F1) sometimes acknowledge this point by emphasizing that forgiveness, like the blameful orientation that it supersedes, is an emotional *stance*. To forgive, or to resent, is to take up an affective mode of seeing or holding one's offender in a thick evaluative light (see, for example, Allais 2008: 52–58). So construed, forgiveness is an interpersonally rich engagement that is not aptly characterized merely as a shift in feelings. On Pamela Hieronymi's account, resentment is a judgment-sensitive emotion that functions as a form of protest against the relevant wrongdoing, and forgiveness can "ratify" the apologetic offender's change of heart — in some sense, changing the significance of her past action (2001: 550).

Some theorists describe the relevant change as a process, the first step of which is a decision to enact the attitudinal shift. This framework allows for the possibility that a victim who made the decision to enact the attitudinal shift or otherwise began the process of enacting the change — even without having completed it — can intelligibly report that she forgave. For my part, whether or not forgiveness involves a process of enacting an attitudinal shift and/or a decision to begin that process, I think we should report that a victim forgave only if she has *completed* the relevant process.[8]

Let's now turn to (F2). On this approach, forgiveness is primarily a matter of altering the offender's socio-normative status. For example, whereas the victim's blame might represent the offender as (partially or wholly) displaced from a moral relationship, forgiveness is sometimes said to involve *re-accepting* the offender into the relevant relationship (Kolnai 1974; Bennett 2003, 2018; Martin 2010). In forgiving, the victim restores the offender's status to a position of equal normative footing with the victim (Murphy & Hampton 1988; Bovens 2008, 2009).

Another way to understand the relevant alteration is in terms of the victim's exercise of a *normative power* whereby she alters the deontic space between herself and her offender (Warmke 2016; Bennett 2018).[9] Consider that on some views, to forgive is (in part) to relieve one's offender of certain compensatory obligations — e.g., her "debt" of apology and amends (Twambley 1976; Murphy & Hampton 1988; Bennett 2003; Nelkin 2013; Warmke 2014). Some theorists add that the victim undertakes an obligation (or commitment) of her own, an obligation to refrain from blaming her offender for the same wrongdoing in the future (Scarre 2016; Bennett 2018).[10]

An adequate version of (F2) should characterize the relevant alteration such that it is not beyond the victim's authority to enact. Insofar as the offender's action distances her from the moral community, the victim may lack the power to unilaterally restore the offender to her former position in the moral relationship. Similarly, it is not obvious which of the offender's obligations the victim has the authority to cancel. The offender may owe it not just to the victim, but to the moral community more broadly, to apologize and to engage in other amendatory activities. These obligations might be beyond the victim's purview to waive.

8. One might disagree, insisting (for example) that forgiveness is a process with no endpoint. This view of forgiveness strikes me as at odds with both the phenomenology of central cases of forgiveness and our typical employment of the term 'forgave'. However, as little hangs on this for present purposes, I will not pursue detailed discussion on the issue here.

9. In a similar vein, although David Owens doesn't identify forgiveness as a normative power, he argues that it nevertheless changes the normative situation between the victim and the offender (and also between other members of the moral community and the offender) by, for example, rendering some forms of blame for the forgiven wrongdoing inapt (2012: 53).

10. Scarre characterizes forgiveness as involving a promise to leave the wrongdoing in the past (2016: 936). On Glen Pettigrove's view, expressed forgiveness is "related" to promising, but the former involves commitments — including, *inter alia* a commitment to forswear and insofar as possible eliminate hostile reactive attitudes — that lack the binding force of promises and breaching them doesn't (without excuse) harm the offender (2004: 385–386).

Proponents of (F2) face the challenge of showing how emotions fit, if at all, into their view. Importantly, many will find dubious any view of forgiveness that cannot accommodate a role for emotions. Suppose, for example, I tell my friend that I forgive her, absolve her of any obligation to me on account of her wrongdoing, and treat her accordingly. Suppose further that a week later, an inauspicious wind blows open my diary to a page that details my unabated resentment toward her since her wrongdoing. My friend might understandably feel that I never truly forgave her. To those for whom this rings true, (F2) will likely seem to be missing something crucial.[11] Consequently, some advocates of (F2) allow for hybrid constructions on which, in addition to the (more fundamental) exercise of a normative power, some forms of forgiveness also involve a change in affective attitude (see, for example, Warmke 2016).[12]

Having examined and elaborated (F1) and (F2), we are now well-positioned to consider what it might mean to un-forgive.

3. To Un-forgive

Given the preceding formulations of forgiveness, we can construct two parallel formulations of un-forgiveness. On the relevant approaches, to say that V un-forgave O for w is to express one or more of the following statements:

11. Not all will balk at views of forgiveness that discount the centrality of emotions. Nicolas Cornell suggests that forgiving needn't involve an emotional shift (2017: 263). Brandon Warmke and Michael McKenna write: "One can… 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" (2013: 203).

12. In discussing the respective challenges that (F1) and (F2) advocates face, I do not to mean to urge that either type of view is untenable. My aim, rather, is to draw out what is at stake in the development of such views in order to illuminate the challenges facing parallel accounts of un-forgiveness. On a related note, I do not intend to argue that a hybrid model of forgiveness is needed — much less, to argue for a particular hybrid model. But I will go on to urge (in §5) that how we understand un-forgiving can inform the construction and assessment of forgiveness theories, including (F1) and (F2) hybrids.

> (U1) V had a shift in affective attitude such that she has come to re-inhabit (roughly) the same resentful emotions toward O for w that characterized her blaming attitude prior to forgiving O for w.
>
> (U2) V returned O's relationship status or normative position to its (or a similar) state prior to forgiveness — e.g., by again displacing O from the relevant relationship with V on account of w or by reinstating certain of O's obligations to V on account of w.

Let's begin with (U1). It is not difficult to imagine a victim whose attitude shifts from resentment toward her offender for a particular act of wrongdoing to a more positive attitude that is (at least largely) devoid of resentment and then back again to an attitude of resentment. One might not think it possible to return to the *same* attitude that one previously held. One's later resentment, for example, might have a different intensity, and at the very least, it would be colored by thoughts and feelings that diverge from those that occurred in the earlier case. This, though, should not concern those who think that we can un-forgive. The more worrying possibilities would be if later episodes of resentment were always best construed as either (1) continuations of one's earlier resentment (suggesting that the victim never *really* forgave) or (2) directed toward an act of wrongdoing other than the one for which the victim forgave. I consider both possibilities in the next section.

For (U1) to represent un-forgiving, it should, like the change represented in (F1), implicate the victim's agency in some robust sense. The attitudinal shift, if not *intentional*, should at least be responsive to the victim's reasons. Furthermore, and again like the change represented in (F1), the shift, rather than a mere shift of feeling, should involve a normatively thick evaluative adjustment in the victim's orientation toward her offender. The victim departs one evaluative mode of engagement and returns to an earlier one.

Now turn to (U2), according to which to forgive is to alter the socio-normative position of the offender. For example, if to forgive is to reaccept the offender into a moral relationship, then presumably to *un*-forgive would be to once more strip the offender of the relevant relationship status on account of the original wrongdoing. On this conception, just as the victim's initial blame disrupted (or reflected a disruption in) her offender's status as a moral peer on equal normative footing, so, too, does her un-forgiveness. Similarly, if to forgive is to waive certain of the offender's obligations or to undertake a new commitment or obligation of one's own to refrain from future blame, then to un-forgive would involve reinstating those obligations or abandoning the relevant commitment, respectively.

Familiarly from above, whether we find this notion of un-forgiving sensible will depend on whether we think the victim has the authority to alter the offender's socio-normative status in the relevant way. And here, the story is more complicated. It is one thing, for example, to waive another's obligation to you, and quite another to (try to) reimpose that obligation after having waived it. Think here of the creditor who cancels your debt — and presumably thereby undertakes an obligation not to demand its repayment — only to later insist that you again owe her the full amount. If there are conditions under which the creditor can successfully alter the norms in the way that she purports to do in this scenario, then one might argue that we can understand un-forgiving in similar terms.

Whether we accept (U1) and/or (U2) as plausible conceptions of un-forgiving, we should interpret them so as to avoid equating un-forgiving with merely "blaming again". Consider that a forgetful victim who ceases to blame an offender because she has forgotten the wrongdoing might later remember it and again blame her offender for it. This, though, is ill-characterized as an act of un-forgiving. Just as forgiveness must be done while retaining the judgment that the offender's act constituted a wrongdoing, so, too, can a victim un-forgive only if she does so in full recognition of her earlier forgiveness.[13] This makes a difference both for the victim who "reverses" her forgiveness and for the offender who is likely to feel not only confused (as she might in response to the merely forgetful victim) but betrayed or otherwise slighted. The point is that un-forgiving has a particular normative shape because it bears on the significance of one's earlier forgiveness. The victim doesn't just blame again, but she "takes back" her forgiveness. And it is this notion of reversal or un-doing that might render it difficult to make sense of a victim un-forgiving her offender.

To see this more clearly, consider that some view forgiveness principally as a speech act.[14] On this view, when the victim says to her offender (under the right circumstances), "I forgive you," she does not

13. There are some interesting questions concerning how best to distinguish the metaphysics of un-forgiving from the epistemology of un-forgiving. If un-forgiving is primarily an attitude change, one might think un-forgiveness can occur below the level of awareness. For example, I might not realize that I have un-forgiven right away but only "discover" later (or perhaps never) that I have done so. I don't want to rule out this possibility, but the phenomena that I'd like to pick out here aren't the sort of things that just "happen to us" in some way that is disconnected from our agency. On my view, forgiveness and un-forgiveness are reasons-responsive and distinct from simply forgetting (either about the relevant wrongdoing or one's earlier attitude toward it). Thanks to Hanna Pickard for helpful discussion on this matter.

14. As this approach would have it, forgiveness is accomplished when the victim says to her offender, "I forgive you" (or something of the sort), and thus performs a particular kind of illocutionary act. Illocutionary acts represent what we *do* in saying something (Austin 1962: 99). For example, when I say "I promise to call you," I not only utter some words, but I do a certain thing i.e., I make a promise. Promises represent what are called *commissives*, a type of illocutionary act that commits the speaker to doing something (Austin 1962: 150; Searle 1976: 11). Pettigrove suggests that the illocutionary act performed when a victim tells her offender, "I forgive you," has commissive force, committing the victim to *inter alia*, forswearing hostile attitudes and retaliation on account of the offender's wrong doing (2004: 385). Forgiveness theorists have identified two other candidates for the type of illocutionary act that one performs in forgiving: *declarations* (Searle 1976: 13) and what J.L. Austin (1962: 83) and John Searle (Searle 1976: 12) identify as *behabitives* and *expressives*, respectively. According to Warmke, forgiveness has declarative force, such that, when said in the proper context, "I forgive you," *makes it the case that* the norms of interaction between the victim and her offender are altered (2016: 698). Pettigrove, following Haber (1991), suggests that in addition to commissive force, communicated forgiveness might also have behabitive force,

merely utter some words, but she *forgives* — perhaps, for example, by adopting or expressing the attitude described in (F1) or by effecting the alteration described in (F2). Since the doing of the deed is contained in the utterance, it might seem difficult to imagine how the victim can undo it. She cannot, after all, *un-utter* her forgiveness-constituting words.

Yet, even when forgiveness is understood as a speech act, I think we can still make sense of un-forgiving. We can make headway here by thinking about how some speech acts can, in a familiar sense, be undone. Consider the speech act "I do" that (said in the right context) binds one to another in marriage. This act can be undone by ending the marriage in certain ways. One temptation is to think of this in terms of annulment, which, for legal purposes, makes it as though one was never married in the first place. I suspect, however, that the better nuptial analogy for un-forgiving is divorce. A divorced couple counts as actually having been married, and this fact may have practical significance that survives the dissolution of their marriage. Yet they are also now unmarried, and at least some aspects of their normative situation have shifted in light of this, taking on a similar character to the one that they stood in prior to marriage (e.g., they can now sever financial ties, marry other people, etc.). Similarly, we needn't think of un-forgiving as making it as though the original forgiveness never occurred but rather as "un-doing" (perhaps, just one of) its central task(s).

While some will find these ideas intuitive, others will insist that forgiveness does not admit of being "undone" in this respect. On some accounts, forgiveness is best construed as *permanent*, and what might initially look like cases of un-forgiving are better described in other terms. Given that a plausible account of un-forgiveness must overcome these obstacles, it will be useful to examine a test case.

insofar as one of its illocutionary points is to disclose (some aspect of) the victim's attitude to the offender (2004: 373).

4. A Candidate Case of Un-forgiveness

Consider the following case:

> *Jane and Mabel*. Jane and Mabel are sisters. Upon his death, their father bequeathed to Mabel a treasured family heirloom, an antique locket, and Jane destroyed it in a fit of jealousy. Mabel, who loved the locket, initially resented Jane for her infraction, but in response to Jane's apology and expressions of remorse, finally told her sister, "I forgive you." Months later, Mabel overhears Jane gleefully telling a friend that she no longer regrets smashing the locket and plans to destroy more of Mabel's "undeserved" possessions. Upon hearing this, Mabel's resentment over Jane's destruction of the locket returned. She tells Jane, "I know that I said that I forgave you for smashing my locket, but I take it back! You need to set things right!"

Here, it seems that Mabel's resentment toward Jane for destroying the locket returned and that she takes herself to be re-imposing amendatory burdens on Jane for destroying it. According to the conceptions of un-forgiving sketched in the previous section, it seems reasonable to think that Mabel un-forgave Jane for destroying the heirloom. But not all would agree. One might argue that cases in which we appear to un-forgive others actually represent something quite different. Let's take a look at how this challenge might go.

First, recall from the preceding section that on U1, or the affect-centered model of un-forgiveness, the returned blameful emotions must target the offender's initial transgression. One might think that Mabel's current blame is best construed as directed not toward Jane's previous wrongdoing, but toward her current remorselessness and shameless plotting. Since she blames Jane for something other than her initial transgression, it is wrong to say that Mabel un-forgives Jane for it. That matter remains forgiven — or so the challenge might go. Yet it seems to me that this challenge fails. In the scenario that we are

imagining, Mabel's attention is clearly redirected toward Jane's earlier act, she obviously seems to resent Jane for what she did back then, and it is *that* act for which she now demands amends. This is not to say that she doesn't also blame Jane for being remorseless or for her new plan, but it seems baseless to rule out the possibility that she resents her for her earlier act as well.

A proponent of this strategy might instead suggest that when Mabel inhabits the new blaming stance with respect to Jane's wrongdoing, she resents Jane and renews her obligation to make amends for the act *under a different description*, and therefore, it is infelicitous to characterize her as un-forgiving Jane for her initial act. H.J.N. Horsbrugh and Geoffrey Scarre each suggest that cases in which one appears to withdraw forgiveness are often better explained in these terms (Horsbrugh 1974: 279; Scarre 2016: 939). As this objection would have it, whereas Mabel initially blamed — and later forgave — Jane for her destructive act qua one-off transgression, she now blames her for the act qua part of a pattern or complex of wrongdoing. Since the objects diverge in this way, there is no un-forgiveness. Jane's original act, qua one-off transgression, remains forgiven.

This route of objection fares better than the first, as it does not arbitrarily rule out the possibility that Mabel now blames Jane for her earlier act of destruction. But still, it misses its mark. To be sure, Mabel might see Jane's act differently now than she did back then, but this is no indication that *what she blames her for* is relevantly different. To see this, consider that one often forgives because one has come to see the relevant act of wrongdoing differently. For example, in the grips of her initial resentment, a victim might initially see the relevant act as a possible symptom of a wicked character, but in response to the offender's contrition, she might begin to see it as an anomalous bit of carelessness and forgive the offender for the act on that basis.[15] But in order to make sense of her forgiveness, we must posit that the act for which she previously resented and the act for which she now forgives are one and the same. Similarly, it makes sense to say that however differently Mabel now regards Jane's earlier act, she holds *that* act of destruction — the one for which she previously forgave — against Jane.

In the preceding section, I raised the worry that un-forgiveness might be easily conflated with incomplete forgiveness — that is, a mere continuation, rather than return, of one's earlier resentment. Drawing on this concern, one might allow that Mabel now blames Jane for her earlier transgression, but still deny that her resentment represents un-forgiveness because her current attitude constitutes evidence that she never *really* forgave Jane in the first place. This approach assumes that in order to count as forgiveness proper, the relevant change must be permanent. One can easily find support for this view in the philosophical literature. For example, Horsbrugh states that it is part of the logic of the term 'forgiveness' that "to forgive is to forgive permanently" (1974: 279). Similarly, David Novitz suggests that forgiveness requires that the changes that constitute it "endure" (1998: 311). Drawing on such views, one might insist that Mabel's current resentment and demand for amends reveal her initial purported forgiveness to be false. And if there was no initial forgiveness, there can be no subsequent un-forgiving.

15. If we take the victim's new view to entail that she now sees the offender as less *blameworthy*, then we might think that what I have described as forgiveness is better construed as (at least partial) excuse. I suspect that it will be difficult to distinguish cases in which a victim's revised view of her offender's wrongdoing supports partially excusing the offender from those in which it supports forgiving the offender. An offender's remorse, for example, might change the victim's initial view of the wrongdoing — making it appear more detached from the offender. And a remorseful offender might well seem less blameworthy than an unremorseful one, but still on many views, the offender's remorse constitutes a reason for the victim to forgive. The larger point, though, is that the mere passage of time will often add layers of context to the victim's perception of the initial wrongdoing, in effect changing how she sees it, but we should take care not to interpret any and all such changes as incompatible with forgiveness. Likewise, as I have argued, we should avoid interpreting any revised view of the offender's act post-forgiveness as incompatible with blame for the same previously forgiven act. Thanks to Dana Nelkin for helpful discussion on this topic. For insightful work on how later actions can change the significance of past events, see Jones (2008). See also Hieronymi's remarks on how the meaning of wedding vows can change when one spouse leaves the other (2001: 547).

It is unclear, though, why we should accept this characterization of the *Jane and Mabel* case. Horsbrugh and Novitz support their respective views by appealing to cases in which it only *appears* that one has forgiven, but resentment — having been only "incompletely" removed or "temporarily occluded" — remains below the surface (Horsbrugh 1974: 279; Novitz 1998: 308). To be sure, sometimes what appears to be a reversion back to a previous attitude is better construed as a case in which one has really held (more or less) the same attitude all along. We do sometimes think we forgave, and on reflection, conclude that we never really did. And as both Horsbrugh and Novitz note, this is understandable, as experience teaches us that it can be both difficult to forgive and difficult to know when one has truly forgiven. Importantly, though, this is not always the case, and there seems to be no compelling reason to impose such an interpretation on Mabel's attitude here.

One might argue that forgiveness is best construed as permanent because its reversibility would undermine its central function of relationship repair (see, for example, Scarre 2016: 936). But just as it is not clear why what appears to be returned resentment is always better construed as a continuation of one's earlier resentment, neither is it clear why we must think that only permanent forgiveness can play a reparative role in relationships. Suppose, for example, that Mabel put forth great effort to overcome her blame toward Jane, doing her very best to view Jane's situation with sympathy and understanding and to restore their relationship. Suppose further that she eventually succeeded in replacing her hostile feelings toward Jane with positive feelings, and that this change endured for months (and would have continued but for recent events), fostering genuinely improved relations between the siblings. Under these circumstances, it is difficult to see why Mabel's current blame would render her previous forgiveness inauthentic or unsuccessful.

To see this more clearly, consider what are arguably forgiveness's sister attitudes, trust and love. They, too, are often difficult to achieve, and we are sometimes mistaken about when and whether we did in fact really come to trust or to love. Often, though, we are correct when we report that we once trusted an agent to perform some act but did not so trust her at a later time, or again, that we fell in and then, later, *out* of love with someone. It would be strange to insist that because I don't trust my friend to dog-sit for me now that I never really trusted her to do so before. Likewise, the fact that my cousin doesn't currently love her ex-husband needn't mean that she never really did. And notice that trust and love, without any guarantee of their continuation, can play restorative roles in damaged interpersonal relationships. Familiarly, just as we can intelligibly ask whether a companion continues to trust, or to love, us — where that trust or love played a healing role in the relationship — we can ask whether she forgives us *still*. And such an inquiry only makes sense if forgiveness is not necessarily permanent. Ideally, trust, love, and of course, forgiveness, would persist; yet things are not always ideal. But even when these attitudes are reversed or revoked, we nevertheless did once inhabit them, and that seems to me a truth we have a stake in preserving. There may be good reason, then, to allow Mabel to own her hard-won forgiveness, even after she un-forgives.

One might say that perhaps Mabel's forgiveness was *conditional on* Jane's continued, sincere remorse. On one interpretation of this claim, Mabel's purported forgiveness would have counted as genuine forgiveness only if Jane was, and continued to be, sincerely remorseful. And here again, Mabel cannot un-forgive if she never truly forgave. On this approach, it is not Mabel's current blame that invalidates her earlier purported forgiveness; rather, it is that one of its conditions for realization was not met. I find this suggestion problematic for the same reasons I adduced in answering the "permanence" objection above — namely, that this conception of forgiveness both conflicts with (many) ordinary experiences of having forgiven someone, and fails to cohere with how we tend to analyze similar attitudes. Oftentimes, even where we acknowledge that we forgave mistakenly or wrongly, we still understand ourselves as having genuinely forgiven. Just as I might regret, without being able to deny, having truly trusted (or praised, blamed, etc.) someone who, as it turned out, didn't really deserve it, so, too,

might I come to realize that I truly forgave someone under false pretenses. My response in this case wouldn't be "I never really forgave you," but rather, "I no longer forgive you."

This analysis brings to the fore an observation from the previous section: Un-forgiving is more like divorce than annulment. Un-forgiving needn't void one's earlier forgiveness, but might "undo it" in another sense – one that allows us to recognize the victim's initial forgiveness as valid and thereby appreciate that it may have a kind of normative significance that survives un-forgiveness. Furthermore, just as we might acknowledge that there are sometimes reasons to "undo" a marriage via divorce, we might understand forgiveness's *conditions* as picking out proper reasons to *un*-forgive. Mabel, for example, might have extended forgiveness with the understanding that Jane's later lack of remorse would constitute a reason to un-forgive her. This interpretation of Mabel's "conditional forgiveness" is friendly to my view, as I take it that we both forgive and un-forgive for reasons.

This line of thought affords us a response to Owens's objection to un-forgiving. Owens allows that a victim might blame an offender for an act of wrongdoing after having forgiven her for the very same one, but he denies that this would constitute un-forgiving. Forgiveness, on his picture, irrevocably renders the victim's resurgent blame inapt (2012: 53). According to this view, when Mabel blames again, she doesn't un-forgive Jane, but her forgiveness — assuming it was valid at the time it was extended — continues to negatively impact the normative status of her current blame. There is something attractive about the idea that Mabel lacks the power to simply reverse the change in the "normative landscape" effected by her earlier forgiveness. Recall from §2 and §3 that the plausibility of socio-normative status alteration views (F2 and U2) depends heavily on whether the victim has the authority to enact the relevant changes. Mabel should not be able to arbitrarily alter the appropriateness of her resentment or to reimpose previously removed burdens on another. But we needn't construe her blame in these terms.

To be sure, the fact that Mabel forgave Jane does make a difference for her current blame. Mabel's earlier forgiveness, for example, likely attenuates the reason-giving force of the initial wrongdoing, thereby reshaping the boundaries of what would constitute appropriate or justifiable blame. Owens's view nicely captures this insight. Still, Jane's current lack of remorse and intent to repeat her offense seem like good candidate reasons for Mabel's return to her previous blaming stance. Thus, Mabel's blame is an apt response to the reasons that she has. In this way, we might say that the current circumstances license her reversal of the normative changes effected by her earlier forgiveness. Forgiveness, then, needn't be irrevocable, and it seems reasonable to construe Mabel as un-forgiving Jane.

I have presented the case of *Jane and Mabel* as one in which the victim is plausibly construed as having un-forgiven her offender for an act of wrongdoing, and I have defended this interpretation against a variety of objections. This is not to deny that some cases that resemble un-forgiving are in fact better construed as cases in which the victim blames her offender for a different act, or again, cases in which the victim never truly forgave her offender in the first place. The preceding analysis suggests, however, that we needn't interpret all such cases in this way. At least sometimes, we can un-forgive our offenders.

5. Theoretical Applications

I have been concerned to offer an analysis and defense of the possibility of un-forgiving.[16] Since few theorists address un-forgiveness directly and there is disagreement about whether we can un-forgive among those who do address it, I take it that these tasks constitute important and necessary first steps toward further fruitful inquiry on

16. Notice that my claim that we can un-forgive is consistent with it being the case that in practice, we rarely do so, or again, that whenever we do un-forgive, we necessarily do so impermissibly. Similarly, my claim is consistent with the view that we should jettison the term 'un-forgive' and replace it with another that adequately captures the same concept that I describe above. I have sought neither to defend the permissibility of the practice, nor to insist on a particular label with which to identify it. My aim, rather, has been to show how we might best make sense of the practice itself.

un-forgiveness. Before closing, however, I want to say a bit more about why we should want to pursue such inquiry — that is, a bit more about why un-forgiving matters.

Once we acknowledge that we can un-forgive, we are better positioned to recognize and to address interesting puzzles regarding the nature and ethics of un-forgiveness. For example, if we do un-forgive, then we have a stake in knowing whether and when we do so *appropriately*. Some acts of un-forgiving may run afoul of moral and/or rational norms, and some attempts to un-forgive may misfire or fail to "come off". Identifying the justificatory and success conditions for un-forgiveness would help to illuminate this under-explored aspect of our interpersonal lives. I cannot take up this task here, but I will offer some brief remarks to show how exploring such puzzles might help us to adjudicate between different accounts of forgiveness.[17]

Let's start by considering the role of the offender in grounding justificatory conditions for un-forgiving. In the case of Jane and Mabel, Jane's lack of remorse and intent to re-offend likely help to justify Mabel's un-forgiveness. Yet we might wonder whether, and to what extent, new information about Jane was necessary to render Mabel justified in un-forgiving her. After all, on many accounts, one can justifiably forgive an offender without the offender having done anything to precipitate or warrant the victim's forgiveness.[18] If no change in the offender's attitude or behavior is necessary to justify forgiving her, then one might think that we can also justifiably un-forgive our offenders without any new interaction with, or new information about, them. Interestingly, socio-normative status alteration accounts of un-forgiving and their affect-centered counterparts may yield different verdicts here.

If forgiveness and its reversal are primarily matters of a shift in affective attitude, then it may not be clear why any behavioral or attitudinal changes on the part of the offender are necessary to warrant the victim's un-forgiveness. One might think, for example, that Jane's initial wrongdoing provides sufficient reason for Mabel to un-forgive, or again, that if further justification is required, those justifying reasons needn't reference Jane's actions or attitudes. Not all will hold this position, but it is not obviously an implausible one.

If, on the other hand, un-forgiving involves altering the offender's socio-normative status, then it might make sense to require weightier, or at least *different*, considerations to justify un-forgiveness. This is especially true for accounts on which forgiveness and its reversal alter the offender's obligations to the victim. Even if we grant that a victim can justifiably cancel the offender's debt to her without the offender's involvement, this might not hold true for cases in which the victim *reinstates* a debt. The latter is more costly to the offender and potentially stands to wrong her if done for no good reason. Had Jane, for example, done nothing (beyond her initial wrongdoing) to warrant it, it might be difficult for Mabel to justify altering her sister's normative position in this way.

It is also worth noting that absent some new attitudinal or behavioral change in the offender (or new information about her past attitudes and/or actions), it may not even be *possible* to un-forgive on status alteration accounts. Victims may not have the authority to unilaterally impose previously cancelled debts or obligations on their offenders. Insofar as victims have such a power, it may be that some feature of the offender's attitude or behavior is necessary to license it. If so, then attempts to un-forgive an offender without any subsequent interaction with, or new information about, her may "misfire" or fail to succeed. Try as the victim might, she may be unable to reimpose the previously cancelled obligations. On status alteration accounts, then, the offender's involvement may be necessary not just to un-forgive justifiably, but to un-forgive at all. Since the success of the victim's affective attitude change doesn't obviously depend upon any feature of the offender, it isn't clear that the same condition would apply to affect-centered accounts of forgiveness.

17. See Wonderly (forthcoming) for a more thorough treatment of the tension between un-forgiving and the commitments internal to forgiveness.

18. For examples of views that defend unconditional forgiveness, see Garrard and McNaughton (2011) and Holmgren (2012: esp. ch. 3).

If affect-centered and status alteration accounts do yield different verdicts regarding the role of the offender in grounding the justificatory and/or success conditions of un-forgiveness, then this difference may be helpful in adjudicating between the respective accounts. For example, those convinced that un-forgiveness is best viewed solely as the victim's prerogative may be less inclined to accept status alteration accounts of forgiveness, which seem to require a more substantive role for the offender. On the other hand, those who hold that some new or previously unrecognized fact about the offender is necessary to justify un-forgiving her (or perhaps to un-forgive her at all) might favor status alteration views, which seem better equipped to accommodate and to explain this requirement. In other words, we might assess accounts of forgiveness based (in part) on how plausible we find their corresponding implications for un-forgiveness.

Importantly, my aim here is not to argue for any particular account of forgiveness. Rather, in this section, I have attempted to show how one might employ the concept of un-forgiveness to begin constructing such an argument. Of course, the individual moves here need further development and defense, but I hope they suffice to show that un-forgiveness is a potentially useful addition to the forgiveness theorist's toolkit.

6. Conclusion

I have attempted to show how we might make sense of un-forgiving an offender for a particular act of wrongdoing. Though the extant philosophical literature offers no ready-made solutions to the puzzles posed by un-forgiveness, I demonstrated that we can use two broad, overlapping conceptions of forgiveness to construct and to critically examine parallel formulations of un-forgiving. I then animated the resulting formulations in a test case in which, as I argued, the victim un-forgives her offender. Finally, I urged that our understanding of whether and how we un-forgive has implications for how we conceive of forgiveness. If we do un-forgive, then this feature of our moral lives doubtless plays an important role in how we navigate our relationships with other agents. Un-forgiveness can also potentially serve as a useful tool for adjudicating between extant accounts of forgiveness and constructing hybrid accounts that yield a more plausible view of how we do (and ought to) respond to those who wrong us.[19]

References

Allais, L. (2008). Wiping the slate clean: The heart of forgiveness. *Philosophy & Public Affairs* 36(1): 33–68.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Bell, M. (2008). Forgiving someone for who they are (and not just what they've done). *Philosophy and Phenomenological Research* 77(3): 625–658.

Bennett, C. (2003). Personal and redemptive forgiveness. *European Journal of Philosophy* 11 (2): 127–144.

—. (2018). The alteration thesis: Forgiveness as a normative power. *Philosophy & Public Affairs* 46(2): 207–233.

Blustein, J. M. (2014). *Forgiveness and Remembrance: Remembering Wrongdoing in Personal and Public Life*. Oxford: Oxford University Press.

Bovens, L. (2008). Apologies. *Proceedings of the Aristotelian Society* 108(1): 219–239.

—. (2009). Must I be forgiven? *Analysis* 69(2): 227–233.

Butler, J. (1897). *The Works of Joseph Butler*: *Vol. 2, Fifteen Sermons* [1726], ed. W. E. Gladstone. Oxford: Clarendon Press.

19. Earlier versions of this paper were presented at the Princeton University Center for Human Values, the Center for Values and Social Policy at the University of Colorado at Boulder, the 2017 American Philosophical Association Pacific Division Meeting, the Zicklin Center Normative Business Ethics Workshop at the Wharton School at the University of Pennsylvania, UC Riverside, UC San Diego, and UCLA. I thank the audiences at these events, as well as Craig Agule, Lucy Allais, David Beglin, Mitch Berman, Gunnar Björnsson, Jeffrey Blustein, Rosalind Chaplin, John Doris, John Martin Fischer, Elizabeth Harman, Pamela Hieronymi, Agnieszka Jaworska, Karen Jones, Coleen Macnamara, Victoria McGeer, Dana Nelkin, Hanna Pickard, Eric Schwitzgebel, Amy Sepinwall, Julie Tannenbaum, Manuel Vargas, and two anonymous referees from *Philosophers' Imprint* for helpful comments and discussion.

Cornell, N. (2017). The possibility of preemptive forgiving. *The Philosophical Review* 126(2): 241–272.

Daniel, E. (1891). *The Grammar, History, and Derivation of the English Language, with Chapters on Parsing, Analysis of Sentences, and Prosody*. London: National Society's Depository.

Forward, S. (1989). *Toxic Parents: Overcoming Their Hurtful Legacy and Reclaiming Your Life*. New York: Bantam Books.

Garcia, E. V. (2011). Bishop Butler on forgiveness and resentment. *Philosophers' Imprint* 11(10): 1–19.

Garrard, E. & McNaughton, D. (2011). Conditional unconditional forgiveness. In Christel Fricke (ed.), *The Ethics of Forgiveness: A Collection of Essays*. New York: Routledge, 97–106.

The Go Go's. (2001). Unforgiven. *God Bless the Go-Go's*. Beyond.

Greenslit, S. (2011). *As If a Bird Flew by Me: A Novel*. Tuscaloosa: The University of Alabama Press.

Griswold, C. L. (2007). *Forgiveness: A Philosophical Exploration*. New York: Cambridge University Press.

Haber, J.G. (1991). *Forgiveness*. New York: Rowman & Littlefield.

Hieronymi, P. (2001). Articulating an uncompromising forgiveness. *Philosophy and Phenomenological Research* 62(3): 529–555.

Holmgren, M. R. (2012). *Forgiveness and Retribution: Responding to Wrongdoing*. New York: Cambridge University Press.

Horsbrugh, H. J. N. (1974). Forgiveness. *Canadian Journal of Philosophy* 4(2): 269–282.

Jones, K. (2008). How to change the past. In Kim Atkins and Catriona Mackenzie (eds.), *Practical Identity and Narrative Agency*. New York: Routledge, 269–288.

Kolnai, A. (1974). Forgiveness. *Proceedings of the Aristotelian Society* 74: 91–106.

Martin, A. M. (2010). Owning up and lowering down: The power of apology. *Journal of Philosophy* 107(10): 534–553.

Milam, P. (2019). Reasons to forgive. *Analysis* 79(2): 242–251.

Murphy, J. G. & Hampton, J. (1988). *Forgiveness and Mercy*. Cambridge University Press.

Nelkin, D. (2013). Freedom and forgiveness. In Ishtiyaque Haji and Justin Caouette (eds.), *Free Will and Moral Responsibility*. Newcastle: Cambridge Scholars Publishing, 165–188.

Novitz, D. (1998). Forgiveness and self-respect. *Philosophy and Phenomenological Research* 58(2): 299–315.

Owens, D. (2012). *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Pettigrove, G. (2004). The forgiveness we speak: The illocutionary force of forgiving. *The Southern Journal of Philosophy* 42(3): 371–392.

Richards, N. (1988). Forgiveness. *Ethics* 99(1): 77–97.

Roberts, R. C. (1995). Forgivingness. *American Philosophical Quarterly* 32(4): 289–306.

Scarre, G. (2016). On taking back forgiveness. *Ethical Theory and Moral Practice* 19(4): 931–944.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society* 5(1): 1–23.

Twambley, P. (1976). Mercy and forgiveness. *Analysis* 36(2): 84–90.

unforgive. (2019, March 14). *Wiktionary, The Free Dictionary*. Retrieved 01:30, August 26, 2019 from https://en.wiktionary.org/w/index.php?title=unforgive&oldid=51809942.

Warmke, B. (2014). The economic model of forgiveness. *Pacific Philosophical Quarterly* 97(4): 570–589.

_____. (2016). The normative significance of forgiveness. *Australasian Journal of Philosophy* 94(4): 687–703.

Warmke, B. & McKenna, M. (2013). Moral responsibility, forgiveness, and conversation. In Ishtiyaque Haji and Justin Caouette (eds.), *Free Will and Moral Responsibility*. Newcastle: Cambridge Scholars Publishing, 189–211.

Wonderly, M. (forthcoming). Forgiving, committing, and un-forgiving. *Philosophy and Phenomenological Research*.