Legal Theory, **18** (2012), 101–137.
© Cambridge University Press 2012 0361-6843/12 $15.00 + 00
doi:10.1017/S1352325212000158

BLAMEWORTHINESS AND TIME*

Jules Coleman
New York University
Alexander Sarch
University of Michigan Law School

Reactive emotion accounts hold that blameworthiness should be analyzed in terms of the familiar reactive emotions. However, despite the attractions of such views, we are not persuaded that blameworthiness is ultimately a matter of correctly felt reactive emotion. In this paper, we draw attention to a range of little-discussed considerations involving the moral significance of the passage of time that drive a wedge between blameworthiness and the reactive emotions: the appropriateness of the reactive emotions is sensitive to the passage of time in ways that attributions of blameworthiness are not. There are a number of ways in which reactive emotion accounts might attempt to accommodate the moral significance of time, however. We consider the most important of these but ultimately find them wanting. Accordingly, we conclude that the prospects for the reactive emotion accounts are bleak. Our argument, if successful, has a range of implications for legal theory, most importantly in providing a novel moral basis for statutes of limitations and in shedding light on new avenues in the theory of criminal law generally.

It is uncontroversial that the blameworthiness of agents for what they have done or failed to do is importantly connected to our practices of blaming—a connection that in one way or another is reflected in criminal punishment, and perhaps in other institutions of liability and responsibility as well. It is far from clear, however, how the relationship between blameworthiness and blame emotions like resentment and indignation is to be spelled out. One natural idea—what we can think of as the standard view—is that the fact that someone is blameworthy for what she has done or failed to do is a *reason* or a *ground* that explains why blaming her or otherwise holding her to account would be justified.

However natural the standard view may be, it is also controversial. Many philosophers resist the idea that the aptness of blame is to be explained by

*We would like to thank Facundo Alonso, Elizabeth Anderson, Zac Cogley, Stephen Darwall, John Martin Fischer, Scott Hershovitz, Dan Jacobson, Shelly Kagan, Gabe Mendlow, Scott Shapiro, Will Thomas, the participants in both the Law and Philosophy Discussion Group and the Philosophy Department Faculty Workshop at Yale University, and two anonymous reviewers for this journal for all their helpful comments on earlier versions of this paper.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

blameworthiness in this way. Instead, they reverse the order of explanation and argue that a person's level of blameworthiness for an act or omission[1] is determined by the blame responses that it would be *correct* (in some sense to be specified) to have toward that person for this act or omission. Taking this view of blameworthiness shifts the focus to two other concepts that require analysis: one is the relevant notion of a *blame response*, and the other is what it is for such a response to be *correct*. Within these accounts, a blame response is understood in terms of the so-called reactive attitudes or emotions[2]—guilt, resentment, and indignation. Because of the prominent role that the reactive emotions or attitudes play in such accounts, we refer to them as "reactive emotion accounts"—or, equivalently, "reactive attitude accounts."[3]

Views of this sort are becoming increasingly popular. Allan Gibbard, for instance, argues that "[t]o say that he would be to blame is to say that it would be rational for him to feel guilty and for others to resent him."[4] Jay Wallace has defended the view that one is morally responsible for some action just in case it would be appropriate to *hold* that person responsible for it, where "holding responsible" is understood in terms of the reactive emotions.[5] While others have endorsed such views as well,[6] these accounts all can claim as authority P.F. Strawson's seminal paper *Freedom and Resentment*. Strawson is widely credited as the originator of the reactive emotion accounts,[7] although, given the sweeping nature of his discussion, it is

1. In this essay we assume that the primary object of blame of interest to the theory of responsibility is an agent: not an agent as such, but an agent for a particular act or omission of his. And insofar as blameworthiness attaches to actions, it attaches to *act tokens* and not to *act types*.

2. The phrase "reactive attitudes" can be traced to P.F. Strawson, *Freedom and Resentment*, 48 PROC. BRIT. ACAD. 1–25 (1960).

3. Throughout the paper we use the phrases "reactive emotions" and "reactive attitudes" interchangeably. Although not all attitudes are emotions (e.g., belief, intention), the mental states we are most concerned with here—resentment, indignation, and guilt—can appropriately be referred to as either attitudes or emotions.

4. ALLAN GIBBARD, WISE CHOICES, APT FEELINGS (1990), at 45; *see generally* 44–47.

5. *See generally* R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS (1994), chs. 3 and 4. On Wallace's view, holding responsible in the relevant sense is either to feel an episode of reactive emotion toward the person in question or to judge that such an emotion would be appropriately felt toward her.

6. Peter A. Graham gives a sustained and interesting defense of the view that "X is blameworthy for φ-ing iff$_{def}$ it is appropriate for a blame emotion—resentment, indignation, or guilt—to be felt toward X in response to X's φ-ing." *See* Peter Graham, A Theory of Blameworthiness. Unpublished paper available at http://people.umass.edu/pgraham/Home_files/Blameworthiness%20%2812–1-07%29%20copy.pdf (2007), at 3. *See also* Michael Zimmerman, *Responsibility, Reaction and Value*, 14 J. ETHICS 103–115 (2010).

7. Strawson, *supra* note 2. *See, e.g.*, Ishtiyaque Haji, *Compatibilist Views of Freedom and Responsibility*, *in* OXFORD HANDBOOK OF FREE WILL (Robert Kane ed., 2002) (claiming that "[i]n Strawson's view, the question about the conditions under which an agent is morally responsible is identified with the question of the conditions under which it is appropriate to hold an agent morally responsible. These conditions, in turn, are explained in terms of susceptibility to reactive attitudes"); T.M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME (2008), at 127 ("One alternative view, suggested by Peter Strawson, identifies blame with reactive attitudes such as resentment and indignation"); and Gary Watson, *Responsibility and the Limits*

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

controversial whether he actually advanced any particular version of the view. In the end, whether Strawson advanced a version of the reactive emotion account of blameworthiness is of little importance. His work remains the inspiration for the vast majority of the accounts of interest to us here.

The differences between the standard view and a Strawsonian reactive attitude account are subtle, but not small. On the standard view, one is blameworthy for an act or omission *in virtue of certain facts about one's agential relationship to the doing or omitting*—for example, the fact that it was the product of a defective character, wicked intentions, a bad will, or some other kind of moral failing of the agent. One's blameworthiness in a sense supervenes on such facts, and so the project is to identify what facts about persons and their actions ground blameworthiness. In contrast, on reactive attitude accounts, one is blameworthy *in virtue of facts about the responses that it would be appropriate for others to have toward one.* The burden of such accounts is to identify the conditions under which blaming responses are appropriate or correct. So on the former view, blameworthiness is what explains the justifiability or rationality of the blame responses, while on the reactive attitude account, blameworthiness is to be explained by the blame responses being correctly held, felt, or otherwise directed toward one.[8]

What might lead one to prefer a reactive emotion account over the standard view? For one thing, reactive emotion accounts promise to explain a relatively mysterious normative concept—blameworthiness—in terms of something of which we have a firmer grasp: namely, feelings of guilt, resentment, and indignation, all of which are familiar features of our lives. As well-socialized adults, we are likely to have a clear sense of the conditions

of Evil, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS (Ferdinand Schoeman, ed., 1987), at 257 (claiming that on Strawson's view the reactive attitudes "are constitutive of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions").

8. Note that reactive emotion accounts of blameworthiness come in two varieties: those that posit a constitutive connection between blameworthiness and aptly felt reactive emotions, and those that posit merely an extensional equivalence between the two notions. The former take it that blameworthiness just is, or consists in, being the target of appropriately felt reactive emotion. By contrast, the latter version takes it that a person is blameworthy to a certain degree *if and only if* it would be appropriate to feel a corresponding amount of reactive emotion toward that person. For more on this distinction, *see* Zimmerman, *supra* note 6, at 111, suggesting that the extensional equivalence view is preferable. By contrast, GIBBARD, *supra* note 4, at 45; WALLACE, *supra* note 5, at 63; and Graham, *supra* note 6, at 3, seem to endorse the constitutive version.

We should emphasize that the criticisms that we develop in this paper threaten both the constitutive and the extensional equivalence versions of the view. We will argue, roughly, that the posited extensional equivalence either does not hold (because the reactive emotions are time-sensitive in a way that blameworthiness is not), or else it provides no illumination of the concept of blameworthiness. Throughout the paper, we will be formulating various versions of the reactive emotion account using the "if and only if" locution, thus suggesting the extensional equivalence type of view. But if we manage to undermine this version of the reactive emotion account, it would of course also undermine the constitutive version of the view. After all, the claim that there is a constitutive connection between blameworthiness and fitting reactive emotion entails an extensional equivalence between the two as well. So by attacking the extensional equivalence claim, we are ipso facto attacking the constitutive claim as well.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

under which our blaming practices permit responses like resentment and indignation, and so it might seem a promising strategy to appeal to them in order to understand the arguably more obscure theoretical notion of blameworthiness.

Second, one might hold that "blameworthy" is like other normative predicates—for example, "valuable" or "desirable"—in being apt for a so-called "buck-passing" analysis.[9] According to buck-passing accounts, to say that something is valuable is just to say that placing a certain kind of value on it is appropriate. Similarly, to say that something is desirable is just to say that desiring it is correct (apt or fit). On these accounts, in saying that something is valuable or desirable, one is saying only that valuing and desiring is an apt, fitting, or more generally correct response to have toward it. It is, in effect, a way of "passing the buck" to the underlying reasons that exist for valuing or desiring while not itself a reason for doing either.

Reactive attitude accounts treat blameworthiness analogously. As we saw, they hold that saying that a person is blameworthy just is a way of expressing the judgment that there are reasons for blaming him—that is, the judgment that blaming him is correct or warranted by reasons. Accordingly, for reasons of theoretical unity, those who are drawn to a buck-passing analysis of value and desirableness might be inclined to adopt the analogous view of blameworthiness, which is no more (and no less) than what the reactive attitude account amounts to. (Of course, one does not have to believe that desirableness and value are apt for a buck-passer account in order to believe that blameworthiness is.)

We now have in place a characterization of the reactive attitude accounts of blameworthiness, a rough idea of some of the important ways in which they differ from the standard view, and some of the reasons why those who pursue reactive attitude accounts are inclined to do so. Interestingly, whereas such accounts are much in vogue among philosophers, they have gone largely unnoticed by legal theorists—even those who are primarily interested in the moral foundations of the institutions of criminal punishment. This is unusual given the strong interest among theorists of the criminal law more generally in exploring the implications of philosophical disputes for criminal law theory—as evidenced by legal theorists' interest in what philosophers have had to say about concepts such as intention, control, voluntariness, and the very idea of action, to name but a few. Reactive attitude accounts are naturally of concern to the theory of criminal law in that they promise to illuminate the concept of the blameworthy that lies at the heart of retributivist theories of punishment. Indeed, part of the attraction of such views to those who hold them is that punishment is an

9. *See, e.g.*, T.M. SCANLON, WHAT WE OWE TO EACH OTHER (1998), at 97; and Wladimir Rabinowicz & Toni Rønnow-Rasmussen, *Strike of the Demon*, 114 ETHICS 391–423 (2004).

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

institutional expression of aptly held blaming attitudes—attitudes that aptly target the blameworthy and not others.[10]

Though there is much to recommend reactive attitude accounts of blameworthiness, we remain unpersuaded that blameworthiness is a matter of correctly held or felt reactive emotions.[11] Several objections to reactive attitude accounts are familiar and the focus of considerable discussion among philosophers. Our focus, however, will be on a distinctive objection that has gone virtually unnoticed, having to do with the ways in which the *passage of time* affects both the moral significance of what we have done and the correctness of our responses to those doings.

We suggest that the passage of time drives a wedge between blameworthiness and the reactive attitudes in a way that both calls into question the viability of the reactive emotion account and opens up new avenues in legal theory generally. In particular, the considerations we draw attention to provide a novel way of thinking about the appropriateness of statutes of limitations, that is, constraints on the period of time in which the state can legitimately prosecute wrongdoers and private parties may press claims against those who have wronged them.

The conventional view is that these time-based limitations on pressing legal claims are justified on epistemic or instrumental grounds. A common thought is that because the passage of time can lead physical proof to decay,

10. It is natural to suppose that those who endorse a reactive attitude account of blameworthiness are committed to a retributivist justification of punishment. After all, on their view, a person is blameworthy if and only if the reactive attitudes are appropriately held or experienced, and punishment is a way of giving expression to the reactive attitudes and is justified because it does and to the extent it does. Thus punishment would be called for if and only if the agent is blameworthy. This, of course, is the essential claim of what we ordinarily think of as strong retributivism.

While one who adopts a reactive attitude account of blameworthiness could be inclined to a retributivist justification of punishment, the fact is that there is nothing in reactive attitude accounts that calls for a retributivist account of punishment. Indeed, there is nothing in the reactive attitude accounts that implies that punishment is justified at all. On the one hand, just because it might be appropriate to bear the reactive attitudes toward someone in a particular case, it does not follow that it is wise or justified, all things considered, to give expression to those attitudes in the form of criminal punishment. After all, many of the wrongs (e.g., breaking promises or being an unfaithful spouse) that warrant the reactive attitudes do not call for the full panoply of state-sponsored criminal sanctions. Thus someone drawn to a reactive attitude account need not endorse the retributivist view that being a proper target of the reactive attitudes (and so being blameworthy) is a sufficient condition (ceteris paribus) for criminal punishment. Of course, one could hold that punishment is justified only if it is imposed on those who are blameworthy for their mischief precisely because punishment is a way of giving expression to aptly held reactive attitudes. Such a person might hold that because punishment expresses indignation, imposing it when the reactive attitudes are not warranted is unfair because it is undeserved or unfitting. But one need not adopt a reactive attitude account of blameworthiness in order to hold that it is unfair to punish those who are not blameworthy for their acts or omissions because doing so would be undeserved. Finally, one can hold a reactive attitude account of blameworthiness and adopt an entirely consequentialist justification of punishment, or maintain that punishment involves the imposition of evils that can never be justified—and so on.

11. This means that we are also skeptical of the adequacy of the buck-passer account of blameworthiness.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

witnesses to become unavailable and evidence generally to become less reliable, it can become inappropriate to prosecute certain criminal offenses or to press certain civil claims after enough time has elapsed.[12] Moreover, the practical reality is that states do not have unlimited resources to devote to punishing wrongs or settling disputes from long ago. To be sure, these epistemic and pragmatic considerations help explain why there are limits on the state's power to press criminal charges and on its willingness to permit others to press civil claims. Considerations of these sorts do not suggest that it would be wrong or inappropriate to press such claims; only that it would be too difficult or wasteful to do so. In what follows, however, we argue that the passage of time can diminish the appropriateness of having or expressing the reactive attitudes (even if it does not similarly impact the blameworthiness of agents for what they have done). Thus there is an argument for statutes of limitations that grounds them in our blaming practices and goes far beyond the epistemic and instrumental considerations that are generally thought to underwrite such limitations.

We begin, however, by setting forth our basic objection to the reactive emotion account of blameworthiness: namely, that the appropriateness of the blaming emotions is sensitive to the passage of time in ways that the blameworthiness of agents for what they have done is not. If we are right, any effort to analyze the latter in terms of the former will face problems that have to this point gone largely unnoticed.

I.

If one is blameworthy now for performing some morally bad action, then one will generally still be just as blameworthy for having performed that action at any later point in one's life. Time, in other words, does not appear to erode one's culpability for what one has done or failed to do.[13]

Of course, it is a good question whether one can diminish one's blameworthiness for a particular action by apologizing for it, atoning, making things right, and so on. It is quite natural to think that I would be less blameworthy today for what I did several years ago if I have sincerely

12. For instance, the Supreme Court has explained that "[s]uch a [statute of] limitation is designed to protect individuals from having to defend themselves against charges when the basic facts may have become obscured by the passage of time." Toussie v. United States, 397 U.S. 112, 114–115 (1970). Among its rationales for time limitations, the Commentary to the MODEL PENAL CODE insists that "[f]oremost is the desirability of requiring that prosecutions be based upon reasonably fresh evidence. With the passage of time memory becomes less reliable, witnesses may die or become otherwise unavailable; physical evidence becomes more difficult to obtain, more difficult to identify and more likely to become contaminated." *See* MODEL PENAL CODE AND COMMENTARIES §1.06, "Time Limitations" (Official Draft and Revised Comments, 1985).

13. A cautionary note: what we are providing right now is just a first pass at the argument, which we go on to refine below. Most importantly, we are using "the passage of time" here as shorthand for "those events that typically come with the passage of time." More on this in Section II.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

apologized for it, undertaken actions to make it right and so on. However, others might object that apology and atonement do not diminish one's blameworthiness for particular actions but instead only reduce one's net level of blameworthiness (e.g., by balancing one's blameworthy act against another act that improves one's overall moral condition).

We need not settle the matter here. For whichever view one takes on the effect that apology, atonement, and redress have on one's blameworthiness for what one has done, it is clear that in the absence of taking any such steps, one would be as blameworthy today for what one has done wrong as one was when that misdeed was originally performed. Nothing would have changed in the meantime that diminishes one's culpability or blameworthiness. Thus the mere passage of time—the mere unfolding of one's life as time passes—does not alter the extent of one's blameworthiness for a particular action or omission.

Matters are different for the reactive attitudes, however. While it may be entirely appropriate *today*, say, to resent someone for a particular misdeed, it will not necessarily be appropriate to resent that person for that misdeed five, ten, or fifty years on; and even when it is still appropriate to feel some resentment or indignation, the extent or depth of the emotions is often appropriately diminished as time goes by—quite apart from the steps the wrongdoer has or has not taken to atone for his mischief.

Thus, if blameworthiness is not affected by the passage of time, while the appropriateness of the reactive emotions intuitively *is*, then considerations about the passage of time drive a wedge between blameworthiness and appropriately felt blaming attitudes, which threatens to undermine any attempt to analyze the former in terms of the latter.

These points are illustrated by the following simple example. Suppose that Megan, a ten-year-old girl, is playing with a ball in her front yard one day, when suddenly the ball rolls over into the flowerbed belonging to the neighbor, Mr. Anderson. No damage is done, but Mr. Anderson rushes over to the girl and lashes out at her, berating her excessively and inappropriately for what was only a harmless error. Megan leaves distraught and in tears. Anderson's action is uncalled-for and warrants the full panoply of the reactive sentiments. Megan would be justified in resenting him; observers would be justified in feeling indignation; and Anderson himself, upon even modest reflection, should be pained by suitable feelings of guilt, even remorse.

Thirty years on, however, (and probably long before then) we should expect matters to have changed substantially. Even supposing that Anderson never apologized, atoned, or sought to make matters right with Megan, it would nevertheless be unreasonable for Megan, now forty years old and (let us imagine) with a family of her own, to feel anything like the same resentment toward him that would have been appropriate when the wounds were still fresh. By this time it is reasonable to expect that Megan would have gotten over, or largely over, what in her youth was no doubt a painful and possibly even traumatic experience.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

The crucial point here is *not* that Megan is likely to be better off this way. Presumably she will be, but whether she is or not plays no role in our argument (for as we explain below, we are prepared to allow that considerations of prudence or personal health and well-being are not reasons that properly bear on the appropriateness of emotions like resentment).[14] Rather, our central claim is that it would be *inappropriate* for Megan to harbor the same resentments she did thirty years ago, because there is something perverse, unreasonable, or confused about dwelling on the past in this way. One would somehow fail to understand the place of the reactive attitudes in our lives were one to dwell on certain forms of wrongdoing in this way (though not all wrongdoing, of course).[15] After all, this kind of childhood experience is something we generally think one should be able to move beyond and not remain hung up on indefinitely—however unpleasant, even traumatic, it might have been at the time. As memories fade and life proceeds, it is appropriate—at least for many sorts of wrongs—for resentment to diminish as well.[16]

Emotional responses, therefore, seem *properly* to diminish in intensity over time in many different kinds of cases. It is natural and *correct* that they do so. Resentment of a certain intensity, *i*, felt by Megan toward Mr. Anderson right after he mistreated her seems to be inherently called for in a way that an episode of resentment of the same intensity felt by her toward him thirty years after the fact would *not* be.

To deny that time can make a difference to the appropriateness of episodes of reactive emotion in this way would be to insist on a view of what it is for these emotions to be apt that bears little resemblance to the way in which blame figures in our moral and social lives. We do not generally regard those whose levels of resentment or indignation toward wrongdoers

14. *See* STEPHEN DARWALL, THE SECOND PERSON STANDPOINT: MORALITY, RESPECT, AND ACCOUNTABILITY (2006), at 15, for helpful discussion of such "reasons of the wrong kind." And though it may well be unfair for Megan to continue to resent him after thirty years just as strongly as she did directly after he mistreated her (indeed, it is our view that it is unfair to hold wrongdoers indefinitely liable to moral or other sanction for all but the most serious wrongs), our argument does not depend on this claim of unfairness—which we are prepared, for the sake of the argument, to follow D'Arms and Jacobson in identifying as a reason of the wrong sort. (*Cf.* Justin D'Arms and Daniel Jacobson, *The Moralistic Fallacy: On the "Appropriateness" of Emotions*, 61 PHIL. & PHENOMENOLOGICAL RES. 65–90.)

15. Just as there is no statute of limitations in the criminal law for some serious crimes, such as murder, perhaps the appropriateness of feeling reactive emotions like resentment or indignation in response to very serious kinds of wrongdoing would not diminish as time passes. This would not undermine our basic point, however, which is that the level of reactive emotion that is appropriately felt *sometimes* diminishes because of the passage of time.

16. At this point, some might object that in this regard resentment and indignation come apart. Even were it inappropriate for Megan to resent Mr. Anderson to the same extent thirty years on, as fellow members of the moral community, we would be entitled to be as indignant and to exhibit as much indignation toward Mr. Anderson thirty years after he mistreated Megan as we would on the day he committed his transgression. The suggestion is unconvincing, however. It would be as unfitting and misguided for members of the moral community to exhibit or express the same intensity of indignation thirty years on as it would be for Megan. It is appropriate for indignation to dissipate with time.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

diminish as time passes as being guilty of any conceptual confusion about the relationship between blameworthiness and resentment and indignation. Nor are we likely to conclude that such people are overly lax in their moral responses. That is, we do not regard people who cease to resent or feel indignant toward wrongdoers as time passes as letting these wrongdoers off the hook too easily—as if they were mistakenly *foregoing* some additional amount of resentment or indignation that they would have been permitted to feel, but which they are passing up merely for practical or extraneous reasons. Thus our experience seems to make room for the possibility that the blame emotions can become less appropriate as time goes on—and not because of reasons of prudence or fairness, or for other intuitively collateral considerations.

The level of reactive emotion it is appropriate to feel in response to a particular action can diminish merely because of the passage of time, whereas the degree to which wrongdoers are *blameworthy* for a misdeed cannot diminish (at least not in the absence of the blameworthy person taking steps to atone or otherwise make matters right). This difference is reflected in the idioms of common parlance. We are prone to say that "time heals all wounds" but say neither that "time renders a culpable person innocent" nor that "time rights wrongs."

II.

In emphasizing the way in which the passage of time impacts the appropriateness of the reactive attitudes, we should not be read as claiming that there is something magical about time as such. Rather, our point is that as time passes, it is inevitable that events will continue to unfold, and in some instances these events will bear on the appropriateness of the reactive emotions while not bearing on the agent's blameworthiness. For instance, it is not the passage of time in itself that explains the diminishing appropriateness of resentment and indignation toward Mr. Anderson for his mistreatment of Megan. Rather, what explains it are the things that happen as time passes: memories fade; Megan becomes an adult and grows up both mentally and emotionally; she moves away from the neighborhood; she takes on the responsibilities of parenthood herself; and so on.

If one's degree of culpability or blameworthiness for a particular action can be diminished at all, this could only be in virtue of certain distinctive events—e.g., atonement, apology as well as other ways of making things right, or a breach in personal identity ("I am literally not the same person now as before"). And while the events that represent the normal course of life as time marches on can bear on appropriately felt *reactive emotion*, they do not have a similar bearing on blameworthiness.

So the more precise way to make our point is this. Rather than saying that the blaming emotions are "sensitive to the passage of time" and that

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

blameworthiness is not, we might say that the kinds of events (if any) that can occur over time that defeat or diminish blameworthiness for a particular action—atonement, rectification, apology, and such—are quite different from those that render the blaming attitudes inappropriate—like moving beyond the events of the past, memories fading, or constructing a narrative of one's life in which the event's significance is reduced, and so on.[17]

In other words, as time goes by, events unfold in the lives of those who are the victims of culpable mischief, the history of the moral community as a whole, and indeed in the lives of wrongdoers themselves that can impact the appropriateness of the feelings of resentment, indignation, and guilt (respectively) that each might have. The claim is not that this is always so; only that the emotions are such that they are appropriately responsive to these features of the passage of time.

The same does not appear to be true of a wrongdoer's blameworthiness for what he or she has done or failed to do. Life unfolding does not in the normal course of events create reasons that diminish or eliminate blameworthiness for a given act or omission. Reductions in blameworthiness for an action (if indeed this is possible at all) are the result of distinctive kinds of events: making repair, atonement, apology, or changing one's identity altogether. Thus blameworthiness and the blaming attitudes are differentially responsive to the passage of time in ways that pose a serious objection to reactive attitude accounts.

This, then, is our basic objection.[18] Is it one that can be met? That is the question we are concerned to investigate in the remainder of this paper.

III.

According to the simplest versions of the reactive emotion account, correctness consists in appropriateness, broadly construed. For an emotion to be appropriate in this broad sense is just for one to have sufficient reason (of one kind or another) to feel or otherwise experience it. The trouble, as a number of philosophers have pointed out, is that not all of the

17. In the main, the kinds of events that arguably can bear on blameworthiness are *wrongdoer centered*: atonement, rectification, and apology, for example; whereas those that bear on the appropriateness of the reactive attitudes are by and large *victim centered*: for example, faded memories, moving on, and the like.

18. Before proceeding, we should stave off a potential source of confusion. Distinguish two claims we might be thought to be making with the objection from time: (1) At t2, P-at-t1 is just as blameworthy for act A as he was at t1; and (2) At t2, P-at-t1 is just as blameworthy for act A as P-at-t2 is. Now, REA clearly has no trouble accommodating the first of these claims and is in tension only with the second. Since we are trying to raise an objection to REA, we do not have much interest in (1)—rather, our interest is in (2). We think that in some cases, (2) is intuitively quite plausible. Indeed, the case of Megan was in part meant to elicit precisely this intuition. That is, we took it to show the plausibility of the claim that in the absence of distinctive blameworthiness-altering events (such as atonement, a loss of the ability to reason, a breach in personal identity), later time slices of a person are just as blameworthy as the time slice that performed the earlier blameworthy act in question.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

considerations that at first might seem to bear on whether to feel one of the blaming emotions are reasons of the "right sort." After all, some considerations might seem to count in favor of or against feeling a reactive emotion in a particular case, even though they do not seem relevant to whether the target of those emotions is blameworthy for what she has done or failed to do. They are instead what Stephen Darwall calls "reasons of the wrong kind."[19]

Take a straightforward example. Suppose your sibling has wronged you in a manner that renders her both blameworthy and a proper target of resentment. It is not hard to imagine circumstances in which displaying or even experiencing resentment toward her would create tension in your relationship, in some cases resulting in complete fracture. Intuitively, the threat of fracture, however real and unfortunate, does not constitute the right sort of reason for determining the appropriateness of experiencing anger or resentment toward her for what she has done. After all, it has no obvious bearing on her degree of blameworthiness. This suggests that it is not enough that we have good reasons for not having or experiencing the reactive attitudes. We need good reasons of "the right sort."

Some considerations are clearly reasons of the wrong sort—those of health, prudence, and etiquette, for example—but not just these. To this list Justin D'Arms and Daniel Jacobson have added a range of moral considerations—including those that bear on one's own or others' well-being, and perhaps fairness as well.[20] Thus there is a wide range of reasons for feeling or withholding the reactive emotions that intuitively have little to do with the blameworthiness of agents for what they have done or failed to do. If such reasons of the wrong kind are allowed to bear on the appropriateness of the reactive emotions, then a reactive emotion account of blameworthiness is doomed from the outset. For allowing them to count would give rise to cases in which the reactive emotions are appropriately felt

19. *See* DARWALL, *supra* note 14, at 15, where he distinguishes the "desirability" of holding someone responsible from the reasons that in themselves warrant the reactive attitudes.

20. Justin D'Arms & Daniel Jacobson, *The Moralistic Fallacy: On the "Appropriateness" of Emotions*, 61 PHIL. & PHENOMENOLOGICAL RES. 65–90 (2000), in discussing what they call "the moralistic fallacy," offer some considerations that might be taken to suggest that fairness is a reason of the wrong kind. It should be noted, however, that their discussion is aimed more at humor than at the blame emotions. And one might think that while considerations of prudence, health, and etiquette are uncontroversially reasons of the wrong sort, it is less clear that the same holds for all considerations of fairness. After all, it is not clear that having wronged another properly renders one indefinitely liable to the victim's wrath or resentment without regard to how it would be fair to treat one. However, in what follows, we take on board D'Arms and Jacobson's suggestive comments and work under the assumption that at least some considerations of fairness fall on the "wrong reasons" side of the ledger already occupied with considerations of health, etiquette, and prudence. For instance, the observation that a wrongdoer has already suffered "more than his fair share" might be a reason to react to him with sympathy, but it is arguably a reason of the wrong kind for blaming him less for his misdeeds. The fact that the wrongdoer suffered unfortunate experiences in entirely unrelated contexts typically does not seem to make his culpable actions any less bad or the reactive emotions any less warranted. Thus some fairness considerations, at least, seem to be properly counted as reasons of the wrong kind.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

toward someone who is not blameworthy, or cases in which these emotions are not appropriately felt toward someone who *is*.

It is plain, therefore, that the reactive emotion account must be reformulated so as to take account of the distinction between reasons of the right kind and reasons of the wrong kind. In particular, the account should be put roughly as follows:

> **Reactive Emotion Account** (**REA**)**:** A person P is blameworthy to degree d for performing action a if and only if feeling the relevant reactive emotions to degree d towards P in response to his doing a is supported by the balance of the reasons *of the right kind*.

Does the distinction between reasons of the right and wrong kinds help to answer the objection from the passage of time? That will depend, of course, on whether one can make out the distinction in a principled way when it comes to the relevant emotions. It is one thing, after all, to *assert* that the passage of time is the wrong kind of reason for the blaming sentiments and quite another to explain *why* that should be so. Unless some account of the distinction between reasons of the right and wrong kinds can be given that explains why time—just like prudence, health, etiquette, and so on—deserves to be included among the "wrong reasons," the present response would appear to be plainly ad hoc.

Pause for a moment to take a look at the laundry list of reasons of the wrong kind compiled so far: prudence, self-interest, etiquette, welfare, morality, fairness, and now considerations that arise in virtue of the passage of time. If a reactive attitude account of blameworthiness is to be sustained, all of these considerations must turn out to fall on the "wrong reasons" side of the ledger. But what, beyond the desire to save a reactive emotion account of blameworthiness, could explain why they are to count as the wrong kind of reasons for the blaming attitudes? Perhaps one does not need a theory in order to explain why a few isolated considerations here and there do not bear on the appropriateness of the reactive attitudes. But once the list of reasons that are ruled out as "the wrong kind" grows as long and diverse as it now has, the plausibility of REA hangs on offering an explanation of why some considerations bear on the appropriateness of the reactive attitudes and others do not.

Given the difficulty of spelling out the distinction between considerations of the right and wrong sort, we would do well first to explore other approaches to meeting the objection from the passage of time. For if the objection can be met without having to characterize what is no doubt a controversial distinction between reasons of the right and wrong sort, then so much the better. As it happens there are several options available, some more promising than others. None, however, are ultimately persuasive—or so we argue.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

IV.

Faced with the objection that blameworthiness and appropriately felt reactive emotion can come apart because of considerations involving the passage of time, one response would be simply to bite the bullet. That is, the defender of a reactive emotion account could argue that it is simply a consequence of his position that blameworthiness is also sensitive to the passage of time in the same way that resentment, indignation, and guilt are.[21]

Without more, this response simply begs the question. After all, if one simply insists that it is a consequence of one's position that blameworthiness must be sensitive to the passage of time in the same way that the blaming attitudes are, then it is one's theory that drives the attitude one is taking toward one's intuitions, and not the other way around. To avoid begging the question, one needs some additional consideration that would justify biting the bullet. As it happens, we can think of at least one such consideration.

Recall that the reactive emotion account of blameworthiness is an instance of the buck-passer account of normative predicates more generally. One could argue that we must rethink our ordinary intuitions about the kinds of considerations that bear on blameworthiness in order to preserve the theoretical advantages of treating blameworthiness in the same way that we treat other normative concepts such as value and desirability. That is, maintaining a certain kind of theoretical unity in the treatment of normative predicates might speak in favor of absorbing the costs involved in abandoning our intuitions about the factors that can bear on ascriptions of blameworthiness for wrongdoing.

In principle, theoretical unity is the kind of consideration that could warrant biting the bullet about whether one's blameworthiness is responsive to the passage of time. Here, however, theoretical unity hangs on the aptness of blameworthiness for a buck-passer account. It is important to recall that a buck-passer account of blameworthiness just is what a reactive attitude account amounts to. And as we saw in the previous section, reactive attitude accounts of blameworthiness are defensible only if the distinction between right and wrong kinds of reasons can be drawn in a principled way.

This does not make buck-passer accounts of blameworthiness unique or otherwise special. A buck-passer account of any normative concept—not just blameworthiness but also value, desirability, and so on—will need to rule out some considerations (e.g., prudence, health, etiquette, and fairness) as reasons of the wrong kind in order to avoid a range of intuitive counterexamples.[22] The difference is that it may be considerably easier to draw that distinction in the case of the desirable and the valuable than it is in the case of the blameworthy. After all, it is not clear that time impacts valuing

21. Thanks to Stephen Darwall for pressing us to consider this line of response.
22. For more, *see* Rabinowicz & Rønnow-Rasmussen, *supra* note 9.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

and desiring in anything like the way it does for resentment, indignation and guilt.

The theoretical unity we would secure by treating blameworthiness in the same way as other evaluative concepts obtains only if the distinction between reasons of the right and wrong sort can be made out in a principled way. Whether this is possible still remains to be seen. Thus, on closer inspection, the strategy of biting the bullet turns out not to be much of an alternative solution to the objection from the passage of time. Instead, it is just another way of raising the concern that any adequate response will require making out the controversial distinction between considerations of the right and wrong sort.

V.

A different sort of answer to the objection from time is this. As formulated in the literature, reactive emotion accounts typically make no explicit mention of time, and so in expressing our concerns about them, we have been satisfied to treat them as they are invariably characterized. One might think, however, that the best way to meet our objection that blameworthiness and blame come apart due to the passage of time is to formulate the reactive emotion accounts in ways that introduce time indexing in a particular way. Specifically, one might think that the objection can be defused if one takes it that a wrongdoer's level of blameworthiness is to be tied to the level of reactive emotion that it is appropriate to feel toward him *at the time of his wrongdoing*.[23] More precisely formulated, the revised view can be stated as follows:

> **REA_t:** P is blameworthy to degree d for an action a he performed at time t iff it would be appropriate at t to feel reactive emotion to degree d toward P in response to his performing a.[24]

23. Thanks to John Martin Fischer for suggesting a response along these lines to us in personal correspondence.

24. This formulation of the view contains a slight technical difficulty that we gloss over in the text for the sake of brevity. In particular, those who think apology and atonement can alter blameworthiness (a view we officially neither endorse nor reject) might object to REA_t in the following way. If I commit a wrong on Monday but then apologize and fully make amends on Tuesday, then intuitively I am no longer blameworthy for the action. However, REA_t would seem to entail that I am just as blameworthy on Tuesday as I was on Monday. For on Tuesday, it is still true that reactive emotion was appropriately felt toward me at the time the action was committed, namely, on Monday. Nonetheless, we think REA_t can probably be modified to handle this problem. In particular, we can take it that P's level of blameworthiness *at a time,* t_1, for a bit of wrongdoing, done at an earlier time, t_0, is equivalent to the level of reactive emotion that would be appropriately felt toward P at t_0—*provided no blameworthiness-altering events (like atonement) have occurred between* t_0 *and* t_1. And in the case where some blameworthiness-altering events *do* occur between t_0 and t_1, then P's blameworthiness would simply have to be recalculated at the time of this new action or event. Formulating the view with this additional complication, we think, would allow the technical problem noted above to be avoided. Strictly speaking, that is, the view is best formulated with *two* time indexes—one for the time at which

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

One might think that adopting REA_t would allow one to accommodate the case of Megan and avoid the problems it brings up. Recall that it would have been entirely appropriate for Megan to resent Mr. Anderson a large amount at the time he harshly and unfairly berated her for upsetting his flowers. But the argument from time insisted that after many years had elapsed, it would be inappropriate for Megan to resent Anderson as much as she was entitled to do on the day he wronged her. However, the defender of REA_t can accept all this, it seems. After all, what REA_t entails is that Mr. Anderson's level of blameworthiness is fixed by the level of reactive emotion it would be appropriate to feel toward him *at the time* of his wrongdoing. And once this level is fixed, there is nothing in REA_t to suggest that Anderson's blameworthiness for his mistreatment of Megan would change, even if the level of reactive emotion that would be aptly felt toward him were to subsequently diminish. Thus, on REA_t, blameworthiness is tethered once and for all to the amount of reactive emotion that the wrongdoer is appropriately subject to at the time of wrongdoing. So it is no problem for the view if blameworthiness and appropriately felt reactive emotion then proceed to part ways as time passes.

As initially promising as the time-indexing response might seem, it faces two distinct problems. First, difficulties seem to arise when considering acts of wrongdoing that do not take place at any identifiable instant but rather extend over a prolonged time period. In particular, we encounter the problem when none of the actions that together constitute culpable wrongdoing would be objectionable if taken on their own. For example, suppose there is an agent—call him Bernie—who engages in an extensive plot to defraud investors. Suppose he performs a range of facially innocuous actions over the course of many years, each of which would be morally permissible considered in isolation. (He slightly downplays cash-flow problems, gets management to release a cookie-jar reserve at the ideal moment, makes use of complex derivative instruments to make risk less apparent to the investor, and so on—all within seemingly normal limits.) However, suppose that the net effect of Bernie's actions is to mislead investors and induce them to part with their money in ways they would not have consented to had they fully understood the investment opportunities offered to them. Suppose further that the investors discover the fraud only after Bernie's plans have already come to a close.

It seems plausible in this case to regard Bernie as extremely blameworthy for his wrongdoing. However, it is not clear how REA_t is able to capture this result. The issue is not that in a case of an extended wrong we cannot make sense of the claim that blameworthiness is to be settled *at the time of the wrongdoing*. For REA_t can just take the relevant time to be the interval during which the extended wrong occurred. Thus blameworthiness during

the agent's blameworthiness is to be assessed (t_1), and one for the wrongdoing in question (t_0).

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

an interval would equal the sum of (or really the integral over) all the momentary levels of appropriate reactive emotion during that interval.

Rather, the problem arises because Bernie's fraud consisted of a string of individual component actions, none of which was morally problematic considered in isolation. Accordingly, the level of reactive emotion it would have been be appropriate to feel toward Bernie at the time of any of these component actions was zero or something close to it. Thus it is not clear how REA_t can generate the result that Bernie is *highly* blameworthy for the entire string of the individually innocuous actions that taken together constitute the fraud. Because wrongdoing can be extended across time and emerge from a string of individually unproblematic actions, it seems mistaken to take one's blameworthiness for some wrongful act to be equivalent to the level of reactive emotion that is appropriate at the time of that wrongdoing, as does REA_t.[25]

The second problem arises because the harm caused by a particular act of wrongdoing might itself materialize only after a great deal of time has elapsed since the initial wrong. To illustrate, suppose you work in a small pharmaceuticals company and on the first day of January 2010, you did your job negligently, with the result that you created a 10 percent risk that a particular person would contract a painful disease ten years later. Suppose that your possible victim turns out to be unlucky and does in fact end up contracting the disease in 2020, thanks to the risk you exposed her to a decade earlier. When the harm actually befalls your victim, it seems she would then be warranted in resenting you a great deal. By contrast, if she were told of your negligence at the time it occurred, in 2010, your then-merely-potential victim would seem warranted in resenting you somewhat—but presumably much less than she would in 2020, when the risk materializes. After all, in 2010, resentment is called for only by the *risk* that you exposed your potential victim to, while in 2020 resentment would also be called for by the actual pain and suffering she endures from the disease. In general, it seems plausible that directly causing a harm of some magnitude, *m*, merits a greater degree of resentment than merely exposing someone to a *risk* of experiencing a harm of that same magnitude *m*. So in the instant case, your victim would seem to be entitled to resent you more in 2020 than in 2010.[26]

25. This objection threatens "additive" versions of the view, according to which blameworthiness during an interval is the sum of all the momentary levels of appropriate reactive emotion for each time during that interval. Because wrongdoing sometimes seems to emerge from a series of individually innocuous actions, we doubt that this additive version of the view can be sustained. *Cf.* Velleman's objection to "additive" theories of well-being in David J. Velleman, *Well-Being and Time*, 72 PAC. PHIL. Q. 48–77 (1991).

26. Of course, this presupposes the existence of a robust kind of moral luck. However, if our argument in this paper is correct, then we may have available a novel way to account for our conflicted intuitions about moral luck. In particular, if blameworthiness and the reactive emotions can come apart, then we might say that blameworthiness is not sensitive to luck, whereas the appropriateness of the reactive emotions is.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

What, then, is the level of resentment that your *blameworthiness* for this negligent act should be set as equal to—the heightened level that would be appropriate in 2020 or the lower level that would be appropriate in 2010? REA_t entails that it would have to be the former. After all, REA_t takes your blameworthiness for an action to correspond to the level of reactive emotion that it would be appropriate to feel toward you *at the time of that action*. However, this invites the worry that REA_t *undervalues* the amount of blameworthiness that accrues in virtue of your negligent behavior. After all, your blameworthiness would be higher if pegged to the level of resentment warranted in 2020 instead of the level that is warranted in 2010. And one wonders what principled reason there could be for not pegging your blameworthiness to the reactive emotion that is warranted at this later time. It thus seems *arbitrary* to privilege the time of your wrongdoing over the time of the resultant harm as the point at which we must look at how much reactive emotion is appropriate in order to determine your level of blameworthiness.

Arguably, one could say that the blame is appropriate to actions under particular descriptions. In that case, the action can be described in ways that do not include its causal upshots or in ways that do; and the level of blame appropriate would depend on the description in question. Setting aside a range of problems that are familiar in action theory for such an approach, it is enough to note here that any such strategy runs the risk of begging the question. After all, one cannot choose which description is apt by looking at its effects on the reactive attitude account, since the very plausibility of that account hangs in the balance. Someone pursuing this line of argument needs to formulate independent criteria of aptness for descriptions of events and then see whether this would yield results about the present case that would save the reactive attitude account.

VI.

Perhaps these difficulties can be avoided if one moves to an ideal observer version of the reactive emotion account instead. The reactive emotion account can be reformulated as follows:

> **REA_i:** P is blameworthy to some degree *d* for some action *a* iff it would be appropriate for an ideal observer—one who is fully informed about the circumstances of *a* and its consequences—to feel reactive emotion to degree *d* toward P in response to *a*.

In the above case, a fully informed observer would know, even at the time of your negligent action (i.e., in 2010), that the end result is that your victim will in fact become ill. Accordingly, the level of reactive emotion it would be appropriate for such an ideal observer to feel at the time of the wrongdoing (i.e., in 2010) would be the same as the level it would be appropriate for

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

her to feel at the time of the resultant harm (i.e., in 2020). Thus REA_i does not face the worry about undervaluing your blameworthiness or the charge of arbitrariness that afflicted REA_t.

Nonetheless, REA_i faces other problems. For one thing, it is unclear what the value is of determining what it would be appropriate for an ideal observer to feel, because no actual person would ever be relevantly like the ideal person. Actual people labor under epistemic limitations about the consequences of their own and others' actions, and this fact of life is deeply embedded in the emotional responses of actual people. But one of the motivations of the reactive emotion account is that it promised to explain a relatively mysterious normative concept in terms of something we have a firmer grasp on—namely, a set of attitudes that we are all familiar with from our own emotional lives. But once we revise the reactive emotion account by introducing the notion of a fully informed idealized agent and appeal to *her* emotional reactions to determine a person's blameworthiness, then this central motivation for the account is lost. After all, the emotional reactions of a fictitious idealized agent are at least as obscure as the underlying normative concept to be analyzed.

Perhaps even more worrisome, REA_i seems to face the underlying objection based on the passage of time with which this article began. Insofar as we are able to get a firm enough grasp on the notion of a fully informed observer to predict what his emotional responses would be, it seems to us that any resentment or indignation that such an idealized person might feel in response to wrongdoing would be sensitive to the passage of time in just the same way as the emotional responses of a real person. Just as it seems appropriate for Megan's resentment toward Mr. Anderson to have diminished somewhat thirty years after her being wronged, so, too, do we suspect that the ideal, fully informed observer would be similarly inclined to bear less reactive emotion toward Mr. Anderson thirty years on. By contrast, as we argue above, Mr. Anderson's blameworthiness for this action would not diminish merely in virtue of the passage of time. Thus we think it is possible for blameworthiness and the level of reactive emotion that an idealized, fully informed observer would feel to come apart as time passes. As a result, REA_i seems to suffer from the initial problem concerning the passage of time that we have been seeking an answer to all along.

Someone drawn to the ideal observer formulation might seek to insulate it from this objection in either of two ways. First, one might hold that the level and intensity of the reactive attitudes that the ideal observer experiences toward wrongdoing in fact would not change with time.[27] After all, while Megan's resentment might understandably diminish with time because her life conditions change and she acquires other more pressing concerns, one might insist that the ideal observer does not change in any such way.

27. Thanks to an anonymous reviewer for pressing us on this point.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

The trouble with this response, however, is that insisting on this threatens to render the ideal observer view too abstract to preserve the purported benefits of the reactive emotion account. In fact, we do not seem to remain equally concerned with every experienced wrong until the end of our days. While some traumatic events might stick with us, a common feature of human psychology is that we become less concerned with many sorts of events as they recede into the past. To abstract away from this feature of human psychology threatens to transform the ideal observer into a demigod whose emotional reactions would be dramatically different from our own and therefore difficult to predict. This would be to deny the reactive emotion account one of its main attractions—namely, the ability to explain blameworthiness in terms of attitudes and practices that we are all intimately acquainted with.

Thus another way in which we might steer clear of these problems is by combining the time-indexing strategy with the ideal observer version of the reactive emotion account.[28] This would yield a version of the view according to which P is blameworthy to degree *d* for an action *a*, performed at time *t*, if and only if it would be appropriate for a fully informed observer, at *t*, to feel reactive emotion to degree *d* toward P in response to *a*. Call this view REA_{it}.

There are several problems with this alternative. First, if the ideal observer's emotional response to a given action can change as time passes (as we suggested above must be supposed to avoid severing the tie to our actual blaming practices), then it seems ad hoc to insist that the ideal observer's reactions at the time of wrongdoing are what determine the agent's blameworthiness, rather than any later reactions the observer might have. Why shouldn't blameworthiness be pegged to the ideal observer's later emotional responses? Is the idea just to preserve the theory by insulating the reactive emotions at issue from any possible relevance that time might have? This would be to beg the question. On the other hand, if one wants to avoid this problem by claiming that the ideal observer's reactions *do not* change over time, then her responses again become divorced from the realities of our actual emotional lives.

What is more, arbitrariness threatens the time-indexed ideal observer view in cases involving genuinely indeterministic processes (just as risk initially made problems for REA_t). Consider your favorite genuinely indeterministic process—a quantum system, say, or a roulette wheel. Suppose it is involves two possible outcomes, A or B. One or the other will occur at t_1, but before that point it is undetermined which one it will be. All it is possible to know in advance is that there is a 10 percent chance of A at t_1 and a 90 percent chance of B at t_1. Suppose further that at some earlier time, t_0, you hook me up to this indeterministic system such that I will be harmed if A occurs at t_1, but not if B occurs. Clearly you are blameworthy for exposing me to

28. Thanks to another anonymous reviewer for pointing out this response.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

this risk. But how blameworthy? Since at the time of your wrongdoing, t_0, it is undetermined whether A or B will later occur, any reactive emotion the ideal observer would feel toward you at t_0 would have to be based just on the risk of harm you exposed me to.

However, if A ends up resulting at t_1 and I get harmed, then a more intense reactive emotion would seem appropriate at t_1 (one that reflects the harm done) than the ideal observer would have felt at t_0 (one that reflects only the risk of harm being done). The time-indexed ideal observer view, REA_{it}, entails that in such a case your blameworthiness must be pegged to the *lesser* amount of reactive emotion the ideal observer would feel at t_0 rather than the *greater* amount she would feel at t_1 when she sees me get harmed. But why is this the right answer, one wonders? REA_{it} thus seems arbitrary.

Again, one might respond that to be truly ideal, the observer must be given not just all the facts that are *available* at the time of your wrongdoing (t_0) but all the facts *period*—even those pertaining to undetermined future events. This means granting her something approaching omniscience, however, and it is far from clear what to make of the emotional responses such a being would have. Would her emotional responses be responsive to all the downstream effects of our actions until the end of time, even those that creatures like us haven't the slightest chance of foreseeing? Thus the cost of going this route, again, is to divorce the ideal observer's responses from our actual blaming practices and strip the reactive emotion account of its ability to shed light on the notion of blameworthiness by grounding it in practices we are familiar with.

Thus the general problem we keep running up against for the ideal observer versions of the reactive emotion account is that they continually face pressure to resort to a mysterious concept in order to explain the notion of blameworthiness: namely, the emotional responses of a creature with mental capacities and affective dispositions that are drastically different from ours. In other words, the worry is that we can preserve the ideal observer strategy but only at the cost of robbing it of its explanatory value. Accordingly, we remain skeptical even of this temporally indexed version of the ideal observer reactive emotion account.[29]

The upshot, then, is that neither the time-indexing strategy nor the ideal observer version of the theory provides a satisfactory response to the basic temporal challenge that reactive emotion accounts of blameworthiness must grapple with. Thus we are left to turn elsewhere for a solution.

29. An additional problem, which seems to threaten all ideal observer views, is that it is not clear what the theory would make of cases in which "reasonable minds might differ." If the amount of reactive emotion a particular ideal observer would feel in response to a particular bit of wrongdoing differs from the amount that another ideal observer would feel, then we face the challenge of finding a principled basis for claiming that one or the other set of responses is appropriate. If we stipulate, however, that all ideal observers would react identically to each act of wrongdoing, then the view again seems to stray into the realm of unhelpful idealization.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

VII.

Recall that all forms of the reactive attitude accounts of blameworthiness rely on a notion of the emotions being correctly held, felt, or experienced. The burden of such accounts is to specify the standard of correctness for the reactive attitudes. Spelling out the notion of correctness in terms of "appropriateness" does not remove the problem to which we are calling the reader's attention. It provides little illumination to say that the attitudes are correctly held if held appropriately, and appropriately held if held for the right kinds of reasons. The ultimate upshot is that there is significant pressure to distinguish between reasons of the right and wrong sort in a principled (non-question-begging) way. And while we see the intuitive appeal of counting health, prudence, and well-being among the class of considerations of the wrong sort (since they are often not related to the moral significance of what we have done and thus to the appropriateness of responding to it in various ways), it is considerably more difficult to see why the passage of time should not bear on what responses are appropriate to what has been done.

We can put the point briskly. The standard of appropriateness calls for a distinction between considerations of the right and wrong sort but itself provides no guidance as to how that distinction is to be drawn. Perhaps an alternative way of characterizing the standard of correctness will offer more guidance.

Just such an alternative has been suggested by D'Arms and Jacobson, who in a series of papers have advanced the view that for emotions, correctness is a matter of what they call "fittingness."[30] Their notion of fitting emotions relies on an analogy with the propositional attitude of belief. Beliefs have objects—the propositions believed—and they represent their objects as being true. Thus beliefs fit their objects, or are fitting, insofar as their objects actually have the property they are represented as having, namely truth. In other words, a belief that p is fit to its object (or fitting) if and only if p.

Similarly, on this view, emotions have objects and represent their objects as having certain properties.[31] Just as beliefs represent their objects as *true*, emotional responses such as envy or fear represent their objects as having certain properties—namely, being enviable and being fearsome, respectively (or something of that sort). An emotion, then, is said to be *fitting*, or *fit to its object* if and only if its object actually has the properties that the emotion represents it as having. Envy, for example, would be a

30. D'Arms & Jacobson, *supra* note 20, at 72. *See also* Justin D'Arms & Daniel Jacobson, *Demystifying Sensibilities: Sentimental Values and the Instability of Affect*, *in* Oxford Handbook of Philosophy of Emotion 585–614 (Peter Goldie ed., 2009).

31. D'Arms and Jacobson claim, e.g., "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." D'Arms & Jacobson, *supra* note 20, at 72.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

fitting emotional response whenever its object is worthy of envy, and fear would be fitting whenever the object of the emotion in question is worthy of being feared (e.g., because it is dangerous). When it comes to the reactive emotions—namely, resentment, indignation, and guilt—the thought is that they are fitting responses only if their objects actually have a certain property, presumably blameworthiness.[32]

With this notion of fittingness in hand, we can now reformulate the reactive emotion account of blameworthiness as follows:

> **REA$_f$:** An agent is blameworthy to a particular degree for an action if and only if that degree of reactive emotion is fittingly felt toward that agent in response to that action.[33]

Substituting fittingness for appropriateness in formulations of reactive attitude accounts of blameworthiness is potentially a significant advance, both in understanding our emotional lives and in providing a defensible version of the reactive attitude accounts of blameworthiness. The great advantage of fittingness accounts is that they obviate the need to provide an independent account of the distinction between reasons of the right and wrong kinds for feeling the emotions.

While it is not obvious, on the appropriateness approach, why considerations involving the passage of time must be regarded as being of the wrong sort for resentment, indignation, and guilt, no similar problem need arise when the reactive emotion account is formulated in terms of fittingness. For on such an approach, emotions are taken to represent their objects as having certain properties and thus provide their own standards of correctness. This implies that if the blame emotions represent their objects as having a property that is not sensitive to the passage of time, then considerations involving the passage of time would have no bearing on whether these emotions are fittingly felt.

VIII.

For all its promise, the concept of an emotion being fit to its object raises problems for the reactive emotion account. We raise two main objections

32. D'Arms & Jacobson in *id.* do not take a firm stand on what property it is, precisely, that the blame emotions represent their objects as having. Some of their comments suggest that they think it might be blameworthiness, but the textual evidence for this is admittedly scarce. We have more to say just below about the question of what property or properties the blame emotions should be taken to represent their objects as having.

33. Note that this approach can indeed make sense of the fittingness of *degrees of an emotion.* As long as the emotion represents its object as having a property that comes in degrees, then it will be easy to see how feeling that emotion to one degree can be more fitting than feeling it to another degree. Feeling fear, for example, to degree d would be fitting iff the object of that emotion has the property of being dangerous to degree d as well (or something of the sort).

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

to the view—one concerning the passage of time, and the other involving a more general worry about the structure of this view.[34]

The problems begin to emerge when we look more closely at the question of what the property is that the reactive emotions are supposed to represent their objects as having. One preliminary concern is that the blame attitudes seem to differ importantly from one another. Resentment is fittingly experienced or expressed only by those with *standing* to resent: namely, victims of culpable wrongdoing. Guilt may be felt only toward oneself for one's own misconduct. And indignation may be felt by all members of the moral community against those who have failed to measure up to the norms of behavior required of them. These facts about the reactive attitudes suggest that they may differ with respect to the property that they represent their objects as having.

Nonetheless, the blaming emotions are obviously related in an important way in that they are felt in response to some kind of culpable misconduct. This suggests that despite their differences, there is indeed some *common* property that all three blame emotions represent their objects as having. And so, for the fittingness version of the reactive emotion account to be defensible, it likely will have to be cashed out in terms of whatever *common* property it is that the blame emotions represent their objects as having. More specifically, the account would be reformulated to say that a person is blameworthy if and only if she possesses the common property, whatever it is, that all three reactive emotions represent their objects as having (even if each also represents its object as having other properties besides).[35]

34. Another worry is this. Many philosophers are drawn to an argument articulated by Nishi Shah and David Velleman that there is an important connection between the views that belief is "transparent" and that belief comes with its own standard of correctness. To say that belief is transparent is roughly to say that the question of whether to believe that p gives way to the question of whether p, in the sense that one answers the former question by answering the latter. The best explanation for the transparency of belief, the argument goes, is that belief entails a standard of correctness: namely, the truth of the proposition believed. (For a fully worked out version of this argument, *see* Nishi Shah, *How Truth Governs Belief*, 112 PHIL. REV. 447–482 (2003).) If this is right, then only evidence for the truth of p would constitute a reason of the right kind for believing that p, because only such evidence is an indicator of truth.

However, it is not at all clear that the reactive attitudes are transparent in the way that belief is and thus it is questionable whether the analogous argument can be made for thinking that the reactive attitudes generate their own standards of correctness. After all, it is not obvious that reasons for (or against) the reactive attitudes just are reasons for (or against) blameworthiness of an agent for what he has done or failed to do. Confronted with all the evidence one needs to establish that an agent is blameworthy for what he has done or failed to do, it does not follow that one *must* experience or have the appropriate blaming attitude—on pain of the charge of otherwise being conceptually confused. (It is a good question to ask whether these remarks hold for each of the emotions considered individually. However this is not the place for addressing it.) Thus, while the transparency of belief makes it plausible to think that beliefs generate their own standard of correctness in the way that the fittingness approach supposes, it is not obvious that the reactive attitudes are transparent in a way that would support the analogous inference about them.

35. An alternative solution might be to select one of the reactive emotions—resentment, say—and formulate the reactive emotion account of blameworthiness only in terms of *it*.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

But now we encounter the first major difficulty with the proposal. Assuming there is some common property that the reactive emotions represent their objects as having, why think that it is not sensitive to the passage of time? This, after all, is precisely what the fittingness version of the reactive emotion account has to suppose in order to avoid the objection. Recall that the original problem was that the reactive emotions seem to be sensitive to the passage of time in a way that blameworthiness is not. To avoid the ensuing problems, the fittingness version of the theory would have to take the reactive emotions to represent their objects as possessing a property that is not sensitive to the passage of time. For only in this way can it ensure that fitting reactive emotion does not part ways with blameworthiness as events unfold with time.

But besides one's desire to save the fittingness version of the theory, what reason could there be for supposing that the reactive emotions represent their objects as possessing a property that is time-insensitive? Why should the reactive emotions not be thought to represent their objects as having a property that can vary in degree as time passes and life goes on? To *stipulate* that the reactive emotions ascribe a time-independent property to their objects simply because that is what it would take to save the account from its temporal problems would be to beg the question.

Above we argue that one cannot rule out temporal considerations as reasons of the wrong kind without some principled explanation of what makes them belong on the wrong-reasons side of the ledger. And now we see that the defender of REA_f faces a similar challenge. To answer the temporal problem in a non-question-begging way, she would have to provide (i) a substantive proposal about what property the reactive emotions actually represent their objects as possessing, and (ii) an explanation of why this property is time-insensitive in the relevant sense. The burden, therefore, is on the defenders of REA_f to explain why the reactive emotions ascribe to their objects a time-insensitive property of the sort that would save their view from its temporal problems.

Even if we suppose that such an argument is in the offing, other problems remain, the most pressing of which is that fittingness versions of REA face an inescapable trilemma: either they are circular or false or they render the reactive emotions otiose.

A.

Begin with the first prong of the trilemma. On D'Arms and Jacobson's view, to say that an episode of reactive emotion is fit to its object would be to say that the person at whom that emotion is directed has a particular property that these emotions in general represent their objects as having. The most

However, for present purposes it does not matter which approach is adopted. Our arguments below apply with equal force if this alternative solution is adopted instead.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

natural suggestion is that the common property that the reactive emotions ascribe to their objects is *blameworthiness*.[36] The account of fitting reactive emotion being suggested here, then, amounts to the following two claims:

(1) Emotions represent their objects as having certain properties and are fitting when their objects have these properties.
(2) The common property that the blame emotions represent their objects as having is that of blameworthiness.

On this view, then, an episode of a blame emotion felt at a time *t* would be fitting only if the person at whom the emotion is directed has the property of being blameworthy at *t*.

However, if one adopts this account in order to formulate a theory of blameworthiness along the lines of the reactive emotion account, then the theory becomes circular.[37] Suppose one begins by claiming that a person is blameworthy to the extent that it is *fitting* to feel or have the blame emotions toward that person. Now, because one accepts (1) and (2), fittingness here would be a matter of the emotion's object possessing the property that this emotion ascribes to it—namely, blameworthiness. But that is unilluminating. The fittingness of an episode of reactive emotion is being cashed out in terms of the emotion's object actually having the property it is represented as having, namely blameworthiness, and then blameworthiness is being cashed out in terms of the fittingness of the blame emotions. This, of course, is circular.[38]

Now, there are ways to avoid this circularity. But as we show below, they run afoul of the other prongs of the trilemma.

36. Though D'Arms & Jacobson, *supra* note 20, do not explicitly endorse this claim about the property ascribed by the blame emotions, it seems to fit well with the analogous things they say about the properties ascribed by other emotions. After all, they claim, for instance, that envy represents its object as being enviable (or worthy of envy) (*cf. id.* at 66 and 71), that shame represents its object as being shameful (or worthy of shame) (*cf. id.* at 67, 69), and that fear represents its object as being fearful (or worthy of being feared) (*cf. id.* at 66, 67).

37. A related circularity worry for other theories is discussed elsewhere. Rabinowicz & Rønnow-Rasmussen, *supra* note 9, at 395, for example, discuss a circularity worry for fitting attitudes analyses of value, which they attribute to W.D. Ross. *See also* the related discussion in Justin D'Arms & Daniel Jacobson, *Sentiment and Value*, 110 ETHICS 722–748 (2000), at 732–733.

38. An anonymous reviewer for this journal has pointed out that one possible response here would be to insist that even if this prong of the trilemma does lead to a circularity, it need not be viewed as a *vicious* one. Perhaps the circle of concepts at issue here is wide enough to be somewhat illuminating. For instance, it might be taken to illuminate the conceptual connections that hold between blameworthiness and the reactive emotions.

Even if this response might be open in principle, we remain unpersuaded. After all, the basic project we are engaged with is to find a plausible and illuminating theory of what makes a person blameworthy for her actions. Even if the fittingness version of the reactive emotion account reveals some helpful conceptual connections, this would not yet be enough to yield an illuminating account of when and why people are blameworthy for their actions. Furthermore, it is not immediately obvious what makes one conceptual circle vicious and another nonvicious, and we think more work would be needed to explain why the one at issue here is of the latter, nonworrisome variety. Thus we continue to doubt that the trilemma can be escaped by claiming that that REA_f, though circular, is only nonviciously so.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

B.

If the account is to make use of the notion of fittingness introduced by D'Arms and Jacobson, then one cannot abandon (1) above. Thus, in order to avoid circularity, one has to reject (2) and replace it with some other claim about which property the blame emotions represent their objects as having. There are only four other possibilities:

(3) The blame emotions represent their objects as having a property that is instantiated in all the cases when blameworthiness is instantiated and also in some additional cases when it is not: call this property blameworthiness+.
(4) The blame emotions represent their objects as having a property that is instantiated only in a proper subset of the cases in which blameworthiness is instantiated: call this property blameworthiness–.
(5) The blame emotions represent their objects as having a property, P, such that (i) P can be instantiated when blameworthiness is not, and (ii) blameworthiness can be instantiated when P is not.
(6) The blame emotions represent their objects as having a property that is instantiated in none of the cases in which blameworthiness is.

All of these possibilities, however, obviously cause problems for the reactive emotion account. If (3) is adopted, such that the blame emotions ascribe blameworthiness+ to their objects, then the blame emotions will be fitting in some cases when blameworthiness is not instantiated. And so the reactive emotion account would face counterexamples in one direction. Similarly, if (4) is adopted, such that the blame emotions ascribe blameworthiness– to their objects, then blameworthiness will be instantiated in some cases when the blame emotions are not fitting. And so the reactive emotion account would face counterexamples in the other direction. The pattern should be clear by now. If (5) is adopted, then the reactive emotion account will face counterexamples in both directions. Finally, it hardly seems necessary to point out that (6) will not work either, given that if it is adopted, blameworthiness would never be instantiated in the cases when the blame emotions are fitting.

C.

Finally, one might decide to conjoin (1) with (2) but avoid circularity by offering an *independent* account of the property that the blaming emotions represent their objects as having, namely, blameworthiness. That is, the defender of the reactive emotion account might seek to identify some other property on which the property of blameworthiness *supervenes*.

This would avoid the charge of circularity, but at too great a cost. Were we able to identify an independent property in virtue of which agents are blameworthy for their actions, we would simply analyze blameworthiness in terms of *it* and be done with the reactive emotion account altogether. More

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

specifically, suppose the defender of the reactive emotion account takes it that blameworthiness consists in fittingly felt reactive emotion and then claims that a reactive emotion is fitting if and only if its object, O, has the property that this emotion represents O as having. Moreover, to avoid the circularity worry discussed above, he proceeds to investigate our emotional responses in order to find some independent property, P (i.e., not just blameworthiness), that the reactive emotions are fittingly felt in response to. However, as soon as he has succeeded in determining what this property P is, he will simply be able to say that blameworthiness is instantiated if and only if P is instantiated, in which case the reactive emotions drop out of the account of blameworthiness entirely. Thus, as soon as this strategy for preventing the reactive emotion account from being circular is successfully completed, it renders the reactive emotions otiose. What we end up with is a theory in which the reactive emotions do no work at all and which thus would not deserve the name "reactive emotion account."

Now, to avoid the charge of otiosity, one might insist that even if the reactive emotions ultimately do drop out of the account of blameworthiness, investigating the reactive emotions could still have *methodological* value.[39] That is, investigating these emotions might help us identify the properties in virtue of which agents, at bottom, are blameworthy—that is, on which blameworthiness supervenes.

We do not deny that the appeal to the reactive emotions might prove to have methodological value in some such way. However, we still do not think that this vindicates the reactive emotion account. Suppose the fittingness version of the account is construed as a methodological strategy for uncovering an illuminating theory of blameworthiness. The trouble is that if this methodological strategy ever succeeds, what it would yield is a theory of blameworthiness in which the reactive emotions themselves end up doing no explanatory work. What would provide illumination of the notion of blameworthiness is the underlying property that blameworthiness is found to supervene on, not the link between blameworthiness and the reactive emotions. Accordingly, we do not think that insisting on the methodological value of the reactive emotions is enough to turn the fittingness version of the account into a satisfying theory of blameworthiness.[40,41]

39. Thanks to Zac Cogley and an anonymous reviewer for this journal for pressing us on this point.

40. Furthermore, to be successful, this methodological strategy, too, would need to be supplemented by some principled way to weed out those considerations that do not bear on whether the reactive emotions are, properly speaking, fit to their objects. Our actual blaming practices are in a sense "noisy" or "unfiltered" and are at times sensitive to considerations that are wholly irrelevant to the fittingness of the reactive emotions. Thus, to make good on the methodological strategy, one would need a reliable way to weed out the emotional "noise" that our account of blameworthiness must ignore—no easy task.

41. An anonymous reviewer for this journal has pointed out another interesting objection to the otiosity prong of our trilemma. Consider the properties of trilaterality and triangularity. They necessarily co-occur, but perhaps only one of them captures the *essence* of what it is to be a triangle. Similarly, it might be that the property of fitting reactive emotion necessarily

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

IX.

With the hopes that the notion of fittingness would provide a defensible version of the reactive emotion account apparently dashed, we are left to reassess our initial skepticism about the prospects for success of accounts on which blameworthiness is a matter of being the target of reactive emotion that is supported by the balance of *reasons of the right kind*. As we saw above, any such account is sustainable only if it is supplemented by a principled way of distinguishing reasons that are of the right kind from those that are not. Accordingly, we now turn our attention to four of the leading attempts to systematically draw this distinction. We argue that they are all inadequate to the task of saving the reactive emotion account of blameworthiness.[42] We begin with Derek Parfit's familiar account of the distinction between reasons of the right and wrong sort.

A.

Parfit maps the distinction between reasons of the right and wrong sort onto the distinction he draws between object-given and state-given reasons.[43] In the case of desire, this is the distinction between reasons for being in the state of having a desire (state-given reasons) and reasons that show that something is in itself worth desiring (object-given reasons). Thus since health, self-interest, and fairness are considerations that bear on whether to be in the *state* of feeling resentment, indignation, or guilt, Parfit's account entails that these considerations are to be regarded as reasons of the wrong kind.

If there is a distinction between considerations of the right and wrong kind, however, Parfit's distinction between object-given and state-given reasons does not track it. For one thing, it seems there can be reasons of the right kind that are not solely object-given. It would be appropriate, for example, for a five-year-old to experience fear upon encountering an angry

co-occurs with the property, P, that these emotions represent their objects as having, but one might nonetheless insist that the former property, and not the latter, is essential to blameworthiness and illuminates its nature.

Though interesting, we do not think this objection succeeds. The reason is that without being able to get a fix on what this property P is, there is no meaningful illumination of the concept of blameworthiness to be had from understanding it in terms of fitting reactive emotion in the way that REA_f does. What does the heavy explanatory lifting in the theory of blameworthiness is one's view of what this property P is—not the appeal to reactive emotions.

42. We could not possibly canvas all the available alternatives and so we do not attempt to establish that there exists no possible way of drawing the distinction that will be adequate to the needs of the reactive emotion account. However, the arguments we adduce here against the leading attempts to draw the relevant distinction do more fully warrant our skepticism about the prospects of success for a reactive attitude account of blameworthiness.

43. *See* DEREK PARFIT, ON WHAT MATTERS (2011), at 50–51. *See also* Derek Parfit, *Rationality and Reasons*, *in* EXPLORING PRACTICAL PHILOSOPHY: FROM ACTION TO VALUES 17–39 (Dan Egonsson, Björn Petersson, Jonas Josefsson & Toni Rønnow-Rasmussen eds., 2001), at 21–22.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

terrier, but not so for a normal adult. After all, the terrier is dangerous to the five-year-old but not to an adult. Thus the child has reasons of the right kind for fearing the dog, but they are not solely object-given. The dangerousness of the terrier is a relational property; it holds between the terrier and the child but not between the terrier and an adult. Thus we have a case of reasons of the right kind that are not solely given by intrinsic properties of the object.

Furthermore, there can be reasons of the wrong kind that *are* object-given.[44] If an evil demon threatens to kill your family unless you manage not to resent him for cheating at cards, you might have an object-given reason (not just a state-given reason) not to resent him. After all, *he* is the object of your resentment in this case. But this does not make the reason provided by the demon in this case a reason of the right kind for resentment. Intuitively, the threat to your family is a paradigmatic example of a wrong kind of reason.

B.

Pam Hieronymi offers a different but equally creative attempt to draw the distinction between reasons of the right kind and reasons of the wrong kind.[45] The problem is that her proposal suffers from the same sort of trilemma that undermined the fittingness version of the reactive emotion account.

Hieronymi's initial thought is that a reason is a consideration that bears on a question, and we can distinguish reasons of the right kind from reasons of the wrong kind by appeal to the questions on which these reasons bear. On her view, some attitudes—such as belief, intention, and resentment—are constituted by settling certain questions for oneself. For example, forming a belief about p simply amounts to settling for oneself the question of "whether p." Thus to believe that p is just to settle the question of "whether p" in the affirmative. Accordingly, reasons that bear on the question of "whether p" are reasons of the right kind for belief. By contrast, any reason that counts in favor of believing p but does not bear on "whether p" is a reason of the wrong kind.

This strategy can be extended to the reactive emotions as well. For resentment, say, there would be some question (or set of questions) the settling of which is constitutive of resenting.[46] For the sake of argument, we accept Hieronymi's claim that the blaming attitudes are constituted by the settling of some question (though some might question this claim, as Hieronymi is well aware).[47]

44. For more discussion of cases of this sort, *see* Rabinowicz & Rønnow-Rasmussen, *supra* note 9. The example in the text is adapted from one they discuss (*see id.* at 407).
45. Pamela Hieronymi, *The Wrong Kind of Reason*, 102 J. PHIL. 437–457 (2005).
46. *See id.* at – n.37 for her substantive proposal about what questions these are..
47. *See id.* at 454.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

We can see that the trilemma that we discussed in connection with fittingness arises in this context as well if we ask what, exactly, the question is, the settlement of which is supposed to constitute the reactive attitudes. The most natural suggestion is that bearing the reactive attitudes toward P simply amounts to settling for oneself the question of whether P is blameworthy in the affirmative. However, if this is right, then we encounter the circularity horn of the trilemma. After all, the reactive emotion account is the theory that P is blameworthy if and only if the balance of reasons of the right kind supports feeling the blaming attitudes toward P, and reasons of the right kind, in turn, are taken to be reasons that bear on the question of whether P is blameworthy. This would amount to a circular account of blameworthiness.

In response to this first horn of the trilemma, one might try to change the question the settling of which is constitutive of the reactive attitudes so that this question is not simply that of whether P is blameworthy. This route, however, would run afoul of the second horn of the trilemma. In particular, it would lead the reactive emotion account to be subject to counterexamples—in just the way we argued the analogous move renders the fittingness version of the reactive emotion account subject to counterexamples. If the relevant question is whether P has some property *other* than blameworthiness, then it would be guaranteed that cases will arise in which there are reasons of the right kind to feel the reactive emotions toward a person even though this person is not blameworthy—or vice versa.

Thus the only other available move is to offer an independent account of what the question "whether P is blameworthy" amounts to. In other words, the thought would be to accept initially that bearing the reactive attitudes toward P is constituted by settling the question of "whether P is blameworthy" but then to go on to give an informative account of what this question amounts to. However, as soon as one succeeds in this endeavor, then the reactive attitudes become otiose. For if we have an informative theory of what the question "whether P is blameworthy" amounts to, then we could just as well formulate our theory of blameworthiness directly, without any reference to the appropriateness of feeling the reactive attitudes.[48]

48. As it turns out, Hieronymi seems to offer just such an independent account of what the question of "whether P is blameworthy" amounts to. As she puts it: "Elsewhere, I have considered whether resentment might be understood as the settling of a complex set of questions—about the genuineness of the wrong, the status of the wrongdoer, one's own status, and the ongoing significance of the wrong done" (*id.* at 455 n.37). If she is right that these are the questions the settling of which is constitutive of the reactive attitudes, then she is in a position to give a direct theory of blameworthiness—one that does not appeal to the reactive attitudes at all. In particular she could say something along the following lines: P is blameworthy at t iff (1) P committed a genuine wrong, (2) P has the status required for being blameworthy (i.e., the status as a responsible agent), and (3) P's wrong has "ongoing significance" at t.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

C.

We turn now to Darwall's account of the distinction between considerations of the right and wrong kind. As he puts it:

> To be a reason of the right kind, a consideration must justify the relevant attitude in its own terms. It must be a fact about or feature of some object, appropriate consideration of which could provide someone's reason for a warranted attitude of that kind toward it. It must be something on the basis of which someone could (and appropriately would) come to hold the attitude as a conclusion of a process of considering whether to do so. In considering whether to believe some proposition p, for example, it is simply impossible to conclude one's deliberation in a belief that p by reflecting on the desirable consequences of believing p.[49]

Thus, on Darwall's account, a consideration, C, is a reason of the wrong kind for feeling attitude A if and only if it is impossible (or "psychologically impossible," as he later says)[50] to come to hold A as the result of deliberating from C about whether to hold A. So since it is impossible to come to believe that p by reflecting on the fact that it would be desirable to believe that p, it follows that the desirability of believing that p is the wrong kind of reason for belief. The same would hold for the blaming attitudes. Since it is impossible to come to resent someone, say, by reflecting on the fact that it would be desirable to do so, desirability is the wrong kind of reason for feeling resentment.

Let us call this the *psychological possibility account* of the distinction between considerations of the right and wrong sorts. The account holds up for a variety of considerations that we are intuitively disposed to treat as reasons of the wrong kind for having an attitude or emotion: considerations, for example, of health and prudence. However, the problem is that it is clearly psychologically possible to adopt or discard the reactive attitudes on the basis of deliberations involving considerations about the passage of time. In other words, it is perfectly possible for someone who has been wronged by another to consider the amount of time that has passed since the wrong and as a result come to feel that resentment would be misguided, unfitting, or otherwise inapt. The psychological possibility of ridding oneself of the reactive attitudes on these grounds thus renders considerations arising from the passage of time reasons of the right kind, on Darwall's account.

Therein lies the rub. If considerations owing to the passage of time count as reasons of the right sort, then they would suffice to drive a wedge between blameworthiness and the reactive attitudes, thus refuting the reactive emotion account. The alternative view is to hold fast to the reactive emotion

49. DARWALL, *supra* note 14, at 16.
50. *Id.* at 287.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

account by insisting that blameworthiness must be sensitive to the passage of time for the reason that the blaming attitudes are.

In conversation this is the response that Darwall himself took to the objection, but it strikes us as problematic; and here is why. It is not just that the response seems question-begging by presupposing that blameworthiness is sensitive to the passage of time because reactive attitudes are, and in precisely the same ways. The deeper worry is that Darwall's proposal, when conjoined with the reactive emotion account, entails that one's blameworthiness for what one has done will depend on facts about what is or is not psychologically possible for *other people*—including one's victim—and not just on facts about what one has done or failed to do.

Recall Megan, who was wronged by Anderson thirty years ago. Whether the passage of time is a reason of the right sort for Megan to resent Anderson depends, on Darwall's account, on whether it is psychologically possible for others, including Megan, to adopt the attitude of resentment on the basis of factors that emerge with the passage of time. Now, the reactive emotion account entails that Anderson is blameworthy only to the extent that reasons of the right kind render it appropriate to bear the reactive attitudes toward him. So whether *this* is so in the present case depends, if Darwall's account is adopted, on what considerations it happens to be psychologically possible for other people, including Megan, to resent Anderson on the basis of. Thus Darwall's proposal, together with the reactive emotion account, entails that a person's blameworthiness depends on what ends up being psychologically possible for others—without more, an implausible result. One would think that my blameworthiness for what I have done would depend on features intrinsic to me and what I have done (or perhaps also on my subsequent acts of atonement or apology). It would be strange for my blameworthiness to depend on the contingencies of how others, including those I have wronged, happen to be psychologically constituted. Yet this is precisely the consequence of the account on offer.

D.

We come finally to the account of reasons of the right kind that Mark Schroeder suggests. Schroeder states his proposal as follows: "The right kind of reasons to do A are reasons that are shared by everyone engaged in the activity of doing A, such that the fact that they are engaged in doing A is sufficient to explain why these are reasons for them."[51]

One advantage of Schroeder's view, as he points out, is that it avoids the following putative counterexample: "Imagine a fantastic case in which an all-powerful demon pledges to grant wealth beyond imagining to any chess player who castles out of check [a prohibited move] (but not to a non-chess

51. MARK SCHROEDER, SLAVES OF THE PASSIONS (2007), at 135.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

player)."[52] Unlike Parfit's view that reasons of the right kind are object-given, Schroeder's view is not threatened by this case. It does not entail that the person in the example has a reason of the right kind to castle out of check, because the fact that he is engaged in the activity of playing chess is not *by itself sufficient* to explain why he has this reason. In order to explain why he has a reason to castle out of check, it is also necessary to appeal to the demon's pledge.

Nonetheless, we think there are significant problems with Schroeder's account of reasons of the right kind. One preliminary concern is that it is not clear how we are to determine what the relevant activity is, the mere participation in which must be sufficient to explain the existence of one's reasons in order for them to count as reasons of the right kind. How finely are activities to be individuated? At what level of generality are they to be described? Perhaps these questions can be given principled answers, but until Schroeder says something about them, his account seems incomplete.

Furthermore, Schroeder's account appears underinclusive. The problem, to continue using the chess example, is that many reasons that intuitively *are* reasons of the right kind to move one's chess pieces this way or that turn out, on Schroeder's account, to be reasons of the *wrong* kind. Here is why. Suppose Linda is playing chess with her friend Larry and she knows that Larry is fond of taking pieces with his knights whenever possible. She sees a way to checkmate Larry in six moves, but there is one move Larry could use to block her stratagem. Had she been playing any other equally skilled player, she would assume that her opponent would see the counterstrategy. However, the counterstrategy here would require passing up the chance to take a few pieces with a knight. Linda calculates that Larry's preference for using his knight will distract him so that he will not see the way to block her stratagem. Accordingly, she decides go ahead and use the stratagem, despite the risk that it will be blocked.

Larry's preference for using his knights is a reason for Linda to go ahead and use the risky strategy in this case. Moreover, it appears to be a reason of the right kind. Considerations about the tendencies and likely tactics of one's opponent, after all, are fair game in chess (like any complex game of strategy). However, Schroeder's account cannot yield this result. His view is that no reason of Linda's to move her pieces one way or another can be a reason of the right kind unless the mere fact that she is engaged in the activity of playing chess is sufficient to explain the existence of this reason. And in this case, the mere fact that Linda is playing chess is *not sufficient* to explain why her opponent's preference for using his knights is a reason for her to proceed with her stratagem. To explain the existence of this reason, we would also have to appeal to her opponent's habits, preferences, and abilities, as well as the positions of the pieces on the board, and so on. Thus Schroeder's account does not seem able to secure the intuitive result that

52. *Id.*

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

Larry's preference for using his knights is a *reason of the right kind* for Linda to go ahead and use her stratagem.[53]

Perhaps the most important problem for Schroeder's account, for purposes of this paper, is that his account is not capable of doing the job for which it is most needed: namely, to save the reactive emotion theory of blameworthiness. For such an account to be at all illuminating, we need to be able to tell which reasons are of the right kind and which are not. Intuitively, being bribed by an evil demon is a reason of the wrong kind to feel the reactive emotions, but getting whacked in the head by a negligent person is a reason of the right kind. We need the account of reasons of the right kind to explain why this is so and to help us figure out, when going forward, which of the new reasons we encounter—for example, the passage of time—are right reasons and which are not. Without such an account, the reactive emotion account is incomplete.

Schroeder's account, however, does not yield substantive results about which reasons for feeling the reactive emotions are reasons of the right kind and which are not. On his view, the right kind of reasons to feel the blaming attitudes toward somebody would have to be reasons that everyone who takes part in the practice of blaming would have (when in the right circumstances), and which, moreover, are such that the *mere fact* that one is engaged in this blaming practice is sufficient to explain why these reasons exist for one. However, to figure out whether some newly encountered reason, R, for feeling the reactive emotions really is a reason of the right kind we would need an account that tells us *specifically which reasons* are shared by everybody who takes part in the blame game merely in virtue of their participation in it. Schroeder's account does not tell us this.

Consider the passage of time. Here is a possible reason for not blaming somebody for whacking you in the head: it happened thirty years ago (and there were no lasting effects). Some philosophers might want to write this off as a reason of the wrong kind for not resenting the guy who whacked you in the head. Others might be happy to say it is a reason of the right kind. How are we to decide who is right? That is exactly what we need our account

53. There is another distinct reason to think that Schroeder's account is underinclusive. In particular, a lot of paradigmatic reasons of the right kind are not going to be shared by everybody engaged in the relevant activity. And so these reasons would fail to count as reasons of the right kind on Schroeder's account. After all, very few chess games are much alike, and so the circumstances in which real chess players find themselves are going to vary dramatically. Thus the various reasons they have will vary dramatically as well. However, Schroeder's account of reasons of the right kind seems to say that any reason that is not shared by everyone engaged in the activity cannot be a reason of the right kind. So lots of reasons that one would have thought were reasons of the right kind end up not being right reasons on his account.

For example, suppose John has a reason to move his knight to B5 because it will get taken if he does not. By contrast, suppose that Jeff has no such reason to move his knight to B5. Instead, his chances of winning will be enhanced if he moves his knight to B3. From the fact that John and Jeff don't both have this reason to move their knights to B5, it would seem to follow, on Schroeder's account, that John's reason cannot be a reason of the right kind. That, however, seems counterintuitive. If anything is a reason of the right kind to move your knight, it seems this would have to be one.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

of the distinction between reasons of the right and wrong kinds in order to determine. But Schroeder's account is not up to the task. To know whether the passage of time is a reason of the right kind for not feeling resentment, on his account, we would need to know two things: (i) whether the passage of time is a reason not to resent that *everyone* engaged in the blame game shares; and (ii) whether the mere fact that person P is engaged in the blame game is sufficient to explain why the passage of time is a reason for P not to resent. But to know (ii), of course, we need to know whether the passage of time *is* a reason that our blaming practice recognizes as properly bearing on whether to feel the blame emotions—that is, whether it is a reason that our blaming practice recognizes as a reason of the right kind. But this is just what we need the account of reasons of the right kind for.

Thus Schroeder's account fails to render the reactive emotion account illuminating. Any reactive emotion account of blameworthiness (like buck-passer accounts of evaluative properties, more generally) is going to need some principled and informative way to distinguish between reasons of the right kind and reasons of the wrong kind. But that is a task for which Schroeder's account is unfit.

X.

The fact that none of these ways of marking the difference between right and wrong reasons for or against the reactive attitudes succeeds does not establish that there is no way of doing so that would be adequate to the task of saving the reactive attitude account of blameworthiness. Establishing an "impossibility theorem" in this context is beyond our grasp. It is, moreover, unnecessary. If the arguments we have adduced to this point are sound, there is little reason to be confident that a sustainable version of reactive attitude account is in the offing. That is the posture we endorse. The reactive emotion account—despite its many attractions—seems to be ultimately unable to provide a defensible solution to our initial objection concerning the passage of time.

Hence our skeptical tone; and in this case a skeptical tone should be followed with a reminder of what the stakes are. There is an intuitive distinction between finding someone guilty of a crime and expressing condemnation, resentment, and indignation toward that person through punishment. For instance, if the relevant statute of limitations has run such that punishment would be inappropriate, we might still have a reason to ask whether the person in question had acted culpably. Is this a distinction that is grounded in our commonsense moral practices or is it a function of pragmatic considerations alone?

If we are right in thinking that blameworthiness and appropriately felt reactive emotion can come apart, then we have a natural way to ground this distinction in our familiar blaming practices. In particular, we can say

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

that to reach a judgment that someone is guilty of a criminal offense is to express a view about the person's blameworthiness, whereas carrying out a punishment centrally involves an expression of reactive attitudes toward the wrongdoer. Since the reactive attitudes are sensitive to the passage of time in a way that blameworthiness is not, we can easily understand why we might find someone guilty even though it has become inappropriate to express resentment and indignation toward the person through punishment.

But if one instead wants to insist that blameworthiness is tightly tethered to appropriate reactive emotion in the way that the reactive emotion account suggests, then it becomes harder to understand the case where we want to judge someone guilty even though the statute of limitations has made it inappropriate to express blame toward the person through punishment. This is not to say that the task is an impossible one. One might appeal to a range of practical considerations to explain it—say, the fact that the evidence might have gone stale or the limited resources of the state to punish crimes long past. But we think there is some value in attempting to view core features of our legal system, such as statutes of limitations, as expressions of our moral practices, where possible. Thus the impact that the passage of time has on the reactive emotions not only poses problems for a well-known theory of blameworthiness but it also promises to open up new theoretical possibilities in the criminal law.

WORKS CITED

D'Arms, Justin, and Daniel Jacobson. (2000) "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61.1: 65–90.

———. (2000) "Sentiment and Value." *Ethics* 110.4: 722–748.

———. (2009) "Demystifying Sensibilities: Sentimental Values and the Instability of Affect." In Peter Goldie, ed., *Oxford Handbook of Philosophy of Emotion*, 585–614 (Oxford University Press).

Darwall, Stephen. (2006) *The Second Person Standpoint: Morality, Respect, and Accountability* (Harvard University Press).

Gibbard, Allan. (1990). *Wise Choices, Apt Feelings* (Harvard University Press).

Graham, Peter A. (2007) "A Theory of Blameworthiness." Available at http://people.umass.edu/pgraham/Home_files/Blameworthiness%20%2812–1-07%29%20copy.pdf.

Haji, Ishtiyaque. (2002) "Compatibilist Views of Freedom and Responsibility." In Robert Kane, ed., *Oxford Handbook of Free Will* (Oxford University Press).

Heathwood, Chris. (2008) "Fitting Attitudes and Welfare." *Oxford Studies in Metaethics* 3: 47–73.

Hieronymi, Pamela. (2005) "The Wrong Kind of Reason." *Journal of Philosophy* 102.9: 437–457.

Parfit, Derek. (2001) "Rationality and Reasons." In Dan Egonsson, Björn Petersson, Jonas Josefsson, and Toni Rønnow-Rasmussen, eds., *Exploring Practical Philosophy: From Action to Values*, 17–39 (Ashgate).

———. (2011) *On What Matters* (Oxford University Press).

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press

Rabinowicz, Wladimir, and Toni Rønnow-Rasmussen. (2004) "Strike of the Demon." 114 *Ethics* 391–423.

Scanlon, T.M. (1998) *What We Owe to Each Other* (Harvard University Press).

———. (2008) *Moral Dimensions: Permissibility, Meaning, Blame* (Harvard University Press).

Schroeder, Mark. (2007) *Slaves of the Passions* (Oxford University Press).

Shah, Nishi. (2003) "How Truth Governs Belief." *Philosophical Review* 112: 4, 447–482.

Strawson, P.F. (1960) "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.

Velleman, David J. (1991) "Well-Being and Time." *Pacific Philosophical Quarterly* 72: 48–77.

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

Watson, Gary. (1987) "Responsibility and the Limits of Evil," in Ferdinand Schoeman, ed., *Responsibility, Character, and the Emotions*, 256–286.

Zimmerman, Michael. (2010) "Responsibility, Reaction and Value." *Journal of Ethics* 14: 103–115.

https://doi.org/10.1017/S1352325212000158 Published online by Cambridge University Press