In the Wake of Wrong:
Essays on the Ethics of Blame, the Reactive Emotions, and Apologies

by

Joseph E. Shin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in the University of Michigan
2022

Doctoral Committee:

Professor Sarah Buss (co-chair)
Professor Brian Weatherson (co-chair)
Professor Daniel Jacobson
Professor Maria Lasonen-Aarnio
Professor Peter Railton
Associate Professor Chandra Sripada

Joseph Shin

joeshin@umich.edu

ORCID iD: 0000-0003-2950-7705

For my wife, Katie.

Acknowledgments

I am indebted to members of my committee: Brian Weatherson, Sarah Buss, Peter Railton, Dan Jacobson, Maria-Lasonen-Aarnio, and Chandra Sripada for their guidance, inspiration, and encouragement over the years. Special thanks to my co-chairs, Brian Weatherson and Sarah Buss for going the extra mile in helping me complete this thesis. Brian Weatherson met with me on a weekly basis and his impact on my scholarship would be difficult to overstate. I also spent a significant amount of time discussing things with Sarah Buss who always pushed me towards greater precision both in thought and word. Peter Railton and Dan Jacobson each made extensive comments on early drafts which improved my work substantially. I am also grateful to Maria-Lasonen-Aarnio for helpful conversations at the earliest stages of this project.

Throughout this journey, my family has been a source of inspiration and support. I would not have had the chance to attend university let alone graduate school, if not for my mother Naomi's hard work, resilience, and immense sacrifice. I am also grateful to my late stepfather Ellis, who we lost along the way, for his love and support. Many thanks to my mother-in-law, Patti Edwards who has always been ready to celebrate my minor successes and to offer encouragement. I am thankful to my dearest friend Tim Lee for steadfastly being in my corner. Last, but not least, I thank my wife Katie who encouraged me to pursue my dreams and has supported me in countless ways through each leg of this journey.

I also benefited from numerous interactions with various students, faculty, staff, and visitors (both present and past) at Michigan. A special thanks to Jamie Tappenden for his help with the job market. I am grateful to Sonya Özbey and David Baker for their mentorship and advice concerning teaching and the job market. I am indebted to Carson Maynard for his administrative super powers and patience. Finally, a special thanks to Johann Hariman, Reza Hadisi, Sara Aronowitz, and Filipa Melo Lopes for their friendship and support over the years.

Contents

Abstract

Moral norms are bound to be broken. When they are broken (and we are aware of it), we respond in characteristic ways. For example, in light of our *own* mistakes, we are inclined to feel guilt (and/or shame and regret), apologize, and seek forgiveness. Similarly, we respond to the wrongdoing of *others* with resentment or indignation, sanctions, and/or by making various changes to our relationship with them.

As it concerns each of these typical responses to wrongdoing, not just anything goes. Some occasions render these reactions to appropriate. Other occasions render them inappropriate. There are even situations in which such responses are *required* of us. In other words, there are norms about how we *should* respond to the fact that an agent has not done as they *should*.

This dissertation is an exploration of some of these norms. Reflection on when we *must* blame, when we *musn't* (continue) resenting, and when we *should* and *need not* apologize, tells us interesting things about the nature of blame, blameworthiness, and normativity more generally. At least that is what I aim to show in the pages to follow. Here, I briefly outline the main ideas of each chapter.

The first chapter develops a new tool by which we can adjudicate between competing theories of blame. It begins with the observation that extant discussions of the nature of blame have tended to focus on blame that is (would-be) directed at others and on the conditions in which such blame would be inappropriate. I argue that we gain important insights by thinking about when it might be inappropriate *not* to blame ourselves and others. I present a series of cases which suggest the following thesis: Whenever one is in a position to appropriately do so, there is a normative expectation to self-blame, but there is not a normative expectation to blame others whenever one is in a position to appropriately do so. I refer to this as the Must-Blame-Asymmetry (MBA). Not only does MBA identify an important aspect of the ethics of blame, but it also provides a *desideratum* for theories concerning the nature of blame. As proof of concept, I consider some prominent theories including, George Sher (2006), TM Scanlon (2008, 2013), and Angela Smith (2013). I argue that each struggles to account for

MBA in different ways. Furthermore, I contend that MBA is evidence for affective theories of blame according to which to blame A for ϕ is to adopt a negative reactive attitude such as guilt, resentment, or indignation towards A for ϕ. I conclude with a discussion on how MBA might likewise inform us about the existence and/or the nature of epistemic blame.

In the second chapter, I argue that whether a wrongdoer must apologize (and self-blame) can depend on her moral beliefs. This will undoubtedly strike many readers as surprising. However, I start with the mundane observation that we occasionally fail to believe that we've done something wrong because we have a mistaken moral belief. As a result, in these situations, we do not apologize (self-blame). I argue that insofar as we are not culpable for this mistaken moral belief, we are not subject to moral criticism for failing to apologize (self-blame). In contrast, agents who are aware that they have done wrong are subject to moral criticism if they do not apologize (in the absence of overriding reasons). That is, there are cases in which what an agent ought to do depends on her moral beliefs. This is to argue for moral internalism concerning the permissibility of (some) actions/omissions. I further employ the observation that an ignorant wrongdoer need not apologize as part of a novel argument that mistaken moral beliefs can be exculpatory. That is, I argue for the controversial claim that agents who fail to believe that they have done wrong because of a mistaken moral belief (for which they aren't cupable) aren't blameworthy for what they have done. This is to argue for normative internalism concerning the blameworthiness of an agent. In this way, I use observations concerning the norms of apologies, in certain non-ideal cases, to argue for two kinds of internalist principles.

The third chapter explores how negative reactive attitudes which are commonly associated with blame, such as resentment, have interesting fittingness conditions. Whether it is fitting for a A resent B for ϕ (at t) can depend on how long how long A has been resenting B in the past for ϕ (relative to t). Such facts concern the resentor's attitudinal history rather than the object of her resentment. Thus, there are situations in which it is fitting for one person to resent A and not fitting for another person to resent A, *keeping fixed the action and time.* In other words, facts concerning the fittingness of object-directed emotions such as resentment are relational facts. I develop this insight into a challenge against theories which (i) aim to

analyze blameworthiness in terms of the negative reactive attitudes and (ii) maintain the standard assumption that whether an agent is blameworthy for ϕ is a non-relational fact. One of the upshots is that either blameworthiness can sometimes be relational (i.e., (ii) is false) or we have some reason to doubt affective theories of blame.

Chapter 1: Must Blame: Self vs. Other

Abstract

I consider two underappreciated phenomena pertaining to moral responsibility, namely, (i) self-blame and (ii) the moral inappropriateness of failing to blame. Reflection on each suggests the following asymmetrical norm of blame: Whenever one is in a position to appropriately do so, there is a normative expectation to self-blame, but there is not a normative expectation to blame others whenever one is in a position to appropriately do so. I refer to this as the Must-Blame-Asymmetry (MBA). Not only does MBA identify an important aspect of the ethics of blame, but it also provides a *desideratum* for theories on the nature of blame. As proof of concept, I consider some recent accounts of blame from George Sher (2006), TM Scanlon (2008, 2013), and Angela Smith (2013). I argue that each struggles to account for MBA in different ways. Further, I contend that MBA is evidence for affective theories of blame according to which to blame A for ϕ is to adopt a negative reactive attitude such as guilt, resentment, or indignation towards A for ϕ. I conclude with a discussion on how MBA might likewise inform us about the existence and/or the nature of epistemic blame.

1 Introduction

Self-blame is as commonplace as blame that is directed at others. Nevertheless, the growing body of literature on the ethics of blame has primarily focused on the norms concerning blame that is (or would be) directed at others. Further, many of the discussions about the norms of blame have to do with the conditions under which blame of a wrongdoer is permissible or appropriate as opposed to when it might be required or inappropriate *not* to blame. In this work, I show that there are interesting upshots to focusing our attention on self-blame as well as on the question, *when must we blame wrongdoers?*

In what follows, I argue that where the blame would be directed at another, there are situations in which we need not blame (even though it would be appropriate to do so). In contrast, I submit that whenever we

are in a position to appropriately do so, we must blame *ourselves*. Not only is this asymmetry interesting in its own right, but we can also put it to work as a *desideratum* for proposals on the nature of blame. To that end, I consider George Sher's (2006) conative theory, TM Scanlon's (2008, 2013) relationship modification view, and Angela Smith's (2013) proposal of blame as moral protest. I argue that each fails to account for the foregoing asymmetry. In contrast, affective theories of blame seem to correctly predict the asymmetry and so we have new evidence in their favor. Additionally, I submit that we can put such a principle to work in adjudicating between competing theories of epistemic blame. I briefly outline how such a project might go in relation to Jessica Brown's (2020a, 2020b) and Cameron Boult's (2021a, 2021b) proposals. Thus, there are theoretical benefits to thinking about the self-blame and the norms concerning when we *must* blame.

2 Background

There has been significant interest in two related questions pertaining to blame. The first is metaphysical— What is the nature of blame? The other is normative and falls under the domain of the ethics of blame. As it is standardly characterized, in asking the normative question, we are inquiring into the conditions that must obtain if blame is to be appropriate. There are two things to note about this construal of the line of inquiry. First, as Coates and Tognazzini (2013) note, 'appropriate' in the present context is equivocal (17). For instance, it could be that in asking whether it is appropriate to blame A for ϕ, we are interested in facts about the (would-be) blamed such as whether A is blameworthy for ϕ. Understood in this light, the central issue pertaining to whether it is appropriate to blame A for ϕ has to do with whether A exemplified certain capacities in ϕ-ing such as being in some sense free. On the other hand, much of the recent literature on the ethics of blame has centered around cases where it is stipulated that the transgressor is blameworthy and yet there is some moral impropriety in a particular subject's blame of them. For instance, it is thought that it can be morally inappropriate for B to blame A (for ϕ) insofar as B is likewise guilty of ϕ.[1] For our purposes, we'll be interested only in this second sense of 'appropriateness,' in relation to blame (or its absence). Thus, we'll be set-

[1] See Angela Smith (2007) and Marilyn Friedman (2013).

ting aside any substantive discussion concerning the conditions that render an agent blameworthy or morally responsible.

A second notable feature about characterizing the ethics of blame in the foregoing manner is that it is partial to the matter of when blame is (or would be) appropriate (or inappropriate) as opposed to when the *omission* of blame might be inappropriate. However, just as it is occasionally the case that blame of the blameworthy transgressor is inappropriate, there seem to be occasions in which the lack of blame is morally objectionable. Indeed, it's quite natural to say of a wrongdoer (who is aware that she is blameworthy), that *ought* to blame herself. It will be this aspect of the norms concerning blame that will be the central focus of the discussion to follow.

The inclination of theorists working in the ethics of blame to focus on situations in which blame (rather than its absence) would be appropriate, is understandable. Plausibly, we have a tendency towards blaming too much rather than too little.[2] Additionally, it may be more costly (in some suitable sense) to blame inappropriately than it is to fail to blame when we should. Thus, there may be a premium on getting the former conditions right. Still, I hope to show that there are payoffs to thinking more about the norms of self-blame as well as those occasions when the lack of blame is (would be) inappropriate.

Throughout this discussion, I will attempt to remain neutral concerning the metaphysics of blame. As it stands, there are currently several families of views, but nothing resembling a consensus. The issues surrounding blame's nature remain live issues and as I've suggested above, it is my hope that the main phenomena of interest in this work will provide a new means by which to compare some of these theories. So as not to prejudge the issue, I will be depending on the reader's raw intuitions about blame. The only substantive claims to which I will be helping myself in this respect is that naïve cognitive theories have got it wrong.

On a naïve cognitive theory, to blame another is nothing more than to judge that the person is worthy of blame. As a number of philosophers have noticed,[3] such views have receded into the background due to coun-

[2] There appears to be an analogous asymmetry of interest in the ethics of belief literature. That is, comparatively more is written on when it is inappropriate to believe as opposed to the conditions under which belief would be required (or the lack of belief is inappropriate).

[3] Angela Smith (2013) and Coates and Tognazzini (2012).

terexamples which many find decisive. I won't rehearse those cases here,[4] but rather will assume that whatever it is to blame A for ϕ-ing, it is not simply to form a belief or judgment that A is blameworthy for ϕ-ing.[5] Secondly, I will take it for granted that whatever blame is, self-blame and other-directed blame are not fundamentally different things. This isn't to deny that there are differences. For instance, it could be that self-blame consists of one set of negative reactive attitudes such as shame or guilt while blame when it is other-directed involves a different set of attitudes such as indignation, or resentment. Despite the differences between these emotions, they are still the same kind of phenomenon—i.e., negative reactive attitudes.

3 Inappropriate Lack of Blame

Consider the following vignettes.

> *Line Cutting:* Fred is running late to work and needs to stop by the bank. Nothing terrible will happen if he is late for work, but he knows that his boss will be a bit annoyed with him and he doesn't want that. Fred gets to the bank and notices the lines are long. He decides to casually cut in front of another patron. The patron confronts him angrily, but Fred decides simply to ignore them. When Fred gets to work, he relays the incident to his coworkers in detail. Sharon, one of the coworkers who hears what has happened, knows that what Fred did was wrong and feels sympathy for the person(s) he slighted. Further, she knows that he is worthy of blame. Hence, she doesn't think there is anything amiss about others blaming him. However, Sharon thinks nothing more of the incident, and in fact, doesn't even blame Fred for his misdeed.

> *Academic Dishonesty:* Michael an undergraduate was caught in the act of academic dishonesty. He knew full well that it was wrong to do so, but rather than write a paper over the weekend, he decided to go on a road trip with his friends. Michael was caught by his instructor and when confronted, admitted that he used a paid service that wrote the paper for him. Michael's friend Jan is aware of all these facts and knows that what he did was wrong. Further, she knows that Michael is blameworthy for what he has done. Thus, she doesn't think there

[4] Kenner (1967), Watson (1987), and Coates and Tognazzini (2012).

[5] In fact, I take it that the central case of this work is a counterexample against such cognitive theories as proponents of such theories of blame would be forced to say that there is something incoherent about my cases in that they include wrongdoers who believe that they are blameworthy for ϕ and yet fail to blame themselves for ϕ. That seems to me a bullet to bite.

is anything amiss about the instructor blaming him. However, Jan doesn't blame him.

Assault: Oliver is walking to work when he sees two of his neighbors (Robin and Patrice) in a heated dispute. Oliver attempts to de-escalate the situation. He comes to learn that Robin is angry because Patrice recently had a party and the attendees parked in front of Robin's house. Patrice suggests that Robin is making too big of a deal out of a minor inconvenience. Robin becomes frustrated and physically assaults Patrice by shoving her to the ground. Oliver is eventually able to break up the altercation and all parties go their separate ways. Upon reflection, Oliver knows that what Robin did was wrong and feels sympathy for Patrice. He also knows that Robin is blameworthy for her actions. Hence, he doesn't think there is anything amiss about Patrice blaming her. However, Oliver doesn't find himself blaming Robin.

Each vignette features an onlooker (Sharon, Jan, or Oliver) who is aware of an agent's wrongful action. Further, despite knowing that they are worthy of blame for the action, each observer fails to blame the respective agent. The type and severity of the wrongs vary. However, it doesn't seem that any of the onlookers are failing in any moral sense. Support for this claim can be found in reflecting on what we would expect if it were the case that our witnesses *must* blame. Plausibly if Sharon, Jan, or Oliver must blame and fail to do so (and there are no excusing conditions for their lack of blame), it would be appropriate for others to criticize them and perhaps blame them solely for their lack of blame. However, it's implausible that they are subject to blame or any other kind of moral opprobrium merely for their lack of blame.

Indeed, if their lack of blame constitutes a moral shortcoming, it would render each of our protagonists the apt target of self-directed reactive attitudes such as guilt. But that strikes me as incorrect. Suppose that following the attack, Oliver is reflecting on the fact that while he feels sympathy for the victim, and knows that Robin is blameworthy, he doesn't blame Robin. It hardly seems fitting for him to feel guilty simply for his lack of blame. If he were to confide in you as a friend that he was struggling with feelings of guilt (over the fact that he did not blame Robin), you would be inclined to try and talk him out of it and to remind him that he has nothing for which to feel guilty. Things might be different if Oliver, despite being confronted with overwhelming evidence of Robin's wrong and culpability, failed to accept that she did wrong or that she was culpable

for doing so.[6] However, as the vignette stipulates, Oliver is aware of the wrongdoer's blameworthiness as are the other onlookers of the other cases.

The foregoing observations become more pronounced when we consider, as a point of contrast, how we feel about the wrongdoers in each case when they fail to self-blame. Suppose that Robin (the attacker) in *Assault* doesn't blame herself. Again, we can stipulate that she is aware of all the relevant facts. That is, she knows that what she did was wrong and that she is worthy of blame for her assault on Patrice. Nevertheless, upon reflection she finds that she simply doesn't blame herself. Unlike Oliver, it seems to me that she is morally criticizable. We are inclined to view her with moral suspicion. It's appropriate for others to blame her not only for the assault, but also for her lack of self-blame. Relatedly, suppose that during a reflective moment, she feels guilty about not blaming herself for the attack. Such a response seems fitting in her case. If she confided in you (as a friend) about her negative feelings, you would not be inclined to suggest that she has nothing for which to feel such things. In fact, one might get a sense of relief knowing that while she didn't blame herself for the attack, at least she feels guilty about that fact. What is more, all of this seems to generalize to the onlookers of our other cases.

Thus, reflection on cases like *Line Cutting*, *Academic Dishonesty* and *Assault*, suggest that there is something morally objectionable about subjects who fail to self-blame despite being in the position to appropriately do so.[7] In contrast, it doesn't appear that the same is true of counterparts who fail to blame others despite being in a position to appropriately blame them. This suggests the following thesis:

> *Must Blame Asymmetry (MBA):* Whenever one is in a position to appropriately do so, there is a normative expectation to self-blame, but there is not a normative expectation to blame others whenever one is in a position to appropriately do so.

MBA presents a "normative expectation." It also appeals to the idea that a

[6] Why might his failing to believe that the agent is blameworthy render him morally criticizable? Plausibly, to deny that an agent is blameworthy might indicate that one excludes the agent from the moral community (or as one that is the sort of being that could be responsible for their actions). When this is not justified, then it may be morally suspect. See Pamela Hieryonmi (2001). Alternatively, it is often an objectionable slight against the victim to deny that any wrong has been done to them.

[7] More on this below.

subject can be *in a position to appropriately blame* someone. Each merits a brief remark.

I use 'normative expectation' here as a term of art. There is a normative expectation that A perform or instantiate ϕ just in case A is subject to moral criticism for failing to ϕ.[8] Moral criticism of the relevant sort may take many forms including but not limited to various kinds of evaluative judgments, the negative reactive attitudes, and/or blame. Thus, it is compatible with MBA that there are full-blown moral requirements or obligations to blame. After all, A could be subject to various kinds of moral criticism for failing to meet a moral requirement/obligation. However, fans of a certain kind of *ought implies can* principle and affective or conative theories of blame might have the following worry. Given that blaming is not within our direct control, it can't be something we are required to do. Whatever one's take on such issues, it is independently plausible that there are occasions when an agent is subject to moral criticism for having certain problematic attitudes (e.g., offensive beliefs), vices and/or traits even when these are not within the agent's direct control.[9] Likewise, it seems we can be subject to opprobrium for lacking certain attitudes, virtues, and traits. For example, failing to have compassion for the suffering or failing to believe that all persons are owed basic respect seems to open one up to moral criticism. Construing MBA in terms of a normative expectation to blame, helps us avoid getting bogged down in these controversies. Note that in referring to the normative expectation to self-blame, I'll often say that we *ought* to, *should*, or *must* blame. This is not only a matter of convenience but also because it accords with common ways of speaking. To reiterate, all I mean when I say that A ought/should/must blame B, is that she is subject to moral criticism if she fails to blame B. That is, in employing these expression, I do not intend to

[8] According to Angela Smith's (2012, 578) "rational relations view," agents can be morally responsible for their attitudes (beliefs, desires, emotions) provided that these attitudes reflect the agent's underlying evaluative judgments. Plausibly, just as it can be morally problematic for an agent to have certain kinds of evaluative judgments (e.g., that non-whites are inferior), it seems as though it can be troublesome for agents to lack certain evaluative judgments. For instance, consider an agent who fails to judge that the rights of others constrain her actions in anyway.

[9] Gary Watson (1996) suggests that there are two "faces" of moral responsibility namely, attributability (which tracks aretaic appraisal) and accountability. Thus even if it doesn't make sense to hold people accountable for things which are beyond their direct control, perhaps we can still hold them responsible in another sense. Also see Macalaster Bell (2013a, 21) who suggests something similar in response to the worry of the lack of control in her discussion concerning the ethics of contempt.

track a requirement or obligation in the strict deontic senses.

Importantly, MBA does not entail that a wrongdoer who fails to blame herself is always morally criticizable. That would be implausibly strong. Instead, according to the thesis, it is *wrongdoers who are in a position to appropriately blame themselves* that are always criticizable, if they fail to self-blame. What is it to be in a position to appropriately blame? By this I mean the following: If the subject were to blame the wrongdoer in question, their blame would not be morally inappropriate. It has been suggested that there are many factors which can render A's blame of B, morally problematic. As we noted earlier some authors[10] contend that if A has also performed ϕ, then A's blame of B for ϕ-ing, will be morally objectionable on pain of hypocrisy. Similarly, some[11] argue that where A is complicit in B's ϕ-ing, it would be morally problematic for A to blame B for ϕ-ing. Furthermore, it could be that a person fails to blame a wrongdoer (perhaps herself) because she is under great and ongoing duress which precludes her from blaming. Or it could be that to blame oneself or another could lead to very bad consequences. In each case, the would-be-blamer may not be in a position to appropriately blame the wrongdoer.

It's also plausible that there are epistemic conditions that can make a difference to whether an instance of blame is appropriate in the relevant sense. Even if a wrongdoer is blameworthy for ϕ, it would be morally problematic for me to blame them unless I am suitably connected to such a fact. Cases in which agents are morally reckless due to poor epistemic positions abound. If I do not have sufficient reason to recommend to my students a particular career path, then I do something wrong in recommending it (even if such a recommendation turns out to be correct). If A truly believes that B is blameworthy for ϕ-ing, but lacks sufficient evidence for the belief, then it seems that A is subject to moral criticism for blaming B for ϕ. Plausibly, for our blame of B to be morally appropriate[12] we must know[13] or at

[10] Angela Smith (2007), Marilyn Friedman (2013), G.A. Cohen (2006), T.M. Scanlon (2008), Kyle Fritz and Daniel Miller (2018).

[11] See Victor Tadros (2009), Anothony Duff (2010), and Gary Watson (2015).

[12] It could be that there are other senses of appropriate that are also impacted by the blamer's epistemic position in relation to such facts.

[13] See Christopher Kelp (2020) for a knowledge norm of blame. It should be noted that while Kelp discusses the conditions under which blame is appropriate, he is not explicit about whether he means moral appropriateness. This appears to be a common omission in the relevant literature.

least justifiably[14] believe[15] that B is blameworthy for ϕ.[16]

Importantly, each of our cases (*Line Cutting*, *Academic Dishonesty*, and *Assault*) stipulate that the would-be-blamer *knows*[17] that the wrongdoer is worthy of blame. Moreover, there's no reason to think that we are reading into these cases that the witnesses of wrongdoing are either complicit in or guilty of the same wrongs as the perpetrators. Nor is there is any reason to think that the situations are unusual in the sense that blaming will lead to serious harms. As such, in each vignette, both the onlooker and the agent is in a position to appropriately blame the wrongdoer.

MBA concerns a normative expectation that wrongdoers self-blame whenever they are in a position to appropriately do so. In contrast, it is not the case that there is the same normative expectation to blame others (whenever we are in a position to appropriately do so). In fact, this way of putting things downplays the extent of the contrast. Situations in which we must blame others are quite rare. This may strike one as initially surprising. However, recall that we're focusing our attention on situations in which the would-be-blamer knows that the wrongdoer has done wrong and is blameworthy. Often we fail to blame the blameworthy because we fail to believe that they are blameworthy. Perhaps we believe that they did the wrong thing but have a sufficient excuse or we believe that they have done no wrong in the first place. Where these judgments are poorly formed or due to favoritism or biases, the resultant lack of blame can be morally objectionable whether the blame would have been self-directed or directed at others. But again, we're setting such occasions aside. Once we've done

[14] See D. Justin Coates (2016) for something like this view.

[15] See Lara Buchak (2014, 299-305) for a view that outright belief is a norm of appropriate blame. Buchak seems to be talking about the moral appropriateness/inappropriateness of blame as she speaks of it being wrong to blame someone when you don't (outright) believe that they are blameworthy. However, she isn't explicit about this.

[16] There are interesting questions here to pursue but which would take us too far afield. For instance, must the agent be blameworthy? What should we say about situations in which B is not blameworthy for ϕ, and yet A has strong evidence that B is blameworthy for ϕ?

[17] I take for granted that we can know that an agent is blameworthy which is to set aside metaphysical worries such as determinism and incompatibilism as well as epistemic concerns as found in Gideon Rosen (2004). Further, by stipulating knowledge of the relevant facts, we set aside worries that the asymmetry in MBA is to be explained away by appeal to an epistemic difference i.e., the view that a wrongdoer is in a better epistemic position with respect to her culpability for her wrongs vs. the culpability of others. In each of our cases and subsequent discussions of them, it is stipulated that both the wrongdoer and the onlooker knows that the latter is worthy of blame for what they have done.

so, it seems difficult to find a case in which an agent is subject to moral opprobrium for merely failing to blame others.

Still there seem to be at least some occasions in which we must respond with other-directed blame. Consider again *Assault*. Imagine that in addition to Oliver (the on-looker) another person has just witnessed Robin shoving Patrice. This other person is Kevin who happens to be Patrice's husband. Assuming that he is privy to all of the facts and judges Robin to be blameworthy for assaulting his wife, it seems plausible that ought to blame Robin. It wouldn't be enough that he merely helps his wife or defends her from further assault. Nor does it seem a notable improvement that he feels disappointed with his wife's attacker and sad for his wife. Kevin *should* feel or do something more specifically towards Robin. Plausibly, this "something more" amounts to or entails blame. It seems reasonable that in the absence of blame, one would have reason to suspect that he doesn't care enough about his partner.[18] Likewise, it would be appropriate for Patrice to feel slighted and to take offense if she were to learn that her husband did not blame her attacker. It seems apt for Kevin to apologize or at least answer in some way for his lack of blame.[19]

Up to this point, in speaking of other-directed-blame, we've been focused on blame that would be from a third-party. However, what of the victims of wrongs? Must they blame their transgressors? Returning to *Assault*, suppose that following the physical attack from Robin, her victim (Patrice) managed to forego blame altogether. Perhaps Patrice finds that

[18] Might this suggest that we're tracking a norm pertaining to what it means to be a good partner here? Suppose it does then we are left with at least two options, neither of which poses trouble for MBA. First, it might be that we are tracking a norm of a particular kind of relationship. Perhaps Kevin is being a bad romantic partner insofar as he doesn't care sufficiently about his wife to blame her attacker and there is nothing more to it. In that case, it might not be morally problematic for him not to blame Robin. But then we've simply uncovered another kind of situation in which one need not (morally speaking) blame others (even though it would be appropriate to do so) which is compatible with MBA. Alternatively, one might think that when someone fails to blame those that harm their loved ones, one doesn't merely violate a norm of a relationship, but in so doing also violates a moral norm. Think about parents who don't care sufficiently about their children for instance. It seems to me that not only are they being bad parents, but in virtue of such bad parenting, they are morally criticizable. Not only is all of this compatible with MBA, but it supports the original point here that there is (under special circumstances) a normative expectation that one blame others.

[19] What about the perpetrators of particularly egregious wrongs? It's plausible that regardless of our relationship to such wrongdoers or their victims, we must at least in some of these cases blame. However, these are special cases.

eschewing blame in such cases makes it much easier for her to move on from the situation. Alternatively, she may simply be overcome with something like disappointment, sorrow, or even compassion for her attacker. In either case,[20] Patrice doesn't seem subject to moral criticism for her failure to blame.[21] She doesn't owe anyone an apology simply for not blaming her attacker and she doesn't have anything for which to feel guilty. We can say the same thing about situations in which one is the victim of a minor wrong such as *Line Cutting*. Patrice is no less permitted to refrain from blaming Robin had the latter simply broken a minor promise as opposed to physically assaulting her. In fact, in general it doesn't seem as though the victims of wrongdoing are under any sort of normative expectation to blame their transgressors whenever they can appropriately do so.[22] Importantly, things might be different in cases where the wrongdoer has committed atrocities or when the would-be-blamer (who may be a victim) is related to (other) victims in a certain way. However, barring such special circumstances, there does not appear to be any normative expectation for second-person blame.[23]

Taking stock, whether or not the wrongdoing is serious, there is a normative expectation that wrongdoers blame themselves (whenever they are in a position to appropriately do so). In fact, it seems rather difficult to encounter situations in which it is appropriate for the relevant kind of

[20] One might wonder whether these amount to overriding moral reasons which undermine the normative expectation that one should blame one's victimizers. However, notice that if Patrice's attacker were to not blame himself for any of these reasons, he would remain subject to moral criticism for not self-blaming. MBA explains this asymmetry in a straightforward way.

[21] What if she never blames her victimizers? Might our victim be morally criticized for being a "door mat"? I don't find this plausible, but I suspect some readers will. Nevertheless, even if we grant this point all that follows is that the victims of wrongs must *sometimes* blame their assailants (when they are in a position to appropriately do so) which is compatible with MBA.

[22] The victims of wrongs appear to have a certain kind of leverage as it concerns blame even as it concerns when it is appropriate for others to blame their victimizers. For instance, some authors have noticed that it seems inappropriate for me (as a third party) to go on blaming a wrongdoer for ϕ, when I know that the victim has forgiven them for ϕ. The victim appears to be in a position to criticize me and cite the fact that they have forgiven their perpetrator and that this is grounds for me to refrain from blaming them.

[23] There is a separate and related class of cases where it seems we must blame others due to considerations of fairness or consistency. For example, suppose that you know that both student A and student B are equally blameworthy for cheating on an exam. Perhaps they were co-conspirators who played an equal role in the misdeed. You, as a casual witness might be under the normative expectation to blame B insofar as you blame A. However, this is compatible with the fact that you needn't blame anyone, full-stop. Hence, these cases do not conflict with our observations so far.

agent to forego self-blame. On the other hand, we must blame others only under special circumstances as when the would-be-blamer bears a special relationship to the victim of a non-trivial wrong. This is just what we would expect if MBA were true.

4 Theories of Blame

We've just seen some initial reason to accept MBA. That is, we've seen evidence of an interesting asymmetry concerning the ethics of blame.[24] Further, it appears that situations in which we must blame others (even when we're in a position to appropriately do so) are special cases. In this section, I employ these observations as a new tool to help us think through some popular theories on the nature of blame.

4.1 Blame as a Response to Frustrated Desire

According to Sher, blame at its core is a belief-desire pair. To blame is to be disposed to react in certain ways (attitudinally and behaviorally) in virtue of a certain belief and strong[25] desire. The relevant belief is that the agent has behaved badly (i.e., violated a moral norm). The relevant desire is that the agent not have behaved badly (i.e., that she had not violated a moral norm). For Sher, when we have the desire that an agent not behave badly and come to believe that they have behaved badly, the desire has been frustrated (105). Further, assuming that the relevant desire is sufficiently strong,

[24] Reflecting on the normative expectation that we blame ourselves also provides other insights in relation to the ethics of blame. For instance, Kyle Fritz and Daniel Miller (2015) contend that what goes wrong in hypocritical blame is that the would-be-blamer instantiates a differential blaming disposition (i.e., they are inclined to blame others for ϕ and not themselves) which is a kind of implicit rejection of the equality of persons. However, it does not seem inappropriate for a subject on pain of hypocrisy to (be disposed to) blaming only herself for something for which multiple wrongdoers are involved. Further, Patrick Todd and Brian Rabern (forthcoming) argue that self-blame is paradoxical. The basic idea behind their argument is that in blaming oneself for ϕ-ing, the blamer is violating the anti-hypocrisy condition on blame. That is, since the blamer in this case is also guilty of the thing for which she is blaming herself, her self-blame is going to be inappropriate. Todd and Rabern maintain that the thing to say here is that we are never in a position to appropriately blame ourselves. However, cases such as *Line-Cutting*, *Academic Dishonesty*, and *Assault* suggest there is no paradox to self-blame. The wrongdoers should blame themselves in which case it can't be morally inappropriate for them to do so.

[25] It is crucial to Sher's theory that the desire is suitably strong. More on this below.

we are disposed to respond to its frustration, in characteristic ways.[26] For example, we will be disposed to respond with reactive emotions, reproach, hostile behavior, and the like (94). Blaming, on this "two-tiered" account is to be disposed to react in such ways to a (strong) desire that has been frustrated.[27]

One of the virtues of Sher's account is that it can make sense of how an apparently disparate set of reactions (a variety of negative reactive emotions, varying behavioral modifications) commonly associated with blame are related to one another. Indeed, Sher takes it that a suitable theory of blame should provide a story about what it is that unifies the set of reactions commonly associated with blame (98). On the present theory, they are each characteristic responses to the frustrated desire that an agent not behave badly. Relatedly, the account allows that blame can be manifested in a variety ways. That is to say, according to Sher, the manner in which an agent responds to the belief that she has done wrong differs from how she might respond to the judgment that another has done so. Despite these virtues, it's not clear that the view can correctly predict MBA.

On Sher's theory, what might explain why in *Academic Dishonesty* Michael (but not Jan) is subject to moral criticism for his lack of blame? There appear to be two options. Firstly, perhaps we read Michael as having the relevant desire and belief (and thus the frustrated desire). However, upon learning that he doesn't blame himself, we judge that he's not disposed to react in characteristic ways to the desire being frustrated. Secondly, perhaps in learning that he doesn't blame himself, we judge Michael (and not Jan) as lacking the relevant desire. That is to say, we judge that Michael doesn't have a strong desire that he not have acted badly. Importantly, for either of these explanations to suffice, they need to make sense of the fact that Michael is subject to moral criticism for his lack of blame even though Jan is not for hers.

Concerning the first option, it isn't clear why Michael would be subject to moral criticism merely for lacking the natural response(s) (or relevant

[26] Provided that we are psychologically typical

[27] Sher doesn't seem to discuss cases in which one has the desire that an agent not have behaved badly, and yet one falsely believes that an agent has done wrong. Surely, we can blame in these cases even if our blame may not be appropriate. However, it doesn't seem apt to say that my desire has been *frustrated* in such cases. Perhaps it's the *appearance* of a frustrated desire that is relevant?

dispositions) that one might have to a frustrated desire. Undoubtedly, it's an oddity, but we aren't subject to moral criticism for mere oddities. Suppose you desire a Boston cream donut and learn that all the local donut shops are sold out. It would be characteristic for you to (be disposed to) feel disappointment, but you're not subject to moral criticism for not feeling disappointed (or for not being so inclined).

Does the second option fare better for Sher? Plausibly, in learning that Michael doesn't blame himself (even though he knows that he is culpable), we suspect that he doesn't sufficiently desire that he not have behaved badly. Here again, one wonders if an agent can be subject to moral criticism merely for lacking such a desire. A fuller look at Sher's account of blame may make this idea more appealing. According to Sher,"...anyone who is fully committed to morality must have the sorts of desires that are constitutive of blame" (126). Further, in recognizing that moral principles are universal, you blame an agent *only if* you desire that the agent not have violated the relevant norm. He writes:

> Thus given that all moral principles apply to all persons, we may indeed conclude that whenever someone accepts a principle as moral (as opposed, say, to embracing it simply as a personal maxim of conduct), he must have not only a motivationally effective desire to obey it himself, but also a variety of motivationally ineffective desires that others obey it as well." (126)

Sher concludes that "...anyone who fully accepts a moral principle must react to wrongdoing and vice by having the relevant blame-constituting desires" (127).

Thus, Sher might say that Michael's lack of blame amounts to a lack of commitment to a moral principle or is evidence of such. It's initially plausible (although far from obvious) that one can be subject to moral criticism for not being committed to a moral principle. This would explain our reaction to learning that Michael doesn't blame himself despite being in a position to appropriately do so. However, we also need to inquire about our reactions to Jan. If Michael's lack of blame constitutes (indicates) a lack of the blame-constituting desire, then plausibly, so does Jan's. Indeed this is what is suggested by Sher's remarks featured above. In turn, if the lack of this desire in Michael constitutes (indicates) a lack of full commitment to morality for which he is subject to moral criticism, then the same is true

of the lack of desire in Jan. Hence, on Sher's account, we'd expect Jan to be subject to moral criticism in just the way that Michael is. If that's right, then Sher's theory would render MBA false. Indeed, as the above passage from Sher suggests, on his theory, we ought to blame ourselves as well as others whenever we're in a position to appropriately do so. That is to say, his account overgenerates.

Notice that the issue for Sher in relation to MBA arises from two of his commitments. The first is that for one to be fully committed to a moral principle in the relevant sense is to have a strong desire that the agent (regardless of whether the agent is oneself or another) not have behaved badly. The second commitment is that this desire is a constituent of blame. Thus, perhaps Sher could deny one of these claims in light of MBA. To deny the second of these claims is to surrender the central aspect of his theory of blame and so it's a non-starter. Can Sher account for our reactions to *Academic Dishonesty* by denying the first of these commitments? That is, perhaps Jan can be fully committed to a moral principle as long as Jan has a strong desire that *she* not violate it. However, Jan need not have a strong desire that *others* not violate the principle for her to be fully committed to it. This would explain why we find something morally problematic about Michael's lack of blame and not Jan's. Only Michael's lack of blame suggests (constitutes) a lack of commitment to the relevant moral principle.

Unfortunately, this is an awkward position for Sher to defend. As we've already noted, Sher argues that there is an intimate connection between a full commitment to a moral principle on the one hand, and the strong desire that agents not violate the principle, on the other. Importantly, he also takes for granted certain "formal features" of morality such that moral principles are universal, omnitemporal, overriding and inescapable (123). For our present purposes we need only focus on the omnitemporal and universal aspects. According to Sher moral principles are universal in that they apply to all persons and across all circumstances. Further, such principles are omnitemporal in that they apply at all times. For Sher, to be fully committed to a moral principle, is to be committed to it in such a way that the desire reflects all of the formal features of morality. For instance, consider the feature of omnitemporality. It appears to be Sher's view that if I merely desire that agents avoid murdering on Tuesdays, then I am not fully committed to the principle that agents should not murder *qua* moral

principle (126). If I am fully committed to the moral principle as such, then I will desire (strongly) that no one murder *at any time (past, present or future)*. Likewise, with respect to universality, if I merely desire that you (but not I) refrain from murdering, then I am not fully committed to the principle *qua* moral principle. Notice that it's also the case that if I merely desire that *I* (but not you) refrain from murdering, then I am not fully committed to the relevant moral principle as such.

Now suppose that we revise *Academic Dishonesty* in the following way. After having committed his wrongful action, Michael′ has been very distracted from reflecting on it. So unlike his twin Michael, Michael′ doesn't ever think about the fact that he is culpable for his wrongful act and so he doesn't blame himself. It is now ten days after his misdeed and it strikes him that what he did was wrong and that he is without excuse. However, Michael′ has peculiar desires. While he has a strong (blame-constituting) desire that he not have behaved badly in the *immediate* past, he has a fairly weak (non-blame constituting) desire that he not have acted wrongly in the past seven or more days. So despite his realization, he doesn't blame himself because he presently doesn't have a sufficiently strong desire that he not have cheated (ten days ago).

Michael′ seems no less subject to moral criticism for his lack of blame than his twin. Sher can account for this by noting that the former's weak desire (which explains why he does not blame himself) amounts to not being fully committed to the relevant moral principle. This is because, according to Sher, to be committed to such a principle *qua* moral principle is to desire that it not be violated *at all times*. In other words, Michael′'s desire does not reflect the omnitemporality of the relevant moral principle (that he not cheat). Importantly, if Michael′ is not fully committed to morality because his desire does not reflect one of the "formal features" of morality to which Sher appeals (i.e., omnitemporality), then it would be surprising that Jan (in the original case) could be fully commmitted to morality even though her desire doesn't reflect one of the "formal features," namely, universality. Recall, that we're presently considering the view that Jan's (and not Micheal's) lack of blame suggests (constitutes) a lack of full commitment to a moral principle. Thus, it seems that Sher cannot account for MBA by adopting the view that Jan's lack of blame-constituting desire is compatible with her

being fully committed to the relevant moral principle.[28] This is true, unless Sher is willing to relax what he argues is a tight connection between the "formal features" of morality on the one hand, and blame-constituting desire, on the other. The upshot is that Sher's theory of blame does not appear to be in a good position to account for MBA.[29]

4.2 Blame as Relationship Modification

For Scanlon (2008,88), a relationship between two rational agents is constituted by a "set of expectations and intentions about how we will behave towards one another" which each party to the relationship (perhaps tacitly) endorses. Moreover, relative to each relationship, there are facts about what sorts of expectations and intentions it would be appropriate for each member of the relationship to have towards the other—what Scanlon calls the "normative ideal". A blameworthy agent is one that violates such norms so as to undermine the relationship and in so doing, alters the kind of expectations and intentions, which would be appropriate for others to have or display towards the wrongdoer. According to Scanlon, you blame someone just in case you (appropriately) modify your intentions, dispositions, and/or expectations in response to the judgment that the wrongdoer has

[28] In fact, Sher explicitly denies this possibility in the following passage. "This asymmetry, if it existed, would be damaging because it would block the conclusion that being fully committed to a moral principle means having desires that support the full range of blame-constituting disposition. However in fact, the asymmetry *cannot* exist; for the possibility that someone who was fully committed to a moral principle could care less about disobedience by some person than by others...is ruled out by variants of our previous appeals to morality's universality..." (128-29).

So why haven't I used the above as a proof text to answer the present worry? In this passage, Sher is considering what might *justify* blame. Sher appeals to the formal features here to explain what can make blame appropriate. So the tension of which he speaks in this quote is between his theory of blame, on the one hand and his theory of what renders blame appropriate, on the other. Given that one could in principle adopt Sher's theory of blame without also accepting his theory concerning what justifies blame, simply citing this passage would not suffice to show the problem with the current proposal.

[29] To be sure, this is to raise a problem for Sher's conative theory of blame (which is the only one of which I am aware). It could be that a different desire-based account of blame could be developed in light of our reflections which avoids the foregoing problems. Perhaps a person can have a blame-constituting desire that a norm not have been violated (so as to be fully committed to morality) without having a strong desire that *all others* not violate it. Or perhaps the matter is about evidence. It could be that Michael's lack of self-blame is sufficient evidence that he lacks the relevant desire whereas the same is not true of Jan's lack of other-directed blame. Still, MBA will have proven useful here in getting us to refine the original proposal developed by Sher.

impaired their relationship with you or others (128-29).

To illustrate, Scanlon (2013) mentions the case of blaming a disloyal friend. Friendships are constituted by a certain set of expectation and intentions. For instance, friends should be disposed to confide in one another and to be trustworthy. Suppose that a close friend has violated your trust and shared what you told to them in confidence. To blame such a person might amount to "suspending one's normal intentions to trust the friend and confide in him or her and assigning a different meaning to one's interactions" (89). What makes blame variable on this account as opposed to identifiable with any particular set of attitudes, behaviors, or dispositions is that what counts as "modify one's relationship appropriately" depends on the nature of the relationship. Thus, for example, blaming a casual acquaintance will likely look quite different from blaming a family member or close friend.

Should we expect a self/other asymmetry concerning whether you must blame on Scalon's account? Consider again *Academic Dishonesty*. For Scanlon, Michael will be blameworthy just in case he has impaired his relationship with others or himself in virtue of his wrong. This means that on-lookers like Jan (who judge him to be blameworthy) have reasons to alter their intentions and expectations in certain ways i.e., she has reasons to modify the nature of her relationship to Michael. Does Michael too have reasons to modify the nature of his relationship to himself? If he judges that he is blameworthy for what he has done, he is judging that he has impaired his relationship with others and this for Scanlon can give him reasons (indeed the very same reasons as Jan's) to modify the intentions and expectations he has towards himself. Scanlon (2008) writes,

> One can make a judgment of blameworthiness about oneself as well as about anyone else, friend or stranger. . . But when the person is oneself, and the judgement is about one's own relations with others specifically about the attitudes they have reason to hold toward one, this gives rise to special concern, regret, and a desire to change things. These responses constitute blame of oneself: because of one's own attitudes toward and the treatment of others, one can no longer endorse one's own feelings and actions, but must instead endorse the criticisms and accusations made against oneself by others (154)

According to Scanlon, in recognizing that I have impaired my relationship with others, I must change the nature of my relationship with myself by

"endorsing the criticisms and accusations made against oneself by others." To put the matter in terms of reasons, in *Academic Dishonesty*, Michael has reasons to alter his relationship towards himself, which is to say that he has reasons to blame himself, just as Jan has reasons to blame him. What is the nature of such reasons? This is unclear.[30] However, for Scanlon's account to have a fighting chance at correctly predicting MBA, we should construe these reasons as moral reasons. To see why, suppose that on Scanlon's view judging someone to be blameworthy (i.e., judging that they have impaired their relationship with others) merely gives one non-moral reasons to blame. Such a proposal would not be able to account for how we can be subject to moral criticism for the lack of blame. That would be of no help in getting to MBA which posits an asymmetry concerning when failures to blame are morally problematic.

Hence, let's suppose that on Scanlon's view, judging someone to be blameworthy gives one moral reasons[31] to blame. Can it account for MBA? I suspect not. If Michael and Jan both judge Michael to be blameworthy, then on the current proposal, they each have the very same moral reasons to blame Michael. Of course, what it looks like for each to blame Michael will vary given the nature of the relationship they each bear to the wrongdoer. However, what we care about is whether each must blame Michael i.e., whether each must respond to the moral reasons they have to modify their relationship to the wrongdoer (i.e., blame).[32]

[30] See Christopher Bennett (2013) for a discussion of the significance of this ambiguity.

[31] Bennett (2013) suggests this sort of disambiguation of Scanlon's view. While Bennett's theory seems to make sense of why it might be wrong to eschew blame altogether (which is his central aim), it also seems to faulter in predicting MBA. On Bennett's account we have ethical reasons to blame wrongdoers whom we judge to be blameworthy. The root of such ethical reasons to blame is that in blaming we value beings with moral status who are the victims of wrongdoing. Hence, to fail to blame those whom we judge to be blameworthy, we fail to sufficiently value others, which is a moral failure. However, on such a view, whether the wrongdoer is myself or another, I should value their victims and so it seems I ought to blame them whenever I'm in a position to appropriately do so. Also see Bennett (2002) for an early articulation of this view.

[32] Might Scanlon respond in the following way? We have something like a default moral reason to blame those whom we know to be blameworthy. However, we aren't subject to moral criticism for failing to blame someone whenever we have such default moral reasons to blame. In contrast, there are special circumstances in which we have further reasons (in addition to the default reason) to blame someone we know to be blameworthy. For instance, when we ourselves are the wrongdoer or when we are close with the victim. Further, we are subject to moral criticism when we don't blame in situations where we have these additional reasons in conjunction with the default reason to blame. In response, the limitation with this approach for Scanlon is that he can't make sense of

To be sure, Scanlon (2008) suggests that the nature of a would-be-blamer's relationship to the wrongdoer can make a difference to whether blame is *appropriate*. For example, he suggests that on some occasions, parents of adult wrongdoers might have a special obligation to sympathize and offer encouragement in response to judging their offspring blameworthy, as opposed to modifying their behavior in a manner that constitutes blame (171). However, this is of no help for Scanlon's proposal in accounting for MBA. In cases like *Academic Dishonesty*, it's not as if Michael and Jan both have moral reasons to blame Michael, and yet Jan has overriding moral reasons to refrain from blaming. It's not *impermissible* for Jan to blame Michael. It's optional for her. Furthermore, insofar as parents can owe it to their children (whom they judge to be blameworthy) to refrain from blaming them, one wonders why an agent like Michael might not also owe it to himself to not self-blame. Thus it seems that Scanlon's relationship modification account of blame struggles to account for MBA.

4.3 Blame as Moral Protest

Angela Smith (2013) offers a theory of blame that is inspired by Scanlon and is intended as an improvement. Smith follows Scanlon (2008) in conceiving of blame as a set of attitudinal modifications in response to the judgment that the agent is blameworthy. However, for Smith blame is intimately connected with a stance or expression of protest yielding the following theory.

> The Moral Protest Account: To blame another is to judge that she is blameworthy (i.e., to judge that she has attitudes that impair her relationship with others) and to modify one's own attitudes, intentions, and expectations toward that person as a way of protesting (i.e., registering and challenging) the moral claim implicit in her conduct, where such protest seeks some kind of moral acknowledgement on the part of the blameworthy agent and/or on the part of others in the moral community (43).

why it is appropriate/permissible for, say, Jan to blame Michael. Scanlon seems forced to say that these "default" moral reasons are sufficient to render Jan's blame of Michael morally appropriate, and yet not enough so that she must. However, I'm skeptical that this is a plausible picture of such reasons in relation to blame. Admittedly, it's plausible to think of supererogatory actions as ones for which we have some moral reason to perform and yet which we are not required to perform. However, it seems odd to suggest that Jan's blame of Michael is supererogatory. That is, it hardly seems like it's a good or admirable thing for her to do or even that it would be better if she were to blame Michael, than not. Thanks to Peter Railton for raising this worry.

As Scanlon sees things, the set of modifications that count as instances of blame are those that would be made *appropriate* from the judgment that the agent is blameworthy. Smith worries that Scanlon's theory is missing a principled way to rule out certain sorts of "appropriate" attitudinal modifications that, intuitively, shouldn't count as blame. For instance, Smith considers a case involving a mother who responds to her judgment that her adult son is blameworthy for a serious wrong, with only empathy and pity. By Smith's lights, such a reaction isn't blame on any viable theory and yet it seems that Scanlon's theory must say that the mother blames her son.[33] Smith's solution is restrict the kinds of relationship modifications which count as blame.

According to Smith, blame is equated with only those attitudinal modifications that count as a way of morally protesting questionable claims (43). Smith's account depends on the idea that certain wrongs (those for which an agent is blameworthy) tacitly make a claim.[34] For example, when a close friend spreads malicious rumors about you, according to Smith, there is an implicit claim of the following sort being expressed by her action: "you aren't worthy of respect." It is this kind of moral claim that blame *qua* protest is intended to challenge. Importantly, that such protest "implicitly seeks some kind of moral acknowledgment" is, for Smith, far from an incidental feature. In fact, Smith refers to it as one of the "constitutive aims" of blame. She writes, "blame by its nature has an expressive point and a broadly communicative aim: it expresses protest, and I submit, it implicitly seeks some kind of moral reply" (39).[35] Additionally, she adds that the more primary "constitutive aim" of blame is to "*register* the fact that the person wronged did not deserve such treatment by *challenging* the moral claim implicit in the wrongdoer's action" (ibid).[36]

[33] We should keep in mind that on Scanlon's view it isn't that just any attitudinal alteration will count as blame, but only those that are "appropriate" (fitting) as determined by the nature of the relationship in some idealized form (i.e., the normative ideal). However, even with this qualification in mind, it isn't clear that the theory will rule out empathy and pity as instances of blame and so Smith's objection remains trenchant.

[34] See Jeffrie Murphy (1988) and Pamela Hieeronymi (2001)

[35] This thesis that blame has a constitutive aim of implicitly seeking some sort of acknowledgment on behalf of the wrongdoer or the moral community seems closely related to Smith's account of moral responsibility as answerability. See Smith (2012).

[36] Smith isn't clear about what sorts of facts determine whether or not a blamer makes some modification as a way of protesting in the relevant sense, but it would be implausible to think that it's solely a matter of the evaluator's intentions. Otherwise, Smith's theory

Given these features, we can see that on Smith's proposal, a would-be-blamer will have moral reasons to blame insofar as they have moral reasons to *protest* certain questionable moral claims that are implicit in the relevant wrongdoing. Applying these considerations to cases such as *Academic Dishonesty*, we see that Smith's theory initially fares better than either Sher's or Scanlon's. If blaming is essentially altering one's relationship to the wrongdoer as a way of protesting a morally objectionable claim which is tacit in the wrongdoing, then we can come up with a plausible story about the kind of asymmetry at the heart of MBA. The transgressor is the one that, *via* her own action (for which she is culpable), expresses the morally objectionable attitude at issue. Hence, plausibly, it is incumbent on her to stand against it or decry it in some manner. This is because if it were not for her conduct, the problematic moral claim (token) would not be "out there" as it were. Moreover, while it may be appropriate for others to weigh in and likewise protest the objectionable claim, it isn't morally inappropriate for them to refrain either, excepting perhaps under special circumstances.

Smith's theory also does well in accounting for those situations in which we must blame others and distinguishing them from situations in which we need not. Recall that when we bear a special relationship to the victim of a non-trivial wrong, we can be aptly criticized for failing to blame the wrongdoer. Plausibly, when a wrongdoer has made a tacit claim that a close family member or friend is not, say, worthy of respect, one should object (insofar as one is in a position to appropriately do so). Indeed, it may suggest a moral deficiency if they fail to stand against such an offensive claim made against their loved ones. On the other hand, we simply may not be expected to protest every objectionable moral claim that we encounter. It seems plausible then that if there is a normative expectation to protest such claims it will be restricted to certain situations. It is initially promising to think that we are only expected to protest such claims when we ourselves are the cause of them, when they are made against our loved ones, or perhaps when their uptake is likely to lead to serious harms. Thus, Smith's theory appears to accord well with some of the observations we considered in the previous section.

would suffer some of same problematic cases, which she thinks plague Scanlon's account. For example, we could think of a person who intends to protest a problematic moral claim in a wrongdoer's behavior in strange ways—say, by intending to trust them more when they have just broken one's confidence—intuitively, this isn't a case of blame.

Despite its initial promise, Smith's theory also has trouble with MBA. As we noticed, there is a normative expectation that a wrongdoer self-blame whenever she is in a position to appropriately do so. However, according to Smith's theory, at its core, blaming consists of adopting certain responses as a way of morally protesting a problematic claim. While it is natural to suggest that wrongdoers should blame themselves, it seems odd to think that wrongdoers must do things which amount to (or aim to) morally protest objectionable claims that are tacit in their own actions. The suggestion seems particularly strange when we think about agents who have, for instance, committed a clandestine wrong. For instance, imagine an agent, who unbeknownst to anyone else, invades the privacy of another. In these cases, there is no risk of an objectionable moral claim receiving uptake from members of the community. Where such wrongdoers are aware of their misdeed (and culpability) they ought to blame themselves. But must they do something which amounts to protesting a morally objectionable claim? If so, what would that look like?

Considering how Smith accommodates the possibility private blame will give us insight into what her take on self-blame (as moral protest) might amount to in the relevant kinds of cases. Of private blame that is directed at *others* she writes,

> The reactive attitudes are clearly one way in which we can register our moral protest of another without outwardly expressing it in any way. Resentment and indignation, in my view, are ways of emotionally protesting the ill treatment of oneself or others. But we can also protest ill treatment privately through the modification of other attitudes, intentions, and expectations. Even if we are not in a position (for whatever reason) to make these attitudinal modifications known, I believe these reactions embody, at a deep level, both moral protest and a desire that the wrongdoer morally acknowledge his wrongdoing" (44).

At least sometimes, merely adopting certain reactive attitudes such as resentment (in response to a judgment of culpability and perhaps with a desire that the wrongdoer acknowledge her wrong) "embody at a deep level," moral protest. Hence, it's plausible that for Smith, self-directed blame can likewise be unexpressed and sometimes be constituted by a self-directed reactive attitude such as guilt. Further, it's plausible that in general, wrongdoers should feel guilt (at least for a time) for culpable wrongs of which

they are aware.

Unfortunately, it isn't clear whether Smith thinks that the reactive attitudes such as resentment arising in such cases (i.e., as a response to the judgment that an agent is blameworthy) always amounts to moral protest. For instance, is it possible that I resent you for ϕ because I believe you to be blameworthy for ϕ-ing, without resenting you *as a way of morally protesting* the claim that is implicit in your wrong? This suggests a more general and related worry. When Smith speaks of "ways of morally protesting" is this a technical notion or are we supposed to have pre-theoretical intuitions about just what counts as the relevant sort of moral protest and what doesn't? If we consider the work the notion is supposed to do, I suspect it is the latter. After all, her view is motivated by the worry that Scanlon's theory does not rule out what *intuitively* does not feel like blame and what *intuitively* is not moral protest. For instance, the mother who responds to her judgment that her son is morally blameworthy with only compassion and deep sorrow (intuitively) neither counts as blaming, nor as morally protesting any claim. We are to have both reactions to this case so as to find Smith's proposal an improvement upon Scanlon's. If this is right, and moral protest of the relevant sort corresponds to some commonsense notion, then it's far from clear to me that having certain reactive attitudes (even as a result of the judgment that someone has impaired their relationship with you and even where you desire some kind of answer for their wrong) should count as moral protest.

The problem becomes more apparent when we focus on such responses in their self-directed forms. Consider the wrongdoer who realizes that he has culpably wronged another party and as a result experiences guilt. Notice that paradigm instances of guilt just don't seem to *aim* at any kind of self-acknowledgement of wrongdoing. In fact, ordinarily, it seems that the very act of acknowledging one's culpable wrongdoing *results* in feelings of guilt. Neither does it make much sense to talk about guilt *as a way of protesting* any claim even to oneself. One might find oneself (at the risk of) endorsing a morally problematic claim (e.g., that an unruly co-worker is not owed respect) and feel guilty about it. However, it seems odd to think that feeling guilt is a way in which one protests such a claim in relation to oneself.

What is more, it's not clear what would count as the kind of reflexive

response which intuitively counts as (i) an instance of moral protest and (ii) is such that wrongdoers are required to adopt it (whenever they are in a position to appropriately do so).[37] The upshot is this: Smith's theory of blame as moral-protest has trouble accounting for the nature of self-blame and MBA.

4.4 Blame as Affect

We've just considered three sophisticated theories of blame which for various reasons seem to have trouble accounting for MBA. Of course, none of this is decisive. Each theory has many virtues and challenges and their relation to MBA is just one of a number of considerations. Still, MBA is useful in helping us get a fuller picture of the reasons for and against a particular account of blame. The last theory that I'll consider is an affective account of blame. I submit that such accounts seem to have the resources to correctly predict MBA.

Affective theories of blame are sometimes attributed to Strawson (1962) and have been developed and defended by several authors.[38] What unites these theories is that in blaming someone we (at least) target them with one of the negative reactive attitudes. Of particular interest for our purposes is that the particular emotion which (in part) constitutes blame is generally taken to be different depending on one's relation to the wrongdoer/wrong. That is, in blaming someone that has wronged you, one resents the wrongdoer. In contrast, third-party blame as when one is merely witness to a wrong befalling another, consists (in part) of indignation. Both are thought to be forms of moralized anger. However, things are different in the case of self-directed blame. A wrongdoer blames herself according to affective accounts by feeling guilt and/or shame.

In *Academic Dishonesty*, Michael (the academically dishonest) should feel bad. In particular, he *should* feel guilty, and/or ashamed. If he doesn't,

[37] One proposal which I will not have room here to pursue would be that there is a kind of self-directed moral anger which counts as a self-directed moral protest. The challenge here for Smith would be to persuade us that this counts not only as blame, but that it is what is missing in cases in which we might fault agents for failing to blame themselves. Plausibly, a wrongdoer that feels guilty and recognizes their blameworthiness/wrong is not subject to moral criticism simply because she fails to feel angry towards herself for her moral failure.

[38] Gary Watson (1987, 1996), RJ Wallace (1994), Susan Wolf (2011) and Leonhard Menges (2017).

he is subject to moral criticism. Even if Michael believes that he's done something wrong and takes positive steps to prevent a repeat offense in the future, to not feel even an iota of guilt/shame suggests something has gone wrong and we'd naturally view him with moral suspicion. This is supported by the fact that it sounds happy for an on-looker who was made aware of Michael's situation as well as his lack of guilt to say something like, "you really *should* feel bad!" or "you *should* be ashamed of yourself!" In fact, it seems suitable to blame Michael should he fail to feel bad for his own conduct. For Michael to escape such tacit moral criticism and/or blame, he would need a good reason or excuse for his lack guilt/shame.

Things are different when we consider the third-person analogue of guilt namely, indignation. Suppose for instance that Jan (Michael's classmate), despite believing that Michael is blameworthy for his academic dishonesty, is not emotionally exercised. She feels nothing like indignation towards him. Jan is hardly the apt target of moral criticism. Indeed, it would be quite odd for a third party to blame Jan for her lack of indignation towards Michael. Nor does it sound felicitous to say, "you *should* feel bad!" or "you *should* feel indignant/angry!" to Jan. Importantly, things might be different if the wrong was particularly egregious or involved harming a close friend or family member. There we might criticize on-lookers who fail to blame. This is good news for the current proposal of blame. As we noted earlier, intuitively, we should blame wrongdoers who perform particularly serious wrongs or when they injure our loved ones.

What about resentment? Resentment is often viewed to be the second-person analogue of guilt and indignation. Hence, on an affective theory of blame, to blame another person is to resent them in response to the judgment that they have wronged us (and are blameworthy for it). Must we resent those that wrong us whenever we're in a position to appropriately do so? I don't think so. Returning again to *Line Cutting*, suppose that Fred (the line-cutter) cuts in front of Jamie. Suppose that Jamie has the quick thought that Fred wronged him and that he is culpable, but he simply doesn't feel anything like anger or resentment towards Fred. Perhaps Jamie has worked consciously on not being angered by what he perceives as trivial slights against himself. So, he calmy mentions to Fred that he has cut in line and reasserts his position in line. However, he is not in the least emotionally exercised. I hardly think that Jamie is subject to moral

criticism. How odd it would be for anyone to say to Jamie, "you *should* feel bad!" or "you *should* be angry/resentful!"

Indeed, I find it plausible that there is something virtuous about persons who foreswear various kinds of moral anger such as resentment and indignation at least when this endeavor is properly restricted. There may be something morally problematic about a person forswearing resentment or indignation altogether, but there is also something good or desirable in the vicinity of such a project. Perhaps the thing to say here is that it is at least sometimes virtuous to not be angered by wrongdoing and/or culpable agents. In fact, I think a person who resolves not to be angered "by the small stuff" is admirable. These reflections lend further support to the claim that affective theories of blame do well in explaining MBA. Importantly, consider how different things are in relation to guilt. I don't find it admirable for an agent to forswear guilt or shame altogether. People should (at least to some extent and for a time) feel bad (guilt and/or shame) about the wrongs that they have committed and for which they have no excuse (whenever they are in a position to appropriately do so).[39]

We've just seen how MBA offers us a new *desideratum* for theories of the nature of (moral) blame. In the remainder of this discussion, I extend our discussions to some recent developments in epistemology. A number of philosophers have defended the claim that there is a distinctively epistemic kind of blame. However, as in the case of moral blame, there are competing proposals. In what follows, I briefly outline how and why I think MBA will be helpful in thinking about the nature of epistemic blame.

5 Epistemic Blame

Recently, Jessica Brown (2020a, 2020b) and Cameron Boult (2021a, 2021b) have argued for and developed theories of epistemic blame. In each case,

[39] Why might there be this asymmetry between the requirement to experience self-directed vs. other-directed reactive attitudes? This is an interesting question which I cannot pursue here. It could be that this is just a basic fact which explains MBA. Alternatively, perhaps as some have suggested the blameworthy deserve to feel guilty in some sense of desert and it is up to the wrongdoer herself to give herself what she deserves.. In contrast, whether or not blameworthy wrongdoers likewise deserve second and third person-reactive attitudes, it may simply not (always) be incumbent on others (who are not the wrongdoer) to give people what they deserve See Randolph Clarke (2016), Andrea Carlsson (2017) for some discussions on desert-based considerations of guilt in relation to blame and blameworthiness.

epistemic blame is modelled after a favored theory of moral blame. Jessica Brown conceives of epistemic blame after George Sher's (2005) conative theory of moral blame. Boult models his theory after Scanlon's (2008, 2013) relationship modification account of moral blame. This should be no surprise as one of the main strands of argument for the view that there is a distinctively epistemic kind of blame is based on the idea that there are important parallels between the epistemic and moral domains. In fact, Boult (2021a, 2021b)[40] presents this as one of the premises in a common kind of argument for the existence of epistemic blame (3). Likewise, Adam Piovarchy (2021) notes, "A conception of epistemic blame that is too distinct from our conception of moral blame risks leaving us without any justification for thinking of the former as a species of blame at all. . . (794)". This suggests that MBA will also be helpful in thinking through proposals on the nature (and existence of) epistemic blame. After all, if MBA is true, then in as much epistemic blame is analogous to moral blame, we have *prima facie* reason to expect the following.

> *Must Blame Asymmetry-Epistemic (MBA-E):* There is a normative expectation that one epistemically self-blame whenever one is in a position to appropriately do so and it is not the case that one must epistemically blame others whenever one is in a position to appropriately do so.

Unlike it's analogue MBA, I find MBA-E to lack the intuitive support of cases. Consider the following.

> *Job Advice:* Sara is having trouble deciding between two job offers, A and B. She has done a good amount of research and careful deliberation about the benefits and costs associated with each, but she just can't seem to make up her mind about which job to take. Sara's cousin Frank suggests that she confer with his life-coach Earl who he swears is clairvoyant. Sara doesn't believe in anything supernatural and so initially scoffs at Frank's suggestion. However, as the deadline to choose draws near she thinks to herself, "well, what's the harm in talking to Earl?" As a result, she calls him. After hearing about her predicament, Earl confidently tells her that he can see into her future and that she will be much happier going with job A than B. Sara, to her own surprise, finds herself persuaded by Earl's remark and comes to believe that she will be happier with job A. However, this belief

[40] Boult (2021a) suggests that such an argument by itself is rather weak. However, he supplements it with the idea that "ought-judgments" in the moral domain are closely associated with blame. He finds something similar to be true in the epistemic domain.

> is temporary. The very next day Sara thinks about how ridiculous it is to believe the words of a purported psychic. "There are no such things!"—she tells herself. Following this thought she reflects on how it was a mistake to believe Earl and even to go to see him in the first place. Further, she thinks that she has no one to fault but herself for her error. Despite these judgments, Sara doesn't blame herself and goes about her day.

As with our moral cases, I am assuming that epistemic blame, whatever it is, consists of more than the mere judgment that one has done (epistemically) wrong and is culpable. If this were not so then *Job Advice* would be incoherent, but that seems implausible. Now the question before us is this: *Is Sara subject to criticism for her lack of self-directed, epistemic blame?* One issue that arises at this point is how it is we that should characterize the normative expectation to epistemically blame, in such cases, if there is one. In our moral cases, we saw that it was morally inappropriate for wrongdoers under suitable conditions to fail to blame themselves for their moral iniquities. Wrongdoers like Michael (who plagiarized) seem subject to *moral* criticism and/or blame for their failure to (morally) blame themselves. We are after an analogue in the case of epistemic blame. Plausibly, the epistemic analogue to a (moral) normative expectation to morally blame would be an *epistemic* normative expectation to blame, epistemically.

Once we've homed in on what it is we're after, we are met with an immediate challenge. Unlike in the case of moral blame, it's unclear what it is that we're to think about here. Moral blame is a perfectly familiar, bit of folk moral psychology. Of course, proponents of epistemic blame will want to suggest something similar about the epistemic analogue. They may insist that even if 'epistemic blame' isn't part of ordinary discourse, it refers to something that is familiar. However, I must admit when I ask myself "Must (in an epistemic sense) Sara (epistemically) blame herself for believing the psychic?" I don't know what to think. I think this by itself is some reason to doubt that there really is something like epistemic blame, but it's far from decisive.

How then should we proceed in making use of MBA-E as a *desideratum* for theories of epistemic blame? I propose that we consider some of the substantive proposals about the nature of epistemic blame and then consider whether agents like Sara are (intuitively) subject to such responses or reactions (as specified under each theory) in virtue of their failing to self-

blame (also as specified by each theory). So, for instance, we might take for granted Jessica Brown's conative theory of epistemic blame according to which epistemic blame is a characteristic response to the frustrated desire that (for instance) agents not form beliefs haphazardly. We can then consider whether the belief that we ourselves have frustrated this desire or whether another person has makes a difference to whether we must (epistemically speaking) respond in such "characteristic" ways. We can think through Cameron Boult's (2021a, 2021b) Scanlonian account of epistemic blame in the same way.[41]

One thing to note is that we've already seen how both Sher's and Scanlon's accounts struggle to make room for MBA, albeit in different ways. Given that the foregoing theories of epistemic blame are modelled closely after these accounts of moral blame, it would be surprising to find the latter doing better in relation to MBA-E. In contrast, we noticed that affective theories of moral blame fare better in this respect. However, some (including Brown and Boult) have expressed skepticism concerning affective theories of epistemic blame.[42] One interesting question which I anticipate here is this: what should we think if the best theory of moral blame turns out to be fundamentally different from the best theory of epistemic blame? That is, should we agree with Piovarchy that this would be bad news for proponents of epistemic blame?

6 Conclusion

Much of the existing literature on the ethics of blame has focused on the conditions under which it is or would be appropriate (permissible) to blame *others*. However, there are situations in which we *must* blame wrongdoers and this norm admits to an interesting self vs. other asymmetry. That is, there is a normative expectation to blame yourself whenever you're in a position to appropriately self-blame, but there is no such expectation that you blame others whenever you're in a position to appropriately do so. There

[41] We might also consider Adam Piovarchy's (2021) theory Agency Cultivation Model which is based on Manuel Vargas' (2013) theory of moral blame. However, as Piovarchy describes it, epistemic and moral blame are fundamentally cognitive judgments and so there will be some question about whether it's possible for one to believe that someone is blameworthy without blaming them (which is a stipulation of my cases).

[42] Others welcome an affective account of epistemic blame. See Conor McHugh (2012) and Linsday Rettler (2017).

are lots of occasions in which blame that would be directed at others seems morally optional. This asymmetry provides a new *desideratum* on some recent theories on the nature of blame. In particular, it raises a new problem for certain theories of blame while making an affective or Strawsonian account of blame more appealing. Furthermore, given that recent proposals on the existence of and the nature of epistemic blame are modelled after prominent theories of moral blame, MBA (or its epistemic analogue) will likely prove helpful as well. Thus, it pays to think carefully about not only self-blame, but also about when we *must* blame.

Chapter 2: Moral Ignorance and Apologies

Abstract

Typically, wrongdoers ought to apologize and blame themselves for their mistakes. In this work, I argue that whether a wrongdoer must apologize (and self-blame) can depend on her moral beliefs. Sometimes we fail to believe that we've done something wrong because we have a mistaken moral belief. As a result, we do not apologize (self-blame). I argue that in some of these cases we are not subject to moral criticism for failing to apologize (self-blame). This is to argue that there are cases in which what an agent ought to do depends on her moral beliefs which is to argue for a kind of moral internalism. Further, I leverage such cases as part of a novel argument for the view that mistaken moral beliefs can be exculpatory against philosophers such as Elizabeth Harman (2011, 2015) and Brian Weatherson (2019). That is, I argue for the controversial claim that agents who fail to believe that they have done wrong because of a mistaken moral belief (for which they aren't cupable) aren't blameworthy for what they have done. Along the way, I consider the relationship between self-blame and apologies as well as the relationship between an agent's culpability and her need to apologize.

1 Introduction

The normative is interestingly distinct from the descriptive in part because in our world the two come apart. That is, we often fail to do as we should or do as we must not. Further, it is a familiar fact that in the face of our own wrongdoing, there are things we must do *qua* wrongdoers. If you tell an insensitive joke at your friend's expense, typically, you should blame yourself and apologize. As Linda Radzik (2009) observes, sometimes wrongdoers must do more including: "[performing]...acts of restitution or reparation; the performance of good deeds that would otherwise be deemed supererogatory; self-punishment..." (5). Radzik is interested in the metaphysics and ethics of "making amends" where this involves repairing relationships that are damaged by wrongful acts (i.e., reconciliation). Offering

an apology under suitable conditions constitutes an important and perhaps necessary step in seeking to reconcile with others in light of our offenses. There may also be occasions when we wrong others in ways that make it inappropriate to seek forgiveness and/or reconciliation.[1] Even in these situations, there are things we must do. We should (in the absence of overriding reasons) apologize[2] and self-blame.[3]

But what happens when a wrongdoer does not track the fact that she has done wrong? In the following, I'll argue that wrongdoers who (nonculpably) don't believe that they've done anything wrong, needn't apologize. This is true even when the reason they don't believe they've done anything wrong is because they have a mistaken moral belief. As I'll suggest below, there is a close connection between the demand to self-blame and to apologize. In particular, the former cannot be satisfied without the latter. Further, when wrongdoers need not apologize (because their actions result from their mistaken moral beliefs), they need not seek to reconcile with others, either.[4] Thus, despite my focus on the norms of apologies, much of what I will say extends to the norms of self-blame and various activities constitutive of seeking to make amends.

There are real payoffs to thinking carefully about the norms of apology (self-blame) in these non-ideal cases. The first is that we find evidence for a kind of normative internalism concerning a subset of our moral demands and one that has been overlooked by authors involved in the debate over normative internalism vs. externalism.[5] That is, I'll present evidence for the position that whether a wrongdoer must (pro tanto) apologize is sometimes a function of her normative beliefs. A second and related up-

[1] What is the relationship between self-blame and apologies (or the demand to self-blame and the demand to apologize)? I discuss this in the next section.

[2] Can we apologize without seeking forgiveness? It seems to me that we can. A may hurt B so deeply that she finds herself unworthy of B's forgiveness. In which case she might say, 'I'm not asking for forgiveness but I'm truly sorry for what I've done.' Such a locution sounds happy in this context.

[3] As I'll argue below, apologies of the sort that I am interested in entail self-blame. Thus if A ought to apologize for ϕ-ing, then she ought to blame herself for ϕ-ing.

[4] At least under typical circumstances.

[5] For arguments that support the presence of internalist norms see Ted Lockhart (2000), Gideon Rosen (2003) (2004), Michael Smith (2006), Andrew Sepielli (2009), William MacAskill (2014) and by Hillary Greaves and Toby Ord (2017). For argument in support of externalist norms see Nomy Arpaly (2002), Timothy Schroder and Arpaly (2014), Maria Lasonen-Aarnio (2010) (2014), Miriam Schoenfield (2015), Brian Weatherson (2019) and Elizabeth Harman (2011) (2015) for externalist views.

shot of reflecting on the ethics of apologies is that it suggests a novel argument that moral ignorance is sometimes exculpatory. An agent who does the wrong thing due to a (non-culpably formed) mistaken moral belief isn't blameworthy or so I'll argue. This is to present a case for a second kind of normative internalism, concerning our evaluation of *agents*.

2 Background

The fact that there are norms about how one ought to respond to one's own moral failures suggests a few things. First, that there are norms which are sensitive to facts about the agent's moral standing. A morally perfect being who never self-blames, apologizes, or seeks to make amends[6] is not subject to moral criticism.

That is, such demands to self-blame, apologize, and seek to make amends seem to be unique to agents who are at least sometimes, wrongdoers. In contrast, there are moral norms concerning what all moral agents must or must not do. That is, one needn't be morally imperfect in order that one should not kill innocents without sufficient justification. However, only someone that has performed some wrong ϕ-ing is required to apologize (and/or self-blame) for ϕ-ing.[7] Setting aside the unimpeachable, how each of us as agents is doing in relation to such norms may be highly variable. If you are prone to moral errors, then you are doing poorly *qua* moral agent in certain respects. If you're disinclined to apologize (self-blame) in light of these errors, you're doing even worse. In contrast, a person that frequently does wrong, but is disposed to respond appropriately will be doing better along such dimensions. In between are agents who either rarely do wrong and are not inclined to apologize (self-blame) or those who rarely do wrong and are disposed to trying to make things right. Hence, consideration of these norms about what you must do when you fail to do as you must complicates our moral evaluations of agents. For instance, we may wonder whether an agent who rarely does wrong and is indisposed to apologize and seek reconciliation is more virtuous than a counterpart who frequently does wrong, but is inclined toward the forgoing responses. Our answer to

[6] Of course, there would be a sense in which the following norm would apply trivially to such beings. "If you ever do wrong, then you should self-blame, seek forgiveness, and make amends." Should such a being have a disposition to make amends nonetheless?

[7] I consider cases in which an agent falsely believes they have done wrong in Section 3.1.

this will depend on what it is that makes it wrong[8] or bad for wrongdoers to fail to attempt to make things right.

Relatedly, the existence of such norms suggests that the worse someone is in certain respects, the more that will be expected of them. If you frequently commit wrongs, then *ceterus paribus* there is more for which you must apologize (self-blame)[9] compared to a counterpart who isn't equally prone to misdeeds and vices. In a way, moral goodness is like wealth—it may be easier to do better, if you're already good at doing the right thing. Undoubtedly, this way of putting things is too coarse. An agent who rarely makes mistakes might find it more difficult psychologically to apologize compared to a counterpart who is quite used to owning up to her own mistakes because she makes so many of them. Still, *other things being equal*, there is a sense in which an agent who is prone to committing wrongs will simply have more to do (morally) compared to a twin who is less inclined towards wrongful acts.

Thirdly, where an agent's failure to respond to her wrongdoing in appropriate ways constitutes a wrong against another, wrongs can compound[10] If you wrong your friend by breaking his trust, and despite realizing what you have done, fail to apologize or seek to make amends,[11] then you further slight your friend. After all, we often cite not only wrongs like being lied to or being stolen from, but also the failure of a wrongdoer to properly apologize and/or compensate the victims of their actions (or to do so in a timely fashion). Likewise, if you found yourself apologizing for the initial wrong after a significant amount of time has passed, it can make sense for you to apologize for taking so long to apologize. You might say, "I'm sorry for ϕ-ing, and I'm also sorry that it took me so long to apologize"

[8] Alternatively: what makes such agents subject to moral criticism for not trying to make things right. I explain this alternative way of framing the discussion below.

[9] In fact, we might expect agents to do increasingly more in these cases since their apologies might start to feel too cheap and insincere.

[10] Linda Radzik (2009) makes a similar observation.

[11] What about the lack of self-blame? Self-blame seems to be a special kind of engagement with the fact that one has wronged another or violated a norm. It's plausible that agents who do not self-blame in such cases do not sufficiently care about morality or their victims. So, under suitable circumstances, an agent that fails to self-blame seems subject to moral criticism for the lack of sufficient care either of the victims or morality. However, the relevant kind of criticism might be one of negative aretaic appraisal as when an agent has a certain vice. The standard view is that while we can owe others an apology for what we do, we can't owe others apologies for the way that we are. See Glenn Pettigrove and Jordan Collins (2011) for a dissenting opinion.

or something of the sort. This isn't to say that we care to keep track of all such wrongs. There are practical and psychological limits to which wrongs are salient or worth fussing over in our day-to-day lives. Nonetheless, at least in principle, it seems that if you wrong someone by ϕ-ing, and then fail to apologize for ϕ-ing, at least in many circumstances, you've slighted them a second time. Thus, you may have proliferated the number of things for which you should blame yourself and for which to apologize. Further, if you fail to respond, you may be subject to moral criticism for three wrongs and so on.

It's worth making explicit that when a wrongdoer, under suitable conditions, fails to apologize (self-blame), they are subject to *moral* criticism. That is, if a friend betrays your trust and despite realizing that she has done so (and is culpable), neglects to offer up an apology (or blame herself), she isn't merely being imprudent or merely being a bad friend. She's failing as a moral agent and/or further wronging you and may be subject to blame for her lack of appropriate response. Thus, there are things we *must* do, *qua* wrongdoers, where the 'must' is understood in a moral sense. This is not to deny that there may be non-moral reasons to respond to our mistakes in various ways. However, my interest lies specifically in the moral sense in which we *ought* to seek to apologize (self-blame).

In speaking of the *demand* to apologize, a few points of clarification are in order. The first pertains to the employment of deontic expressions. I find it natural to speak of a wrongdoer who must/should/ought to apologize and even to speak of the demand or requirement to apologize. The same is true of self-blame. Nevertheless, some doubt that there can be strict requirements or demands to do whatever is not within our direct control (under some suitable understanding of this notion). Provided that it is beyond one's control to self-blame or even offer up apologies (at least of a suitable sort[12]), one might worry that there can't be demands on one to do so. We can sidestep such controversies by speaking instead of a normative expectation that one apologize (self-blame). I mean this as a bit of jargon. There is a normative expectation for A to ψ *just in case* A is subject to moral criticism if A does not ψ. So, in speaking of the demand that someone apologize (for ϕ-ing) or in saying that A must/should/ought to apologize or self-blame (for ϕ-ing), I can be understood as stating that there is a norma-

[12] More on this below.

tive expectation that A apologize (for ϕ-ing) without doing violence to the central arguments in this work.[13]

Secondly, our interest concerns apologies (self-blame) *for actions or omissions* rather than, say, for the outcomes of one's actions (omissions) or states of affairs more broadly. However, we offer what appear to be apologies for a wide variety of phenomena. Often in saying, 'I'm sorry' the pertinent object is a state of affairs which negatively impacts others.[14] In that case, we're not offering an apology (as I am using the expression), but rather offering condolences or expressing sympathy. Note how in saying 'I'm sorry for your loss' in these contexts one neither expresses nor implicates moral culpability. Plausibly, in apologizing specifically for one's conduct, one expresses or at least implicates moral responsibility. We also utter what appear to be apologies in situations where we harm or inconvenience others through no fault of our own. If another person pushes me into a bystander, I am naturally inclined to say, 'I'm sorry'. Here too, insofar as I have not been negligent in anyway, it seems odd for me to apologize (and self-blame) *for an action (omission)*. Notice how in these cases, it seems fitting for me to regret the harm that has befallen you and perhaps my causal role in it. However, it doesn't seem fitting for me to regret *my conduct*. If I must apologize in these situations, it is not *for a wrongful action*.[15] That is to say, I am not subject to moral criticism in these sorts of cases simply for not

[13] Personally, I'm inclined to deny the relevant "ought implies can" principle, but will not depend on this view in the discussion to follow.

[14] A test for distinguishes these cases from apologies in the proper sense is that the latter can be translated using the word 'apologize' and cognates with propriety. For instance, 'I'm sorry (for ϕ-ing)' when it is an apology can be translated with, 'my apologies (for ϕ-ing)' or 'I apologize (for ϕ-ing)' without sounding infelicitous. In contrast, 'I apologize (for ϕ-ing)' or 'my apologies (for ϕ-ing)' would not be a suitable way of offering condolences for some unfortunate circumstance. Consider how odd it would be to assert 'I apologize for your loss' at a funeral (assuming that you are not culpable for the death).

[15] It would be inaccurate to construe all cases which we are inclined to called "accidental harms" in common parlance as ones in which the agent is entirely free from moral culpability. In at least some of these cases, the agent may actually be partially culpable perhaps for not being more careful not to accidentally harm others. In those cases, there is an action (or omission) for which one ought to apologize. However, what should we say about cases in which a person is causally responsible for a harm for which they are entirely free of moral culpability? Must they apologize? And if so, for *what* should they apologize? It could be that we should sometimes apologize *for harming others* even when it isn't our fault (and we know this). Fortunately, such possibilities are compatible with the central arguments presented in this work. I take up some of these issues in Section 5·3

apologizing *for my conduct*.[16]. Relatedly, the kinds of apologies of interest in this work are unconditional apologies. Some authors[17] have suggested that there are such things as conditional apologies of the form, "*If* I ϕ-ed, then I'm sorry." Hereafter, when I say that wrongdoers in certain situations ought (pro tanto) to apologize, I mean that they must (pro tanto) offer an unconditional apology for some action, as opposed to a conditional one.

In speaking of the demand for apologies (self-blame), I've included a *pro tanto* clause. This is the third point of clarification. In saying that some wrongdoers ought (pro tanto) to apologize, I mean that there are moral reasons for them to do so which can be overridden. If an agent has *pro tanto* reasons to ϕ, then she ought to ϕ *in the absence of overriding reasons for her to not-*ϕ. Similarly, to say that there is a normative expectation for A to apologize for ϕ-ing, is to say that she would be subject to criticism for failing to apologize for ϕ-ing, *in the absence of overriding reasons not to apologize*. There may be occasions when offering an apology may do more harm than good or when offering it might come at too high a cost to the apologizer (or others). In these sorts of cases, while the wrongdoer has moral reasons to apologize, they are overridden by other normative reasons.

The *pro tanto* demand that we apologize *qua* wrongdoers concerns *genuine*[18] apologies. As I noted, the mere utterance of 'I'm sorry' doesn't suffice to count as an apology let alone a genuine one. Importantly, even apologies for one's conduct can be deficient. Suppose that Jon and Eugene have long been arguing about whether what Jon said about Eugene's parents was disrespectful. Furthermore, Jon, despite remaining fully committed to the belief that he has not be disrespectful, eventually says, "I'm sorry for being disrespectful" because he's simply exhausted with the conflict and would like to move on. We can criticize his apology as being in some sense disingenuous even if we think that it expresses or implicates culpability. What is missing here? Plausibly, we want apologizers to say "I'm sorry" while also *being* sorry for what they have done (or failed to do).[19] Indeed,

[16] This is compatible with the view that I can be subject to moral criticism for not apologizing *per se*. See Section 5.3 for more

[17] See Kristie Miller (2014) and Peter Baumann (2021) for discussions on the nature and semantics of conditional apologies.

[18] Authors working on theories of apologies (and the ethics of) commonly draw such a distinction. See Luc Bovens (2008), Adrienne M. Martin (2010), and Jeffrey S. Helmreich (2015) for example.

[19] I suspect we also want an agent to apologize *in virtue of* being sorry.

I take it that we have moral reasons against offering insincere apologies of this sort since apologies (plausibly in virtue of what they express/insinuate) tend to lead others to forgive us of wrongs. As a result, insincere or non-genuine apologies risk misleading others in significant ways.[20]

What constitutes *being* sorry or apologetic in the relevant sense? I submit the following.

> Self-Blame Analysis: A *is* sorry for ϕ-ing *if and only if* S blames herself for ϕ-ing.

Evidence for this account of *being sorry* comes from reflecting on cases when we're skeptical that a genuine apology has been offered. Consider public apologies from public figures. We are sometimes skeptical about whether such persons are sorry for what they have done as opposed to merely offering "the apology," say, because they were caught (i.e., in order to save face). Plausibly, what we suspect to be missing in these cases is that the agent blames herself for the wrongful act/omission. Indeed, it seems incoherent that A does not blame herself for ϕ-ing, and yet *is* sorry for ϕ-ing.[21] So it seems, the fact that A blames herself for ϕ-ing is a necessary condition for A's *being sorry* for ϕ-ing[22] Similar considerations suggest that self-blame is also sufficient for one's *being sorry*. It seems bizarre that someone could not be sorry for ϕ-ing and yet blame herself for ϕ-ing. Likewise, it is strange to suggest that A can blame herself for ϕ-ing and yet her apologizing for ϕ-ing is not genuine.

[20] There may be a variety of different reasons why non-genuine apologies of the relevant sort are morally problematic. Perhaps there's a kind of "phoneyness" which we take to be vicious (Thanks to Sarah Buss for this suggestion). It could also be that we have moral reasons to abstain from deception more generally, of which non-genuine apologies are an instance. Alternatively, as I've hinted here, apologizing when one isn't sorry for one's wrongful conduct risks making it the case that the victim of the wrong will forgive one under false pretenses. Given that forgiveness as a response to an apology is characteristically associated with a kind of relationship restoration, to apologize without being sorry is to potentially mislead another into re-entering a relationship with oneself. Plausibly, we have moral reasons to avoid doing so.

[21] Importantly, ϕ-ing is an action or omission. Notice this is compatible with the idea that one can be sorry for some unfortunate state of affairs without blaming oneself for that state of affairs.

[22] What should we say about wrongs in the very distant past? Suppose you insulted a person 30 years ago and that you blamed yourself for it then and for some time after. Must you go on blaming yourself forever? What about "being sorry?" Must you be sorry for what you have done, forevermore? I think the answer to both these questions will coincide which is what we'd expect given the Self-Blame Analysis and for her apology for ϕ-ing to be genuine.

Thus, excepting unusual circumstances, culpable wrongdoers ought to apologize genuinely and unconditionally. That is, we aren't looking for or satisfied with mere lip service. Further, one cannot apologize genuinely for ϕ-ing without being sorry for ϕ-ing and the latter entails that one blames oneself for ϕ-ing[23]. Thus, when I say that wrongdoers ought (pro tanto) to apologize, this entails that they must (pro tanto) self-blame as well. Just what self-blame (or blame more generally) amounts to is a matter of much controversy. However, our observations concerning the relationship between genuine apologies and self-blame suggests a new way to elucidate the nature of self-blame. We have judgments about what is missing in cases of apologies which we deem to be counterfeit. Provided that A's apology for ϕ-ing is genuine *if and only if* A blames herself for ϕ-ing, we can gain insight into the nature of self-blame by reflecting on what we feel to be lacking in non-genuine apologies. For example, Luc Bovens (2008) suggests that genuine apologies consist of certain cognitive, affective, and conative elements. Supposing this is right, we might infer that self-blame likewise consists of these features.[24] I won't have space to take up such matters here and so will leave self-blame unanalyzed. However, nothing that I argue will depend on any particular theory of self-blame.

3 Moral Ignorance and Apologies

There are occasions when each of us has done wrong without realizing it. We may eventually become aware of our misdeeds perhaps after some reflection or due to being confronted by others. However, this isn't always the case. There are likely to be many moral mistakes in our past that we have failed to recognize as such. It is this aspect of the ethics of apologies (self-blame) that will be of central interest in the remainder of this discussion. I think that reflecting on these norms tells us interesting things about

[23] Extant literature on what is required for genuine apologies in conjunction with the foregoing observations may suggest a promising way to produce a theory of (self) blame. If A's apology for ϕ-ing is genuine *if and only if* A blame herself for ϕ-ing, then we can consider independently what is it that seems missing in insincere apologies. In turn that might inform us about the nature of self-blame (or blame more generally). For instance, Bovens (2008) considers that sincere apologies have affective, cognitive, and conative elements. If that is right, then perhaps self-blame is constituted by these features as well.

[24] Bovens (2008) discusses the proposal that apologies that are genuine express a complex of attitudes. Hence, self-blame might consist of the attitudes themselves.

the nature of normativity. To that end, consider the following case.

> *Ryan's Tickets:* Ryan promised his co-worker Stan a pair of concert tickets to a sold-out show. Ryan and Stan get along, but neither would describe the other as a friend. The tickets are currently sitting on Ryan's desk, but he plans to take them into work next week and give them to Stan. Just then Ryan receives a call from his close friend Pamela and learns that she is hoping to go to the very same concert. Pamela mentions that she has been unable to find any tickets. Ryan thinks that while it will be unfortunate for Stan, he'd rather make his friend Pamela happy. As a result, he breaks the news to Stan and gives the tickets to Pamela. Upon reflection, Ryan comes to believe that what he has done was wrong—he should have kept his promise to Stan. Nevertheless, Ryan shrugs it off. He doesn't blame himself nor does he apologize to Stan.

Ryan is subject to criticism in more than one way. Not only is his instance of promise-breaking objectionable, but his failure to apologize to Stan (and blame himself) opens him up for further reproach. Consider things from Stan's perspective. It makes sense for him to feel slighted by the broken promise as well as by the fact that Ryan hasn't apologized.[25] Indeed, not only does Ryan owe Stan an apology for not giving him the concert tickets, he owes Stan an apology for not apologizing for the broken promise. Thus, it appears that it's simply not enough that Ryan realizes that he's done something wrong. He must respond to this realization in certain ways. For comparison consider the following.

> *Oscar's Tickets:* Oscar promised his co-worker Kevin a pair of concert tickets to a sold-out show. Oscar and Kevin get along but neither would describe the other as a friend. The tickets are currently sitting on Oscar's desk, but he plans to take them into work next week and give them to Kevin. Just then Oscar receives a call from his close friend Angela and learns that she is hoping to go to the very same concert. Angela mentions that she has been unable to find any tickets. Oscar thinks that while it will be unfortunate for Kevin, he ought to break his promise to Kevin in order to make his close friend very happy. As a result, he breaks the news to Kevin and gives the tickets to Angela. However, since he doesn't believe he's doing anything wrong, he doesn't blame himself nor does he apologize to Kevin.

Oscar's act of promise-breaking is impermissible just like Ryan's. However, that's not how things look by Oscar's lights. Oscar falsely believes that

[25] Nor would it appease him if he knew or suspected that Ryan said "I'm sorry" without actually being sorry (which on my view is coextensive with self-blaming).

making his close friend happy overrides the reasons he has to keep his promise to a coworker. As a result, he mistakenly believes that he ought to break his promise to Kevin. Let's also stipulate that this mistaken moral belief (essentially an endorsement of a partiality principle) is not the result of the mismanagement of his opinions. That is, it isn't the result of anything like wishful thinking or motivated reasoning.[26] Suppose that Oscar was raised in a household and community where partiality of the relevant sort was taught as a kind of moral ideal. Further, Oscar didn't just take what he was taught for granted. He has reflected on it and even questioned the principle from time to time. In particular, he thought a lot about it when he took an introductory ethics course in college. Despite his moments of questioning, and hearing what some moral philosophers have said about the matter, he remains convinced that morally speaking, the happiness of his friends matters a great deal more than the happiness of, say, co-workers. Hence, let's suppose that Oscar isn't culpable for this mistaken belief.

While Oscar, like his counterpart, is subject to moral criticism for his misdeed, he's not doing further wrong in not apologizing. This is reflected in the fact that while he seems the appropriate target of reproach for breaking his promise to Kevin, the same is not true concerning his lack of apology (and self-directed blame). We can imagine Kevin initially resenting Oscar for both the broken promise and for failing to apologize. However, once Kevin learns of Oscar's mistaken belief, his resentment (over Kevin's lack of apology) would naturally fade even if he continues resenting him for the broken promise. Unlike in Ryan's case, Oscar doesn't plausibly owe Kevin an apology for not apologizing. To be sure, Kevin might find it frustrating and perhaps undesirable that Oscar fails to realize that he's done any wrong. Still, it would be odd for him to feel further slighted or disregarded[27] by the fact that Oscar has not apologized (and has not blamed himself).

[26] See Michele M. Moody-Adams (1994) and William Fitzpatrick (2008) for discussions of affected ignorance in relation to wrongdoing and culpability.

[27] Might Oscar be subject to moral criticism for not apologizing even though his lack of an apology doesn't constitute a further wrong against Kevin? For instance, we might worry that he is calloused if he offers no apology whatsoever. I find this idea initially plausible. However, we should consider what he must apologize for if he is to be free from moral criticism for not apologizing. Notice that it seems sufficient for him to say something like, 'I'm sorry for the inconvenience' which is not to apologize for his conduct. In contrast, it would not suffice for Ryan to offer such an apology to Stan. I discuss this further at the end of Section 6.

The difference between the two cases is plausibly explained by the fact that Ryan truly believes (knows or has the justified belief) that he has done something wrong while Oscar does not. Notice that Oscar may *positively believe* that he has not done anything wrong as opposed to merely *lacking the belief*. But that makes no difference to the judgments regarding the case. Hence, reflection on this pair of vignettes suggests that when a wrongdoer truly believes (knows or has the justified belief) that she has done something wrong, she *pro tanto* ought to apologize (and self-blame). On the other hand, there is no similar demand on a wrongdoer who (mistakenly) does not believe she has done wrong, provided that she is not to blame for this lack of belief. So, reflection on our cases suggests the following regarding the norms of apologies.

> For any wrongdoer A and wrong action ϕ,
>
> *True Belief is Sufficient for Apology Norm (TBS):* If A truly believes (knows or has the justified belief) that she has done wrong in ϕ-ing, then A ought (pro tanto) to apologize for ϕ-ing.
>
> *True Belief is Necessary for Apology Norm(TBN):* If, through no fault of her own, A does not truly believe that she has done wrong in ϕ-ing, then it is not the case that A ought to apologize for ϕ-ing.

Our main interest hereafter will be in some version of TBN. Nevertheless, before setting TBS aside, I want to say a word about the parenthetical disjunction contained in the antecedent. I think it's an interesting question whether merely having the true belief that you have done wrong is sufficient to make it the case that you must apologize. If Ryan correctly believes he's wronged Kevin and this belief is haphazardly formed, does Ryan owe Kevin an apology? Or perhaps anything short of knowledge that one has done wrong is insufficient to make it the case that a wrongdoer owes anyone an apology. These are interesting questions which I will not have room to explore here. In fact, our interest for the rest of this discussion will be on the other principle, TBN. I leave TBS as is to indicate my lack of commitment on such matters.[28]

[28] Of course, if having a true belief that one has done wrong is sufficient to make it the case that one ought (pro tanto) to apologize, then it follows that knowing and having a justified true belief that one has done wrong are likewise sufficient conditions.

3.1 Falsely Believing You've Done Wrong

According to TBN, the *true* belief that one has done wrong (and is culpable) is a necessary condition for one being required (pro tanto) to apologize. What about agents who mistakenly believe that they have done something wrong? Must they apologize? Consider the frequently discussed case of Huck Finn. Often said to be a case of inverse akrasia,[29] Huck does nothing wrong in helping his friend Jim escape enslavement (he does what is morally required). However, due to a mistaken moral belief, he falsely believes he's acted wrongly. Someone in his position is inclined to apologize (and blame himself). Still, it doesn't appear that he *must*. Suppose you were to meet with Huck moments after he has helped Jim escape. He reports to you that he blames himself and asks for your help in coming up with a suitable apology to Ms. Watson. It seems that the correct thing to tell him is that he has nothing for which to blame himself nor anything for which to apologize because he hasn't done anything wrong. Indeed, how strange to think that Huck owes Ms. Watson an apology.[30] Hence, it appears that whether an agent must apologize or self-blame for ϕ-ing doesn't simply track the agent's beliefs about her conduct (or her belief about the permissibility of ϕ-ing). That is, merely believing that one has done something wrong doesn't entail that one ought (pro tanto) to apologize.[31] It is for this reason that according to TBN, if you lack the true belief (or are not disposed to truly believe) that you have done wrong, then it is not the case that you must apologize.

Before moving on, there's just one bit of tinkering to do with TBN. Recall, that Ryan (in *Ryan's Tickets*) has the true belief that he's wronged Stan. Suppose that Ryan's twin Bryan likewise has the (true) belief that promise-breaking in situations like the one he finds himself in is wrong. However, despite this belief and his belief that he has broken a promise

[29] See Jonathan Bennett (1974) for the earliest discussion of Huck Finn in this connection. I believe Nomy Arpaly and Timothy Schroeder (1999) coined the expression.

[30] Consider also some cases of gaslighting where the victims are lead to the false belief that they are actually victimizing their oppressors. It hardly seems apt to say that in such cases, the victims of the gaslighting ought to self-blame or apologize.

[31] Bernard Williams (1981) in discussing agent-regret cases suggests that we would be prone to talk the lorry driver out of his feeling of agent-regret. However, he also argues that we'd be suspicious of him, if he were not to experience agent-regret despite recognizing that he is not to blame for the tragedy. In comparison, I don't think we have reason to view Huck with suspicion if he were to refrain from apologizing or feeling guilty.

to Dan, he is unlike Ryan in the following way. It never occurs to Bryan that what he did on that occasion was wrong. What should we say about these irregular occasions? I suspect it will depend on the reason why our wrongdoer, despite his other beliefs, lacks the belief that they have done wrong. In at least some of these cases, it seems to me that the wrongdoer would be subject to moral criticism if they did not apologize (self-blame). Suppose I believe (truly) that you have a peanut allergy, that what I am cooking you contains peanuts and also that I should not serve you a dish that would harm you. However, I serve you the dish because I fail to draw the necessary inference(s) and so don't believe that I should abstain from serving you *this* dish. If I harm you as a result, I am blameworthy unless there are extenuating circumstances which explain either why I failed to draw the inference(s) or why I was not moved to abstain from serving the dish in virtue of my other beliefs.

Following Robert Audi (1994) we might account for such cases by saying that I was disposed to believe that my dish would harm you or that I ought not to serve it. A disposition to believe is distinct from an occurrent and dispositional belief in that the former is not an attitude. Additionally, having the disposition to believe a proposition is also distinct from merely having the capacity to believe a proposition. Presumably, we have the capacity to believe all manner of propositions and yet there's something distinct about (for instance) propositions which are entailed by things we already believe. Thus while Oscar is not disposed to believe that he has done something wrong, there is a sense in which Bryan (insofar as he accepts that promise breaking in his situation is impermissible and that he has broken a promise), is disposed to believe that what he has done is wrong – he need only put "two and two together." And it's plausible that at least in some cases, one can be culpable for failing to do so. We can recast our principles to make room for such possibilities in the following way.

> For any wrongdoer A and wrong action ϕ,
>
> *True Belief is Necessary for Apology Demand* (TBN*):* If, through no fault of her own, A does not truly believe (and is not disposed to truly believe) that she has done wrong in ϕ-ing, then it is not the case that A ought to apologize for ϕ-ing.

3.2 Being Unsure of Wrongdoing

An agent like Oscar fails to believe that he has done wrong as a result of a mistaken moral belief. Neither is he disposed in the relevant sense to truly believe that he has done wrong. But we might imagine his twin Oscar′ who differs from Oscar in the following respect. Oscar′ doesn't outright believe he has done something wrong (and doesn't outright believe he's culpable) because he's unsure about the correct moral principles. He thinks that if only if morality makes room for partiality for one's close friends, then he's done nothing wrong in breaking his promise to Kevin′. But he's not sure moral partiality of the relevant sort is correct. Suppose then that Oscar′ assigns .7 credence to the proposition that he's wronged Kevin′ (and is culpable) and .3 credence that he hasn't. Unlike in the case of Oscar (who believes he has done wrong) it strikes me as a mistake to say that Oscar′ must apologize in the absence of overriding reasons. Here is the argument.

> (1) If A is unsure that she has done wrong (or is culpable) in ϕ-ing, then in apologizing to B for ϕ-ing, A potentially misleads B.
>
> (2) It is *pro tanto* wrong to potentially mislead others.
>
> (3) One cannot be *pro tanto* required to do what is *pro tanto* wrong to do.
>
> (4) Thus, if A is unsure that she has done wrong (or is culpable) in ϕ-ing in relation to B, then it is not the case that A must apologize to B for ϕ-ing.

I take premise (2) and (3) to be fairly uncontroversial. Premise (1) in contrast needs motivating. Why should we think that it in apologizing for an action *sans* the outright belief (that one has done wrong and is culpable), the apologizer potentially misleads others?

In the first place, it is commonly thought that an (unconditional) apology for ϕ-ing, typically gives the recipient of the apology reason to forgive the apologizer for ϕ-ing. In virtue of what might an apology provide the victims of a wrong reasons to forgive or reconcile with the wrongdoer? Plausibly, the apologizer, communicates the fact that she believes that she has violated a norm and that she is culpable on the basis of her apology. If you were the victim of a wrong and learned that the other person had their doubts about whether they had wronged you or were culpable, you would

not feel like you have good reason to forgive them. Further support for this idea comes from reflecting on the infelicity of the following locutions.

> # 'I'm almost certain that it was wrong for me to ϕ and so I'm sorry for ϕ-ing.'
>
> # 'I'm sorry for hurting you but there's a chance that I didn't hurt you.'
>
> # 'I'm not certain that I wronged you, but there's a good chance that I did, so I apologize.'
>
> # 'I'm not sure that I'm the one that hurt you, but I'm sorry that I hurt you.'
>
> # 'I'm not sure that I'm to blame for ϕ-ing, but I apologize for ϕ-ing.[32] [33]

[32] One might not be sure that one has done wrong and is culpable for various reasons. A might believe that ϕ-ing is wrong and that he has performed ϕ-ed and yet have doubts that he is culpable for ϕ-ing. Alternatively, A might have doubts that ϕ-ing is wrong or even about whether he is the one that ϕ-ed. If merely have a high credence that one has done something wrong and is culpable were sufficient to make it the case that one ought to apologize, then we would expect such a demand on agents who are unsure for each of these reasons.

[33] Since there is a common usage of 'I'm sorry for' which expresses sympathy it's helpful to replace instances of it with 'I apologize for'. In fact, I think the locutions sound considerable worse (more clashy) when this is done. Notice that saying 'I'm sorry for your loss' to someone that has experience a death in their family is importantly different from saying, 'I apologize for your loss.' Only the latter is infelicitous unless one is responsible in some suitable sense for the death.

In each case, the relevant qualification expressing uncertainty clashes with the (apparent) apology. This suggests that there is something incoherent about apologizing for ϕ-ing when you've got doubts about whether ϕ-ing is wrong or whether you're culpable for ϕ-ing. Plausibly, when a wrongdoer (unconditionally) apologizes for an action, she represents herself as (at least) being sure that she has done wrong and is culpable. This is why, as we noted earlier, apologies of the relevant kind typically give their recipients reasons to forgive. It is also for this reason that apologizing unconditionally for ϕ-ing when you're not sure that ϕ-ing is wrong, potentially misleads others. Recipents of the apology are likely to infer that you haven't the relevant doubts and therefore, may come to believe they have reason to forgive you. If that's right, then to insist that agents like Oscar′ who don't believe they've done wrong (despite having a high credence that they have) must apologize is to insist that they must do what will potentially (and likely) mislead others including those who may have been negatively impacted by the agent's conduct. Indeed, saying "I'm sorry" when one is not sure that one has done anything wrong feels a lot like a "pseudo-apology" as when someone asserts, "I'm sorry but R" where R purports to be a justifying or excusing reason. So it seems merely having a high credence that one has done wrong (and is culpable) is not enough for one to be *pro tanto* required to apologize. Importantly, all of this is compatible with the idea that there may be versions of the above locutions which express a conditional apology that are felicitous. For instance, these sound at least a bit better.

> 'I'm almost certain that it was wrong to ϕ and so if I ϕ-ed, I'm sorry.'
>
> 'There's a chance I didn't hurt you, but if I did, I'm sorry'
>
> 'I'm not certain that I wronged you, but there's a good chance that I did, and if I did, I apologize.'
>
> 'I'm not sure that I'm the one that hurt you, but if I am, I'm sorry that I hurt you.'
>
> 'I'm not sure that I'm to blame for ϕ-ing, but if I am, I apologize for ϕ-ing.'

If these are improvements to their counterparts, then perhaps wrong-

doers like Oscar′ should (pro tanto) offer a conditional apology of the form, "If I've wronged you, then I'm sorry." While I have my doubts about even this suggestion, we need not settle such matters here. As we noted at the outset, we're interested solely in the demand on wrongdoers to apologize not only genuinely but also unconditionally. Thus it seems, that the outright (true) belief that one has done wrong is necessary.[34]

Taking stock, sometimes, a wrongdoer ought (pro tanto) to self-blame and apologize. However, such a demand applies to an agent only if she truly believes (or is disposed to truly believe) that she has committed a wrong. That is to say, falsely believing that one has done wrong does not place one under the demand to apologize. Hence, the relevant norm(s) are sensitive to a condition that is external to the agent. On the other hand, a wrongdoer who does not track the fact that she has done wrong (either by falsely believing that she has done no wrong, or merely failing to believe that she has) does no further wrong by refraining from apologizing (unless she is disposed to believe that she has done wrong). It seems then that whether a wrongdoer must (pro tanto) apologize (and self-blame) can be sensitive to facts that are internal to the agent (i.e., her beliefs).

4 Normative Internalism

While TBN* concerns wrongdoers who lack the belief that they have done wrong, it is silent about what might explain the absence of belief. A wrongdoer may fail to track her own wrongdoing because she has some mistaken beliefs and yet these beliefs may either concern a moral principle or not. Importantly, it seems to make no difference to our initial judgments. Wrongdoers who fail to truly believe that they have done wrong either in virtue

[34] What about self-blame? While this may be ruled out by some theories which build into blame the judgment that someone is blameworthy, intuitively, it's possible to blame someone despite not believing that what they did was wrong or that they are culpable (I say, so much worse the theories of blame that rule this possibility out). Lara Buchak (2014) argues that blaming someone for a wrongful act without outright believing that they have committed the act is inappropriate. If that's right, and the relevant kind of inappropriateness here is moral, then perhaps one could argue that self-blame is likewise morally inappropriate if the blamer is not sure that she has done something wrong. Further, it's plausible that insofar as self-blame in such cases is morally inappropriate, it can't be the case that one must (even pro tanto) self-blame. However, there are controversial premises here which I will not have room to explore and so I remain neutral on whether you ought (pro tanto) to self-blame when you have a high credence in the proposition that you've done wrong (and are culpable).

of a mistaken moral belief or a non-moral one, are not required to apologize (self-blame). The protagonist in *Oscar's Tickets* fails to believe that he has done wrong in virtue of the fact that he accepts a mistaken moral principle. He falsely believes that he may set aside his promissory obligation to Stan just to provide his friend a minor benefit.[35] But a wrongdoer may likewise fail to realize that she has done wrong as a result of a mistaken belief that does not concern a moral principle. For instance, A may make B ill by offering her food that contains peanuts, simply because A falsely believes that B has no dietary restrictions.[36] Insofar as A has and remains in the grip of her mistaken belief through no fault of her own, she wouldn't be subject to moral criticism for not apologizing to B. Thus, there is a parity between the two kinds of mistaken beliefs in relation to the norms about how wrongdoers ought to respond to their wrongful acts.

The foregoing parity between mistaken moral beliefs and mistaken non-moral beliefs relates to an ongoing controversy. There has been much discussion concerning the question, *does moral ignorance exculpate?* The disputants here are concerned with whether an *agent* who performs a wrongful act because of a mistaken *moral* belief, is *blameworthy*. Gideon Rosen (2003, 2004) and Michael Zimmerman (2008) are among those who answer in the affirmative (provided that the agent is not culpable for her mistaken moral belief).[37] Both authors do so on grounds that we should treat mistaken factual beliefs and moral ones symmetrically in our appraisals of wrongdoers. Furthermore, it is taken as common ground between both sides of the controversy that (non-culpable) factual ignorance is exculpatory.[38] In contrast, Elizabeth Harman (2011, 2015) and Brian Weatherson (2019)[39] argue that we should not treat mistaken factual and moral beliefs symmetrically and (albeit for differing reasons) submit that mistaken moral beliefs are not excul-

[35] If you find yourself sympathetic to Oscar's way of thinking as the case is described, you can simply alter it so that the harm to Stan is more than a minor inconvenience (and Oscar knows it).

[36] Many philosophers seem to accept this verdict but there are dissenters. Brian Weatherson (2019) for instance, takes it that agents like A don't do anything wrong in the first place despite doing something that leads to harm.

[37] For early proponents of this view see Cheshire Calhoun (1989) and Sarah Buss (1997). Also see Neil Levy (2009).

[38] Weatherson (2019: 29) doubts this. He argues that in the relevant cases of factual ignorance, the agents are justified in doing what leads to a bad outcome i.e., they haven't done anything wrong.

[39] Note Weatherson is open to the idea that moral ignorance can provide, in rare cases, a partial excuse.

patory. Note TBN* does not concern a wrongdoer's status as blameworthy. Instead, it is about whether a certain kind of omission is permissible for an agent, given her belief about her own conduct. Still, the cases that support the principle suggest a novel argument for the view that mistaken moral beliefs can render a wrongdoer, blameless. I pursue such an argument in Section 5. But as I stated at the outset of the present section, there is also a controversy about whether the permissibility of an *action* or *omission* can be a function of the agent's own normative beliefs. It turns out that reflection on cases like *Ryan's/Oscar's Tickets* provides insight into this related debate to which we turn our attention.

4.1 Action Internalism

As we noted, Weatherson (2019) denies that mistaken moral beliefs[40] are (fully) excusing. However, this is part of a broader effort to argue against what he refers to as normative internalism. Roughly, an internalist in the relevant sense is committed to the position that "we should be guided by norms that are internal to our own minds, in the sense that our beliefs, and our (normative evidence) is internal to our minds" (1). Furthermore, given that moral evaluations pertain to the rightness (wrongness) of actions, the culpability of agents, and the goodness (badness) of states of affairs, there is space for different kinds of internalist theories. As for Weatherson, not only does he argue against the kind of internalism defended by Rosen and Zimmerman (concerning blameworthiness), but he also rejects the view that an action (or omission) can be permissible for an agent on the basis of a mistaken moral belief.

Normative internalism concerning the rightness or wrongness of actions[41] naturally arises in relation to consequentialist theories of the Right.

[40] It's not always clear what it is that counts as a mistaken moral belief. Suppose that Oscar falsely believed that his promise to Stan wasn't binding or didn't provide for him sufficient reasons to keep his promise. Is that a mistaken moral belief? Or what if Oscar had strange beliefs about the uptake conditions of promises (i.e., the conditions under which one was under promissory obligations to make ϕ happen in virtue of saying, 'I promise to ϕ')? Are such mistaken beliefs moral in the relevant sense or merely factual? Addressing these difficult cases seems crucial for externalists such as Weatherson who want to argue that we should treat ignorance of the moral variety differently from that of the non-moral sort.

[41] Harman (2015) presents Actualism which appears to contrast with the kind of internalism suggested here. However, it isn't so with respect to the letter of the law though it may be in spirit. Actualism as Harman defines it is the view that "A person's moral beliefs

Consider the following kind of case.

> Anne has just poured poison into Bill's tea. However, the only reason that Anne has done so is because she falsely believes that she is spooning sugar into Bill's cup. Anne is not culpable for her false belief. Someone spiked the sugar container and Anne has no reason to think anything is amiss. Bill gets very sick as a result of her action[42].

A number of philosophers[43] take it that agents in Anne's situation are not blameworthy and yet have done something wrong. In turn, some[44] have argued that even if the mistaken belief concerns a moral principle, things would be no different. Weatherson (2019) agrees that characters like Anne are not blameworthy but also argues that this is because Anne hasn't done anything wrong in the first place. This is because Weatherson finds it plausible that an agent's mistaken factual beliefs can make a difference to whether a particular course of action is permissible for her (29). Weatherson suggests that agents like Anne should maximize expected value as opposed to maximizing value.

This difference in views connects to a dispute among objectivists, prospectivists, and subjectivists about rightness (wrongness) or obligations (permissions). The various camps have different takes on what an agent in the face of uncertainty or incomplete information should or should not do. Roughly, objectivists[45] think agents should maximize value while their prospectivist[46] counterparts contend that they should maximize expected value,[47] where what is "expected" is sometimes cashed out in terms of

and moral credences are usually irrelevant to how she (subjectively) should act. How a person subjectively should act usually depends solely on her non-moral beliefs and credences; her moral beliefs and credences are relevant only insofar as they provide warrant for beliefs and credences about what her non-moral situation may be" (Italics added, 58). Given that Harman puts things in terms of what is "usually" the case, it isn't clear that the kind of internalism that I am presenting conflicts with her Actualism. This is because for all I have said (and will argue) it is only a subset of moral norms (those concerning how we should respond to the fact that we are blameworthy and have done wrong) which are sensitive to the agent's own moral beliefs in the relevant way. By the same token, Harman does not explain what she means by "usually" in this account.

[42] This is based on a case presented by Harman (2011).

[43] Rosen (2003, 2008), Graham (2014), and Harman (2015).

[44] Rosen (2003).

[45] G.E. Moore (1912), W D Ross (1930) and Peter Graham (2010).

[46] Elinor Mason (2017), Michael Zimmerman (2006, 2008, 2014).

[47] The dispute is usually couched in terms of what it is that certain consequentialist views of rightness say about what such agents should/should not do.

the agent's evidence[48]. Finally, subjectivists[49], contend that such agents should do what they believe will bring about the most value. Importantly, disputants have tended to focus entirely on cases in which the agent fails to track (or has a mistaken belief about) some non-moral fact. However, Weatherson, who falls in the prospectivist camp, contends that mistaken moral beliefs and uncertainty concerning the truth of moral principles does not alter an agent's moral obligations or permissions. That is to say, when it comes to our evaluations of actions (and agents), only non-moral mistakes or uncertainty can make a difference to whether some action or omission is permissible for her. However, it is just this sort of view that I suggest TBN* calls into doubt.

Returning again to *Oscar's Tickets* we saw that his failure to track the fact that he has done something wrong makes a difference to whether he must respond in certain ways. In contrast, his counterpart in *Ryan's Tickets*, truly believes that he's wronged another, and intuitively, must apologize (self-blame). This is just the idea that is summed up in TBN*. But is Oscar failing to track a non-moral fact or is his ignorance of the normative kind? Plausibly, mistaken moral beliefs (and moral uncertainty) of the kind that Weatherson is interested in concerns the failure to track a correct moral principle or theory. Strictly speaking, failing to believe that you have wronged another person, despite having some moral content, doesn't itself concern a failure to track a moral principle. However, recall that the reason that Oscar fails to track the fact that he has done wrong is due to his mistaken belief that it is permissible to break a promise to provide a minor benefit to a friend. This is a mistaken belief concerning a moral principle if anything is. Nevertheless, Oscar need not apologize (self-blame) in the way that Ryan must. Thus, it seems that at least with respect to some actions or omissions, their rightness/wrongness depends on the agent's normative beliefs—in particular, mistaken moral beliefs can make an otherwise impermissible action or omissions, permissible for the one that has the mistaken belief.

[48] There are a various way this may be fleshed out for prospectivists. For instance, it could be that A's actual evidential state is the relevant standard or it could be some idealization such as her "available" evidence.

[49] H.A. Prichard (1932) and W D Ross (1939).

Too Indirect for Interalism?

For all I've said so far, one might worry that there is something peculiar about the foregoing cases which precludes them from being evidence for internalism of the relevant sort. To see this, let's consider a straightforward internalist norm about the permissibility of an action/omission.

> *No Belief Not Required (NBNR):* For any agent A and action or omission ϕ, if A does not believe that they must ϕ, then it is not the case that A must ϕ.

Notice that NBNR posits a direct connection between the thing not believed and the thing that agent is not required to do. TBN* is unlike this. Returning again to Oscar's Tickets, Oscar fails to believe that he has done anything wrong (in breaking his promise to Stan). Yet, our focus has been on whether it is permissible for him to refrain from self-blame and attempting to make amends for what he has done. One might worry then whether TBN* is indeed internalist in the relevant sense. Instead of being of the form "if you don't believe you must ϕ, then you are not required to ϕ" we have something of the following: "if you don't (truly) believe that you have ϕ-ed, then you are not required to ψ."

It isn't obvious that TBN* is not internalist simply because it purports an indirect connection between the relevant moral belief, on the one hand and the relevant action/omission, on the other. Still, exploring this worry will help us get a better sense of what is at issue between internalists and externalists of the relevant kind. In the course of distinguishing between the two camps Weatherson (2019) considers an example (attributed to Derek Ball) in which an agent's mistaken moral belief entails that he ought to do something. He writes: "if you believe that it is permissible to murder your neighbors, then you ought to seek therapy (8)." Here is a case where whether an agent should ψ, depends on what she believes about the permissibility of ϕ-ing. However, Weatherson does not classify such a case as suggesting any kind of internalist norm. He adds,

> Sometimes normative beliefs change the normative significance of other actions. So the externalist claim I'm defending is a little weaker than this general independence claim. It allows that a normative belief B may change the normative status of actions and beliefs that are not part of the content of B (8, Italics added).

In my view a norm like TBN* is internalist despite the fact that what it posits is that "a normative belief B may change the normative status of actions... that are not part of the content of B". Oscar's belief that promise-breaking is permissible in his situation is about promise-breaking, but it impacts whether it is permissible for him to refrain from blaming himself and seeking to make amends for his wrong. Still, I'll argue that there is an important difference. If successful, I'll have addressed the foregoing worry and shown that there is an interesting version of normative internalism which Weatherson has set aside, but which it would be worthwhile to consider.

One driving motivation for normative internalism is the view that norms, if they are to be genuine or important, must be action-guiding.[50] While there is more than one way to flesh out this notion, I submit that it's promising to think of it in the following way. For a norm to be action-guiding in the relevant sense is for it to issue actions that are intelligible to the agent in question, given her beliefs.[51] For instance, in discussing a case in which an agent, Bonnie, steals a cab because she mistakenly (and non-culpably) believes it is permissible to do so, Rosen (2003) writes the following.

> Here is Bonnie. She blamelessly thinks that she has most reason to steal the cab. What do you expect her to do? To set that judgment aside? To act on what she blamelessly takes to be the weaker reason? To expect this is to expect her to act unreasonably by her own lights. This is certainly a possibility, but is it fair to expect it or demand it? Is it reasonable to subject an agent to sanctions for failing to exhibit akrasia in this sense (79-80)?

A natural way to interpret Rosen here is this. It would be strange to expect/demand that Bonnie go against her judgment about what is permissible for her to do. This is because to do so would be to expect/demand that Bonnie do what does not make sense to her *by her own lights*. Plausibly,

[50] See Ted Lockhart (2000), Michael Smith (2006), Andrew Sepielli (2009), William MacAskill (2014) and Hillary Greaves and Toby Ord (2017).

[51] Peter Railton (1984) discusses (and aims to assuage) the problem of alienation in relation to consequentialist theories which may be germane. That is, there is a concern that consequentialist theories of right action require that agents be alienated in some sense from moral principles in some sense. This is because even in a world where one ought to act in such a way as to bring about the best consequences, it might be that the best way to do so is to not be guided by a principle that one ought to perform actions with the best consequences.

internalist norms will be ones that respect this basic idea. That is, cases which support internalist norms will be ones where intuitively, what the agent should or should not do, is that which it would be intelligible for her to do by her own lights.

Clearly, an agent who believes it is permissible to kill her neighbor (without sufficient reason) ought to seek therapy. However, notice that for an agent who believes that there is nothing wrong with murdering one's neighbors, the thought that they should go to therapy (to correct this belief) would be bizarre (assuming they don't have bizarre views about when to seek therapy). From their point of view, it would amount to going to a therapist to either talk them out of a correct moral principle or to coach them into being *akratic* against a true moral belief. It is for this reason that in the case that Weatherson discusses, we find a dependence of an agent's moral requirements on her moral belief(s), but the dependence isn't the sort that captures what the internalist is aiming to capture with her theory.

The situation depicted in *Oscar's Tickets* is quite different. From Oscar's perspective, he has done no wrong and so the thought of apologizing or even blaming himself would not make sense from his point of view (provided of course that he doesn't have strange views about when things like apologies and self-blame are required). Indeed, this is what seems to explain the intuition that he needn't (while remaining in the grip of his mistaken moral belief) apologize or self-blame. On the other hand, from Ryan's point of view, it looks to him as though he has done something wrong. Hence, what makes TBN* an internalist principle is that it concerns cases which suggest that if something is to be required of an agent, then it must be intelligible from that agent's perspective. For all I've said, the kind of normative internalism concerning rightness/wrongness at issue here is fairly narrow in character. It concerns a subset of norms specifically having to do with responding to our wrongs. That is, it doesn't follow that all of our moral duties or even most of them are internalist in this way.[52]

[52] We might also consider whether and how a wrongdoer's beliefs (or lack of belief) about whether she owes others an apology (and should self-blame) might impact whether she ought to do so. For instance, suppose that Bob lies to Kevin and it's wrong for him to do so. Furthermore, suppose that Bob believes that he's done something wrong and that he's culpable, and yet has strange beliefs about when apologies are owed so that he believes he need not apologize. Is Bob subject to moral criticism for not apologizing (self-blaming)? There are a number of cases here to explore. However, I must admit I don't have stable convictions concerning them and so I have not featured them here.

Explaining the Asymmetry

Before moving on, I want to consider what may strike many as a tempting explanation for why it is that wrongdoers with mistaken moral beliefs (for which they are not culpable) are not *pro tanto* required to apologize. I'll explain why, despite initial appearances to the contrary, it won't work, before briefly sketching a more promising account. Plausibly, to self-blame and genuinely apologize requires that one believe[53] that one has done something wrong. Hence, it's simply impossible for wrongdoers such as Oscar insofar as they don't believe that they have done any wrong, to do any of these things. Assuming that "ought implies can," it may be said that they aren't obligated to perform them either. As such, what makes it permissible for certain agents with mistaken moral beliefs to refrain seeking to make amends or self-blaming is simply that it's impossible for them to do so (given their mistaken moral beliefs).

Despite its initial appeal I don't think this suffices as an explanation of the phenomenon. After all, as we noted at the outset, wrongdoers are not merely required to apologize and self-blame. In many cases, wrongdoers ought also to pay restitution and take steps to reform their behavior. Importantly, it is not impossible for wrongdoers while in the grip of mistaken moral beliefs to perform at least some of these things. Regardless of what he believes about the permissibility of his action, Oscar could in principle perform overt actions such as offering Kevin tickets to a future concert or taking a moment to reevaluate similar decisions in the future. However, despite the fact that each of these are things that Oscar in his current state can do, while in the grip of his mistaken moral belief (and his failure to believe that he's done any wrong), he doesn't seem required to do these things. Indeed, the thing to say here is that it would be unintelligible from Oscar's perspective that he must make anything up to Kevin. For this reason, he need not do so. Thus, it seems that what is driving our judgments is something other than consideration of "ought implies can."

While I won't have space here to explore the matter in sufficient

[53] What about merely being disposed to (truly) believe that you have done wrong? If agents who are merely disposed to believe that they have done wrong must self-blame and seek to make amends, then the current proposal will not be able to explain such cases. This is because it doesn't seem as though you can sincerely apologize if you are merely disposed to (truly) believe that you've done wrong.

detail, I briefly consider what I think is at least more initially promising account of the data. The account is suggested by reflecting on the phenomenology of one who feels slighted specifically by the lack of an apology. Think about a time when you felt that someone owed you an apology and yet they never reached out to formally apologize. What were you (or are you) concerned about? It seems to me that the most bothersome aspect of such a situation, from the vantage point of the wronged, is that the person that owes the apology is being disrespectful (or insufficiently respectful) in virtue of not apologizing. In cases of this sort, it's natural to think about how the agent wronged you and yet either can't be bothered to express their remorse or are too proud to acknowledge to you that they have made a mistake. However, notice that if you somehow learned that the reason the agent had not apologized is that (through no fault of their own) they were not aware or did not believe they had done anything wrong, the feelings or concerns of being disrespected (in relation to the lack of apology) dissipate.

Additional support for the current sketch come from reflecting on the relationship between disrespect of the relevant sort and its relation to the agent's beliefs. Consider how we think about an agent who invades the personal space of another because she is not aware (through no fault of her own) that there is another person in her vicinity. In comparison, consider our attitudes about an agent who knowingly engages in such behavior. In either case, the harm might be the same and the agents could be committing the same type of wrong. Even so, it's natural to say of the agent who is aware of the presence of the other (and the local norms of personal space) and yet stands too close to the latter that she disrespects the other party. In contrast, our first agent who stands too close to another because she is (through no fault of her own) unaware of the other's presence, does not seem to disrespect the other. Thus, we have some initial evidence that a theory concerning disrespect may prove to be a promising account of the phenomena in question.

5 Moral Ignorance is Exculpatory

We've just seen how the cases in support of TBN* suggests that there are internalist norms concerning our moral evaluations of actions (omissions). In particular, the demand for wrongdoers to respond to their misdeeds

in certain ways seems sensitive to their normative beliefs in a surprising way. Thus, normative externalists concerning norms of actions/omissions will need something to say about these cases. In this section, I want to suggest that our considerations so far also recommend a novel argument for the view that mistaken moral beliefs can excuse (i.e., an internalist principle concerning our evaluations of agents). That is to say, not only is there evidence of internalist norms concerning the permissibility of actions/omissions, but also concerning the culpability of agents.

Before proceeding, I should note that the kind of blameworthiness at issue in the discussion to follow is often referred to as that of accountability rather than attributability. The distinction traces back to Gary Watson (1996) who distinguishes between "two faces of responsibility". An agent can be responsible (or blameworthy) in the attributability sense for ϕ-ing, if ϕ-ing is expressive of the agent in some suitable sense. For instance, ϕ might be some action or attitude which is suitably connected to A's deep or true self. In contrast, for A to be responsible (blameworthy) for ϕ-ing, in the accountability sense is for A to be subject to sanctions for ϕ-ing. It is this latter kind of blameworthiness or moral responsibility that seems to be at issue in debates about whether moral ignorance (or mistaken moral beliefs) can be exculpatory.

Rosen's (2003) approach to arguing that mistaken normative beliefs[54] provide an excuse is to consider cases in which there is pressure to grant that an agent is non-culpably mistaken about some normative issue and behaves badly as a result. For instance, he asks us to imagine a "run of the mill American sexist circa (say) 1952" (66). This father saves money for his son's college education but does not extend that treatment to his daughter. This is because the father does not believe that such behavior is impermissible. Importantly, Rosen suggests there is a coherent telling of the story in which the father hasn't mismanaged his opinions in

[54] Rosen and a number of others put things in terms of moral ignorance but as Harman (2011, 2015) has suggested this is not quite apt. At least insofar as ignorance is intended to contrast with knowledge, an agent who performs a wrong ϕ because she merely has the truth belief that she ought not to ϕ, is not off the hook for her failure to know that she ought not to ϕ (i.e., due to her moral ignorance). Similarly, there are cases in which an agent is merely uncertain concerning whether ϕ-ing is permissible or not. However, it seems accepted by all sides of the debate that such uncertainty (even if non-culpable) doesn't itself provide an excuse insofar as the agent also has (and is aware of) a morally safer option.

any way. Further, and as we noted, Rosen argues for a parity thesis that we should evaluate agents who do wrong as a result of mistaken moral beliefs just as we evaluate their counterparts who err as a result of mistaken non-moral beliefs. Finally, Rosen along with his critics accepts that (non culpable) factual ignorance is sometimes exculpatory. Putting these considerations together, we have an argument that normative ignorance (or mistaken moral beliefs) is (are) sometimes excusing.

Resistance to Rosen's line of argument has taken a number of forms. As we noted, Weatherson doubts the symmetry thesis.[55] Harman (2011, 2015) has argued that cases such as that of the 1950's sexist father are not instances of non-culpable moral ignorance because we have a moral obligation to believe certain moral truths and characters like the sexist father have culpably failed to meet such requirements. However, TBN* and our observations concerning the cases that lead us to the principle suggests a new way to argue that mistaken normative beliefs can excuse. Here is the argument.

> (1) For any agent A, and wrong ϕ, if A is blameworthy for ϕ-ing, then A *pro tanto* ought to apologize.
>
> (2) There are situations in which A has performed a wrong ϕ because of a mistaken normative belief and it is not the case that A *pro tanto* ought to apologize for ϕ-ing.
>
> (3) So, there are situations in which A has performed a wrong ϕ (because of a mistaken normative belief) and A is not blameworthy for ϕ-ing.

(3) is just the view that mistaken moral beliefs can excuse.[56] Premise (2) sums up our verdicts about cases such as *Oscar's Tickets*. Hence, it is (1) that is need of defending. Admittedly, in its current form it's not particularly attractive at least when we think of matters diachronically. Suppose that Ashley is blameworthy at t_2 for stealing from her coworker at t. Further imagine that at t_1, Ashely apologized to her victim. It follows that by t_2, there may be nothing more that she must do in response to her ϕ-ing at t_1. However, it could still be the case that she remains blameworthy at t_2

[55] Importantly, Weatherson also presents other arguments to doubt internalist principles.

[56] Provided that in the relevant cases, it is *in virtue of* the agent's moral ignorance that she is off the hook.

and thereafter. If that's right, then it seems we have a counterexample to (1). That is, Ashley is blameworthy at t_2 for ϕ-ing.and it is not the case that she must(pro tanto) apologize for ϕ-ing. Fortunately, we can easily accommodate such cases in the following way.

> (1)* For any agent A and action ϕ, if A is blameworthy for ϕ-ing at t, then A *pro tanto* ought to apologize at t, unless A has already done so.

(1)* is plausible. After all, there seems to be an intimate link between an agent being blameworthy for a wrong, on the one hand, and her needing to and/or having reasons to apologize, on the other. Indeed, it is plausible that the facts that render an agent blameworthy for what she has done are the very facts that give her good reason to self-blame, apologize and the like. Furthermore, there is linguistic data to suggest that (1)* is true. Consider the following locutions.

> # 'You don't have anything for which to apologize (and never have), but you are blameworthy.'
>
> # 'She has no reason to apologize (and never has), but she is blameworthy.'
>
> # 'They are to blame for ϕ but they don't owe anyone an apology (and never have) for ϕ.'
>
> # 'You are blameworthy, but you needn't be sorry.'
>
> # 'I'm culpable for ϕ-ing and I don't have any reason to apologize for ϕ-ing.' [57]

None of these locutions sounds happy and (1)* explains their infelicity. That is, if a wrongdoer is not (and has never been) required (pro tanto) to apologize, then she is not blameworthy for ϕ-ing. Thus, there is some initial evidence to suggest (1)* is true and by replacing (1) with (1)* we get a new argument for the view that mistaken moral beliefs can excuse wrongdoers.[58]

[57] Thanks to Sumeet Patwardhan for asking me to consider third-person and first-person locutions.

[58] Harman (2011, 2015) suggests that we must consider the following kinds of cases if we are to determine whether false moral beliefs can exculpate.

> Max works for a Mafia "family" and believe he has a moral obliga-

5.1 Challenges and Replies

Before concluding our discussion, I want to briefly consider a couple of objections and respond to them. The objections stem from reflecting on the relationship between blameworthiness on the one hand, and apologies and forgiveness on the other. So, in addressing them, I hope to shed some light on just how these related notions might interact with one another.

Externalists may raise doubts about the foregoing argument by challenging premise (1)*. In the first place, externalists may ask us to consider the relationship between forgiveness and blameworthiness. One thing to note is that in cases such as *Oscar's Tickets* and Rosen's case of the sexist father, it seems as though the wronged parties have something for which they can forgive the agents. That is, it appears that the daughter who is harmed by the sexist father's inequitable treatment and Kevin who is slighted by Oscar can each forgive their offenders regardless of whether the agents see their conduct as wrong. Furthermore, it's initially plausible that A can be forgiven for ϕ-ing at t *only if* A is blameworthy for ϕ-ing at t. In other words, it's tempting to think that one cannot be forgiven for that which one is not blameworthy.[59] If that's right, then our sexist father is blameworthy

> tion of loyalty to the family that requires him to kill innocents when it is necessary to protect the financial interests of the family. This is his genuine moral conviction, of which he is deeply convinced. If Max failed to "take care of his own" he would think of himself as disloyal and he would be ashamed.
>
> Gail is a gang member who believes that she has a moral obligation to kill a member of a neighboring gang as revenge after a member of her own gang is killed, although her victim was not responsible for the killing This is her genuine moral conviction, of which she is deeply convinced. If Gail failed to "take care of her own" she would think of herself as disloyal and she would be ashamed.

Harman (2015, 66) takes it that if they act wrongly in accordance with the relevant moral beliefs, neither Max nor Gail would be off the hook. However, I suspect intuitions will vary here. Interestingly, if we consider instead whether Max and Gail would be subject to moral criticism if they did not apologize to their victims (and/or self-blame), it seems clear that the answer is, no.

[59] Jeffrey Murphy (2003:13) writes, "To regard conduct as excused (as in the insanity defense, for example) is to admit that the conduct was wrong but to claim that the person who engaged in the conduct lack substantial capacity to conform his conduct to the relevant norms and thus was not a fully responsible agent... resentment of that person would make no more sense than resenting a sudden storm that soaks me. Again, there is nothing here to forgive.

for his conduct even though he is morally ignorant (and even though he need not apologize or self-blame while in the grip of his mistaken beliefs).

I'll pursue two kinds of responses to the present challenge. According to the first, there are counterexamples to the principle that A can be forgiven for ϕ-ing at t, only if A is blameworthy for ϕ-ing at t. According to the second response, contrary to initial appearances, there is not something for which wrongdoers like Oscar can be forgiven. Instead, there is merely something for which they can be *excused*.[60]

5.2 Blameless Forgiveness

Are there situations in which someone can be forgiven for ϕ-ing despite not being culpable for ϕ-ing? Espen Gamlund (2011) argues in the affirmative based on cases in which agents face a moral dilemma.[61] For example, Gamlund discusses a case in which a politician must decide between letting thousands of innocent persons die *via* an explosion or allowing a single person to be tortured for pertinent information. Gamlund contends that in such cases, reasonable disagreements can emerge about which course of action is justified. Suppose that the right thing to do for such an agent is to torture the one innocent person in order to save the thousands and the politician decides to do so based on these reasons. Gamlund suggests that persons could reasonably reject such reasons as justifying the politician's action. He writes, "Even though a bystander to the case accepts the politician's justification for action, the victim may still reasonably reject his justification and feel wronged... a case can be made for the claim that she may legitimately claim[62] that there is something to forgive (or not to forgive)" (115). Insofar as one cannot be blameworthy for doing what is in fact the right thing to do (i.e., what one is *all things considered* justified in doing)[63], if Gamlund is

[60] A third response? It can be appropriate/fitting for affected parties to resent agents who are not blameworthy but that are merely causally responsible for harms that they have endured. It is this kind of resentment which characters like the daughter can forego/overcome which we're tracking when we think that there's something for which they can forgive.

[61] More precisely, Gamlund labels such cases "tragic dilemma cases" or a case of "dirty hands."

[62] I take it that what Gamlund means in saying that the victim can "legitimately claim to forgive" is that they make no mistake in claiming that they are forgiving. That is to say, there is something to forgive in these cases

[63] Johann Frick (draft) suggests that we can do what is all things considered not wrong (permissible) and at the same time perform a wrong against a specific party. He thinks

right, then the mere fact that A can be forgiven for ϕ-ing does not entail that A is blameworthy.

We find a second kind of case in which an agent who is blameless can be forgiven in Bernard Williams (1981). In discussing an instance of bad moral luck, Williams considers the case of a lorry driver, who through no fault of his own, runs over a child. As Williams sees it, there is a special kind of regret which the driver is naturally inclined to experience that is intimately linked to his realization that he is causally responsible for the harm (28). Importantly, Williams contends that it would be inappropriate for the driver not to experience this special form of regret which he labels "agent-regret." This is so even though the driver is aware the he is not culpable for the harm. Moreover, at least in some cases, agent-regret is said to be "expressed" by certain actions (or dispositions to act) which "constitute or at least symbolizes some kind of recompense or restitution" (28).

Suppose that the parents of the victim were to meet with the lorry driver in the aftermath. It does not sound out of place or presumptuous for the parents to tell the lorry driver that they forgive him. The most straightforward explanation here is that there is something for which he can be forgiven even though *ex hypothesi* he is blameless. If so, then that

this is what is occurring in cases such the trolley case variant which features the large man on the footbridge. If you push the large man, provided that there are a significant number of innocent lives you're saving in the process, you do what is permissible and you wrong the large man. Hence, perhaps the thing to say about the case that Gamlund discusses is that the agent does the right thing in torturing the innocent and also performs a wrong against them. However, even if one can wrong a particular person in this way, it remains to be seen whether one is blameworthy for it. Crucially, our interest in entertaining the current objection is concerns whether such an agent is blameworthy. Frick appears to take for granted a Scanlonian account of blameworthiness according to which A is blameworthy for ϕ-ing *just in case* A has impaired their relationship with another in ϕ-ing. Further, Frick intuits that if you push the large man to save the others, you impair your relationship with him (and perhaps others). However, even spotting Frick that Scanlon's picture of blameworthiness is right, I don't share Frick's conviction that in pushing the large man in this scenario, you thereby have impaired your relationship with them. To impair your relationship with another in the relevant sense is to give them reasons to alter their own intentions and expectations of you (i.e., to modify their relationship with you). It is one thing to perceive that I am pushed to my death by another out of malice or disregard for my safety and quite another to perceive that I have been pushed to my death in order to save a large population of innocents. The latter does not necessarily give one reason to distrust or downgrade one's relation to the agent. Curiously, Frick discusses a case from Joel Feinberg (1978) in which an agent trespasses and destroys the property of another (to build a fire) in order to survive. Frick does not think that in doing so the agent impairs her relationship with the property owner (and so is not blameworthy). It seems to me that setting aside the difference in the seriousness of the harms either both have impaired their relationship with others or neither have.

there is something for which A can be forgiven does not entail that A is or was blameworthy for any action.

On a related note, one might even begin to wonder whether expressions of forgiveness should be taken at face value at least in the kinds of contexts which the lorry diver finds himself[64]. As Joanna North (1987: 501) notes, we often say things like, 'you should forgive him-you know he is very young'. In asserting such things we seem to express the idea that one should excuse rather than forgive. [65]

This brings us to a second type of response which depends on distinguishing between *excusing* an agent and forgiving her. According to this response, it's only apparently the case that, say, the daughter who is harmed by the 1950's father's inequitable treatment has something to forgive. She may of course have some hostile attitudes like resentment to forego or overcome. However, once she recognizes her father's moral ignorance (and the fact that he is not culpable for it), it would be a mistake for her to *forgive* him. Instead, there is something for which he can be excused and his (non-culpable) moral ignorance provides such grounds for excuse. Indeed, it strikes me as initially plausible that the daughter upon reflecting on the father's mistaken moral beliefs would naturally find her resentment dissipating which is often what happens when learn of an excusing condition.

5.3 No Longer Ignorant, Must Apologize?

At this juncture, externalists may wish to pursue a different line of objection having to do with the responses we expect of our wrongdoers once they have realized their errors. Suppose some time after Oscar slights Kevin, he comes to realize that promise-breaking in such circumstances is wrong and thus comes to believe that he has wronged Kevin. In the absence of overriding reasons to do so, shouldn't he now apologize to Kevin (and blame himself)? Likewise if Rosen's "run of the mill" 1950's sexist father comes to learn that he wronged his daughter, it seems as though he ought (pro tanto) to apologize. Additionally, the following is initially plausible.

[64] David Sussman (2008) contends that it's only quasi-forgiveness that is granted in such cases.

[65] Sussman (2008:788) similarly writes, "When I say that I'm sorry that I couldn't pick you up at the airport, I may be starting to offer not so much an apology, but rather an explanation and perhaps an excuse." Sussman distinguishes these from "real apologies".

> *No Retroactive Blameworthiness (NRBW):* Where A performed ϕ at t_{-1}, If A is not blameworthy for ϕ-ing at t_{-1}, then A is not blameworthy for ϕ-ing at t .

According to NRBW, if you're blameworthy *now* for something you did (or did not do) in the past, then you must have been blameworthy for it *at the time of the action (omission)*. Putting these considerations together it follows that an agent who performs a wrongful act (at t), as a result of a mistaken moral belief, is blameworthy (at t).

In response, I submit that on some occasions agents ought (pro tanto) to apologize even when they are not culpable for what they have done. That is to say, even if we spot my critic NRBW and the premise that the disabused (and once morally ignorant) wrongdoer must presently apologize, it doesn't follow that they are blameworthy. It may be helpful at this juncture to be clear about what I am committed to concerning the relationship between blameworthiness on the one hand, and the *pro tanto* demand that (some) wrongdoers apologize, on the other. My argument for the claim that agents like Oscar (i.e., those with mistaken moral beliefs which lead to wrongdoing) are not culpable depends on the following principle.

> Blameworthiness is Sufficient for Apology Norm (BWSA): If A is blameworthy for ϕ-ing, then A ought (pro tanto) to apologize for ϕ-ing (providing A has not done so already).

Such a principle is compatible with the claim that agents like Oscar (despite not being blameworthy) should (pro tanto) apologize. While I endorse BWSA, I deny the following related principle which is incompatbile with the view that Oscar could be blameless and yet owe an apology.

> Blameworthiness is Necessary for Apology (BWNA): A is blameworthy for ϕ-ing *only if* A ought (pro tanto) to apologize.

In accepting BWSA and denying BWNA, I am arguing that every instance of culpable wrongdoing[66] comes with a *pro tanto* demand for an apology. However, not every instance in which an agent owes an apology is one where the agent is blameworthy for some action (omission). Why should we think that the second of these is true (i.e., that BWNA is false)?

[66] More cautiously: every wrongdoing against another person. I don't think we owe animals apologies. However, we ought to self-blame and in some cases make restitutions if we wrong animals.

As I noted earlier, it is common ground to parties of the present debate that mistaken non-moral beliefs can be excusing. That is, if Anne pours cyanide in Bill's coffee because, through no fault of her own, she falsely believes it is sugar, then she is not blameworthy. Normative externalists such as Weatherson and Harman grant this (albeit for different reasons). It's just that they think mistaken *moral* beliefs are different. Suppose that Anne's action harms Bill. In the aftermath, assuming she is aware of the outcome, it is natural that she would feel terrible and moved to apologize to Bill. In this way, this case is like agent-regret cases we find in Williams (1981). Indeed, another feature that such cases have in common with agent-regret cases is that there is a sense in which the agent ought (pro tanto) to feel a special kind of regret or perhaps guilt. What is more, we would view the lorry driver with some suspicion if he weren't at least inclined to apologize to the victim's parents. Likewise, we would view Anne with suspicion, if she were not at least inclined to apologize to Bill which indicates that there is a normative expectation that she apologize (in the absence of overriding reasons).[67] Indeed, a lack of apology in these situations to the affected parties seems to constitute or at least indicate something like callousness[68] on the part of the agent. However, *ex hypothesi*, Anne (just as the lorry driver) is blameless.

Importantly, we should also consider what the object of Anne's apology should be. It seems to me insufficient for Anne to merely offer up condolences for a bad state of affairs. For instance, she would be subject to moral criticism if (in the absence of overriding reasons) she were simply to assert, "I'm sorry that this happened to you" or "I'm sorry that you were poisoned." What appears problematic about these kinds of locutions in the present context is that they don't acknowledge the agent's causal role in bringing about the harm. This is similar to the manner in which as Williams (1981) points out, the lorry driver should feel more than the kind

[67] Williams (1981:28) contends that the lorry driver should not only experience agent-regret, but also to be inclined to compensate the surviving family members of the victim. However, strangely, he does not speak of the need for an apology. Still, it seems to me that insofar as we would view such an agent with suspicion, if he were not to experience agent-regret (or even guilt), we would feel similarly about him if he were not inclined to apologize.

[68] D'Arms and Jacobson (ms.) explicate agent regret cases in the same way. As they see it, it is *unfitting* for the lorry driver to experience guilt and yet it would be *morally inappropriate* if he did not experience guilt.

of regret that third-party observers are inclined to experience in response to the accident (i.e., spectator-regret or guilt). In contrast, it seems felicitous for her to say something like, "I'm sorry for poisoning you" or "I'm so sorry for pouring poison in your cup." Moreover, if Anne were to apologize in these ways she does not seem subject to moral criticism for not apologizing.[69] Thus, we have reason to think that A can be blameless for ϕ-ing, even though she ought (pro tanto) to apologize (for ϕ-ing).[70] Indeed, the same is true of situations in which (once) ignorant wrongdoers must apologize upon realizing the error of their ways. Oscar, upon learning that he was mistaken, would be subject to moral criticism if he did not apologize to Kevin. In particular, he should (pro tanto) apologize for inconveniencing Kevin, or for not giving him the tickets as promised. Merely asserting "I'm sorry you were inconvenienced" seems insufficient.[71] However, as we've just seen it doesn't follow that Oscar is blameworthy. Thus, reflection on cases in which an agent causes harm and is not morally responsible for her conduct casts doubt on BWNA undermining the present worry.[72]

[69] I take it that such apologies could also be genuine or non-genuine depending on the agent's attitudes. Plausibly, she should also feel guilt or something like agent-regret and apologize in the foregoing way in virtue of these emotions.

[70] David Sussman (2018) argues that there is something special about the kind of apologies that agents like the lorry driver must offer. He calls them "quasi-apologies" and contends that in contrast to apologies, quasi-apologies "need not express any sort of change of heart, any resolution to act differently in future (806)." One might wonder whether agents like Anne merely have reasons to given a quasi-apology (which does not entail blameworthiness) whereas, agents like Oscar (upon being disabused) must (pro tanto) offer an apology (which does entail blameworthiness). However, the basis of Sussman's distinction is far from obvious. In the first place, it's not clear that the kinds of apologies that are owed in paradigm cases of culpable wrongdoing must *express* a change of heart or a resolution to behave differently. Often, saying something like, "I'm sorry for lying to you" will suffice to serve as an apology and there will be nothing more that is required of the wrongdoer. Naturally, things might be different if the agent has made it a habit to lie and responds merely with the same apology. In that case, it's plausible that the recipient of the apology has good reason to doubt that the apologizer really is sorry. Of course, if the lorry driver makes a habit of running over persons (even due to bad moral luck), merely offering an apology may no longer suffice. Sussman also alleges that a distinguishing mark of a quasi-apology is that when it is appropriately offered, the recipients "should respond by telling him that none is needed" (ibid). This purportedly contrasts with the case of apologies which when appropriately offered can either be accepted (so that forgiveness is granted) or rejected. However, it's not clear that the recipient of say, the lorry driver's apology, *must* respond by telling him that no apology is necessary. In fact, it seems appropriate for the parents of the child that has been injured to *accept* the apology and even forgive the driver.

[71] Indeed, he seems subject to moral criticism if he merely felt third-person disappointment rather than guilt and/or agent-regret.

[72] Does an apology for one's action in these contexts express or implicate culpability? I'm

6 Conclusion

There are things that wrongdoers must do in response to their wrongdoing. That is to say, wrongdoers who fail to self-blame or apologize for their wrongs can be open to moral criticism. However, a peculiar thing about these norms is that they are sensitive to the wrongdoer's beliefs about the valence of her conduct. Agents who have in fact done something wrong may fail to believe that they have because they have mistaken factual beliefs or mistaken moral beliefs. Of course, we might find it morally objectionable that they have these mistaken beliefs as well as that they fail to track the fact that they have fallen morally. However, while in the grip of their ignorance, it seems they are not required to apologize or self-blame. Not only does this suggest that some norms are internalist, but it suggests a novel argument that moral ignorance can be exculpatory.

Finally, I end this discussion with a suggestion about future work. While I have focused in this chapter on what wrongdoers must do, and in particular, on the norm of apologies, there may be things we must do *qua* epistemic agents who fail epistemic norms. Indeed, recently a number of philosophers[73] have argued that there is blame that is distinctively epistemic. In that case, it could be that there is a normative expectation that we blame ourselves, epistemically for say, forming beliefs haphazardly. Some authors have also suggested that there can be epistemic norms concerning various activities such as evidence-gathering[74], preventing epistemic bads from occuring[75] and attention allocation.[76] Further, some[77] speak of epistemic injustice and epistemic harms. Thus, insofar as we can fail to adhere to such norms, we may be called to respond in certain ways. That is to say, perhaps there are some things that epistemic agents must do when they

not sure. While I think that it is frequently if not paradigmatically the case that apologies for one's conduct expresses/implicates culpability, I am not committed to anything stronger. Perhaps in abnormal contexts such as agent-regret cases, apologies for one's conduct does not express/implicate that one is culpability but does express/implicate one's causal role in the harm.

[73] See Jessica Brown (2020a, 2020b), Cameron Boult (2021a, 2021b), and Adam Piovarchy (2021). Also see Antti Kauppinen (2018) for a discussion on Epistemic Accountability).

[74] Richard Hall and Charles Johnson (1998), and Alex Worsnip (2019).

[75] Jennifer Lackey (2018) and (2020).

[76] Susanna Siegel (2017)

[77] Miranda Fricker (2007), Christopher Hookway (2010), David Coady (2010), and Kristie Dotson (2011).

fail to do as they (epistemically) must including, but not limited to self-blame and offering apologies. As a result, the explorations in this work suggest interesting questions for the epistemic domain which are not only interesting in their own right, but also bear on the nature of relationship between the epistemic and moral. For instance, are epistemic agents who flout epistemic norms (due to accepting false epistemic principles for which they are not culpable) required to self-blame (epistemically)? If they don't self-blame, are they subject to criticism that is distinctively epistemic? What of the epistemic agent (who through no fault of her own) commits an epistemic injustice against another? Must she apologize or otherwise seek to make amends? If she does not, is she deficient in some way *qua* epistemic agent? And are there any cases in which whether an agent is (epistemically) blameworthy for flouting an epistemic norm can be a function of her beliefs concerning epistemic principles? While I will not be able to address these questions here, I anticipate that further investigation will reveal interesting analogies and disanalogies between the moral and epistemic domains.

Chapter 3: Fittingness, Relationality, and Blameworthiness

Abstract

I argue that "being the fitting target of the negative reactive emotions" is a relational notion i.e., there are occasions when single object can be the fitting target of one subject's negative reactive attitudes and not that of another. In contrast, a standard assumption concerning blameworthiness is that it is not relational in this way. That is, on the orthodox view, an agent cannot be blameworthy *in relation to* one person and not another. If we accept this standard picture concerning blameworthiness, then we should reject the theory that 'A is blameworthy for ϕ' is coextensive with 'A is the fitting target of the negative reactive emotions for ϕ'. More generally, this is some reason to doubt that we can analyze blameworthiness in terms of the negative reactive attitudes. This is because replacing the notion of fittingness with other notions such as *desert* either leads to similar problems or threatens to be uninformative. The upshots for those committed to the non-relationality of blameworthiness are two-fold. First, this casts doubt on the widely endorsed view that 'being blameworthy' is coextensive with 'being the *appropriate* target of the negative reactive attitudes.' Second, it raises doubts for affective theories of blame according to which blaming *just is* targeting one with the negative reactive attitudes.

1 Introduction

A number of philosophers[1] accept the thesis that A is blameworthy for ϕ *if and only if* A is the appropriate target of the negative reactive attitudes for ϕ. One initially promising way to flesh out such an analysis is to construe the appropriateness at issue in terms of fittingness. It has been suggested[2] that object-directed emotions can be fitting or unfitting depending on facts about the object in question. In that case, perhaps reflecting on the conditions under which an agent is the fitting target of reactive emotions such as

[1] See for instance, Allan Gibbard (1990), R. Jay Wallace (1994), Michael Zimmerman (2010), John Martin Fischer and Mark Ravizza (1998), Peter Graham (2014), and Jada Twedt Strabbing (2019).

[2] Justin D'Arms and Daniel Jacobson (2000a, 200b).

resentment may provide insight into the nature of blameworthiness. However, this approach faces a significant obstacle or so I argue. The problem is that there appears to be a structural difference between an agent's status as blameworthy on the one hand (at least as standardly conceived) and an agent's status as the fitting target of certain emotions like resentment, on the other.

It is part of orthodoxy that whether a wrongdoer is blameworthy for what she has done depends solely on facts about her or the situation under which she acted wrongly. In that case, an agent can't be blameworthy *in relation to* one person while not being blameworthy in relation to another. However, I argue that there are situations in which an agent is the fitting target of resentment in relation to one would-be-resentor and not in relation to another. Exploiting such cases, I argue that the fittingness conditions of the negative reactive attitudes are importantly different from the conditions that make an agent blameworthy. This calls into question the theory that 'being blameworthy for ϕ' is coextensive with 'being the fitting target of the negative reactive attitudes for ϕ'. Furthermore, I suggest that the kinds of cases seem to generalize to neighboring notions so as to cast a more general doubt on the prospects of analyzing blameworthiness in relation to the negative reactive attitudes.

2 The Fitting Target of Emotions

Despite having many proponents, I will argue that the following proposal is unpromising.

> Appropriateness Account of Blameworthiness(AAB): A is blameworthy: (at t), for ϕ *if and only if* A is the appropriate target of the negative reactive attitudes (at t) for ϕ.

However, given that there are many senses of 'appropriate,' I will restrict our discussion to one formulation of the above principle which construes appropriateness in terms of fittingness or correctness. In doing so, I'm following the suggestion made by D'Arms and Jacobson (2000a and 2000b).[3]

[3] Jada Twedt Strabbing (2019) also follows this suggestion. For Strabbing, being the appropriate target of the negative reactive emotions is coextensive with being blameworthy. She writes, "In other words, an agent is blameworthy for an action if and only if a negative reactive attitude is appropriate toward her on account of it…(3122)" In turn, the

At some point, I'll diverge with them on some of the finer details of what they take to be a requirement of such fittingness conditions. However, for now we can ignore such differences. My point in adopting this framework is merely to focus our attention on a narrower set of conditions which can make an attitude rational or irrational in a particular way as distinct from broader issues about the morality or prudence of such attitudes.

According to D'Arms and Jacobson (D&J hereafter), we can speak about object-directed emotions in terms of their fittingness in a way that is analogous to the manner in which beliefs can be correct. They write, "Emotions present things to us as having certain evaluative features. When we ask whether an emotion is fitting, in the sense relevant to whether its object is Φ, we are asking about the correctness of these presentations... In this respect, the fittingness of an emotion is like the truth of a belief" (72). D&J in speaking of emotions in this way are concerned with what they deem to be the "moralistic fallacy" which is committed when someone conflates the moral, prudential, or all things considered inappropriateness of an attitude with its being unfitting.

As a number of writers working in the ethics of blame[4] have noticed, there seem to be situations in which it is inappropriate for some individuals to blame an agent that is blameworthy (and whom they judge to be so). For example, Angela Smith (2007), distinguishes between an agent being morally accountable[5] for ϕ on the one hand, and it being morally appropriate for an individual to hold the agent responsible for ϕ, on the other. Likewise, Marilyn Friedman (2013)[6] draws a distinction between an agent's being blameworthy for ϕ on the one hand, and a would-be-blamer being blam*er*-worthy (i.e., entitled to blame), on the other. It is important to note that a common thread in these discussions is the idea that there are facts which can vary from one would-be-blamer to the next, which can make a difference to whether it is in some sense, appropriate for each subject to blame one and the same agent, for one and the same action. Moreover, in each of these cases, it is supposed by the writers that the relevant agent is

appropriateness at issue is fittingness or accuracy. She adds, "In this paper, I assume that we should understand 'appropriateness' as 'accuracy.'

4 See Justin Coates and Neal Tognazzini (2012).

5 For Smith, holding someone morally accountable/responsible for wrongdoing consists in active blaming, where the latter is paradigmatically to target the agent with one of the negative reactive attitudes.

6 Also see Patrick Todd (2019).

nonetheless blameworthy for what she has done (and that the would-be-blamers judge this to be so).

Notably, there appears to be a parallel between various senses in which blame may be appropriate or inappropriate and the various senses in which objected-directed emotions may be as well. Thus, my choice to talk about fittingness in relation to the negative reactive attitudes at this juncture, is intended to pick out the set of conditions (which in a narrower sense) render the agent the appropriate target of such attitudes. It is meant to be analogous to the manner in which an agent's blameworthiness is distinct from facts about whether it is morally or prudentially appropriate for anyone to blame her. As such, and for the sake of clarity, we can replace the occurrence of 'appropriate' with 'fitting' to yield the following.

> Fittingness Account of Blameworthiness (FAB): A is blameworthy (at *t*)[7], for ϕ *if and only if* A is the fitting target of the negative reactive attitudes (at *t*) for ϕ.

The gist of FAB is that 'A is blameworthy for ϕ-ing' is coextensive with 'A is the fitting target of the negative reactive attitudes for ϕ-ing'. Indeed, some philosophers have defended just this kind of thesis.[8]

There are various reasons to be attracted to a theory of this sort. In the first place, one might find the Strawsonian inspiration a virtue. That is, there may be grounds for thinking that blaming just is the adopting of negative reactive attitudes say, in response to a perceived lack of proper regard for others. Where being blameworthy amounts to being the fitting target of blame, we would expect the conditions which make an agent blameworthy to be just those conditions that render her the fitting target of the negative reactive emotions.[9] Granted, one need not adopt this Strawsonian account

[7] While I take it that both blame and the negative reactive attitudes can come in various degrees, for the sake of simplicity, in what follows I omit the additional index that would reflect this fact. In order to appreciate the complexity, note how proponents of FAB can remain open to the nature of the correspondence between the degree of blame that is fitting to target A with (for ϕ) on the one hand, and the degree of say, resentment it is fitting to target A with (for ϕ), on the other. For example, it could be that when some degree d of blame is fitting in relation to A for ϕ, it follows that it is fitting to target A for ϕ with some degree c of resentment, but where d is not equivalent to c.

[8] See Jada Twedt Strabbing (2018). Michael Zimmerman (2015) also speaks of fittingness, but intends it to pick out something like the sense in which the blameworthy agent is deserving of certain negative reactive attitudes.

[9] Note that in order for blaming to be identified with targeting someone with the negative reactive attitudes, insofar as one accepts that the relevant emotions can be fitting or

of blame in order to accept FAB. After all, there might be some further set of features which accounts for an agent's being the fitting target of the negative reactive attitudes, and for her blameworthiness, which in turn, grounds FAB. But I suspect a major motivation for FAB is a commitment to a theory of blame in which blame is identified with some sort of emotional response. The significance of this point is that problems with FAB suggest problems for affective theories of blame which are inspired by Strawson (1962). We can put the point in the following way. If Strawsonian-affective theories of blame are correct, then FAB is true. That is, if blaming someone *just is* adopting certain negative emotions towards them (in response to a perception that the agent has failed to show proper regard for others), then it would be surprising to learn that being blameworthy amounted to something different than being the fitting target of the negative reactive emotions i.e., it would be surprising to learn that FAB was false. Thus, problems for FAB cast doubt on such affective theories of blame.[10]

Alternatively, even if one does not identify blameworthiness with being the fitting target of the negative reactive emotions, one might think that the latter notion is somehow more familiar to us, or that it is explanatorily more basic, and thus FAB represents a promising approach to elucidating the conditions under which an agent is blameworthy. We see a similar motivation for so-called fitting attitude theories of value, which purport to account for the goodness of a thing, in terms of that which it is fitting to bear some pro-attitude towards.[11]

The plan for remainder of the paper is as follows. In Section 3, I present a case in which it is fitting for one subject to target an agent with the negative reactive attitudes (at t) for ϕ, but not fitting for another subject to do so. Section 4 is dedicated to answering a worry that our intuitions concerning the vignette featured in Section 3, are tracking something like the moral propriety or impropriety of the negative reactive attitudes rather than their fittingness or unfittingness. In Section 5, I make explicit just how the situation in Section 3 raises a problem for FAB. Further, I conclude that

unfitting (in contrast to morally or prudentially appropriate), one must also grant that blame likewise can be fitting or unfitting.

[10] More precisley, it would raise problems for the conjunctin of such a theory with the standard assumption that an agent can't be blameworthy in relation to one person and not another.

[11] A.C. Ewing (1947), Franz Brentano (1969), and John McDowell (1985) among others.

section with a brief discussion about why we should think my case against FAB casts doubt about the prospects of the more general analysis AAB. Section 6 addresses a potential roadblock to my main argument rooted in the manner in which D&J (2000a and 2000b) conceive of the fittingness conditions of emotions. Finally, in Section 7, I conclude with some general upshots of the exploration.

3 Inappropriate for You and Not for Me

Consider the following.

> *Frank's Theft*: During their final semester at university, Janice, Sarah and Frank embarked on an online business venture together. However, due to their inexperience as well as many economic factors beyond their control, the business failed only after a month. Soon thereafter, upon graduating, the three former colleagues went their separate ways losing touch with one another, entirely.

With the basic story in mind, let's fill in some additional details. First, suppose that shortly after the demise of the venture (a month or so), Sarah is combing through the financial records and learns that Frank cheated the others of some minor profits. While Sarah has no way of contacting Frank (or Janice), she feels resentment towards Frank for what he has done. Supposing that Frank wasn't coerced, brainwashed, or ignorant that what he was doing was wrong (in an excusable manner), Sarah's resentment of Frank seems fitting. However, let's suppose that Sarah hangs onto this resentment. That is, let's skip ahead to 20 years from the time of her initial discovery. Imagine that Sarah goes on resenting Frank even though so much time has passed since she initially felt resentment towards him. Surely, she is subject to criticism for her persistent attitude. At some point, it no longer is fitting for her to resent Frank. In contrast, suppose that Janice is ignorant of Frank's misdeed because she lost all of the financial records during her relocation. In fact, she doesn't learn about Frank's minor theft until two decades following the demise of the business. It now (20 years later) seems fitting for Janice to resent Frank upon learning of what he has done, despite the fact that it was a long time ago.[12]

[12] Interestingly, it doesn't seem as though Janice gets to resent Frank for as long a period of time as Sarah does, before her attitude becomes excessive. Perhaps what this suggests is that the fittingness of one's resentment (in terms of the degree of the emotion) is not

4 From Appropriatness to Fittingness

It might be thought that there is something *morally* vicious about a person who resents a wrongdoer for too long, even if the target continues to be the fitting target of such reactions. We've already noted that an analogous distinction has been made in the case of blame—viz., that it can sometimes be morally inappropriate (on pain of say, hypocrisy) for a particular would-be-blamer to blame an agent who is blameworthy, even when it is appropriate for another to do so. Thus, a proponent of FAB might argue that cases like *Frank's Theft* pose no threat to their analysis. After all, FAB concerns the *fittingness* of the negative reactive attitudes as opposed to broader kinds of appropriateness. So why should we think that in *Frank's Theft*, the inappropriateness of Sarah's continued resentment of Frank 20 years later, and the appropriateness of Janice's resentment at the same time, is tracking the *unfittingness* and *fittingness* of resentment, respectively? In the remainder of this section, I present three separate arguments for the view that we are (at least) tracking differences in the fittingness of resentment in *Frank's Theft*.

4.1 Being Grateful for Too Long

Strawson (1962) suggests that resentment and gratitude are a "usefully opposed pair" (77).[13] While Strawson enumerates a range of both positive and negative reactive emotions, resentment in particular is said to be appropriate for a subject to have towards the wrongdoer, only if the subject is among the wronged. Analogously, gratitude is an appropriate response for a subject, *only if* she is the benefactor of the agent's good deeds in contrast to admiration or pride. Supposing this analogy holds, we can generate an analogous situation as that presented in *Frank's Theft*—let's call it *Tom's Gift*—which helps us appreciate that we're responding to a difference in the fittingness/unfittingness of resentment in the former case.

> *Tom's Gift*: Jeff, Sonny and Tom embarked on an online business venture together during their final semester at university. However, due to their inexperience as well as many economic factors beyond their control, the business failed only after a month. They each incurred a

only a function of how long one has been resenting, but also how long ago the offense occurred.

[13] See Justin Coates (2019) for a dissenting view.

> minor debt of $5. Soon thereafter, upon graduating, the three former colleagues went their separate ways losing touch with one another.

Suppose further that without a word Tom paid off the remaining minor debt so that both Sonny and Jeff would be off the hook. Sonny learns about it a few weeks later (fill in the backstory however you like) and is grateful towards Tom, but Jeff doesn't learn about the matter until about 20 years later (call this time t). It would appear that as in the case of *Frank's Theft*, at time t, it is appropriate for Jeff to feel grateful towards Tom for what he has done because he's just learned about the minor favor.[14] On the other hand, supposing that Sonny has been grateful towards Tom, since the moment he learned about the favor, it seems inappropriate (in some sense) for him to continue to be grateful at t (some 20 years later). What is more, it's implausible to suggest that the difference between our assessments of Sonny and Jeff has anything to do with moral propriety or impropriety. Instead, the best thing to say here is that it is unfitting even if Jeff's reaction at t, is fitting.[15] Sonny's strange and enduring gratitude towards Tom isn't immoral or vicious. Insofar as gratitude is analogous to resentment, we should say the same thing about *Frank's Theft*—Janice's resentment of Frank, 20 years removed from the wrongdoing is fitting, even when Sarah's is not.

It is important to note that for all I've said so far, the possibility remains that there are some acts of beneficence which would make fitting gratitude that is everlasting or at least that would outstrip our lifetimes.

[14] Suppose at t, Sonny is considering doing something nice for someone in his life, and he is reminded of the small favor that Tom paid him 20 years ago. Is it inappropriate (unfitting) for Sonny's current decision to be guided by the fact that Tom gave him the small gift in the past? Further, perhaps Sonny is continuing to be grateful in doing so. Might this pose a problem for my argument? In response, it isn't clear that Sonny's being guided by the relevant fact in this manner entails that he is grateful towards Tom at that time. Assuming again that resentment is a suitable analogue of gratitude, it doesn't seem as though merely taking into account the fact that someone has wronged you as you deliberate about what to do, counts as resenting them. For instance, someone might have shown themselves untrustworthy by past behavior. However, it seems possible for me to forgive and forgo resentment of them despite also trusting them less in my future dealings with them. Thanks to Peter Railton for pressing this worry.

[15] Is it plausible that we are responding to the fact that Sonny's enduring gratitude is an indication of a lack of self-worth? If so, that might be morally objectionable. Thanks to Jesse Holloway for this suggestion. Here's one reason to think not. There are people that have resolved to be grateful for the "minor blessings" in life. Such persons might be inclined to feel grateful for minor favors, for longer than most. However, it hardly seems that this indicates that they have a lack of self-worth. We can stipulate that Sonny is just this kind of person and yet his ongoing gratitude still strikes me as inappropriate in some sense. Thanks to Sarah Buss for providing this example which inspired this response.

Perhaps it's fitting for us to currently feel gratitude towards agents who lived and acted long before our time because what they have done was momentous or because it continues to benefit us today. I suspect something similar is true about wrong actions in relation to resentment as well. But presumably, there are also minor acts of charity as in the case of Tom's Gift which make continued gratitude (at least of a certain degree) unfitting at some point in time (and likewise for wrongdoing and resentment).[16]

4.2 The Vice of Resenting too Long

Further support of the claim that we are tracking a difference in the fittingness of resentment in *Frank's Theft*, is suggested to us by the following case.

> *Timothy's Rumor*: At t, Timothy comes to learn that his coworker Elliott has spread a minor yet unflattering rumor about him to some mutual acquaintances at t_{-1}. As a result, Timothy has been resenting Elliott since t. Elliott has since moved on from the company and the two have lost touch. It is now tn, which is ten years removed from the time that Timothy first learned about Elliott's misdeed. Incidentally, at the same time, Timothy also learns that another one of his (now former) coworkers, Luis, also spread the very same rumor about him to some mutual acquaintances at t_{-1}.

Provided the rumor isn't too serious, it's clear that it would be excessive for Timothy to go on resenting Elliott at t_n (i.e., ten years from the time that he first learned of the wrong). It may of course, also be imprudent or morally vicious for him to go on resenting Elliott. However, there doesn't seem to be any sort of temptation to say that there is anything rational or proper about his continued resentment of Elliott at t_n.

Things are quite different when we consider Timothy resenting Luis at t_n, for spreading a rumor about him at t_{-1}. It seems in some sense that his resentment of Luis is okay, but also a sense in which it suggests some sort of impropriety. The fact that we feel ambivalence concerning Timothy's resentment of Luis at t_n, but no such ambivalence about his continued resentment of Elliott at t_n, is in need of explication. Fortunately, we can make

[16] The phenomena may extend to other positive reactive attitudes. For instance, there may be something unfitting about someone feeling pride for too long. Take for instance, the trope of a middle-aged person reliving their glory days on the high school sports team. Importantly, it doesn't seem morally inappropriate for such a person to relieve their glory days, but there's something amiss. Thanks to Brian Weatherson for this example.

sense of the situation by suggesting that it is unfitting at t_n, for Timothy to go on resenting Elliott for spreading the rumor about him, but it would be fitting for him to resent Luis at t_n. The asymmetry is due to the fact that the fittingness conditions of attitudes like resentment are sensitive to the subject's attitudinal history. That is, whether or not it is fitting for Timothy at t_n, to target Elliott with the negative reactive attitudes, depends in part on whether he has held such attitudes (and for how long) towards Elliott in the past. The same is true regarding the fittingness of his resentment towards Luis. Moreover, there is something morally questionable (and perhaps imprudent) about Timothy hanging onto resentment (whatever the target) for so long. This explains why we feel ambivalence about Timothy's resentment towards Luis at t_n. Plausibly, what might make resentment imprudent or morally vicious in relation to duration (at least in situations like *Timothy's Rumor*), doesn't hinge on facts about the object of one's resentment. For instance, ongoing resentment might be imprudent for Timothy because it's psychologically damaging to him to go on resenting for ten or more years. It doesn't seem less psychologically damaging for Timothy to equally distribute his resentment between two agents so that he is resenting only Elliott for half the time, and only Luis for the remainder. The same is true when we consider situations where it might be morally vicious for Timothy to go on resenting for very long. Plausibly, the viciousness here is a failure to move on with one's life. Timothy would be just as guilty of this vice if he were to equally distribute his ongoing resentment between two subjects, as he would be if his resentment targeted only one agent the entire time. Thus, *Timothy's Rumor* suggests that a subject's attitudinal history can make a difference to whether their current resentment is *fitting*.

4.3 Attitudinal History and Fitting Emotions

Another reason to think we are tracking a difference in the fittingness of resentment in *Frank's Theft*, comes from the fact that this seems to be a part of a more general phenomena. Recently, Oded Na'aman (2021) has argued that there are such things as "rationally self-consuming emotions." Such emotions are said to be "less fitting the longer they endure" (3). Na'aman's aim is not to weigh in on the nature of blameworthiness or anything of the sort. Instead, he wants to draw our attention to a peculiar feature of

emotions in relation to their fittingness conditions. In his discussion he includes among the relevant class of attitudes, the negative reactive emotions, but also others such as anxiety, grief, fear, amusement and the like. As Na'aman notices, there appears to be a class of object-directed emotions which have the following feature: whether it is fitting for a particular subject to have them at a given time can depend on what attitudes one had in the past. He writes,

> Crudely put, we can say that relief is not fitting without a history of frustration or anxiety; desperation is not fitting without a history of hope; satisfaction is not fitting without a history of desire or longing. In short, it is a mistake to assume that we can have a good grasp of the evaluative content of an attitude independently of its relation to the agent's mind over time. The fittingness of an emotion at a time may be partly explained by its broader diachronic context, which includes a properly evolving response to one and the same object with one and the same relevant set of evaluative properties (13).[17]

In each case, whether we're dealing with desperation, relief, or longing, it's easy to come up with situations that are analogous to *Frank's Theft* such that it is appropriate for one subject to, say, feel desperation regarding some state of affairs, but not for another to feel the same. What is more this difference can be based on what sort of attitude(s) each subject had in the past. But why think that the appropriateness at issue in these cases concerns the fittingness of the attitudes?

As in the case of our discussion on gratitude, when we reflect on the details of a relevant case, I contend that it becomes implausible to suggest the appropriateness at issue anything other than fittingness. Consider Na'aman's remark concerning the feeling of relief which is appropriate only in relation to some past anxiety. Suppose that Kris and Jim are about to give a group presentation and Kris has severe anxiety over public speaking, while Jim has no such aversion. In that case (and barring other related anxieties), it can be appropriate for Kris to feel relief that the presentation is over, even if it would be inappropriate in Jim's case. But as in the case of gratitude, it seems implausible to think that the appropriateness at issue

[17] Na'aman also briefly mentions a case similar to *Frank's Theft*, but that involves grief instead of resentment. He writes, "But consider a case where one receives a letter bearing the sad news of a friend's death long after the death had occurred. In such a case, I submit, grief might fittingly diminish later in time, compared to an episode of grief that begins immediately after the loss" (9).

here is merely moral. Jim doesn't do anything immoral if he feels the extent of relief that Kris does in virtue of the presentation being completed.

Furthermore, it's quite easy to stipulate details so that in both Kris and Jim's case, it would be prudentially appropriate to feel relief and yet it would remain the case that there's something nonetheless awry for Jim to experience relief. We can further support these points by reflecting on the sort of judgments we might make of Jim should he feel relief at some state of affairs without the related prior worry. The temptation is to think that Jim doesn't quite understand how worry is supposed to feature in the economy of emotions. Indeed, it appears that someone who resents for too long, (such as Sarah would be if she went on resenting Frank at t, in *Frank's Theft*) is subject to a similar criticism. Thus, I think it most plausible that the appropriateness/inappropriateness at issue in this case involving the emotion of feeling relief, concerns the fittingness of the emotion.

The upshot of the last two sections is the following. There is good reason to think that a single wrongdoer can be the fitting target of the negative reactive attitudes (resentment) in relation to one person and yet not another (at the same time and for one and the same action). That is to say, there seem to be situations in which there is no simple fact about whether A is the fitting target of the negative reactive attitudes for ϕ. Instead, in these situations, A is the fitting target of such attitudes *in relation to* some and *not in relation* to others.

5 What About FAB?

In this section, I aim to make explicit how the observations we have gathered so far are problematic for FAB. To reiterate, my present target is the following.

> Fittingness Account of Blameworthiness (FAB): A is blameworthy (at t), for ϕ *if and only if* A is the fitting target of the negative reactive attitudes (at t) for ϕ.

We have just seen that there are cases in which it is fitting for some and not others to target an agent for a wrongful act. Given FAB, such an agent will be blameworthy *if and only if* they are "the fitting target of the negative reactive attitudes." But given that it is fitting for Janice and unfitting for Sarah to resent Frank for his indiscretion, the question is this: *is Frank the fitting*

target of the negative reactive attitudes for his theft? Notice that on the standard view, Frank is either blameworthy or he is not. This is because as it is commonly conceived, blameworthiness is not a relational status—an agent can't be blameworthy for a single action in relation to one person and not another. Hence it would be a concession at this point for a proponent of FAB to suggest that there is no non-relational answer to our query. However, as we'll see, there appears to be no plausible way out of this concession, either.

In order to account for the relevant kinds of cases (e.g., *Frank's Theft*), the proponent of FAB will need to supplement their analysis with the following of what it means to be the fitting target of the negative reactive attitudes.

> Fitting Target For Anyone: (FT-Any): A is the fitting target of the negative reactive attitudes (at t) for ϕ *if and only if* it is fitting for any B to target A for ϕ with at least one of the negative reactive attitudes (at t).

FT-Any gives us the verdict that in *Frank's Theft*, Frank is not the fitting target of the negative reactive attitudes. This is because it is unfitting for Sarah to resent him (at t for ϕ). In accordance with FAB, that would mean that Frank is not blameworthy (at t for ϕ). Nevertheless, as we've seen, it is fitting for Janice to resent him (at t for ϕ). That means, Frank is neither blameworthy nor the fitting target of resentment and yet it is somehow fitting for Janice to resent him. Something has clearly gone wrong, here. Thus it seems FT-Any will not do. An alternative for FAB advocates is the following.

> Fitting Target For Someone: (FT-Some): A is the fitting target of the negative reactive attitudes (at t) for ϕ *if and only if* it is fitting for some B to target A for ϕ with at least one of the negative reactive attitudes (at t).[18]

According to FT-Some, in *Frank's Theft*, Frank is the fitting target of the negative reactive attitudes at (t for ϕ). This is because there is at least one person (Janice) for whom it will be fitting to resent him (at t for ϕ). FAB in turn informs us that Frank is blameworthy (at t for ϕ). Hence, FT-Some seems more promising than FT-Any. It doesn't force FAB advocates into the view that it can be fitting for someone to resent an agent that is nei-

[18] Thanks to Sumeet Patwardhan and Sarah Buss for asking me to consider this alternative.

ther blameworthy nor the fitting target of the negative reactive emotions. However, it too comes at a steep cost for those committed to FAB.

As we have just observed, the conjunction of FAB and FT-Some permits situations in which (i) A is the fitting target of the negative reactive emotions (at t for ϕ), and (ii) A is blameworthy (at t for ϕ). But it also permits cases in which in addition to (i) and (ii), it is unfitting for some to target A with the negative reactive emotions (at t for ϕ). But just what sorts of facts could make it unfitting for some and not others to resent A, despite her being blameworthy? Crucially, they must be the sorts of facts which can vary from one would-be-resentor to the next. These might include facts about the moral conduct of each would-be-resentor, about the nature of their relationship to the wrongdoer, or as in the case of *Frank's Theft*, facts pertaining to their attitudinal histories.

Now consider twins A and B. Suppose that B performed an identical wrong as A at t, and under identical conditions, only in a different world. B's world differs from A's only in one detail: every person in B's world has whatever property it is that renders it unfitting for some of the subjects in A's world to resent A (t for ϕ). That is, in this world, it is not the case that there is at least some person for whom it would be fitting to resent B (at t for ϕ). What is the proponent of FAB and FT-Some to say here? She seems committed to the view that A is blameworthy for ϕ, but B is not. However, this is to endorse the view that facts about everyone but B (the wrongdoer), somehow renders B not blameworthy. This is to deny a standard assumption that whether a wrongdoer is blameworthy for an action depends centrally on facts about the agent and the circumstances under which she performed the wrong. So it looks like FT-Some comes at a cost for those committed to FAB.

For good measure, I'll consider two more ways in which the proponent of FAB can attempt to account for the relevant cases.

> FT-Sometime: A is blameworthy (at t) for ϕ *if and only if* there is some earlier time t_n at which it is fitting for some B to target A for ϕ with at least one of the negative reactive attitudes.

> FT- In Principle: A is blameworthy (at t) for ϕ *if and only if* it is fitting in principle to target A for ϕ with at least one of the negative reactive attitudes (at t).

According to FT-Sometime, as long as there is some point in time t_n (in the past) at which A was the fitting target of the negative reactive attitudes for ϕ, A will now (at t) be the fitting target of someone's negative reactive attitudes (for ϕ). In turn, according to FAB, she will now (at t) count as being blameworthy for her past performance of ϕ. Thus, Frank would be the fitting target of resentment (and thus blameworthy), 20 years after his misdeed because there is a prior time at which he was the fitting target of resentment. The problem with this view is that it entails that we can be blameworthy for our moral mistakes, forever. Consider what this means for Frank's situation. Suppose 40 years have passed since his minor misdeed and it is now t_2. By t_2 it's neither fitting for Janice nor Sarah to target him with resentment. Indeed, we might even stipulate that his misdeed became widely known and that everyone has resented him at this point, for too long. That is, at t_2, it is unfitting for anyone to target Frank with any of the negative reactive attitudes for his 40-year-old theft. According to FAB and FT-Sometime, he is nonetheless the fitting target of the negative reactive attitudes and blameworthy at t_2 for his minor theft 40 years ago. This strikes me as implausible.

Finally, FT-In Principle attempts to abstract away from facts about whether it is fitting for some or every person at a particular time to target the wrongdoer with any of the negative reactive attitudes. Thus it avoids all of the troubles facing the other proposals. The problem with FT-In Principle is that it is no longer elucidating. What does it mean for a wrongdoer to be the "in principle" fitting object of the negative reactive emotions (as distinct from what any of the other proposals specified)? It is important not to lose sight of a central motivation for an analysis like FAB, namely, the hope that it will shed light on the opaque notion of blameworthiness in more familiar terms. This suggestion betrays this aim.

Taking stock, advocates of FAB need a way to account for cases such as Frank's Theft. We've just considered four attempts and found them each wanting. In fact, the best thing to say about such cases is that the wrongdoer is the fitting target of the negative reactive attitudes in relation to one subject, but not in relation to another. However, given that an agent's blameworthiness is not similarly a relational fact,[19] we have reason to doubt

[19] To be sure, FAB advocates may wish to dispense of this assumption. However, this would be a significant departure from orthodoxy.

FAB.[20]

For all I've said about FAB, at the outset of our discussion, I noted that my aim was to also cast some doubts concerning the following more general thesis.

> Appropriate Target Account of Blameworthiness (AAB): A is blameworthy (at t) for ϕ *if and only if* A is the appropriate target of the negative reactive attitudes (at t) for ϕ.

How does casting doubt on FAB, call AAB into question? Recall that in considering the former, we set aside various senses in which an agent might be the appropriate or inappropriate target of emotions like resentment. Notice that if, for instance, the proponent of AAB were to suggest that we should understand the relevant sense of 'appropriate' as moral or prudential, such an analysis would suffer the very same kind affliction as FAB. This is because whether or not A is the (morally or prudentially) appropriate target of resentment can be a relational fact—i.e., it can be appropriate in both of these senses for some to resent A for ϕ, and not others. What is more, FAB was initially promising because it made use of a different dimension of evaluation distinct from the moral or prudential, which seemed to readily apply to emotions. However, this was found wanting. The challenge then, for proponents of AAB, is to suggest some other sense in which an agent can be the appropriate target of the negative reactive attitudes, which is distinct from the moral, prudential, and even the fittingness sense[21] Moreover, such an analysis should pick out a status or property which does not admit to troubling cases such as *Frank's Theft* and yet manage to be elucidating[22] in terms of the nature of blameworthiness. I'm doubtful that such a theory

[20] Jules Coleman and Alexander Sarch (2012) (hereafter, C&S) present a similar case against AAB. C&S offer a purported counterexample directly against AAB involving a single resentor that has been resenting a wrongdoer for too long. However, the success of their counterexample depends crucially on the assumption that once an agent is blameworthy for ϕ, she is forever blameworthy for ϕ (for any ϕ). While there may be some wrongs for which we are forever blameworthy (particularly heinous ones), it seems implausible that we are forever blameworthy for even minor mistakes.

[21] It is not uncommon to read/hear philosophers speaking about an agent being susceptible to, deserving of, or it's being fair to target her with the negative reactive attitudes. However, I contend that these notions are either too opaque to provide a promising theory, or else they will track something very much like "being the fitting target of" and thus will be prone to the same instability. For instance, it seems plausible that there are situations in which A is deserving of B's resentment and not C's (for ϕ at t).

[22] As we noted, a strong motivation for a thesis like AAB is the idea we have a fairly good grasp of when an agent is the appropriate target of the negative reactive attitudes.

is forthcoming.[23]

6 Emotional Presentations and Fittingness

At the outset of our discussion, I noted that while I will be appropriating D&J's (2000a and 2000b) notion of fittingness, I would eventually be parting ways with them in a certain respect. In this section, I will present our difference(s) in terms of another potential challenge to my claim that we are tracking the fittingness/unfittingness of resentment in our original vignette. After having answered the worry and explaining where D&J and I seem to part ways concerning the fittingness conditions of the emotions, I suggest how my discussion extends beyond the likes of FAB to a more general analysis of blameworthiness (AAB).

According to D&J (2000b), for an object-directed attitude like amusement to be fitting, is for there to be a match between the presentation of the object as given by the emotion on the one hand, and the way the world is, on the other. To remind the reader, they write, "Emotions present things to us as having certain evaluative features. When we ask whether an emotion is fitting, in the sense relevant to whether its object is ϕ, we are asking about the correctness of these presentations (72)". Additionally, in discussing the fittingness conditions of amusement as distinct from broader appropriateness conditions, they add, "The only relevant considerations are those reasons that speak to whether an emotion correctly represents its object (66)."

By D&J's lights, the emotion envy for instance, presents its object as having some feature which one desires, but also lacks. In turn, envy of a person (for having a trait) will be fitting only on condition that the person actually has the feature *given in the presentation*. That is, on their account of fittingness (in relation to object-directed emotions) when we ask whether or not an attitude like resentment is fitting, we are asking whether or not the object of resentment (e.g., a particular agent) has certain features (which render her fitting of resentment), which coincides with what the attitude

[23] Gideon Rosen (2015) defends an analysis which he refers to as the Alethic view. On his proposal, it is appropriate to target A with the negative reactive attitudes for ϕ, just in case the ingredient thoughts of resentment are true. Strabbing (2018) understands this as a kind of fittingness which is narrower than what I have in mind and argues for a similar view. In particular, it's the kind of fittingness which is imagined in the objection considered in Section 6. There we saw that while this might save the likes of FAB/AAB, we lose a simple way to explain our reactions to a range of cases.

presents its object as having. Notice that on such an view, there appears to be little room for the kind of relationalism[24] for which I have been arguing in this work. In *Frank's Theft*, whether or not the presentation of Frank as having certain properties (as given by resentment) reflects reality, depends on Frank rather than on anything concerning Sarah or Janice's attitudinal history. Thus, if D&J's (2000b) account on the nature of the fittingness conditions of emotions is right, then my argument seems like a non-starter.

However, the objection has bite only inasmuch as we grant that the fittingness of emotions is closely analogous to the manner in which attitudes like belief can be correct. Indeed, D&J are attempting to capture the way in which an emotional response can be suitable in some sense (despite being imprudent or immoral) and which is also familiar to commonsense.[25] This leads them to consider the manner in which beliefs can be correct (whether or not they are prudent or immoral). On popular ways of characterizing the representational content of beliefs, the belief *that* P presents the world as a P-world. Furthermore, insofar as P is true, such a belief will be correct or accurate. Importantly, there's nothing more to it. Hence, insofar as the fittingness conditions of emotions are to be analogous, we might expect something similar when it comes to the negative reactive attitudes.

By understanding what is underwriting the current worry, we can readily address it. We have come to a theoretical choice point. Either we can insist that the facts pertaining to the fittingness of emotions are closely analogous to the correctness conditions of beliefs *as it concerns being given entirely by the representation* or we can loosen the analogy between the two and be more inclusive. At this juncture, I contend that the scale tips in favor of the latter for the following reasons. In the first place, beliefs and emotions seem different from the start, when it comes to the fittingness or correctness conditions. Indeed, this is true even by D&J's lights. I say this

[24] In more recent work, D'Arms and Jacobson (manuscript) allow that whether an attitude is fitting for a particular person to have can depend on facts about her context which I take to mean that they permit a kind of relationalism about fitting attitudes.

[25] D&J write, "Talk of the fittingness of emotions may sound recherche, but 'fittingness' is simply intended as a technical term for a familiar type of evaluation. Endorsement and criticism of emotions on grounds of fit is a crucial tool of our ordinary thought about them, and of our folk psychology. For instance, the homily that "The grass is always greener on the other side of the fence" warns against the common tendency to overrate the value of things we don't possess." We thus use this proverb to criticize ourselves and others for feeling envy (or mere longing) when the rival's possession isn't really enviable."

because paradigmatic object-directed emotions such as envy and fear are such that whether they are fitting responses in various circumstances turns out be a relational fact. The same is not true of belief. So, for instance, D&J write, "...envy portrays a rival as having a desirable possession which you do not, and it presents this circumstance in a specific negative light" (72, italics added).

Notice that among what D&J take to be the representation content of envy, are facts about the envier. What this allows is that it can be fitting for one subject to envy A for F and unfitting for another to do so. After all, we don't all lack or want the same things. If you have a promotion that I already have or don't even desire, then my envy (entirely apart from whether it would be moral or prudent) would be unfitting. But we can't generalize from that to say that it would likewise be unfitting for others to envy you for the very same promotion. This point extends to other paradigmatic object-directed emotions.[26] In contrast, beliefs are different. The accuracy or correctness of a belief depends entirely on the truth value of its object, namely, a proposition. Given that propositions can't be true for you and false for me[27], there's something fundamentally different about the fittingness of emotions on the one hand, and the correctness or accuracy of beliefs on the other. Thus, whether or not an emotion is a fitting response to a situation/object can be a relational fact, whereas the same is not true of belief.[28] As such, the analogy between beliefs and object directed emotions in terms of their fittingness conditions is from the start, a tenuous one.

Moreover, by allowing that the fittingness of an emotion can depend in part on a subject's attitudinal history, we can explicate cases such as *Frank's Theft, Tom's Gift, Timothy's Rumor,* as well as the kinds of examples presented by Na'aman (2021). Recall, that in each of these cases, it seems

[26] For instance, it can be fitting for A to fear the chicken pox and unfitting for person B. Suppose person A has never had the chicken pox before while person B has.

[27] Perhaps *de se* beliefs can be relational in this way. However, it remains to be seen what to say about the accuracy conditions of such beliefs and whether they would provide a promising model of the fittingness of emotions. Thanks to Peter Railton for mentioning this possibility.

[28] What should we say about the suspension of belief? Jane Friedman (2013) makes the persuasive case that suspending judgment about P is a propositional attitude, which has as its object P. If we permit that such an attitude can be fitting or unfitting, then we must allow that facts beyond the truth value of the object can make a difference to whether the attitude is fitting. Arguably the same goes for other propositional attitudes such as endorsing that P or accepting that P.

implausible to suggest that we were tracking mere disparity in the moral or prudential appropriateness of certain emotions. By suggesting that there are differences in the fittingness of the various emotions in each of these cases, we can explain what is happening in each circumstance, without having to proliferate additional dimensions of evaluation. Thus, we have good reasons to deny that the fittingness conditions of any emotion must be given entirely by its representational content.

7 Conclusion

In this work, I have argued that if we accept certain widely held assumptions pertaining to blameworthiness, there are reasons to deny a theory which analyzes an agent's blameworthiness in terms of her being the fitting target of the negative reactive attitudes. Moreover, this suggests that the prospects of the following analysis are dim. (AAB): A is blameworthy (at t), for ϕ *if and only if* A is the appropriate target of the negative reactive attitudes (at t) for ϕ. One option for those that are committed to either of these analyses is to allow that an agent can be blameworthy (for ϕ at t) relative to one would-be-blamer and not another. As we noted earlier, this would be heterodoxy. While I'm not in principle, opposed to such a move, I take it that this will strike many as far less attractive than abandoning the likes of FAB or even AAB.

Another possibility which I will not have room enough here to explore is that proponents of FAB/AAB may wish to analyze blameworthiness in terms of guilt. Notice that we've focused our attention on resentment as our paradigm reactive attitude and our observations seemed to generalize to emotions such as indignation which are also other-directed emotions. Self-directed reactive attitudes such as guilt seem importantly different. This is because there are not going to be cases in which A is the fitting target of B's guilt but not C's given that guilt is necessarily self-directed. Perhaps then proponents of FAB could construe an agent's blameworthiness in terms of her *being the fitting target of guilt*[29] or other self-directed negative reactive attitudes.

Our observations are not only telling about the nature of blamewor-

[29] See Randolph Clarke (2016) and Andreas Carlsson (2017) for recent proposals in this vein.

thiness, but also present a metaphysical upshot. This is because, as we noted earlier, one motivation for FAB/AAB was the metaphysical thesis that to blame an agent *just is* to target her with one or more of the negative reactive attitudes. Indeed, it's difficult to see how FAB/AAB could be false, given a theory of blame which identifies it with certain negative reactive emotions. Hence, the arguments presented in this work against the foregoing analyses, provides some evidence to suggest that blaming A for ϕ is something other than targeting A with one of the negative reactive attitudes for ϕ. Crucially, this usphot is only for those that deny that whether A is blameworthy for ϕ can be a relational fact. Hence, champions of affective theories of blame may wish to part with this standard assumption.

The denial of FAB/ABB also suggests a way to account for a problem related to agent-regret.[30] Briefly, the puzzle stems from the idea that it is rational for an agent to feel regret over an action that has caused a great deal of harm, even when she is not blameworthy (say, because there was bad moral luck involved). Where we understand this to be a puzzle about how regret might be fitting, the problem of agent regret seems to be most serious on views according to which experiencing regret over one's action is tantamount to blaming oneself for that action. This is because according to such theories, the relevant situations of agent-regret are those where an agent's blame of herself is appropriate (fitting), even when she is not blameworthy. However, the denial of the foregoing analyses opens up the possibility that there are situations in which an agent is not blameworthy and yet nevertheless, it can be fitting/appropriate for some to target her with the negative reactive attitudes.[31]

Finally, inasmuch as an agent's blameworthiness for ϕ at t, can come apart from her susceptibility to the negative reactive attitudes for ϕ at t, I think this has important implications for our everyday lives. In recent years, largely due to the internet's impeccable memory and ability to proliferate

[30] Bernard Willaims (1981)

[31] In the cases that draw out the phenomena of agent-regret, it is normally a stipulation that the agent is not blameworthy and is aware of this fact. If we think of such cases diachronically, we might say that the relevant agent who feels reactive attitudes like regret, also (truly) judges that she was never blameworthy for the pertinent action and yet we may think it appropriate (fitting) for her to feel regret for her involvement in a tragedy. In contrast, notice that in cases such as *Frank's Theft*, it seems as though the agent, Frank was at least blameworthy at one time for the relevant action. Thus, one might worry that even the denial of FAB/AAB will not suggest a straightforward solution to the puzzling cases of agent-regret. Thanks to Sarah Buss for this insight.

information, we have come to learn about the past wrongs of many public figures. At least in some of these cases, where the wrong is not very serious (although not entirely trivial either) and where a sufficient amount of time has passed, there is some ambivalence about whether (and to what extent) the agent now should be blamed for what she did in the distant past. Nevertheless, many stand behind their expressions of emotions such as anger and indignation which presumably has the agent now (or her current time-slice) as the target. Perhaps the thing to say about at least some of these cases is that the current agent is no longer blameworthy, but she remains (at least in relation to some) the fitting target of the negative reactive attitudes.

References

Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.

Arpaly, N., & Schroeder, T. (1999). Praise, Blame and the Whole Self. *Philosophical Studies*, *93*(2), 161–188. doi: 10.1023/A:1004222928272

Arpaly, N., & Schroeder, T. (2014). *In Praise of Desire*. Oxford University Press.

Audi, R. (1994). Dispositional Beliefs and Dispositions to Believe. *Noûs*, *28*(4), 419–434.

Baumann, P. (2021). Sorry if! On Conditional Apologies. *Ethical Theory and Moral Practice*, 1–12.

Bell, M. (2013). *Hard Feelings: The Moral Psychology of Contempt*. Oxford University Press.

Bennett, C. (2013). The Expressive Function of Blame. *Blame: Its Nature and Norms*, 66–83.

Bennett, J. (1974). The Conscience of Huckleberry Finn. *Philosophy*, *49*(188), 123–134.

Boult, C. (2021a). Epistemic Blame. *Philosophy Compass*, *16*(8), e12762.

Boult, C. (2021b). There is a distinctively epistemic kind of blame. *Philosophy and Phenomenological Research*, *103*(3), 518–534.

Bovens, L. (2008). XII—Apologies. In *Proceedings of the Aristotelian Society* (Vol. 108, pp. 219–239).

Brentano, F., Chisholm, R. M., & Kraus, O. (1969). *The Origin of Our Knowledge of Right and Wrong. Edited by Oskar Kraus*. Routledge & K. Paul.

Brown, J. (2020a). Epistemically Blameworthy Belief. *Philosophical Studies*, *177*(12), 3595–3614.

Brown, J. (2020b). What is Epistemic Blame? *Noûs*.

Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, *169*(2), 285–311.

Buss, S. (1997). Justified Wrongdoing. *Nous*, *31*(3), 337–369.

Calhoun, C. (1989). Responsibility and Reproach. *Ethics*, *99*(2), 389–406.

Carlsson, A. B. (2017). Blameworthiness as Deserved Guilt. *The Journal of Ethics*, *21*(1), 89–115.

Clarke, R. (2016). Moral Responsibility, Guilt, and Retributivism. *The Journal of Ethics*, *20*(1), 121–137.

Coady, D. (2010). Two Concepts of Epistemic Injustice. *Episteme*, *7*(2), 101–113.

Coates, D. J. (2016). The Epistemic Norm of Blame. *Ethical Theory and Moral Practice*, *19*(2), 457–473.

Coates, D. J., & Tognazzini, N. A. (2012). The Nature and Ethics of blame. *Philosophy Compass*, *7*(3), 197–207.

Coates, D. J., & Tognazzini, N. A. (2013). *Blame: Its Nature and Norms*. Oxford University Press on Demand.

Coates, J. (2019). Gratitude and Resentment: Some Asymmetries. *The Moral Psychology of Gratitude*, 160.

Cohen, G. A. (2006). Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists? *Royal Institute of Philosophy Supplements*, *58*, 113–136.

Coleman, J., & Sarch, A. (2012). Blameworthiness and Time. *Legal Theory*, *18*(2), 101–137.

D'Arms, J., & Jacobson, D. (n.d.). *Rational Sentimentalism.*

D'Arms, J., & Jacobson, D. (2000a). Sentiment and Value. *Ethics*, *110*(4), 722–748.

D'Arms, J., & Jacobson, D. (2000b). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research*, *61*(1), 65–90.

Dotson, K. (2011). Tracking Epistemic Violence, Tracking Practices of Silencing. *Hypatia*, *26*(2), 236–257.

Duff, R. A. (2010). Blame, Moral Standing and the Legitimacy of the Criminal Trial. *Ratio*, *23*(2), 123–140.

Duggan, A. (2018). Moral Responsibility as Guiltworthiness. *Ethical Theory and Moral Practice*, *21*(2), 291–309.

Ewing, A. (1948). *The Definition of Good*. Routledge.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge university press.

FitzPatrick, W. J. (2008). Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge. *Ethics*, *118*(4), 589–613.

Frick, J. (n.d.). *Dilemmas, Luck, and the Two Faces of Morality.*

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Friedman, J. (2013). Suspended Judgment. *Philosophical Studies*, *162*(2), 165–181.

Friedman, M. (2013). How to Blame People Responsibly. *The Journal of Value Inquiry*, *47*(3), 271–284.

Fritz, K. G., & Miller, D. (2018). Hypocrisy and the Standing to Blame. *Pacific Philosophical Quarterly*, *99*(1), 118–139.

Gamlund, E. (2013). Forgiveness Without Blame. In *The Ethics of Forgiveness* (pp. 115–137). Routledge.

Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*.

Goldberg, S. C. (2017). Should Have Known. *Synthese*, *194*(8), 2863–2894.

Graham, P. A. (2010). In Defense of Objectivism about Moral Obligation. *Ethics*, *121*(1), 88–115.

Graham, P. A. (2014). A Sketch of a Theory of Moral Blameworthiness. *Philosophy and Phenomenological Research*, *88*(2), 388–409.
Greaves, H., & Ord, T. (2017). Moral Uncertainty about Population Axiology. *J. Ethics & Soc. Phil.*, *12*, 135.
Hall, R. J., & Johnson, C. R. (1998). The Epistemic Duty to Seek More Evidence. *American Philosophical Quarterly*, *35*(2), 129–139.
Harman, E. (2011). Does Moral Ignorance Exculpate? *Ratio*, *24*(4), 443–468.
Harman, E. (2015). The Irrelevance of Moral Uncertainty. *Oxford Studies in Metaethics*, *10*. doi: 10.1093/acprof:oso/9780198738695.003.0003
Helmreich, J. S. (2015). The Apologetic Stance. *Philosophy & Public Affairs*, *43*(2), 75–108.
Hieronymi, P. (2001). Articulating an Uncompromising Forgiveness. *Philosophy and Phenomenological Research*, *62*(3), 529–555.
Hieronymi, P. (2004). The Force and Fairness of Blame. *Philosophical Perspectives*, *18*, 115–148.
Hookway, C. (2010). Some Varieties of Epistemic Injustice: Reflections on Fricker. *Episteme*, *7*(2), 151–163.
Kauppinen, A. (2018). Epistemic Norms and Epistemic Accountability. *Philosophers*, *18*.
Kelp, C. (2020). The Knowledge Norm of Blaming. *Analysis*.
Kenner, L. (1967). On Blaming. *Mind*, *76*(302), 238–249.
Khoury, A. C., & Matheson, B. (2018). Is Blameworthiness Forever? *Journal of the American Philosophical Association*, *4*(2), 204–224.
Lackey, J. (2020a). The Duty to Object. *Philosophy and Phenomenological Research*, *101*(1), 35–60.
Lackey, J. (2020b). Epistemic Duties Regarding Others. In *Epistemic Duties* (pp. 281–295). Routledge.
Levy, N. (2009). Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick. *Ethics*, *119*(4), 729–741.
Lockhart, T. (2000). *Moral Uncertainty and its Consequences*. Oxford University Press.
MacAskill, W. (2014). *Normative Uncertainty* (Unpublished doctoral dissertation). University of Oxford.
Markovits, J. (2010). Acting for the Right Reasons. *Philosophical Review*, *119*(2), 201–242.
Martin, A. M. (2010). Owning Up and Lowering Down: The Power of Apology. *The Journal of Philosophy*, *107*(10), 534–553.
Mason, E. (2012). Objectivism and Prospectivism about Rightness. *J. Ethics & Soc. Phil.*, *7*, iv.
Mason, E. (2015). Moral Ignorance and Blameworthiness. *Philosophical Studies*, *172*(11), 3037–3057.
McDowell, J. (1985). Values and Secondary Qualities.

McHugh, C. (2012). Epistemic Deontology and Voluntariness. *Erkenntnis*, *77*(1), 65–94.
McHugh, C., & Way, J. (2016). Fittingness First. *Ethics*, *126*(3), 575–606.
Menges, L. (2017). The Emotion Account of Blame. *Philosophical Studies*, *174*(1), 257–273.
Miller, K. (2014). Conditional and Prospective Apologies. *The Journal of Value Inquiry*, *48*(3), 403–417.
Moody-Adams, M. M. (1994). Culture, Responsibility, and Affected Ignorance. *Ethics*, *104*(2), 291–309.
Moore, G. E. (1912). *Ethics* (Vol. 52). H. Holt.
Murphy, J. G. (2003). *Getting Even: Forgiveness and Its Limits*. Oxford University Press.
Na'aman, O. (2021). The Rationality of Emotional Change: Toward a Process View. *Noûs*, *55*(2), 245–269.
Nelkin, D. K. (2008). Responsibility and Rational Abilities: Defending an Asymmetrical View. *Pacific Philosophical Quarterly*, *89*(4), 497–515.
Pettigrove, G., & Collins, J. (2011). Apologizing for Who I Am. *Journal of Applied Philosophy*, *28*(2), 137–150.
Piovarchy, A. (2021). What Do We Want From a Theory of Epistemic Blame? *Australasian Journal of Philosophy*, *99*(4), 791–805.
Radzik, L., et al. (2009). *Making Amends: Atonement in Morality, Law, and Politics*. OUP USA.
Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy & Public Affairs*, 134–171.
Rettler, L. (2018). In Defense of Doxastic Blame. *Synthese*, *195*(5), 2205–2226.
Rosen, G. (2003). IV—Culpability and Ignorance. In *Proceedings of the Aristotelian Society (Hardback)* (Vol. 103, pp. 61–84).
Rosen, G. (2004). Skepticism About Moral Responsibility. *Philosophical Perspectives*, *18*(1), 295–313. doi: 10.1111/j.1520-8583.2004.00030.x
Rosen, G. (2015). The Alethic Conception of Moral Responsibility. In R. Clarke, M. Michael, & A. Smith (Eds.), *The Nature of Moral Responsibility: New Essays* (pp. 65–88). Oxford University Press Oxford.
Ross, W. D. (1930). *The Right and the Good: Some Problems in Ethics*. Clarendon Press.
Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Harvard University Press.
Scanlon, T. M. (2013). Interpreting Blame. *Blame. Its Nature and Norms*, 84–99.
Sepielli, A. (2009). What to Do When You Don't Know What to Do. *Oxford Studies in Metaethics*, *4*, 5–28.
Sher, G. (2006). *In Praise of Blame*. Oxford University Press.
Siegel, S. (2016). *The Rationality of Perception*. Oxford University Press.
Singer, D. J., & Aronowitz, S. (in press). What Epistemic Reasons Are For:

Against the Belief-Sandwich Distinction. In B. Dunaway & D. Plunkett (Eds.), *Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard*.

Smith, A. (2007). On Being Responsible and Holding Responsible. *The Journal of Ethics*, *11*(4), 465–484.

Smith, A. (2013). Moral Blame and Moral Protest. *Blame: Its Nature and Norms*, *27*, 48.

Smith, M., et al. (2006). Moore on the Right, the Good, and Uncertainty. *Metaethics after moore*, 133.

Strabbing, J. T. (2019). Accountability and the Thoughts in Reactive Attitudes. *Philosophical Studies*, *176*(12), 3121–3140.

Strawson, P. (1962a). Freedom and Resentment. In J. M. Fischer & M. Ravizza (Eds.), *Perspectives on moral responsibility* (pp. 1–25). Cornell University Press.

Strawson, P. (1962b). Freedom and Resentment. *Proceedings of the British Academy*, *48*, 187–211.

Sussman, D. (2018). Is Agent-Regret Rational? *Ethics*, *128*(4), 788–808.

Tadros, V. (2009). Poverty and Criminal Responsibility. *The Journal of Value Inquiry*, *43*(3), 391–413.

Todd, P. (2019). A Unified Account of the Moral Standing to Blame. *Noûs*, *53*(2), 347–374.

Todd, P., & Rabern, B. (2022). The Paradox of Self-Blame. *American Philosophical Quarterly*, *59*(2), 111–125.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.

Watson, G. (1987). Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In J. M. Fischer & M. Ravizza (Eds.), *Perspectives on Moral Responsibility* (pp. 119–148). Cornell University Press.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, *24*(2), 227–248.

Watson, G. (2015). A Moral Predicament in the Criminal Law. *Inquiry*, *58*(2), 168–188.

Weatherson, B. (2019). *Normative Externalism*. Oxford University Press.

Williams, B. (1981). *Moral luck: Philosophical Papers 1973-1980*. Cambridge University Press.

Worsnip, A. (2019). The Obligation to Diversify One's Sources: Against Epistemic Partisanship in the Consumption of News Media. In C. Fox & J. Saunders (Eds.), *Media Ethics: Free Speech and the Requirements of Democracy* (pp. 240–264). London: Routledge.

Zimmerman, M. J. (2006). Is Moral Obligation Objective or Subjective? *Utilitas*, *18*(4), 329–361.

Zimmerman, M. J. (2008). *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge University Press.

Zimmerman, M. J. (2010). Responsibility, Reaction, and Value. *The Journal of Ethics*, *14*(2), 103–115.
Zimmerman, M. J. (2015). Varieties of Moral Responsibility. *The Nature of Moral Responsibility: New essays*, 45–64.