



# Kent Academic Repository

**Couto, Alexandra (2021) *The Reactive Attitudes Account of Forgiveness and Repeated Offences*. The Journal of Applied Philosophy . ISSN 0264-3758.**

## Downloaded from

<https://kar.kent.ac.uk/91825/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1111/japp.12562>

## This document version

Author's Accepted Manuscript

## DOI for this version

## Licence for this version

UNSPECIFIED

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

# Forgiveness and the Problem of Repeated Offences

## Abstract

According to the dominant account of forgiveness, to forgive is to overcome the reactive attitudes warranted by a wrongdoing. On one version of this ‘reactive attitudes’ account, forgiveness involves cognitive dissociation, while on another it involves affective dissociation. In this paper, I will argue that reflection on cases of repeated offences—where a wrongdoer is forgiven but then keeps repeating an offence—raises two challenges to this account of forgiveness. First, I will argue that, on either way of developing the account, it has the implausible implication that those who forgive must avoid considering the forgiven wrongdoing in their deliberations. Second, I will argue that the reactive attitude account of forgiveness is vulnerable to the objection that it is epistemically vicious. But the reactive attitudes account of forgiveness needn’t be abandoned: I will propose several ways in which the account can be revised or clarified in order to address the problem of repeated offences.

**Keywords: Forgiveness, Reactive Attitudes, Emotion, Wrongdoing, Domestic Abuse**

## I. The Reactive Attitudes Account of Forgiveness

### 1.1 Introduction

In the year ending in March 2019, according to the Crime Survey of England and Wales, 1.6 million women in the UK have suffered domestic abuse.<sup>1</sup> Given the fact that only a minority of victims report domestic abuse to the police, these statistics about domestic abuse are striking. Domestic situations are complex and there are often many reasons why domestic abuse victims stay with those who abuse them. But one of them is that forgiveness is widely perceived as a virtue irrespective of the circumstances. This may lead victims of domestic abuse cases forgive their abusers, and to believe that their abusers will redeem themselves despite evidence to the contrary. As one domestic abuse survivor says: “[when] he would apologise... and the next day say he was sorry. I always believed him.”<sup>2</sup>

Domestic abuse stands as a paradigmatic example of what I will be calling the problem of repeated offences. But this phenomenon is much more widespread and can occur in many other contexts—between family relations, friends, colleagues and so forth. Forgiveness often is appropriate, and virtuous, but in some cases, repeated forgiveness enables a cycle of exploitation and abuse. A desideratum for an adequate account of forgiveness is thus that it makes it possible for victims of wrongdoing to distinguish between wrongdoings that warrant forgiveness and those that do not. If we want individuals to be able to stand up to repeat offenders, it is crucial that our account of forgiveness doesn’t portray forgiveness as unconditionally virtuous, encouraging victims to tolerate repeated offences in this way.

---

<sup>1</sup>

<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2019>

<sup>2</sup> [http://safelives.org.uk/practice\\_blog/babs-story](http://safelives.org.uk/practice_blog/babs-story)

According to the dominant account of forgiveness, which I will call the *Reactive Attitudes Account*, forgiveness is constituted by the overcoming of reactive attitudes warranted by the wrongdoing (Butler 1897; Murphy 1990, Darwall 2006; Bovens 2009; Zaragoza 2010; Allais 2008). This paper argues that this account faces two challenges. First, the account, as currently formulated by Allais and others, entails a highly counterintuitive implication which is highlighted by cases of repeated offences. I will suggest two ways of modifying the account to avoid this implication. Second, the case of repeated offences highlights a more fundamental problem: that this account renders forgiveness epistemically vicious. In response to this objection, I will propose that proponents of the reactive attitudes account should understand appropriate forgiveness to be conditional on the belief that the wrongdoer won't commit the same offence again.

## 1.2 Forgiveness: The Concept

According to the Reactive Attitudes Account, forgiveness is identified with the overcoming of negative reactive attitudes.<sup>3</sup> This view has had many defenders (Murphy 1990, Darwall 2006; Bovens 2009; Zaragoza 2010; Allais 2008a). As a working conception, I will focus on one of the most developed versions of the Reactive Attitudes Account, that defended recently by Lucy Allais (2008a). This account can be stated as follows:

**To forgive a wrongdoer for his wrongdoing** is to overcome the negative reactive attitudes towards the wrongdoer that this wrongdoing warrants, while still holding the wrongdoer responsible for his wrongdoing.

In other words, by forgiving, the victim refrains from holding against the offender a certain wrongdoing, and goes back to seeing him affectively as her friend, relative or lover. There are two distinctive features to Allais' version of the Reactive Attitudes Account. First, Allais is offering an account of what is involved in forgiving someone *for a specific* wrongdoing (Allais L., 2008a, p. 51). Second, instead of narrowly taking forgiveness to be the overcoming of *resentment*, she widens it to include the overcoming of all retributive reactive attitudes warranted by the wrongdoing (contempt, anger, mistrust, etc.). I will use the phrase 'reactive attitudes' to refer to what Allais describes as 'retributive reactive attitudes'. Before proceeding, let me emphasize one important condition for genuine forgiveness: forgiveness can only occur if the wrongdoer is still held responsible for his wrongdoing. This is an important feature of forgiveness, as this allows us to distinguish *forgiving* someone from *excusing* someone.

## 1.3 The Strawsonian Insight

In his classic article 'Freedom and Resentment', Strawson argued that to hold people responsible for wrongdoing is to be *disposed* towards the negative reactive attitudes (Strawson, 1974). I'll call this claim the Strawsonian Insight:

**Strawsonian Insight:** Human Beings are disposed to experience negative reactive attitudes towards other agents who have wronged them as long as three conditions are fulfilled:

- i) The wrongdoer is a "psychologically normally developed" agent;
- ii) The wrongdoer was aware of his causing harm;
- iii) The wrongdoer was not coerced.

---

<sup>3</sup> This account is also sometimes referred to as the Standard View (Kekes, 2009, Warmke, 2011).

Although the Strawsonian Insight is clearly a background assumption of the reactive attitudes account of forgiveness, proponents of the reactive attitudes account rarely spell it out in detail. However, the Strawsonian Insight could be interpreted in very different ways. I'll start by describing three interpretations I *won't* be endorsing: a conceptual, normative and strong naturalist interpretation.

First, one could understand the Strawsonian Insight to be making a conceptual claim. On this interpretation, provided certain conditions are fulfilled, to hold someone responsible *entails* experiencing the reactive attitudes towards that person (Strawson, 1974, Gibbard, 1992, Wallace, 2006). There is no question whether we should have these reactive attitudes toward wrongdoers: to have these attitudes *just is* what it is to hold someone responsible for acting wrongly.

Second, on a normative interpretation of this claim, even if it were possible to decide whether or not to retain our reactive attitudes; we *ought* not to abandon them, because this would involve losing important dimensions of human life (Shabo, 2012, Wallace, 1994, Nichols, 2007).

Third, on the strong naturalist reading, our disposition to experience the reactive attitudes is a deep natural disposition and that, even if we wanted to, we *could* not change our disposition to experience the reactive attitudes (Eshleman, 2001; for criticism, see Strawson, 1986, Pereboom, 2007, Nichols, 2007).

In what follows I will assume only a qualified version of the naturalist interpretation, namely the much more modest and fairly uncontroversial empirical claim that human beings have a *highly fixed* disposition towards experiencing negative reactive attitudes when a morally responsible sane adult wrongs them. A disposition is *highly fixed* if it is difficult to change the strength of that disposition (Russell, 2009). The claim would then simply be that we have a strong natural disposition to experience the reactive attitudes. Since we are dealing here with a psychological disposition, I will assume a simple probabilistic account of dispositions. On such an account, to say that an individual is disposed to experience the reactive attitudes when he is wronged is simply to say that it is *highly probable* that this individual will experience the reactive attitudes in these conditions.<sup>4</sup>

What is crucial to the argument I will develop here is that, according to this interpretation of the Strawsonian Insight, the disposition towards forming retributive reactive attitudes is *entirely blocked* only if the offender is not morally responsible (i.e. the person is excused in some way) or if the offender is not a psychologically normally developed agent (i.e. the offender is a child or insane). So the Strawsonian Insight implies that reactive attitudes can't be entirely blocked in cases in which forgiveness might be relevant (for the details of this argument, see section 1.6).

#### 1.4 The Problem and the Solutions Put Forward Thus Far

The problem is thus to explain how, on the Reactive Attitudes Account, one can forgive someone, given the Strawsonian Insight. Recall that, on the Reactive Attitudes Account,

- (1) **To forgive x for y** is to refrain from holding the reactive attitude towards x that wrongdoing y warrants, while still holding y morally responsible for x.

Yet an implication of the Strawsonian Insight is that

- (2) **To judge x responsible for wrong y** is to be disposed to experience reactive attitudes against x.

---

<sup>4</sup> Dispositions are often analyzed in terms of counterfactual conditionals (Martin, 1994). Alternatively, one could adopt a habitual analysis of dispositions (Fara, 2005). On both views, serious difficulties arise if one tries to identify the conditions in which the disposition fails to be manifested in the presence of the relevant conditions.

Although (2) doesn't necessarily require the actual experiencing of reactive attitudes, to say that there is such a disposition is at least to say that the person *is likely to* experience the reactive attitudes *if* certain stimuli are present. And if the person does experience the reactive attitudes in question, this would be incompatible with forgiving. Therefore, one needs to find a way to *reliably block* the disposition from manifesting itself. The most promising route for the disposition to be reliably blocked is to assume that the wrongdoing can somehow be dissociated from the wrongdoer.<sup>5</sup>

Two kinds of dissociation have been suggested in the literature. Overcoming the negative reactive attitudes is taken to be feasible via either

**Affective Dissociation:** The victim excludes the wrongdoing from the possible grounds on which her attitude towards the wrongdoer is based.

Or

**Cognitive Dissociation:** The victim holds the belief that the wrongdoing is not representative of the character of the wrong-doer.

In order to give a better idea of what affective dissociation amounts to, let me quote Allais:

“holding something against someone involves lowering the way you affectively esteem or regard her as a result of her action ...when you forgive the perpetrator your attitude towards her as a person is no longer the negative one that her wrongdoing supports; in other words, the act is disregarded in your ways of regarding and esteeming her, and in this sense, the slate is wiped clean, and the act is not held against her..... when you forgive her, you dissociate her wrongdoing from the way you feel about her, and cease to have this attitude towards her.” (Allais, 2008a, p.56-57)

Allais claims that, because reactive attitudes have a different relation to evidence than beliefs have, we can somehow decide to *change our focus* and ground our attitude towards the wrongdoer on, say, other aspects of the wrongdoer without necessarily changing our beliefs about the wrongdoer's character (Allais, 2008a, p.51).

In contrast, Pamela Hieronymi has argued that forgiving involves a cognitive dissociation, as the victim commits herself to a *belief*, more specifically the belief that the wrongdoing is not representative of the character of the forgivee (Hieronymi, 2001, p.550). Several other philosophers have defended a similar view (Hampton, 1990, p.83-84, Kolnai, 1973-1974, Holmgren, 1993, p.348).

## 1.5 Forgiveness as a Speech-Act and its Implicit Commitment

For simplicity's sake, I will restrict my argument to the paradigmatic case of interpersonal forgiveness, in which the victim announces to the wrongdoer “I forgive you”. When the victim announces, “I forgive you”, she performs a speech-act. As such, it amounts not only to the *description* of the victim's overcoming of negative reactive attitudes (with respect to the particular forgiven wrongdoing) but also to a *commitment* to overcome them in the long-term and to strive to do so successfully.<sup>6</sup> When I claim to forgive you for, say, spreading false rumours about me behind my back, I'm not merely reporting the temporary absence of resentment; I commit myself to refrain from

---

<sup>5</sup> Lucy Allais takes this dissociation to be constitutive of forgiveness (Allais, 2008a, p. 51). Others have also taken this dissociation to be necessary for forgiveness to be even possible (Murphy, 1990, p. 24-25).

<sup>6</sup> It is possible that such a commitment is present even when no speech-act is articulated. Indeed, forgiveness seems to be more than the mere description of the temporary overcoming of resentment but a *decision* to do so consistently in the long term (see Warmke, 2016). However, I won't discuss this possibility here and restrict my argument to the specific instance of forgiveness articulated as a speech-act.

basing my attitude towards you on that episode in the future. So, if we wanted to sum up what the speech-act of stating “I forgive you” amounts to, we could say that:

(F) To say “I forgive you” to wrongdoer x is to commit oneself not to hold the negative reactive attitudes towards x that his wrongdoing y warrants.

And this commitment in turn implies further commitments to avoid doing things that may compromise the primary commitment (I will return to this below). In the following, I will assume that this commitment is analogous to a promise.<sup>7</sup>

## 1.6 The Problem Returns: The Argument from Commitment and the Overlooked Implication

Let me briefly state some assumptions of the argument I will now set out. First, in order to commit myself to do something, I need to believe that, unlikely circumstances notwithstanding, I will be able to fulfill that commitment. Moreover, unlike a commitment to do something, a commitment to forgive someone for something (and thus to overcome resentment) is not a commitment limited in time: one doesn’t commit oneself to forgive someone for half an hour. My first assumption is thus that a commitment to forgive presupposes that one believes that one is able to *reliably* and *consistently* fulfill one’s commitment, unlikely circumstances notwithstanding.

Second, although one can strive not to feel resentment—and even attempt not to feel resentment *while* considering the wrongdoing—one cannot *reliably will* not to experience resentment when considering the wrongdoing in the future. Emotions are generally not under our direct control (Moller, 2003, Landau, 2004, Drake, 2011, Marušić, 2013, D’Cruz and Kalef, 2015).

Third, recall all the conditions necessary for the manifestation of the disposition to resent: (a) the consideration of (b) a serious moral wrong done against me (c) by a sane adult (d) who is aware of causing harm and is not forced to do so. On the Reactive Attitudes Account, forgiveness requires all the conditions necessary for the manifestation of the disposition to obtain barring (a). However, forgiving as a speech-act requires that the victim commits herself not to hold the reactive attitudes warranted by the wrongdoing. But, if one can’t reliably will not to experience resentment (when the conditions for the manifestation of the disposition obtain), it seems that the only thing left for the victim to do is to ensure that condition (a) doesn’t obtain. Thus, once the victim has forgiven, she commits herself to avoid considering the wrongdoing, as it is a situation in which she would be disposed toward the reactive attitudes.<sup>8</sup>

Stated explicitly, the steps of what I will call *the Argument from Commitment* go as follows:

---

<sup>7</sup> Note that, in order to forgive the wrongdoer, the victim needs to still *be able* to recall the wrongdoing, as otherwise she would have simply forgotten about it rather than forgiven it. We need to have the capacity to retrieve the memory of the wrongdoing to avoid having entirely forgotten about it.

<sup>8</sup> Even if you hold that one might occasionally succeed in blocking the manifestation of this disposition while considering the wrongdoing, it would be hard to contend that we are able to reliably and consistently will ourselves out of resentment. Thus even if the manifestation of a disposition may occasionally be blocked by an act of will or diversion of attention, this won’t be sufficient since a commitment presupposes that one believes that one is able to *reliably* and *consistently* fulfill one’s commitment.

- (1) **The Reactive Attitudes Account of Forgiveness.** To forgive x for y is to refrain from holding the reactive attitude towards x that wrongdoing y warranted while holding y morally responsible for x.
- (2) **Forgiveness as a Speech-Act (F):** To say “I forgive you” is to commit oneself not to hold the negative reactive attitudes towards x that y warrants.
- (3) **The Strawsonian Insight.** To judge x to be responsible for wrong y is to be strongly disposed to experience reactive attitudes against x.
- (4) **The Probabilistic Analysis of Disposition.** To be disposed to experience the reactive attitudes towards others is to be likely to experience these attitudes when certain conditions obtain.
- (5) **Not Consistently and Reliably Under Our Control:** When the conditions for the manifestation of the disposition towards the reactive attitudes obtain, the experience of the reactive attitudes is not consistently and reliably *under our control* given that we have a strong natural disposition to experience them.
- (6) **Necessary Conditions for Resentment:** The conditions that need to obtain for someone to experience the reactive attitudes are (a) the victim considers (pays attention to) the wrongdoing committed against her (b) the victim believes that this was a serious moral wrong committed against her (c) the victim believes that the wrongdoer was aware of causing harm and was not forced to do so (d) the victim believes that the wrongdoer is a sane adult.
- (7) One can’t reliably and consistently block the disposition to the reactive attitudes if the conditions enumerated in (6) are present.
- (8) **Necessary Conditions for Forgiveness:** Forgiveness presupposes the presence of all conditions barring (a).
- (9) In order to reliably block the disposition, the only thing left for the victim to do is to prevent condition (a) from obtaining.

Therefore,

- (10) **The Overlooked Implication.** To forgive, the victim must generally avoid considering the forgiven wrongdoing.

Although this conclusion might sound rather strong, this overlooked implication just follows from the Strawsonian Insight and the Reactive Attitudes Account, as it is currently formulated by some of the proponents of the affective dissociation version.<sup>9</sup>

But it is very counterintuitive to suggest that when we forgive, we are committing ourselves to avoid considering the wrongdoing in question altogether. Furthermore, and more importantly, this commitment would have negative implications for the value of forgiveness. In order to highlight this latter point, I’ll consider the case of repeated offences.

---

<sup>9</sup> One might remark here that we might not have full control over whether we *consider* the wrongdoing in the first place. But if one can’t even control whether to consider the wrongdoing, then this would make the Reactive Attitudes Account of forgiveness even less feasible.

## II. The Case of Repeated Offences

What happens when the victim is confronted with a similar case of wrongdoing committed by a wrongdoer she forgave the first time around? I have argued that an overlooked implication of the Reactive Attitudes Account was that the forgiver commits herself not to consider the previous wrongdoing. But if that is the case, then this commitment still binds the forgiver when the second offence takes place. The previous wrongdoing cannot be taken into account as a relevant factor in order to make the decision to forgive or not to forgive the second offence. This implies that, from the point of view of the forgiver, there can't be anything different the second time around.

To see this, consider the example of two friends, Francine and Julie. Let us assume that Francine has broken a norm of friendship, say by failing to render an important service to Julie, but that this is an isolated event. Francine has recognised her fault and apologised, and Julie has explicitly forgiven Francine. A month later, Francine commits another wrong, say by failing again to render an important service to Julie at a time in which she is vulnerable (and we are assuming that there are no exculpatory circumstances in Francine's life). Julie, when she realizes that, shouts out: "So typical, you can't think of anyone else but yourself."

We might think that Julie is justified in having this attitude *because it is the second time that Francine wrongs her*. However, if the overlooked commitment is true and if Julie has truly forgiven Francine the first time around, Julie has committed herself to refrain from considering the previous wrongdoing.<sup>10</sup> Therefore, it would seem that there is *nothing different* the second time Francine wrongs Julie and that she might thus not be justified in her reaction. More precisely, the fact that it is the *second* time around presumably cannot be taken into account.

A mistaken objection would be to claim that the description of the second wrongdoing would include reference to the fact that it's the second time. This would certainly be the case if one were to describe the situation *from an external perspective*. However, the perspective that is adopted here is the perspective of Julie who has forgiven Francine. Practical reasons are reasons possessed by specific *agents*.

If there is nothing different the second time around, there are two possible problematic implications. Either you think that Julie isn't committed to forgiving Francine, but that, from her deliberative perspective, there is nevertheless nothing different the second time she is wronging her. Or you think that Julie *has* to forgive Francine again because her decision to forgive is subject to a requirement of consistency. Whatever your view is on whether or not we are subject to such a requirement of consistency, I'll argue here that if the overlooked implication holds, this will have implausible implications.

On the first, more plausible view, Julie isn't required to forgive the second time around, as she isn't subject to a requirement of consistency. However, if Julie cannot take into account the fact that

---

<sup>10</sup> Pettigrove (2012) highlights a similar issue: a kind of paradox due to the fact that it looks both appropriate and inappropriate for the forgiver to reconsider the first offence. However, he suggests that we can avoid the paradox by expanding the account of forgiveness to include traits of character, instead of mere actions. However, I don't think this could solve the problem I raise here. Pettigrove's strategy isn't available to a supporter of Allais's account, given that Allais insists on describing her account as distinctive because it focuses on the reactive attitudes warranted by the *specific* action and *not* the character of the individual. (Allais, 2008, p.57).



it is the *second* time Francine commits a wrongdoing, as she can't refer to the previous wrongdoing, there is *nothing different* when the wrongdoing is repeated. This result is disturbing enough as it means that *no ground* can be provided to judge that it might *not* be reasonable to forgive repeatedly (even in repeated wrongdoings of a *severe* kind). This effectively robs the victim of the ability to invoke the repeated nature of the wrongdoing as determining her refusal to forgive.

On the second view, which relies on the more controversial assumption that we are subject to a requirement of consistency, the reasons that are relevant for Julie's decision whether to forgive ought to be similar to the ones she invoked the first time around (the reasons that are relevant here are presumably how important the service was to Julie, what reasons there were for Francine not performing this service, whether Francine has apologised or regretted her wrongdoing, etc.). Assuming that Julie is rational, uses reliable mechanisms of reasoning and is subject to a requirement of consistency, she ought to reach a similar decision when confronted with a similar moral wrong. If we assume that the second wrongdoing is similar to the first in these respects, then Julie presumably cannot now claim that, say, every single failure to perform a service would undermine a friendship.<sup>11</sup> Recall that it's a similar moral wrong precisely because, having forgiven Francine, Julie cannot take the first wrongdoing as relevant to her current decision. So in cases involving wrongdoing of the same kind and severity, it seems that Julie is committed to be a perpetual forgiver, a Christian icon of sort. It might be objected to that view that forgiving is morally supererogatory and that, accordingly, Julie is not morally required to forgive Francine the second time around. I agree that there is no *moral* requirement to forgive Francine the second time around. Whether it's *rationaly* required of Julie to forgive the second time around depends on whether or not we believe that she is *required* to be consistent when making this kind of decision.

Both possible implications seem to me implausible, even paradoxical. So let us go back to the overlooked implication and see if there are ways to avoid it.

### III. How Not to Avoid the Overlooked Implication

Below I'll examine attempts to avoid the overlooked implication, which, I believe, fail to do so.

#### 3.1. First Reply: Too Demanding

One might think that the challenge fails because it attributes to the Reactive Attitudes Account the far too demanding claim that forgiveness requires a stringent commitment to *not* consider the previous wrongdoing. But surely Julie could forgive Francine and still keep in mind that Francine did something wrong?

However, if there is excessive demandingness, it resides in the Reactive Attitudes Account of forgiveness, at least as it is usually understood, and not in the challenge presented here, which merely highlights its apparent and demanding implications.

But someone might insist: why couldn't I promise to overcome resentment without committing myself not to consider the wrongdoing? After all, in other cases of promising, the fact that some factors are beyond your control doesn't necessarily imply that you need to take every step

---

<sup>11</sup> The consistency I have in mind here is consistency in the evaluation by the forgiver of a range of scenarios in which they are justified to forgive. I would be inconsistent, for instance, if I were holding a grudge against someone for a trivial wrongdoing, say they borrowed my pencil without asking but forgave immediately the most terrible betrayal by someone else. It is plausible to say that, if someone were to witness this inconsistency, they could rightly point out that I am treating these wrongdoers unfairly. So forgiveness does seem subject to a requirement of consistency, including, I suggest, the basic idea that like cases should be treated alike.

to avoid failing to uphold the promise. For instance, I can promise you to return your book safely without committing myself to hide it from everyone so that nobody could be tempted to steal it.

In order to show why this approach doesn't work, I need to elucidate the difference between committing oneself to forgive and other commitments. Promising to forgive (i.e. to overcome certain reactive attitudes) is a case of a promise to *feel*. And promising to feel is distinctive because emotions are less under our control than actions (Moller, 2003, Landau, 2004, Drake, 2011, Marušić, 2013, D'Cruz and Kalef, 2015).<sup>12</sup> In a typical case of promising, my succeeding or failing to uphold a promise generally depends on many factors, only a minority of which lies entirely beyond my control (the traffic, the absence of storm, etc.). However, in the case of promising to feel, the main factor that will determine the success or failure of my commitment is not under my direct control. Because of this, promises to feel often come with a significantly higher probability of failure.

In other words, because of the nature of the relation between will and emotions, in the case of promising to feel, promises come with a further commitment to undertake specific acts that will reduce the likelihood that the promisor fails to uphold her promise. Consider the case of promising to wait for one's partner to return from war (where he/she would stay two years with the obvious risk of never returning) (Moller, 2003, Landau, 2004, Drake, 2011, Marušić, 2013). If the evidence shows that many have failed to uphold similar promises, this shows that the promise will be hard to fulfil and that we might need to take specific actions to avoid breaking the promise (Marušić, 2013, p.310). Going out every evenings in single bars or using a dating app would certainly seem to contradict the promise. It is thus not too demanding to interpret the commitment to forgive as requiring avoiding to put oneself in situations where the disposition to experience resentment would be strong.

### 3.2 Second Reply: Revising (F) so that it doesn't refer to the Reactive Attitudes Account

Let us consider a radically different interpretation of the speech-act "I forgive you". David Novitz has argued that the only thing that this speech-act involves is relinquishing the claims that the forgiver has towards the wrongdoer (Novitz, 1998, p.302). In opposition to the reactive attitudes account of forgiveness considered in this paper, some have indeed argued that forgiveness should instead be understood as the deliberate refusal to punish the wrongdoer (Zaibert, 2009, Hobbes, 1961). After all, if one can't overcome resentment at will, maybe the only thing we can mean when we say "I forgive you" is that we will refrain from *acting* in certain ways. Accordingly, consider

(F\*\*) To say "I forgive you" is only to commit oneself to *refrain from exacting revenge or punishment* from the wrongdoer.

This is usually taken to constitute an alternative account of forgiveness, but I want to suggest that there might be a way of granting the validity of both accounts by attributing them different *remits*. Although the Reactive Attitudes Account would be retained as an account of forgiveness, its scope would be restricted to the internal experience of forgiveness, whereas an alternative account would be used in order to describe the utterance "I forgive you". This revision might rescue the Reactive Attitudes Account, but it would in effect greatly restrict its scope in a way that is unlikely to be attractive to proponents of this account.

### 3.3 Third Reply: Revising (F) to Soften the Commitment

---

<sup>12</sup> Some have even argued that promising to feel (in particular promising to love), is generally impossible (Drake, 2011).

Let us now consider the possibility that we could soften the commitment to overcome resentment. In the argument from commitment, I assumed

(F) To say “I forgive you” to wrongdoer x is to commit oneself *not to hold the negative reactive attitudes* towards the wrongdoing that his wrongdoing warrants.

But let us consider an alternative interpretation of the commitment made by this speech act

(F\*) To say “I forgive you” to wrongdoer x is to commit oneself *to take reasonable steps* towards not holding the negative reactive attitudes towards x that his wrongdoing y warrants.

If I am only committed to take reasonable steps, then I don’t need to do *everything possible* not to experience the reactive attitudes. Moreover, the victim would be allowed to consider the wrongdoing in exceptional circumstances. However, I believe this proposal still faces several problems.

To begin with, this interpretation doesn’t seem to reflect what people mean when they say “I forgive you”. It just seems too weak an interpretation of the commitment involved in forgiving: it turns out that all we are really saying is that we will *try* to forgive the wrongdoer, not that we do forgive them.<sup>13</sup> This is worsened by the fact that *what* would constitute reasonable steps (and what won’t) remains unclear and at the discretion of the forgiver.

Second, this response might be self-defeating. If the forgiver needs to reflect on what constitute reasonable steps, on whether to consider the wrongdoing and on what constitutes exceptional circumstances, she might end up dwelling on the forgiven offence more often than is compatible with true forgiveness.

Third, even if we accept this interpretation of the speech-act, the overlooked implication still holds. Recall that one cannot *reliably* block the disposition towards the reactive attitudes while considering the wrongdoing. Taking reasonable steps (or striving) not to hold the negative reactive attitudes would still require from the forgiver to avoid putting herself in situations (or states of mind) in which she would be *strongly* disposed to experience the negative reactive attitudes as she can’t reliably block that disposition. Given that the consideration of the wrongdoing is a state of mind in which the forgiver would be strongly disposed to experience the reactive attitudes, even this weaker interpretation of the speech-act involves a commitment to avoid considering the wrongdoing.

### 3.4. Fourth Reply: It is up to the forgiver

Another possible way to avoid the repeated offences problem is to argue that the Argument from Commitment, presented in section 1.6.1, seems to depend on a view in which we, as agents, are alienated from our own emotional life. Although it is true that, from a theoretical standpoint, we may possess empirical evidence about what may affect our future emotions, it is only through practical reasoning, not theoretical reasoning, that we can decide what to do in a way that is guided by what *we* value (for such an approach, see Moran, 2001). In other words, practical reasoning should guide our commitment practice. Let me explain below what this implies for forgiveness.

On this view, when we commit ourselves to forgive (or more generally commit to feel a certain way), we should not adopt a theoretical stance towards what we are likely to achieve. In my proposed Argument from Commitment, the theoretical consideration that emotions are not under our direct control constitutes an important premiss. On this approach, although it is true that, from an observer’s perspective, there is a substantial likelihood that the forgiver will experience resentment when

---

<sup>13</sup> For the claim that promising to try might amount to inadequate commitment see Marušić (2013, p.297) and D’Cruz and Kalef (2015).

considering the wrongdoing, this evidence comes from the wrong perspective. Giving this evidence due weight amounts to treating the experience of resentment as if whether or not the forgiver will experience resentment and break her commitment were not up to *her*. Given that it is still (in some sense) up to the forgiver whether or not she breaks her commitment, deferring to the evidence that she is disposed to experience resentment amounts to dismissing her agential powers.<sup>14</sup>

This response faces one significant problem however: there are cases in which not deferring to the available empirical evidence is very counter-intuitive (D’Cruz and Kalef, 2015). Consider the case of the very forgetful doctor who promises his hospitalised patient to check on him before going home (instead of asking another doctor to do it). If the doctor is past a certain level of forgetfulness, it would be irresponsible for the doctor to make such a promise, given that it invites the reliance of others.

## **4. Avoiding the Overlooked Implication but still facing the Repeated Offences Problem**

### **4.1. Revising the Affective Dissociation version of the Reactive Attitudes Account of Forgiveness**

One way to successfully avoid the Overlooked Implication, I will now argue, is to revise the affective dissociation version of the Reactive Attitudes Account (but this still won’t address the deeper problem raised by the case of repeated offences). Recall premiss 1, which stated that ‘to forgive x for y is to refrain from holding the reactive attitudes towards x that wrongdoing y warranted while holding y morally responsible for x’. So far, we have assumed that to ‘refrain from holding the reactive attitudes’ was to be interpreted, following Allais, as referring to the overcoming of the retributive reactive attitudes. But another interpretation, favored by Strawson is to interpret forgiveness as the forswearing of the retributive reactive attitudes. The most important difference for our purposes is that forswearing the reactive attitudes doesn’t entail committing oneself to *overcome* the reactive attitudes but merely to *disavow* them. The mere experience of resentment and the other retributive reactive attitudes does not necessarily undermine the commitment to forgive, as long as those reactive attitudes remain unbidden and disavowed. The Overlooked Implication doesn’t follow from such a revised account, as resentment and other retributive reactive attitudes deriving from the wrongdoing can thus return. This revision is thus necessary as it allows proponents of the affective version of the Reactive Attitudes Account to avoid a highly counter-intuitive implication, the Overlooked Implication.

However, this revision doesn’t allow us to escape the problem of repeated offences. Even if the forgiver committed herself not to endorse these experiences of the reactive attitudes, after the second wrongdoing, it is still constitutive of the account of forgiveness that she can’t ground her overall attitude towards the wrongdoer on the first forgiven wrongdoing. Only the second wrongdoing can provide a ground for the retributive reactive attitudes. Going back to our illustration, Julie can’t endorse the reactive attitudes that are based on the first wrongdoing so she can only base her reactive attitudes towards Francine on the second wrongdoing. If Julie’s attitude towards the wrongdoer can’t be based on the first wrong-doing, then it seems that there is nothing different- affectively- the second time around. The problem of repeated offences remains and requires another revision of the account, which I will provide in section 5 below.

### **4.2. Choosing the Cognitive Dissociation version of the Reactive Attitudes Account**

---

<sup>14</sup> A similar argument has recently been made by Berislav Marušić in defence of the claim that we can promise against the evidence (and that we don’t need to promise merely to try) in cases where the chances of failing to uphold our promise is substantial (Marušić, 2013).

On the cognitive dissociation view, the overcoming of the reactive attitudes towards the wrongdoer is done through the acquisition of the belief that the wrongdoing is *not representative* of the wrongdoer's character. Cognitive dissociation gives us a *reason* to not be disposed to experience resentment and might thus be taken to enable us to overcome the disposition to experience the reactive attitudes even *while* we are considering the wrongdoing—thereby allowing us to avoid the Overlooked Implication.

But again, although choosing the cognitive dissociation version allows us to avoid the counter-intuitive Overlooked Implication, it still doesn't address the deeper problem raised by the case of repeated offences. The forgiveness conferred via cognitive dissociation is given on the assumption that the wrongdoing is *not representative* of the character of the wrongdoer. Once I have said to someone that I have forgiven her, however, this assumption also becomes a commitment to not further question this belief. The forgiven act has been judged to be not representative of the offender's character and as such cannot figure as a *ground* in future evaluation of her character. Although I can on this view *reconsider* the wrongdoing, I cannot *use it* to evaluate the wrongdoer's character at a later point. This is precisely what forgiving requires us not to do on the cognitive dissociation view. So the problem of repeated offences remains a challenge even if the victim *can consider* the wrongdoing. Avoiding the Overlooked Implication doesn't allow us to get rid of the problem of repeated offences.

To go back to our earlier illustration, if Julie uses the previous wrongdoing as evidence about Francine's character, then she hasn't really forgiven Francine, given that on the cognitive dissociation view, to forgive is to not use the forgiven wrongdoing as grounds for assessing the character of the wrongdoer. So Julie cannot invoke the previous wrongdoing as *evidence* that her friend is not a good friend when she fails to perform an essential duty of friendship the second time around. In this case too, there is nothing different the second time around.

The deeper more fundamental problem of repeated offences can be articulated in the following way: without qualification or revision, the current formulation of the account entails an impairment in the *future* epistemic and normative assessment of wrongdoers.

In the illustration we examined, Julie can't use the first wrongdoing in order to either ground her reactive attitudes or justify her judgment of Francine in the future. Yet this result is odd. Forgiving is often justified on the basis that the offender might have changed. Yet, because forgiving commits us to certain epistemic restrictions, it seems to prevent us from appropriately assessing future cases in which forgiving is (or isn't) warranted. There is thus a chasm between fulfilling what is conceptually required for forgiveness and fulfilling the conditions that could justify it.

To sum up, avoiding the overlooked implication is necessary for a plausible account of forgiveness, but it doesn't yet allow us to escape the normative impasse of the problem of repeated offences on either the cognitive or the affective dissociation version of the reactive attitudes account of forgiveness. In both cases, Julie can't use the previous forgiven wrongdoing as a ground for her reactive attitudes or as evidence for her evaluation of Francine's character. The case of repeated offences highlights that forgiveness arguably involves a kind of wilful self-deception—it almost seems to involve a kind of *epistemic vice*, since one commits oneself also to avoid using relevant available information in forming beliefs and making decisions.<sup>15</sup>

In order to address the second deeper problem raised by the case of repeated offences, I will argue that proponents of the reactive attitudes account have only one option: to understand forgiveness as conditional.

## V. Forgiveness as Conditional

---

<sup>15</sup> In the context of unfair blame, Pamela Hieronymi has already argued that one would be justified in using some means of self-deception or belief manipulation, so as not to consider a particular judgement when there are practical second-order reasons to do so (Hieronymi, 2004, p. 128).

In order to avoid the problem raised by the case of repeated offences, proponents of the reactive attitudes account need to understand the commitment to forgive as conditional on the truth of certain crucial beliefs.<sup>16</sup> Note that by describing forgiveness as ‘conditional’ I do *not* mean the more familiar idea that the *forgivee* must do something in order to earn forgiveness. What I am referring to is a condition related to a belief that the *forgiver* (not the forgivee) endorses when they forgive. One variant of this suggestion would be to argue that forgiving is conditional on the *belief* that the forgivee will not wrong the forgiver again. If the promise to forgive is conditional on this belief, the forgiver would not be bound by her promise, were the forgivee to wrong the forgiver again. But this proposal faces an immediate objection: it seems unjustified to rely on such an implausible belief, given that it’s unlikely that one can completely avoid wronging someone in the future.

In order to avoid this problem, we could specify the condition as the belief that the forgivee will not do a wrongdoing *of a similar type* again. If the promise to forgive is conditional on this belief, the forgiver would not be bound by her promise, were the forgivee to commit a wrong *of the same type*. This would allow the commitment to forgive to be dissolved in the case of repeated offences.

This is a promising suggestion but two problems would need to be addressed. To begin with, this suggestion would face an issue about identifying what counts as a wrongdoing of the same type. If I lied to you, I did something that broke your trust. But what would qualify as a similar wrongdoing, lying again or breaking your trust? The former might run the risk of not including enough wrongdoings whereas the latter might run the opposite risk of including too many possible wrongdoings.

Another issue would consist in ensuring that making forgiveness conditional isn’t at odds with the core idea of the reactive attitudes account of forgiveness. On the reactive attitudes account, the wrongdoing ought not to influence the way you regard the wrongdoer. But now with a conditional account, the forgiver essentially regards the forgivee as a possible recidivist. These two ideas might still be in tension. And treating forgiveness as conditional on the belief that the forgivee will not offend again would certainly be in tension with the metaphor favored by Allais in talking about forgiveness, the metaphor of ‘wiping the slate clean’. So this solution might not be entirely attractive to proponents of Allais’s account.

## **VI. Conclusion: What does the case of repeated offences tell us about forgiveness?**

I have argued the Reactive Attitudes Account of forgiveness has implausible implications when it is coupled with a naturalist interpretation of the Strawsonian Insight. In this paper, I raised two challenges against the reactive attitudes account of forgiveness. The first challenge is that current formulations of the reactive attitudes account seem to entail that, when we forgive, we commit ourselves not to consider the forgiven wrongdoing. I considered different possible responses and argued that only two of these were successful. One response would be to endorse the cognitive dissociation version of the account as this would allow the commitment in the speech-act ‘I forgive you’ to be restricted to the use of the wrongdoing as evidence (and not to the experience of the reactive attitudes themselves). The other response would be, for those who endorse the affective dissociation strategy, to take the commitment to be about *forswearing* the negative reactive attitudes and not overcoming them. I argued that, whereas a commitment to overcome the reactive attitudes might lead to the Overlooked Implication, a commitment to disavow such experiences is fully compatible with considering the wrongdoing again. However, I argued that these responses to the Overlooked Implication do not, as such, allow us to escape the deeper problem highlighted by repeated offences. This more fundamental problem is the epistemic vice that seems to be entailed in committing not to either ground our attitudes on the first wrongdoing or to use the first wrongdoing as evidence to

---

<sup>16</sup> For a similar suggestion with respect to the promise to love in marriage, see Moller (2003).

assess the character of the wrongdoer in the future. In order to address that deeper problem, I argued that proponents of the account needed to be more explicit about the fact that forgiveness ought to be granted under the condition that the forgiver believes that the offender isn't going to repeat their wrongdoing.

## Bibliography

- Allais L., 'Wiping the Slate Clean: The Heart of Forgiveness', *Philosophy and Public Affairs*, 2008a.
- Allais L., 'Dissolving Reactive Attitudes: Forgiving and Understanding', *The South African Journal of Philosophy*, 2008b.
- Bovens L., 'Must I be Forgiven', *Analysis*, 2009.
- Butler J., 'Fifteen Sermons', in Gladston W.E. (ed.), *The Works of Joseph Butler*, vol. 2, Clarendon Press, Oxford, 1897.
- Darwall S., *The Second Person Standpoint: Morality, Respect and Accountability*, Harvard University Press, Cambridge (Ma), 2006.
- Drake E., *Is Divorce Promise-Breaking? Ethical Theory and Moral Practice*, 2011.
- D'Cruz J. & Kalef J., 'Promising to Try', *Ethics*, 125, 2015.
- Eshleman A., *Moral Responsibility*, Stanford Encyclopedia of Philosophy, 2001.
- Fara M., 'Dispositions and Habituals', *Noûs*, 2005.
- Fischer J. M. and Ravizza, M. (eds.), *Perspectives on Moral Responsibility*, Cornell University Press, 1993.
- Gibbard A., *Wise Choices, Apt Feelings*, Clarendon Press, Oxford, 1992.
- Hampton J., 'Forgiveness, Resentment and Hatred', in *Forgiveness and Mercy*, Hampton J. and Murphy J., Cambridge University Press, 1990.
- Hieronymi P., 'Articulating an Uncompromising Account of Forgiveness', *Philosophy and Phenomenological Research*, 2001.
- Hieronymi P., 'The Force and Fairness of Blame', *Philosophical Perspectives*, 2004.
- Hobbes T., *Leviathan*, Menston, England: The Scholar Press Limited, 1961.
- Holmgren M.A., 'Forgiveness and the Intrinsic Value of Persons', *American Philosophical Quarterly*, 1993.
- Hurley E. and McNamara C., 'Beyond Belief: toward a Theory of the Reactive Attitudes', *Philosophical Papers*, 2010.
- Kahn L., 'Moral Blameworthiness and the Reactive Attitudes', *Ethical Theory and Moral Practice*, 2011.
- Kekes J., 'Blame versus Forgiveness', *The Monist*, 2009.

- Keltner D. and Haidt J., 'Social Functions of Emotions', in Mayne T. & Bonanno G.A. (Eds), *Emotions: Current Issues and Future Directions*, Guilford Press, New York, 2001.
- Kolnai A., 'On Forgiveness', *The Proceedings of the Aristotelian Society*, 1973-1974.
- Landau I., 'An Argument for Marriage', *Philosophy* 79, 2004.
- Lazarus R., *Emotion and Adaptation*, Oxford University Press, 1991.
- Lewis M. & Ramsey D., 'Infant Emotional and Cortisol Responses to Goal Blockage', *Child Development*, 2005.
- Martin C.B., 'Dispositions and Conditionals', *Philosophical Quarterly*, 1994.
- Marušić B., 'Promising against the Evidence', *Ethics*, vol.123 (2), 2013.
- McKenna M.S., 'The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism', *Journal of Ethics*, 1998.
- Michaelian K., 'The Epistemology of Forgetting', *Erkenntnis*, 2011.
- Moller D., 'An Argument Against Marriage', *Philosophy* 78, 2003.
- Moran R., *Authority and Estrangement*, Princeton University Press, Princeton NJ, 2001.
- Murphy J., 'Forgiveness and Resentment' in Hampton J. and Murphy J. (ed.), *Forgiveness and Mercy*, Cambridge University Press, 1990.
- Neblett W., 'Forgiveness and Ideals', *Mind* 83 (330), 1974.
- Nichols S., 'After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes', *Philosophical Perspectives*, December 2007.
- Norvin R., 'Forgiveness', *Ethics*, 1988.
- Pereboom D., 'Hard Incompatibilism and Response to Fisher, Kane and Vargas', in Fischer J., Pettigrove, G. *Forgiveness and Love*, Oxford University Press, 2012.
- Kane R., Pereboom D. & Vargas M., *Four Views on Free Will*, Blackwells, 2007.
- Russell L., 'Dispositional Accounts of Evil Personhood', *Philosophical Studies*, 2009.
- Shabo S., 'Where Love and Resentment Meet: Strawson's Intrapersonal Defence of Compatibilism', *Philosophical Review*, 2012.
- Strawson G., *Freedom and Belief*, New York, Oxford University Press, 1986.
- Strawson P. F., *Freedom and Resentment and other Essays*, Methuen & co, London, 1974.
- Wallace J., *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford, 2006.
- Wallace J., *Responsibility and the Moral Sentiments*, Harvard University Press, 1994.
- Warmke B., 'Is Forgiveness the Deliberate Refusal to Punish', *The Journal of Moral Philosophy* 8, 2011.



Warmke B., The Normative Significance of Forgiveness, *The Journal of Australasian Philosophy*, 2016.

Zaibert L., 'The Paradox of Forgiveness', *The Journal of Moral Philosophy*, 2009.

Zaragoza K., 'Forgiveness and Standing', *Philosophy and Phenomenological Research*, 2012.