Philosophical Studies (2024) 181:1167–1190
https://doi.org/10.1007/s11098-024-02138-2

In defense of genuine un-forgiving

Anna-Bella Sicilia[1]

Accepted: 15 March 2024 / Published online: 29 April 2024

Abstract
Despite much philosophical attention on forgiveness itself, the phenomenon of un-forgiving is relatively neglected. Some views of forgiveness commit us to denying that we can ever permissibly un-forgive. Some go so far as to say the concept of un-forgiving is incomprehensible—it is the nature of forgiveness to be permanent. Yet many apparent cases of un-forgiving strike us as both real and justified. In what follows, I will address the latter view, that genuine un-forgiving is impossible or incomprehensible as a phenomenon, advanced by a character I will call the "Un-Forgiving Denier." I address two views which purport to describe candidate un-forgiving cases in alternative ways and deny that any candidate un-forgiving cases are truly cases of un-forgiving: the "epistemic invalidation" and "new forgiveness opportunity" views. In creating problems for those views, I hope to defend the possibility of genuine un-forgiving. Even if it's possible to respond to the "Un-Forgiving Denier," a defender of genuine un-forgiving still faces the "Un-Forgiving Critic," who insists that un-forgiving, while possible, is morally indefensible. Against this view, I argue that un-forgiving enables an ideal of forgiveness wherein victims hold wrongdoers accountable for their moral development and allows certain opportunities for relational repair. I conclude that there is good reason to think un-forgiving is both possible and permissible. Embracing genuine un-forgiving puts constraints on how we should theorize about forgiveness itself and gives us an additional tool for understanding and navigating our relationships.

Keywords Ethics · Forgiveness · Un-forgiving · Moral development · Responsibility · Self-respect

> *Hotheaded Hannah* | In January, your close friend Hannah loses her temper when you accidentally spill her coffee. She causes quite a scene and you feel justifiably resentful. You leave the exchange feeling unsure that you can see your friendship with Hannah in the same light. But, after Hannah has calmed down, she apologizes profusely. "I know I have a problem with my temper,"

✉ Anna-Bella Sicilia
absicilia@arizona.edu

[1] Department of Philosophy, University of Arizona, Tucson, AZ, USA

> Hannah admits. "I had a frustrating day and took it out on you. I'm terribly sorry, and I promise to work on it. I know it will be difficult, but I'm really committed to working on my anger issues and becoming a more patient person." You take Hannah to be sincere, and you trust that she is indeed committed to changing. You forgive her for the outburst in the hope that she will learn to become more patient. Happily, your relationship feels more equitable again.
>
> However, this peace does not last long. In February Hannah loses it again, this time over your failure to offer her a ride to a friend's house. Though you are disappointed, you remain hopeful that she can change and forgive her once again. By the time Hannah has a *third* outburst in March, you have lost hope. Hannah has now *repeatedly* yelled at you over silly mistakes and things are clearly not getting better. You un-forgive Hannah for her outbursts.

What does it mean to say you un-forgive Hannah? Despite much philosophical attention on forgiveness itself, the phenomenon of un-forgiveness is relatively neglected.[1] Some views of forgiveness commit us to denying that we can ever permissibly un-forgive. They offer alternate explanations for why I may be angrier at Hannah after repeated outbursts; after all, on some views of forgiveness, you renounced the prerogative to blame her for these past misdeeds or forswore resentment. Some go so far as to say the concept of un-forgiving is incomprehensible—it is the nature of forgiveness to be permanent.[2] Yet the act of un-forgiving in *Hotheaded Hannah* strikes many of us as both real and justified.

In what follows, I address the latter view, that genuine un-forgiveness is impossible or incomprehensible as a phenomenon, advanced by a character I will call the "Un-Forgiving Denier." I address two views which purport to describe candidate un-forgiving cases in alternative ways and deny that any candidate un-forgiving cases are truly cases of un-forgiving: the "epistemic invalidation" and "new forgiveness opportunity" views. In creating problems for those views, I hope to defend the possibility of genuine un-forgiving.

Even if it's possible to respond to the Un-Forgiving Denier, a defender of genuine un-forgiving still faces the Un-Forgiving Critic, who insists that un-forgiving, while possible, is morally indefensible. Against this view, I argue that un-forgiving enables an ideal of forgiveness wherein victims hold wrongdoers accountable for their moral development and allows certain opportunities for relational repair. I conclude that there is good reason to think un-forgiving is both possible and permissible, a significant conclusion for at least two reasons. First, it has important implications for how we theorize about forgiveness itself; if un-forgiving is a real phenomenon and is sometimes permissible, any account of forgiveness unable to accommodate this must be rejected. Second, seeing that un-forgiving is (morally) on the table will

[1] Exceptions include Wonderly (2021a, 2021b), Scarre (2015), and Bash (2007, 2015). Warmke (2014) very briefly discusses un-forgiving on 585.

[2] Scarre (2015) offers such a view.

allow us to tackle various and complex questions about the ethics of un-forgiving, opening discussion of when and how we ought to un-forgive.

1 Defending un-forgiving against the Denier

1.1 The "epistemic invalidation" view

Geoffrey Scarre rejects genuine un-forgiving on the basis that apparent instances of un-forgiving are just the revelation that the so-called "forgiveness" offered was never genuine to begin with. He claims that these apparent instances of forgiveness are simply revealed to have been granted in error—either because of the wrongdoer's deception, the victim's hastiness in forgiving, or the revelation of other facts about the wrong (such as its harmful impact on other parties). On this interpretation, forgiveness can be *invalidated* when the forgiver realizes that the forgiveness was never, truly granted, but it cannot be granted, then later undone.[3] Permanence is essential to the nature of forgiveness, according to Scarre: "forgiveness, if truly granted, cannot be taken back."[4] David Owens expresses a similar view, writing that "forgiveness is irrevocable," and that retracting forgiveness must be distinguished from "asserting the invalidity of one's earlier act of forgiveness."[5] On this view, those who try to take back forgiveness fundamentally misunderstand the nature of forgiveness.

We might be tempted by this view because of the popular idea that forgiveness is a response to a wrongdoer meeting certain conditions at the time forgiveness is offered. Some forgiveness theorists defend the view that forgiveness can only have a positive moral status when offered for the right, wrongdoer-dependent reasons (sometimes called "wrongdoer-dependent conditionalism").[6] Stronger still, some theorists endorse the constitutive claim that "forgiveness" offered for incorrect wrongdoer-dependent reasons *cannot count as forgiveness at all*. On this "thick" conception of forgiveness, "the conditions for positive moral status for an act of forgiveness are built into the very constitutive conditions for forgiveness itself."[7] One committed to this view may want to deny the possibility of un-forgiving for the following reason: if victims sometimes offer "forgiveness" which does not count because it is later discovered that the wrongdoer had not satisfied

[3] Scarre (2015) Scarre also uses the language of "withdrawing" forgiveness in epistemic invalidation cases, which I will avoid for the sake of clarity.

[4] Scarre (2015), 933.

[5] Owens (2012), 53; footnote 13.

[6] Kolnai (1974), Lang (1994), Murphy and Hampton (1988), Novitz (1998), Richards (1988), Swinburne (1989), Wilson (1988), and Milam (2019) argue for wrongdoer-dependent conditions on forgiveness. Hieronymi (2001), Novitz (1998), Holmgren (1993), and others have pointed out that there may be victim-dependent conditions on forgiveness as well. I remain neutral on both wrongdoer- and victim-dependent conditionalism.

[7] Hughes and Warmke (2017).

wrongdoer-dependent conditions at the time, they do not *un*-forgive but rather learn that they had never forgiven in the first place.

Most versions of thick, wrongdoer-dependent conditionalism do not say something this strong. For instance, Per-Erik Milam defends a thick conception of forgiveness according to which various gestures—"apology, remorse, repentance, atonement, and making amends"—give us the right reasons to forgive insofar as they offer evidence that the wrongdoer has undergone the required "change of heart."[8] According to Milam,

> [T]hese actions support the would-be forgiver's belief that the offender is no longer willing to act as they did, that they have proper respect for our interests, that the threat implied by their action is withdrawn, or that their wrongdoing no longer reflects their commitments, cares, or practical identity.[9]

On his account, forgiveness can only take place if these conditions for moral development appear to have been met ("it is difficult to coherently imagine forgiving for no reason"[10]). When they do not appear to have been met, we may relinquish blame—perhaps we "let it go"—but we do not forgive.[11] On this plausible version of thick wrongdoer-dependent conditionalism, forgiveness can only be truly offered to wrongdoers who appear to meet certain conditions at the time of forgiving (in short, their attitudes, beliefs, and identity must appear to be "worthy" of forgiveness in some way). Thus, being forgivable is a matter of making oneself forgivable in the eye of the forgiver.

A wrongdoer's apparent change of heart is one kind of wrongdoer-dependent condition. But perhaps, on another plausible view, forgiveness takes place in response to a *real* change of heart, not merely an *apparent* one. This condition seems to be operating in the background of Scarre's insistence that forgiveness granted in error (because of a *merely apparent* change of heart) is no forgiveness at all.[12] On this stronger view, there is some fact of the matter about whether the wrongdoer has undergone the requisite change of heart, and the victim can (at least in principle) acquire sufficient evidence for this at the time of forgiving. If the victim turns out to be mistaken (perhaps because the wrongdoer's "commitment, cares, and practical identity" had not sufficiently shifted after all), later evidence of this simply allows the victim to correct a mistake by calling attention to the fact that her earlier forgiveness was never genuinely offered, and that both victim and wrongdoer had been proceeding on an incorrect understanding of what had transpired. A wrongdoer's

[8] Milam (2019, 247).

[9] Milam (2019, 247–248).

[10] Milam (2019, 243).

[11] Milam (2019, 246). See also Brunning and Milam (2022).

[12] Milam briefly discusses the possibility of "reversing" or "withdrawing" forgiveness but seems to have something more like epistemic invalidation in mind; see page 603 and footnote 18 in Milam (2022). If he indeed endorses the weaker view I've identified here, he can also allow for cases of genuine un-forgiving. Thank you to an anonymous reviewer for *Philosophical Studies* for encouraging me to clarify this and pointing me to this passage.

re-offense—indeed, the fact that the wrongdoer is capable of reoffending—may provide sufficient evidence that no change of heart took place, thereby revealing the invalidity of the prior act of apparent forgiveness. In other words: if it's not proper forgiveness, it's no forgiveness at all. And if it wasn't proper forgiveness, there's nothing to un-forgive.

Call this interpretation of apparent un-forgiving cases the "epistemic invalidation" view, as these cases involve the victim learning some piece of information which leads them to realize their apparent forgiveness was invalid. Scarre writes that these cases involve a "significant cognitive adjustment to [the victim's] appraisal of either the offender or the offence, believing that her original forgiveness was granted in error."[13] The epistemic invalidation view says the wrongdoer must exhibit certain features which either do or do not reflect the right sort of change at the time of forgiving to render forgiving possible. Victims must correctly assess the wrongdoer's attitudes, identity, and beliefs to truly forgive. (Indeed, Scarre points out that victims are obligated to make efforts at discerning the "offender's attitude," given the morally weighty nature of forgiveness.[14]) If they are wrong, the view says, their apparent act of forgiveness does not take, even when both parties are convinced forgiveness has occurred.[15]

Epistemic invalidation |

1. Forgiveness can only be truly offered to wrongdoers who meet certain conditions at $time_1$ (in short, their attitudes, beliefs, and identity must be "worthy" of forgiveness in some way).
2. Re-offense at $time_2$ always demonstrates that the wrongdoer had not sufficiently shifted her practical identity so as to be worthy at $time_1$.
3. Thus, all candidate un-forgiving cases are cases in which new evidence reveals that the initial act of apparent forgiveness was offered due to wrong or incomplete evidence about the wrongdoer's worthiness at $time_1$; thus, the initial act of apparent forgiveness was never real forgiveness.[16]

[13] Scarre (2015, 933).

[14] Scarre (2015, 932, footnote 1).

[15] Griswold (2007) sometimes wavers between the normative claim (that forgiveness offered for the wrong wrongdoer-dependent reasons lacks a positive moral status) and the descriptive, constitutive claim (that such forgiveness is not forgiveness at all). He writes that "forgiveness has not been given or received simply because one believes or feels that it has been […] regardless of the level of subjective conviction" (xv). In other words, we are sometimes mistaken about whether we have given or received forgiveness.

[16] This includes cases where the victim more carefully considers the wrong, realizing that he offered forgiveness too hastily. Scarre (2015) offers a case in which Miles forgives Jules's cruel practical joke, convinced in the moment that she intended nothing malicious. But upon more careful and cool-headed consideration, he reconsiders the "joke" and determines Jules was never worthy of forgiveness—her behavior was simply cruel (240). This is a case of epistemic invalidation because Miles's more thorough deliberations allowed him to understand the wrong more clearly and therefore realize that his assessment of Jules's attitudes, beliefs, and so on was mistaken. So, re-offense is not the only way to get better evidence about a wrongdoer's worthiness at $time_1$.

Contrast this with:

> *Genuine un-forgiving* | Some candidate un-forgiving cases involve genuine un-forgiving, where forgiveness was offered and is later revoked, reversed, or taken back.

I am happy to concede that some (perhaps even many) candidate un-forgiving cases are really cases of epistemic invalidation. But supposing that *all* candidate un-forgiving cases involve epistemic invalidation requires that all the relevant information about a wrongdoer's present state would deliver a verdict about whether they are worthy or unworthy of forgiveness, and that forgiveness of the unworthy is impossible.[17] In other words, we always and only forgive with an implied clause that says something like "I forgive… *under the assumption that you have such-and-such features demonstrating your worthiness*." This is far from obvious.

We can apply some pressure to both 1 and 2. Can we only coherently forgive when the wrongdoer is "worthy," and what constraints must we put on such worthiness? Is it true that a re-offense at $time_2$ would always reveal Hannah's lack of commitment at $time_1$? That depends on how we fill in some details of the case. Here is one way to do so:

> *Deception* | A few weeks after Hannah's most recent blow-up, you overhear her phone call. "No, I'm not *actually* going to anger management. I just told her that to get her off my back. I said a whole thing about how I have a 'real bad problem' and am going to change, blah blah. I should get an Oscar!" It turns out Hannah merely *seemed* sincere during her apology, but her commitments to changing were dishonest. Predictably, she continues to lose her temper repeatedly.

In this case, you were seriously misled about Hannah's attitude toward her outburst (and probably misled about her personality). It seems fair to say that your apparent forgiveness was offered on false pretenses, and so you never forgave her in the first place. I am happy to concede to a staunch Denier that *Deception* may be a case where the epistemic invalidation framework is applicable, given the extent to which Hannah misled you about her worthiness at $time_1$.[18] But consider another version of the case:

[17] Later I will discuss the view that says forgiveness of the "unworthy" is not impossible but is always impermissible.

[18] *Deception* shares many features with the case of James and Peter in Scarre (2015, 931–932). It is also similar to Wonderly's *Jane and Mabel* case (Wonderly 2021a, 7). Wonderly thinks *Deception*-type cases, in which "one forgives while operating under certain false assumptions about—or, again, an incomplete understanding of—the relevant wrongdoing," may be cases of genuine un-forgiving, though this epistemic condition is neither necessary nor sufficient for genuine un-forgiving (6). *Contra* Wonderly, I think it is plausible that we should reserve a distinction between *Deception*-type cases, in which forgiveness is plausibly invalidated, and genuine un-forgiving cases. Because Hannah seems to have exhibited an apparent (though not actual) change of heart at the time of forgiving in this case, theorists who think an *apparent* change of heart is what makes a wrongdoer forgivable may concur with Wonderly's assessment of the case.

Akrasia | After Hannah's January blowup, she feels genuinely remorseful about her behavior and tells herself "This is important, I've got to stick to it." She starts strong and has perfect attendance at anger management for the first few months. But Hannah has an akratic streak, and sometimes overestimates her ability to commit to things. Against her better judgment, she begins slipping, neglecting her daily breathing exercises, and eventually sliding back into old, harmful habits. It is after this slip that she yells at you a third time, prompting your un-forgiving.

Does the epistemic invalidation framework make sense of this version of the case? There are two things we can say here. First, it might lead us to deny or seriously revise 1; even if Hannah had not undergone a thorough and settled change of heart by $time_1$, that does not seem to render forgiveness *impossible*. Perhaps she was on the cusp of serious moral change but had not undergone it quite yet. After all, the cultivation of patience requires the development of a stable disposition, a process which takes time. Nevertheless, Hannah's state in January seems *compatible* with eventual moral development—it was not foolish to hope for her development at that time. If this is true, we should either reject 1 or amend it so that we are more permissive about the kinds of wrongdoers who are "worthy" of forgiveness.[19] Secondly, we might reject 2; that is, we might reject the idea that Hannah's re-offense at $time_2$ demonstrates that she had not sufficiently shifted her practical identity, beliefs, and so on at $time_1$ so as to be worthy of forgiveness. Her re-offense demonstrates that she struggles with maintaining her commitments, as many of us do, but this doesn't seem to render your forgiveness hollow in the way the outright lies in *Deception* do. Thus, if we want to retract forgiveness later, it will not involve denying the reality of the initial act of forgiveness.

Lastly, consider:

Self-deception | After Hannah's January blowup, she feels genuinely remorseful about her behavior and engages in some deep reflection. There's a long history of hot tempers in her family, and she recalls her father's lifelong battle with his temper. Despite this, Hannah believes she is different from her parents, and is convinced it won't be as difficult to overcome her anger as it was for her father – she thinks, for instance, that with some serious effort, it won't be something she struggles with for her whole life. She might even think that the few outbursts she's had recently were flukes. Unfortunately, Hannah is a bit self-deceived. Her anger issues are more deep-seated than she realizes, and her lack of self-awareness about the seriousness of her issue means that her hard

[19] Of course, most theorists do reject 1. It is much more common to argue that wrongdoer-dependent conditions affect the moral status of forgiveness rather than its conceptual coherence. Furthermore, others reject any wrongdoer-dependent conditions for the moral status of forgiveness, arguing instead of "unconditional forgiveness"; see Garrard and McNaughton (2003).

work in trying to overcome the issue isn't enough to stave off future outbursts. One such future outburst prompts you to un-forgive her.[20]

Again, I don't think it's clear that Hannah was unworthy of forgiveness in *Self-deception*. Recall Scarre's belief that cases of apparent un-forgiving involve a "significant cognitive adjustment to [the victim's] appraisal of either the offender or the offence, believing that her original forgiveness was granted in error."[21] There is no cognitive adjustment necessary following *Akrasia* or *Self-deception,* as they are both compatible with your hopeful assessment of Hannah's ability to improve during your initial act of forgiveness. The trajectory of her moral stagnation is now complete, revealing a less sunny outcome than you'd hoped for. Nonetheless, she was plausibly worthy of forgiveness at the time, if not ultimately capable of deep, lasting change. Furthermore, her re-offense need not lead you to conclude that you were *wrong* about her cares, commitments, and practical identity at $time_1$. Indeed, I think it is possible to possess a constellation of features compatible with a trajectory of moral improvement, stagnation, and degradation, such that forgiveness may be appropriate in a state of uncertainty about one's future development. If this is true, one's eventual failure to morally develop does not always reveal the prior forgiveness to be hollow.

Proving that *Akrasia* and *Self-deception* are not epistemic invalidation cases requires a richer picture of Hannah's features—her beliefs, commitments, cares, and identity—at the time of forgiveness. I hope to demonstrate that people like Hannah possess traits compatible with a trajectory of moral improvement (as well as backsliding) at the time of the act of forgiveness. To do so, I appeal to Andrea Westlund's suggestion that many features which constitute a wrongdoer's journey of moral development, stagnation, or degradation—her "attempts at moral renewal"—are what Karen Jones calls "trajectory-dependent properties."[22]

According to Jones, trajectory-dependent properties display an "openness to future contingency." Because trajectories are not "wholly present at a time," whether or not some trajectory-dependent property obtains depends on temporally-extended truthmakers; whether or not we can ascribe some trajectory-dependent property depends on what happens "else*when*".[23] Jones cites "journeys" (including intellectual and political journeys) and "quests" as paradigmatic trajectory-dependent properties.[24] As such, we might think of the trajectory relevant for our purposes

[20] We might wonder if this is a case in which Hannah is incapable of controlling her anger due to, for instance, a behavioral disorder. If this is so, *excusing* her rather than *forgiving* her might be the appropriate response. Furthermore, un-forgiving her might be inappropriate insofar as she lacks the control required to keep up her promise. I am leaving aside difficult questions about the degree or kind of control required for responsibility here, but readers should assume for the sake of clarity that this is a case in which Hannah has the sort of control required to be an appropriate target of forgiveness. Thank you to Danielle Guzman and Michael McKenna for help in thinking through the nuances of this sort of case.

[21] Scarre (2015, 933).

[22] Jones (2007); see Westlund (2009, 519–520).

[23] Jones (2007, 271–272).

[24] Jones (2007, 270, 273).

as a "journey of moral development." Whether or not one's current commitments, beliefs, cares, and practical identity constitute part of a journey of moral development depends on what happens else*when*—it depends on whether they are able to cultivate good habits, exhibit new behaviors, and begin to experience certain morally-appropriate emotional responses. It may take a rather long time for one's features to stabilize such that we may describe them as changed; nonetheless, her early efforts at compassion, honesty, patience, and so on can be said to partially constitute her journey of moral development *if* this stabilization comes to pass. Her journey begins as soon as she begins to orient herself toward moral improvement, perhaps even long before she can manifest her commitment to improvement in truly virtuous action.

Notice that a failed or attempted journey of moral development, one which ends in tragic backsliding and bitterness, may begin with the same orientation toward moral improvement. Someone may possess the same constellation of beliefs, commitments, cares, and so on, and these features may either represent a journey of moral improvement or stagnation depending on their temporally-extended truthmakers. If this is so, it may sometimes be difficult or impossible to tell what sort of journey one is on, given their present features only; each of these journeys begins with a commitment to change. Scarre's suggestion that forgiveness "can be made conditional on the fulfilment of certain conditions by the offender" is complicated by the fact that those conditions cannot always be fixed by her features at some particular time slice.[25]

This gives us good reason to reject the first premise of the *epistemic invalidation* argument above, and to cautiously avoid theorizing about forgiveness in ways that presuppose its truth. We must be able to forgive wrongdoers taking their first steps on the journey of moral development, wrongdoers whose features at $time_1$ are merely *compatible* with (yet unrealized) moral improvement. A reasonable conception of forgiveness should allow that forgiveness can be genuinely offered in such cases without counting as something else, like the mere relinquishment of blame.

Of course, we can sometimes be quite sure that one's promises to change are empty, making it impossible to forgive on a view that requires any kind of wrongdoer-dependent conditions. On the other hand, sometimes we can be quite confident that someone will change in the intended way—perhaps each time they've wronged us in the past, they've taken their commitment to changing seriously and found success. But we are often forced to navigate our relationships—and our offers of forgiveness—in a state of uncertainty about the future. As Westlund puts it, "underlying realities may often legitimately be taken to support *either* forgiving or not forgiving"; we must determine how to proceed despite lacking conclusive evidence.[26]

[25] Scarre (2015, 938).

[26] Westlund (2009, 521). As such, Westlund's view of forgiveness involves the restoration of goodwill in the wrongdoer which sometimes requires a "leap of faith." Aurel Kolnai (1974) has also pointed out that we often forgive before a wrongdoer has "undergone an obvious and credible change of heart." According to him, one of the most important considerations for a victim is whether the wrongdoer "is engaged in an 'upward' movement or struggle or is on the contrary gliding down the slope" (100).

There may be *moral* problems with forgiving in a state of uncertainty about a wrongdoer's trajectory of moral development—perhaps, for instance, this involves a problematic sacrifice of self-respect. Perhaps there is something morally important about waiting until a wrongdoer has manifested their journey of moral improvement in consistent action. Nonetheless, we can answer the Denier: it seems *possible* to do so and still call your act one of genuine forgiveness, and the act can even meet some reasonable, less demanding wrongdoer-dependent conditions for forgiveness. This means that some candidate un-forgiveness cases cannot involve invalidation due to wrong or incomplete evidence. I agree with Monique Wonderly that "Un-forgiving is more like divorce than annulment."[27] Not all candidate un-forgiving cases are cases of epistemic invalidation.

1.2 The "new forgiveness opportunity" view

The Denier may still insist that candidate un-forgiving cases are best explained another way. She may insist that we mistake a refusal to forgive Hannah for her March outburst with un-forgiving her January one, and that all candidate un-forgiving cases are similarly misunderstood. This view, unlike the epistemic invalidation view, denies that a candidate act of un-forgiving makes essential reference to the initial wrong. Scarre insists that some apparent cases of un-forgiving are really what I am calling "new forgiveness opportunity" cases: "what may initially appear to be a withdrawal of forgiveness for an offence turns out to be something different, namely the advent of an unforgiving stance towards a different offence (or, more precisely, the offence differently contextualised)."[28] For instance, "When Sue's husband Paul is unfaithful to her for a second time, the painful memories this revives do not amount to a withdrawal of her forgiveness for his first offence. That episode was dealt with when Sue, seeing Paul's genuine contrition, put away her indignation and determined to act towards him with her former good will."[29] Here is a strong version of this view, contrasted again with the genuine un-forgiving view:

> *New forgiveness opportunity* | All candidate cases of un-forgiving are just cases in which the victim is presented with an opportunity to forgive the wrongdoer for a *distinct* wrong and chooses not to forgive a second (or third, etc.) time. Thus, the victim does not undo the initial act of forgiveness, but rather declines to forgive for the new offense.
> *Genuine un-forgiving* | Some candidate un-forgiving cases involve genuine un-forgiving, where forgiveness was offered and is later revoked, reversed, or taken back.

Defenders of the new forgiveness opportunity view may want to, more precisely, constrain opportunities for new forgiveness to those cases which do *not* meet the

[27] Wonderly (2021a, 2021b, 10).
[28] Scarre (2015, 938).
[29] Scarre (2015, 942).

conditions of epistemic invalidation. In other words, these two alternative explanations of candidate un-forgiving cases can be combined, presenting the defender of genuine un-forgiving with the following dilemma:

1. Forgiveness can only be truly offered to wrongdoers who meet certain conditions at $time_1$ (their attitudes, beliefs, and identity must be "worthy" of forgiveness in some way).
2. Re-offense at $time_2$ *either* (a) demonstrates that the wrongdoer had not sufficiently shifted her practical identity so as to be worthy of forgiveness at $time_1$ *or* (b) does not demonstrate that the wrongdoer had not sufficiently shifted her practical identity so as to be worthy of forgiveness at $time_1$.
3. If (a), then the initial act of apparent forgiveness was never real forgiveness and cannot be revoked, since it never existed in the first place.
4. If (b), then the re-offense at $time_2$ constitutes a *distinct* wrong that provides the victim with a new opportunity to forgive and has no bearing on the initial act of forgiveness at $time_1$.
5. Thus, re-offenses cannot serve as opportunities to un-forgive.[30]

We saw from the arguments in Sect. 1.1 that premise 2 is true, thereby denying the strong version of epistemic invalidation. But this dilemma makes the burden of proof for the defender of un-forgiving clear: we must show that cases of (b)—which are *not* explained by epistemic invalidation—are also not all explained by new forgiveness opportunities. If some of these cases are genuine cases of un-forgiving the Denier's view must be rejected.

Wonderly discusses the "new forgiveness opportunity" view, finding it implausible insofar as the victim sometimes just *does* direct her attention towards, and demand amends for, the initial act rather than the re-offense in some candidate un-forgiving cases.[31] In other words, she insists that paying attention to our emotional lives reveals that candidate un-forgiving cases make essential reference to the initial wrong. The revival of Sue's painful memories will, in some cases, be accompanied by the revival of negative emotions associated specifically with Paul's initial unfaithfulness.[32]

I agree with Wonderly that it is far more psychologically plausible that we redirect negative emotions toward some wrongdoers for their initial wrong, in addition to subsequent wrongs, than for *just* the subsequent wrong(s). This is powerful, as the "standard view" of forgiveness aligns forgiveness with the overcoming of negative emotions associated with the wrong (paradigmatically resentment,[33] but sometimes more expansive negative emotions including anger, contempt, indifference,

[30] Thanks much to Hannah Tierney for suggesting this dilemma on behalf of the Denier.

[31] See Wonderly's case of Jane and Mabel on page 7–8 of Wonderly (2021a, 2021b). (Nor, she argues, is it plausible to argue that the victim fails to un-forgive because she resents the wrongdoer for the same act *under a different description,* what Scarre calls "the wrong different contextualized.").

[32] This is so whether such a revival reflects a moral failing or not; recall that we are still concerned only with the *possibility* of un-forgiving, not its permissibility.

[33] See Murphy and Hampton (1988).

and sadness[34]). If those negative emotions are redirected toward the initial wrong, it is difficult for subscribers to the standard view to deny that un-forgiving has taken place (permissibly or not).[35] This may be especially clear in cases where the initial offense is quite severe, and the subsequent offense is relatively mild. If rather powerful negative emotions arise after the mild offense, and if we think this is fitting, this calls out for an explanation. The fact that we are, in fact, un-forgiving the initial, serious offense provides such an explanation.[36] For instance, some serious instance of dishonesty—an act of infidelity, say—may prompt strong feelings of resentment and anger. The fact that a later, rather minor instance of dishonesty ("You said you had already folded the laundry!") prompts a resurgence of these *strong* feelings is well explained by an act of un-forgiveness of the first offense.

I think there is yet another powerful reason to doubt the Denier can defend the view that all candidate un-forgiving cases are new forgiveness opportunities: these cases also seem to involve the withdrawal of *positive* emotions which were offered *as part of the initial act of forgiveness*. Many theorists of forgiveness have pointed out that forgiveness involves more than the overcoming of negative emotions; it involves the extension of positive attitudes like goodwill, solidarity, love, trust, or faith in the wrongdoer.[37] In un-forgiving Hannah, I retract any such positive attitudes. And because they were extended specifically in light of my initial act of forgiveness, it would be rather odd to suggest that their retraction does not make essential reference to this first act. Imagine, for instance, that my first act of forgiveness involved adopting trust in Hannah to work on her anger. Is it more accurate to say I found a *new* reason not to trust Hannah, or that I *lost* my ability to trust in her—that she "broke my trust," as is sometimes said? If the latter provides a better explanation of at least some cases, we have another good reason to doubt the new forgiveness opportunity view. And since this is true of cases that are not best explained by epistemic invalidation (in *Akrasia* or *Self-deception,* for instance), we can avoid the dilemma posed by the Denier.[38]

[34] Richards (1988). Murphy (2003, 59). See Bell (2008) for a defense of forgiveness as overcoming contempt and not merely resentment.

[35] This is not to be confused with forgiving as the *foreswearing* of negative emotions, sometimes invoked as part of the "standard view," which involves a normative claim.

[36] Thank you to Michael McKenna for pointing out the significance of these asymmetric offenses.

[37] See Kolnai (1974), Westlund (2009), Dillon (2001), Stump (2018), and Garrard and McNaughton (2003, 2010).

[38] There may be other plausible alternative explanations for candidate un-forgiving cases. For instance, we might think these cases involve forgiving a *person* rather than a particular *act,* a distinction that might bear on either the possibility or permissibility of un-forgiving. (Thanks to Maggie Shea for this suggestion.) For the sake of space, I do not entertain all possible alternative explanations here.

2 Defending un-forgiving against the Critic

I have argued that not all candidate un-forgiving cases are simply epistemic invalidation cases, nor are they new forgiveness opportunity cases. These arguments make space for the possibility of genuine un-forgiving, but more is required to prove its permissibility: perhaps we do, in fact, un-forgive but there is always something morally problematic about doing so. After all, on many attractive views of forgiveness, we renounce the prerogative to resent or blame the wrongdoer for their initial wrong after forgiving her; perhaps we alter the normative landscape such that these attitudes are henceforth impermissible.[39] Insofar as un-forgiveness involves a resurgence of blame or resentment, this view says we cannot undertake it without doing something morally wrong. Call the defender of the claim that un-forgiving is never morally permissible an Un-Forgiving Critic (or "Critic" for short). I will cast some doubt on the attractiveness of the Critic's view by demonstrating how it fails to permit certain goals and ideals of forgiveness. This should lead us to consider un-forgiving not only permissible, but sometimes morally recommended.

2.1 Faith, un-forgiving, and moral development

The Critic's position may fall short of capturing the moral dimensions of our forgiveness practices by excluding a certain ideal of forgiveness, one in which forgiveness enables and encourages the moral development of the wrongdoer.[40] Westlund's conception of "fideistic forgiveness," wherein the victim exhibits a "readiness to interpret the offender as worthy of one's restored goodwill even though the underlying realities are indeterminate," serves as my inspiration here, though I have a more specific subset of cases in mind.[41] In some (though certainly not all) cases of fideistic forgiveness, the act of forgiving is offered *specifically in light of the wrongdoer's commitments to improve* and enables this improvement.

To see this, imagine that my act of forgiving Hannah is undertaken with uncertainty about her improvement. She promises to work on her temper, and I adopt a hopeful interpretation of her potential, one which both acknowledges her firm intention to become a better person and provides it with social uptake. I might forgive her by saying something like, "I forgive you, but you have to really work on this," or "I forgive you, but you need to promise not to do it again." The fact that my

[39] See Warmke (2016) for an articulation of this view. See Bennett (2018, 226–227) for a brief reply to the suggestion that a view on which forgiveness alters the normative landscape cannot accommodate un-forgiving. See also Wonderly (2021a, 2021b, 13), which I discuss in Sect. 2.2.

[40] I discuss one ideal of forgiveness made possible by un-forgiving here, but I do not mean to suggest that this is the *only* form of un-forgiving with a positive moral status. I do not discuss, for instance, whether we can permissibly un-forgive for reasons other than a wrongdoer re-offending in a similar way. Nor do I discuss the moral status of forms of un-forgiving that result from changes in the *victim.* There may be cases in which victims undergo transformations that alter their relationship to the previously-forgiven wrong, and some such cases may be good candidates for permissible un-forgiving. Thanks to Lel Jones and Hannah Tierney for raising this sort of case.

[41] Westlund (2009, 523).

forgiveness is offered "with conditions" so to speak *facilitates* and *encourages* her efforts.[42] Notice: the knowledge that moral backsliding may lead me to justifiably un-forgive her provides Hannah with the accountability she needs to improve—a kind of social infrastructure. My forgiveness, offered *with the explicit condition that it will be revoked with serious re-offense,* communicates to Hannah that I expect better of her. In doing so, she receives the necessary encouragement to become the person that I see her as capable of becoming. I will refer to this sort of forgiveness, which is offered in this specifically conditional way that leaves open the possibility of future un-forgiving, as "faithful forgiveness."

One might wonder whether there is a real difference between faithfully forgiving (in a way that allows for un-forgiving) and merely exhibiting an *openness to forgiving*.[43] After all, a victim faithfully forgiving and a victim being open to forgiving both leave the wrongdoer with the unfinished business of moral improvement. We might therefore wonder what faithful forgiveness offers a wrongdoer which is not already offered by the openness of their victim. To explain this difference, we can return to the idea that the act of forgiving changes the normative landscape, altering the norms governing interactions between victim and wrongdoer. Whether or not we embrace the norm alteration view of forgiveness (which says these alterations are *constitutive* of forgiveness), it is true that forgiving always or typically has normative effects. Faithful forgiveness involves adopting new norms so long as—and *as* long as—the wrongdoer lives up to their goals of moral improvement. Being open to forgiving, on the other hand, has no normative effects. Analogously, being open to marrying someone is not the same as actually marrying them.[44]

Another important difference between faithful forgiveness and exhibiting an openness to forgiving is that the latter likely involves absolute discretion about when and how to resume blaming the wrongdoer. On the other hand, if I have faithfully forgiven you, the status of being forgiven (and its attendant norms) is assured to the wrongdoer *unless* they fail to live up to specified expectations of moral improvement. Because forgiveness is assured so long as they live up to these expectations, faithful forgiveness is likely to offer a stronger incentive to the wrongdoer to morally improve than mere openness to forgiving.[45]

For a real life case, consider the story of Katie Kitchen's efforts to secure the release of Joseff Deon White, convicted of murdering her father in a robbery, from prison.[46] Kitchen became motivated to reach out to White after wondering "if he

[42] By "with conditions," I do not mean to invoke the kind of wrongdoer-dependent conditionalism mentioned in Sect. 1.1; I just mean a commonsense understanding of conditions: "You're forgiven if/as long as you never do it again.".

[43] Thanks to an anonymous reviewer for encouraging me to discuss this, and to Andrew Lichter for conversations about this distinction.

[44] To carry the analogy further, some marriages end in divorce. But (holding aside, perhaps, sacramental marriage) we don't think the possibility of divorce undermines the legitimacy of the altered norms during the period of marriage.

[45] See related remarks in Sect. 2.2 about the distinctive moral and narrative contours of cases in which a victim un-forgives, and cases in which a victim simply declines to forgive for a new offense.

[46] The following is summarized from a *New Yorker* article entitled "A Daughter's Quest to Free Her Father's Killer" (Orbey 2022).

had become a good person," eventually leading her to establish a relationship with him and advocating successfully for his release.[47] Kitchen is described as consistently viewing White in a sympathetic and positive light: she writes in a pretend letter to him, "I have to believe that you never set out that night to hurt anyone." White describes their first meeting as "all love,"; he and Kitchen spent over four hours learning about one another, after which White completed the rehabilitation program that enabled his release. After release, White describes feeling an ongoing sense of obligation to be the sort of person Kitchen believed him to be. Describing her as "a central part of his support system," White tells Kitchen, "You're on the list of people I look at and say, 'I refuse to let you down.'"[48] Indeed, we are led to believe from the story that White's turning over a new leaf—holding a steady job and avoiding criminal activities, in addition to maintaining an attitude of gratitude for his early release—is attributable at least in part to Kitchen's willingness to forgive him. My suggestion is that the moral improvement of wrongdoers like White may depend, importantly, on the awareness that forgiveness is offered *with the expectation of improvement,* and that such forgiveness may be appropriately revoked.[49] Recall White's commitment to not letting Kitchen down; could this really do the work of holding White appropriately accountable if he thought Kitchen's patience was boundless, and if he felt no real risk of being un-forgiven? I suspect not.

Some might still deny that faithful forgiveness can ever be permissible in these cases, let alone part of a moral ideal. Those who insist on wrongdoer-dependent conditions on forgiveness may argue that, insofar as faithful forgiveness makes any wrongdoer-dependent conditions on forgiveness relatively weak (after all, we can forgive when it is *indeterminate* that the wrongdoer has undergone any change of heart), it involves a problematic sacrifice of self-respect or condones bad behavior. Objectors might also protest that faithful forgiveness is invasive and (problematically) paternalistic. Why would anybody seek faithful forgiveness if it invites a forgiving victim to constantly observe you, paying attention to your efforts at moral improvement and assessing their success? Lastly, one might also characterize faithful forgiveness as undesirable because risky. Because it threatens us with the possibility of un-forgiveness, faithful forgiveness is too fragile. Other forms of

[47] Orbey (2022).

[48] Orbey (2022). I have presented a simplified version of this case here, leaving out details that may complicate the moral status of Kitchen's actual forgiveness. For instance, there seems to be lingering uncertainty about whether White himself killed Kitchen's father or merely accompanied the killer. Also, Orbey sometimes casts the willingness of Kitchen (who is white) to forgive as related to a broader (perhaps racialized) condescension toward White (who is Black) ("In her narrative, the murder was a terrible accident, and White, because of systemic injustices, had been as much a victim as her father"; Kitchen herself concedes that this might be "disrespectful"). I believe what I have said about the case, or cases like it, holds true nonetheless. Thanks to Andrew Lichter for sharing and discussing this case with me.

[49] See Basu (2023) for an interesting discussion of the ethics of expectations. She points out that expectations can play a role in shaping or influencing our behavior, and rightly claims that such influencing (when it conflicts with our own self-understanding) can be deeply harmful. I am interested in cases in which the "expecter" and "expectee" share a vision for how the "expectee" ought to change. See also Breakey (2022), Martin (2010, 542–546), Horgan and Timmons (2023), and Mellema (1998, 2004) on the normativity of expectations.

forgiveness, so the objection might go, are much more robust and therefore more valuable. As Scarre puts it, "Recipients of forgiveness […] would take less satisfaction in being forgiven if they thought that it might later be taken back. If forgiveness is really as unstable, and potentially revisable as this, it scarcely deserves the praise that moralists have conventionally bestowed on it."[50] When we seek forgiveness, we seek to put the matter behind us and gain a more stable social standing, not to put ourselves in a socially vulnerable position.

I will address the worry about self-respect first. Many have worried that a victim who forgives "too readily" fails to properly demonstrate self-respect; they "underestimate their own worth and fail to take their projects and entitlements seriously enough,"[51] or even fail to "care about the very rules of morality."[52] Perhaps I've laid out a form of forgiveness that makes any wrongdoer-dependent conditions far too weak, and is therefore compatible with a victim failing to respect herself when she forgives.

Plausibly, there is a vice associated with forgiving too readily, but there is no reason to think acts of faithful forgiveness necessarily exhibit this vice. A faithfully forgiving victim does not fail to take the wrong they experienced sufficiently seriously; to the contrary, in hoping for the wrongdoer's moral improvement, they are clear-eyed about the seriousness of his moral failings and invested in preventing future wrongs of the same kind. The way in which faithful forgiveness alters the normative landscape confers respect upon the victim by recognizing the wrongdoer's obligation to live up to her expectations for improvement—it shows that her expectations *matter*. Some have argued that "imperfect" forgiveness, wherein the victim continues to resent the wrongdoer but forgive nonetheless, better traces our real forgiveness practices and allows victims to enable the kind of relational transformations that forgiveness can achieve while preserving an important stance of moral protest (often against injustice).[53] Preserving the sense that the wrongdoer must continue her efforts at improvement may strike a similar balance. Any vice associated with too-ready forgiveness is plausibly manifested by the kind of victim who forgives completely unconditionally, without any lingering resentment or expectations for improvement.

In reply to the worry about invasiveness, the first thing to note is that faithful forgiveness is only appropriate in some circumstances. I am happy to concede that many acts of forgiveness are not suitable candidates for faithful forgiveness. Indeed, faithful forgiveness cannot be offered unilaterally, without the participation of the wrongdoer, so in many cases this kind of forgiveness is simply off the table. When I forgive a stranger for stealing my laptop, I cannot decide to forgive him faithfully if he makes it clear that he wishes no further interaction with me. I cannot invest myself in his development and keep tabs on how well he resists the temptation to

[50] Scarre (2015, 943).
[51] Novitz (1998, 301).
[52] Murphy and Hampton (1988, 18). See also Swinburne (1989).
[53] Maclachlan (2009), Westlund (2009).

steal moving forward. We have no relationship, so to track or even have faith in his development would be invasive (and a bit creepy).[54]

But there are many circumstances in which faithful forgiveness is appropriate and mutually sought-after. Moral improvement is integral to the healthy functioning of many close friendships, partnerships, and family relationships; it is very much the business of those with whom we relate most closely. In fact, one of the highest compliments we can pay our loved ones is that they make us better people. Under those circumstances, the characterization of faithful forgiveness as invasive and paternalistic is unfair. It's true that faithful forgiveness involves dependency, but there are many contexts in which dependence on others is enriching, intimate, necessary, and even autonomy-enhancing rather than invasive and (problematically) paternalistic. There are some people I want caring about my life in close and personal ways, often even holding me accountable. Holding our loved ones accountable via faithful forgiveness may be an extension of an existing, valuable form of dependency. Morally improving due to the (real or imagined) expectations of those I care about is a relational good faithful forgiveness helps us achieve.

The objector to faithful forgiveness may also characterize it as impermissible because it is excessively risky for the wrongdoer. Scarre worries that the possibility of un-forgiving makes the wrongdoer vulnerable to a "see-saw alternation of good will and resentment."[55] Why would we accept faithful forgiveness, which is conditional on our future behavior, rather than accepting low-stakes, permanent forgiveness? It's true that the expectation to morally improve may be seen as a burden if we are not independently interested in improvement. Of course, wrongdoers often seek forgiveness without seeing this as an investment in the project of moral improvement. Perhaps the wrongdoer is hoping that forgiveness will offer relief from the heavy burden of guilt or will repair a damaged relationship. Maybe she is genuinely remorseful about the previous incident but has very little confidence in her ability to morally improve, so wishes to leave the idea of moral improvement out of the act of forgiveness. Surely there are genuine cases of forgiveness with these features, and I concede that under certain conditions, they may be the most appropriate kind of forgiveness on offer.

But the objection that faithful forgiveness is always undesirable because risky suggests that we never independently desire to morally improve. In many—perhaps even most—cases of forgiveness, that is rather unlikely. In the moment of asking for forgiveness we are often guilty or ashamed, eager to distance ourselves from the person we were. Usually, faithful forgiveness offers accountability and support in a difficult project we have already endeavored to undertake. Whenever that is the case

[54] Of course, many valuable forms of forgiveness *can* be offered unilaterally and without any ongoing relationship with the wrongdoer—including the "fideistic forgiveness" that I take my inspiration from here. In Westlund's leading case, the parents of a murdered child forgive the perpetrator on faith without meeting or speaking with him (2009, 507). While such cases do not count as cases of faithful forgiveness on my account, they are still admirable instances of forgiveness. Thank you to Andrea Westlund for encouraging me to clarify this.

[55] Scarre (2015, 936).

(and, as I've stressed, within the context of certain relationships), the riskiness of faithful forgiveness is a benefit rather than a downside.

The objector suggests that faithful forgiveness is never desirable. Lastly, and perhaps most importantly, I ask the reader to consider whether we, in fact, desire it. Pleading for forgiveness is something we sometimes do by promising to change; an apology quite naturally involves the promise to change. Consider what Adrienne Martin says about apology: "[T]here is a performative element of apology that goes beyond demonstrating that resentment's expectation has been satisfied-namely, a *pseudo-contract* with the recipient; a second-personal, remorseful taking of ownership; and a *commitment to self-repair* (which thereby becomes a commitment to you-me repair, or repair of our relationship)."[56] While it would be reductive to think of apology or forgiveness as involving all the features of a contract, the idea of a *pseudo-contract* is helpful here insofar as contracts are tools for converging on mutually satisfactory expectations. The pseudo-contract framing is also helpful because contracts can be broken, and we think the violation of contractual expectations is grounds for dissolving the agreement (in this case, some agreement of mutual support or involvement). That a common performative element of apology involves approximating this contractual structure provides evidence for the point that we desire faithful forgiveness in the wake of wrongdoing. I think we crave the accountability that this kind of forgiveness affords—very often we want someone in our corner on the journey to self-improvement. Perhaps we fear that we can't do it alone.

Sometimes, I suspect that these pleas are offered hastily, without the considered intention to secure faithful forgiveness. "It will never happen again" seems like the "right thing to say" when we hope to be forgiven, even when there's a decent chance it will happen again. Why would this be a tempting thing to offer our victims if we did not consider faithful forgiveness to be a valuable kind of forgiveness? Perhaps sometimes we ambitiously take aim for faithful forgiveness, expecting to land somewhere slightly less desirable. In doing so, we aspire to hear that, despite what we've done, we are worth the faith, attention, and investment of our victim.

Finally, not only do we sometimes desire forgiveness that is conditional on our future behavior; we might also specifically find forgiveness that places no constraints on the wrongdoer's present or future features *un*desirable. As David Beglin points out, unconditional forgiveness can involve "settl[ing] a matter of interpersonal significance *without regard for the other person's perspective on that matter*," and this can disrespect the wrongdoer insofar as it suggests that their perspective doesn't *matter*.[57] If you offered Hannah forgiveness without the expectation of improvement, this may fail to take her seriously as a moral "participant," someone whose treatment of others matters. Even if we are not worried that this manifests a

[56] Martin (2010, 547 (my italics)).

[57] Beglin (2021, 260 (my italics)). Here, Beglin uses "unconditional" to refer to a kind of forgiveness that places no constraints on a wrongdoer's features at the time of forgiveness. I am suggesting that similar worries apply to forgiveness that places no constraints on a wrongdoer's *future* features or behavior.

lack of *self*-respect, it might manifest a lack of respect for *Hannah.* Thus, faithful forgiveness is one way to demonstrate "participant respect" for wrongdoers.[58]

2.2 Relational repair

I have argued that the Critic must take a valuable form of faithful forgiveness off the table. Second, the Critic may not be able to make relational transformation sufficiently central to his account of forgiveness. Un-Forgiving Deniers and Critics have argued that un-forgiving *undermines* the ability of forgiveness to enable relational repair; if forgiveness is not necessarily permanent, it cannot allow the victim and wrongdoer to move past the wrong in the way characteristic of forgiveness.[59] But, while it's true that permanence may be important for *some* acts of forgiveness to achieve the desired kind of relational repair, there is no reason to assume that permanent acts of forgiveness are most conducive to long-term repair and flourishing in all relationships and with all acts of forgiveness.

If the temptation to genuinely un-forgive crops up occasionally in our moral lives (and I believe it does), we should ask what response to this temptation is most conducive to the health of our relationships. Imagine, for instance, that you find yourself the victim of Hannah's third hotheaded outburst. You resent her anew, feel reparations are justified, and feel strongly that the initial forgiven outburst is centrally relevant to these sentiments. You have roughly two options: first, you might bite your tongue about these feelings. Perhaps you make sure to only demand reparations for her most recent outburst, carefully avoiding mention of your revived resentment of the earlier incidents. You may attempt to suppress or distract yourself from the negative emotions that have reemerged. You feel it's morally important to do so because those earlier incidents were forgiven.[60]

Second, you might communicate openly with Hannah about how you feel, and the connection between your current resentment and the initial outbursts. Sometimes this communication may involve a guilty confession that you find yourself unable to suppress emotions that you do not endorse, but other times it may involve a deliberate communicative act of un-forgiveness. To refrain from any such communication would force the confrontation to be artificial and stilted and would preclude an honest discussion about where your relationship stands. When you *feel* that you un-forgive Hannah, it may sometimes be morally important that you make the act official.

No doubt, careful avoidance of any un-forgiving talk will sometimes be the best option. This is especially likely if you and Hannah are not particularly close and if your relationship can return to a workable (suitably "repaired") equilibrium without

[58] Thanks to Monique Wonderly for suggesting the possibility of unconditional forgiveness being condescending to wrongdoers, and to Andrew Lichter for many helpful discussions of "participant respect," a term I borrow from him here. See Satne (2016, 1046–1047) for discussion of a closely related idea in Kant.

[59] See Scarre (2015, 934–937).

[60] On some understandings of forgiveness, forgiveness is a matter of what happens in your heart (not what is articulated to the wrongdoer); if un-forgiving can operate in the same way, biting your tongue may not be able to prevent it. See Warmke and McKenna on "private forgiveness" (2013, 198).

complete honesty about your emotions. But in the space of a close relationship, where would the withholding of the truth about your attitudes leave your friendship? Things are going to be awfully uncomfortable with Hannah moving forward if you continue to suppress the shape of your resentment, and you risk wronging her with an act of dishonesty (imagine this is the kind of relationship where you share everything with one another). Plausibly, withholding your emotions may be more toxic to your relationship in the long term than being forthright and un-forgiving Hannah. The former freezes her out from participation in a moral dialogue and undermines a basic tenant of your relationship: honesty. And while the latter is likely to result in short-term relational turmoil, it involves participation in a dialogue which opens the possibility of future opportunities for repair. Only by making your resentments clear can you allow Hannah to ever atone for them.

At this point the Critic may protest: isn't there something morally criticizable about being tempted to un-forgive in the first place? Perhaps it's true that, once gripped by this temptation, the best way to proceed is to follow through, but only the morally immature or vicious person (the Critic might insist) will ever experience such a temptation. Thus, when we conceive of our forgiving practices in relation to some moral ideal, we can leave out the vicious temptation to un-forgive entirely.[61]

There are two ways to respond to the Critic here. First, even if the fact that I fail to possess the ability to overcome my temptation to un-forgive speaks to some moral failing, it does not follow that the act of un-forgiving itself represents an additional moral failing. Where the temptation arises due to some criticizable character, we may think of the virtue associated with appropriate un-forgiving as what Julia Driver calls a *coping virtue*—a virtue which allows us to overcome or counteract certain character deficits.[62] In some cases, un-forgiving plausibly allows someone to overcome the vice associated with the excessive suppression or discounting of one's own emotions. Embracing this kind of coping virtue is especially attractive given how many of us are socialized by oppressive norms to excessively suppress or discount our emotions in ways that fall along lines of race, gender, and so on.

Second, it need not be the case that the temptation to un-forgive represents any kind of moral failing. Not only can such temptation reflect a concern with the wrongdoer's moral development (as I discussed in Sect. 2.1), but it can also represent an assertion of admirable self-respect. Theorists of forgiveness often struggle to balance the value of forgiveness with the close association between ongoing resentment and self-respect.[63] While we often assume it is virtuous to forgive, too-willing forgivers may fail to take the wrong sufficiently seriously—and thereby fail to take *themselves* sufficiently seriously as moral agents whose interests matter. Suitable accounts of forgiveness strive to balance these considerations by placing certain conditions on forgiveness such that forgiving only has a positive moral status

[61] Thanks to Max Kramer for pressing me on this.

[62] Driver (2015).

[63] See Murphy and Hampton (1988), Novitz (1998), Hieronymi (2001). See Dillon (2001) on the role of self-respect in forgiving oneself.

when it does not problematically undermine self-respect.[64] If self-respect sometimes requires that we refuse to forgive, perhaps it can also require that we revoke forgiveness when a wrongdoer has betrayed our trust in certain cases. If this is so, the temptation to un-forgive may not arise because of some character deficit; rather, it may arise from the accurate appreciation of how a wrongdoer's behavior disregards one's own moral worth and the recognition that maintaining a forgiving attitude would threaten one's self-respect.

In a similar spirit, Wonderly argues that even a view that treats forgiveness as the alteration of the normative landscape can accommodate un-forgiving. In rare cases where un-forgiving is appropriate, a victim must communicate about why she has defied the reasonable (but defeasible) norm of foreswearing resentment after forgiving. She writes,

> [A] victim who un-forgives at least owes her offender an explanation for why her (unrepudiated) blame has returned. Otherwise, the offender would be understandably confused and frustrated. Notice, though, that this explanatory burden is born not of the forgiver's commitment to refrain from future blame, but of the norms of moral communication. In expressing forgiveness, the victim communicates to the offender that she no longer blames him (at least in the same way) for the offense. The offender might reasonably rely on this information in navigating the relationship and without further explanation, would likely find expressed blame for the previously forgiven wrongdoing unintelligible. *The victim who expresses forgiveness, then, faces an onus to explain any renewed blame to her offender, qua moral interlocutor, so that the offender can revise his expectations accordingly and make sense of their relationship.* We have reason to think that instances of un-forgiving that run afoul of this requirement are normatively problematic.[65]

Thus, it's misleadingly simplistic to say all relationships are best repaired by permanent acts of forgiveness; rather, we are best served by allowing space for a kind of un-forgiving that fulfills the norms of moral communication—norms which are importantly tailored to the relationship in question. We can embrace permanence as a typical feature of forgiveness while accepting that long-term relational flourishing will sometimes require communication about attitudes or norms like the ones constituted by un-forgiving. The Critic, then, must provide better reasons for thinking that un-forgiving is incompatible with relational repair, or must offer an account of forgiveness without relational repair at its center.

[64] But see Garrard and McNaughton (2010) for a defense of the idea that unconditional forgiveness is consistent with self-respect.

[65] Wonderly (2021a, 2021b, 13 (my italics)).

3 Conclusion

These remarks suggest that permanent forms of forgiveness are not the only way to achieve the moral benefits forgiveness promises. A final objection from the Critic remains, however. Why think genuine un-forgiving is necessary for the moral benefits discussed above, when the kind of new forgiveness opportunities discussed in Sect. 1.2 seem to offer many of these same benefits? For instance, Hannah may receive sufficient encouragement in her efforts to morally improve if she knows you will resent her anew for subsequent offenses. Furthermore, fresh resentment for future offenses may secure the sorts of opportunities for relational repair discussed in Sect. 2.2. This objection gains plausibility if we think genuine un-forgiving and declining to forgive for new offenses are often difficult to distinguish. In Sect. 1.2, I argued that we clearly direct moral attention to the initial wrong in candidate un-forgiving cases. But we may think resenting for subsequent offenses can involve directing moral attention at the initial offense, perhaps as a way to understand them as part of a *pattern,* without involving the sort of moral attention that involve un-forgiving for the initial offense. If declining to forgive for new offenses gets us all the benefits and explanatory power of un-forgiving, why accept genuine un-forgiving as a legitimate part of our moral lives?[66]

While I agree that these cases may sometimes be difficult to distinguish, I think there are morally-important reasons to make space for both. It's true that declining to forgive for new offenses will sometimes involve directing moral attention at the initial wrong, but there may be important constraints on the *nature* and *degree* of attention. In genuine un-forgiving cases, I may be able to protest the initial wrong in relatively uncomplicated ways, while my attitudes and behaviors toward that wrong may require more nuance in cases where I merely mean to invoke it as a way of protesting a pattern.[67] In short, declining to forgive a new offense and un-forgiving may involve morally-important differences in how we relate to the initial wrong. They may also involve *narratively*-important differences, as these actions allow participants in relationships to understand their moral histories in distinct ways. (Notice how the story of your relationship with Hannah differs when you un-forgive her as opposed to declining to forgive her for her March outburst.) So, while declining to forgive new offenses may, in fact, offer many overlapping moral benefits, this redundancy is not itself a reason for rejecting genuine un-forgiving.[68] I am skeptical of the desire to explain away non-permanent forgiveness, since un-forgiving offers us an additional tool to understand and shape our moral practices. If un-forgiving is a part of how we understand our relationships with others, we ought to embrace this, rather than revise or criticize it.

[66] I am very grateful to Hannah Tierney for pressing me on this, and for her clear articulation of the reasons for this objection.

[67] It is also worth flagging that some accounts of forgiveness prohibit us from protesting a previously-forgiven wrong in this way. See Couto (2022) for discussion and criticism of this feature of some accounts.

[68] Thanks to Andrew Lichter for pointing out that the redundancy of moral practices is not itself a reason to reject one.

I have argued that we can make sense of un-forgiving as the genuine retraction of an earlier act of forgiveness. Against the Un-Forgiving Denier, apparent un-forgiving is not easily explained away as epistemic invalidation or a new forgiveness opportunity. Furthermore, I've offered some reasons to doubt the Un-Forgiving Critic's claim that un-forgiveness will always be impermissible. The Critic may be forced to restrict his conception of forgiveness undesirably, he must either explain why un-forgiving is incompatible with relational repair or decenter relational repair in his account of forgiveness, and he may close off an ideal of forgiveness which enables valuable forms of moral development.

Acknowledgements I am extremely grateful to Michael McKenna, Monique Wonderly, Andrea Westlund, Travis Quigley, and Andrew Lichter for their feedback and discussion about versions of this paper. I am also grateful for a wonderful audience at the Rocky Mountain Ethics Congress in August 2022, where I presented a version of it. Thanks especially to Hannah Tierney for her excellent comments and feedback. Members of the University of Arizona Writing Support Group helped me think through many of the ideas here; thanks to Robert Lazo for organizing this group. Thanks to Danielle Guzman for helpful conversations about forgiving and un-forgiving that aided my thinking on the topic, and to two anonymous reviewers of *Philosophical Studies* for their constructive comments.

Funding No funding.

References

Bash, A. (2007). *Forgiveness and christian ethics*. Cambridge University Press.
Bash, A. (2015). *Forgiveness: A theology*. Wipf and Stock Publishers.
Basu, R. (2023). The ethics of expectations. In M. Timmons (Ed.), *Oxford studies in normative ethics.* Oxford University Press.
Beglin, D. (2021). Unconditional forgiveness and normative condescension. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility.* (Vol. 7). 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. (2018). The alteration thesis: Forgiveness as a normative power. *Philosophy & Public Affairs, 46*(2), 207–233.
Breakey, H. (2022). When normal is normative: The ethical significance of conforming to reasonable expectations. *Philosophical Studies, 179*, 1–25.
Brunning, L., & Milam, P.-E. (2022). Letting go of blame. *Philosophy and Phenomenological Research, 106*(3), 720–740.
Couto, A. (2022). Forgiveness and the problem of repeated offences. *Journal of Applied Philosophy, 39*(2), 327–345.
Dillon, R. (2001). Self-forgiveness and self-respect. *Ethics, 112*, 53–83.
Driver, J. (2015). Virtue and moral deference. *Etica & Politica*, *17*(2), 27–40.
Garrard, E., & McNaughton, D. (2003). In defense of unconditional forgiveness. *Proceedings of the Aristotelian Society, 103*(1), 39–60.
Garrard, E., & McNaughton, D. (2010). *Forgiveness*. Routledge.
Griswold, C. (2007). *Forgiveness: A philosophical exploration*. Cambridge University Press.
Hieronymi, P. (2001). Articulating and uncompromising forgiveness. *Philosophy and Phenomenological Research, 62*(3), 529–555.
Holmgren, M. R. (1993). Forgiveness and the intrinsic value of persons. *American Philosophical Quarterly, 30*(4), 341–352.
Horgan, T., & Timmons, M. (2023). The expected, the contra-expected, the supererogatory, and the suberogatory. In D. Heyd (Ed.), *Handbook of supererogation.* Springer.
Hughes, P. M., & Warmke, B. (2017) Forgiveness. In Zalta E. N. (ed.) *The Stanford encyclopedia of philosophy* (Summer 2017 Edition). https://plato.stanford.edu/archives/sum2017/entries/forgiveness/

Jones, K. (2007). How to change the past. In C. Mackenzie & K. Atkins (Eds.), *Practical identity and narrative agency.* Routledge.
Kolnai, A. (1973–1974). Forgiveness. *Proceedings of the Aristotelian Society, 73*: 91–106.
Lang, B. (1994). Forgiveness. *American Philosophical Quarterly, 31*(2), 105–117.
Maclachlan, A. (2009). Practicing imperfect forgiveness. In L. Tessman (Ed.), *Feminist ethics and social and political philosophy: Theorizing the non-ideal.* Springer.
Martin, A. (2010). Owning up and lowering down: The power of apology. *The Journal of Philosophy, 107*(10), 534–553.
Mellema, G. (1998). Moral expectation. *Journal of Value Inquiry, 32*(4), 479–488.
Mellema, G. (2004). *The expectations of morality*. BRILL.
Milam, P.-E. (2019). Reasons to forgive. *Analysis, 79*(2), 242–251.
Milam, P.-E. (2022). Forgiveness. In D. K. Nelkin & D. Pereboom (Eds.), *Oxford handbook of moral responsibility.* Oxford University Press.
Murphy, J. (2003). *Getting even: Forgiveness and its limits*. Oxford University Press.
Murphy, J., & Hampton, J. (1988). *Forgiveness and mercy*. Cambridge University Press.
Neblett, W. (1974). Forgiveness and ideals. *Mind, 83*(330), 269–275.
Novitz, D. (1998). Forgiveness and self-respect. *Philosophy and Phenomenological Research, 58*(2), 299–315.
Orbey, E. (2022). *A daughter's quest to free her father's killer.* New Yorker. https://www.newyorker.com/magazine/2022/01/24/a-daughters-quest-to-free-her-fathers-killer
Owens, D. (2012). *Shaping the normative landscape*. Oxford University Press.
Richards, N. (1988). Forgiveness. *Ethics, 99*(1), 77–97.
Satne, P. (2016). Forgiveness and moral development. *Philosophia, 44*(4), 1029–1055.
Scarre, G. (2015). On taking back forgiveness. *Ethical Theory and Moral Practice, 19*, 931–944.
Smith, N. (2008). *I was wrong: The meanings of apologies*. Cambridge University Press.
Stump, E. (2018). *Atonement*. Oxford University Press.
Swinburne, R. (1989). *Responsibility and atonement*. Oxford Scholarship Online.
Walker, M. U. (2008). *Moral repair: Reconstructing moral relations after wronging*. Cambridge University Press.
Warmke, B. (2014). The economic model of forgiveness. *Pacific Philosophical Quarterly, 97*(4), 570–589.
Warmke, B. (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 I. Haji & J. Caouette (Eds.), *Free will and moral responsibility.* Cambridge Scholars Publisher.
Warmke, B., Nelkin, D., & McKenna, M. (Eds.). (2021). *Forgiveness and its moral dimensions*. Oxford University Press.
Westlund, A. (2009). Anger, faith, and forgiveness. *The Monist, 92*(4), 507–536.
Wilson, J. (1988). Why forgiveness requires repentance. *Philosophy, 63*(246), 534–535.
Wonderly, M. (2021a). Can we un-forgive? *Philosopher's Imprint, 21*(6), 1–13.
Wonderly, M. (2021b). Forgiving, committing, and un-forgiving. *Philosophy and Phenomenological Research, 104*, 474.