

## The paradox of self-blame\*

Patrick Todd & Brian Rabern

*University of Edinburgh*

It is widely accepted both in theory and in practice that there is what has been called a *non-hypocrisy* norm on the appropriateness of moral blame. In the terminology of the recent literature on these topics, one has *standing to blame* only if, as a first approximation, one is not guilty of the very offence one seeks to criticize. Our acceptance of this norm – or one like it – is embodied in the common retorts to criticism, “Who are *you* to blame me?”, and “Look who’s talking!” If I regularly fail to reply to your emails on time, for instance, I’m in no moral position to criticize you for not replying to *my* emails on time – crucially, even if you are indeed blameworthy for this failure, and, crucially, even if someone *else* can appropriately make this criticism. Precisely how to formulate and motivate the non-hypocrisy norm on the standing to blame is a complicated affair. But the following is uncontroversial: if there is a standing-norm on blame at all, then there is some suitable non-hypocrisy norm on standing to blame.<sup>1</sup>

But there is a paradox lurking behind this seemingly commonplace norm in the ethics of blame. And this is that this is a condition, it would seem, that we necessarily cannot meet with respect to ourselves. Consider a (colourful) character we might call Jones the Adulterer:

Caught in adultery for the third time, someone asked Jones how he felt about his moral failings. He replied, “Well, I don’t feel I’m in position to cast any blame. Unfortunately, I’ve been known to commit adultery myself.”

Jones’ stance here is certainly surprising. And yet: *prima facie*, it would seem to be a straightforward application of the non-hypocrisy norm on the standing to blame. Jones is simply reasoning – impeccably, as far as it goes – as follows:

---

\* We first presented a draft of this paper at the “Who are we to blame?” workshop at the University of Edinburgh in April 2019 (generously supported by the Royal Society of Edinburgh); we wish to thank audience members for their helpful comments, especially Jessica Brown, Antony Duff, Dana Nelkin, Samuel Rickless, Matthew Talbert, Elinor Mason, Justin Coates, and Maggie O’Brien. For helpful discussion and/or comments on previous drafts of this paper, we wish to thank Mark Balaguer, Neal Tognazzini, Philip Swenson, Andrew Bailey, and anonymous referees for this journal.

<sup>1</sup> For differing defences of a non-hypocrisy norm on standing, see, e.g., Cohen 2006, Wallace 2010, Fritz and Miller 2018 and 2019, Herstein 2016, Isserow and Klein 2017, Todd 2019, and Rossi 2019. For skepticism about the non-hypocrisy norm, however, see Bell 2013 and Dover 2019; in this paper, we set these skeptical views aside.

Adulterers can't blame adulterers. (Instance of non-hypocrisy norm)

I'm an adulterer. So,

I can't blame myself.

And yet it seems that something must be wrong with Jones' argument; for just as it would seem to be ethical common-sense that those who violate a given norm are not in moral position to criticize violations of that norm, so also it seems to be ethical common-sense that we are often, sadly, in position (indeed, excellent, privileged position) to blame *ourselves* for our own moral failings. And thus we have a paradox. In short, we have a conflict between the inappropriateness of hypocritical blame, and the appropriateness of self-blame:

Inappropriateness of hypocritical blame (IHB): For all subjects  $x$  and  $y$  and norms  $N$ , if  $x$  is a violator of  $N$ , then it is inappropriate for  $x$  to blame  $y$  for violating  $N$ .

Appropriateness of self-blame (ASB): For some subject  $x$  and norm  $N$ , if  $x$  is a violator of  $N$ , it is appropriate for  $x$  to blame  $x$  for violating norm  $N$ .

While both of these principles are *prima facie* plausible, it is plain enough that both cannot be true. For let subject  $s$  be a witness for a true instance of ASB. Then  $s$  is a blameworthy violator of some norm and  $s$  appropriately blames  $s$  for the offence. But by IHB, when  $s$  is both the blamer and the blamee (that is, when  $x = y$ ), since  $s$  is a violator of the relevant norm, it is inappropriate for  $s$  to blame  $s$  for the offence. Contradiction.

As with any paradox of this kind—a set of principles that are each individually plausible yet which seem to be jointly inconsistent—the general options for resolving the paradox are clear. Either demonstrate that one of the seemingly plausible principles is false, or demonstrate that the apparently inconsistent principles are in fact consistent. In this paper, we consider several promising ways of resolving the paradox which attempt to vindicate the claim that common sense cases of self-blame are often appropriate. In the end, however, we contend that none of these ways are as defensible as a position that simply accepts it: we should never blame ourselves. To a defence of this comforting conclusion we now turn.

## 1. *Standing toward the self*

Now, one might think the diagnosis here is simple: standing to blame is not a relation one could lack towards oneself, and thus the consequent of the non-hypocrisy condition must build in the disjunct “unless the blamer is the blamee” (i.e., unless  $x = y$ ). This would, of course, immediately block the application of the non-hypocrisy norm to *oneself* – and thus immediately block our paradox from arising; the revised IHB principle would be consistent with ASB. But matters are not so simple. The principle that *standing is not a relation one could lack towards oneself* is open to counterexample.

Consider the following case – inspired, of course, by John Perry’s famous case pertaining to what he called the “essential indexical” (Perry 1979):

*Two messy shoppers:* Sam is shopping and notices a trail of sugar around the supermarket; some inattentive shopper who should know better is making a mess. It is Julia. Sam finally catches up to Julia and says, “Hey, watch what you’re doing!” Julia then points out that Sam too has been making a mess in exactly the same way; a trail of sugar can be seen behind *his* cart. Julia says, “Look who’s talking!”

Was it appropriate for Sam to walk around the store (even internally) blaming Julia? More to the point: did Sam have the *standing* to make the criticism he made of Julia? And the intuition here – which one might contest, but which we shall not contest – is that he didn’t. Intuitively, Sam is in no position to blame Julia – for his own behavior reveals that he too is a violator of precisely the norm at issue. Accordingly, he is in no position to criticize Julia – a fact that Julia recognizes in her familiar retort.<sup>2</sup>

But now consider what is in essence Perry’s original case:

*The messy shopper:* Anand is shopping and notices a trail of sugar around the supermarket; some inattentive shopper who should know better is making a mess. He begins to feel anger and annoyance with the messy shopper, pushing his cart around the supermarket

---

<sup>2</sup> Note: it is crucial to our case that the relevant messy shoppers are *culpably unaware* that they are making a mess. In other words, although the shoppers don’t know that they are making a mess, they *should know* that they are making a mess. Thus, we assume that the relevant “epistemic condition” on moral responsibility is in fact met in these cases; in particular, we assume that the epistemic condition on responsibility should not be stated in such a way as to require that a blameworthy agent *knows* that she is doing wrong. It is enough that the relevant agent *should know* that she is doing wrong (Sher 2009: 71-87; for discussion, see Rudy-Hiller 2018).

in an attempt to find the shopper and confront him or her over this transgression.

Anand finally catches a glimpse of the messy shopper – in the mirror. It is himself. He’s been making the mess.

Our claim is simple. Just as Sam lacked the standing (even internally) to walk around the supermarket criticizing Julia, so Anand lacked the standing (internally) to walk around the supermarket criticizing the messy shopper – which is to say that Anand lacked the standing to walk around the supermarket criticizing Anand. And that is to say that one *could* lack standing to blame oneself for violating a given moral norm.

Such a result, of course, does not immediately entail that *no one ever* has the standing to blame himself. Such a result only demotivates a conception of standing that makes it a conceptual truth that one cannot lack standing to blame oneself. In other words, this result makes the first-pass suggestion to append “unless  $x = y$ ” to IHB unduly *ad hoc*. For that modification is motivated by the sheer contention that identity with the wrongdoer is sufficient for standing to blame that wrongdoer. And, as we have seen, that contention is false.<sup>3</sup>

But perhaps there is a related fix – that is, perhaps there is another way of formulating a principle about “standing to blame” relevantly similar to IHB that is *consistent* with ASB. For perhaps one might concede that the case of Anand is a case in which someone lacks standing to blame himself. But one might still insist on the following thought. Yes, identity between blamer and blamee is not sufficient for standing. However, *known* identity *is* sufficient for standing; that is, if a given subject *knows* that he is blameworthy for violating a given norm, then that subject has the standing to blame that subject for violating that norm. And since, in the usual case, we know when we are violators of the relevant norms, in the usual cases of self-blame, we will have standing to blame ourselves. The suggestion here is thus to append something like, “unless  $x$  knows that  $x = y$ ” to IHB, thereby rendering the modified principle consistent with ASB.<sup>4</sup>

---

<sup>3</sup> One philosopher who independently agrees with this limited conclusion is Hannah Tierney; responding to Lippert-Rasmussen’s (2021) invocation of “hypercrisy” – wherein one blames *oneself* for violating a certain norm, but not others – Tierney (forthcoming) contends that “while the hypocrite has standing to blame others but not herself, the hypocrite possesses standing to blame himself but not others.” Our conclusion, however, is that *no one* has standing to self-blame.

<sup>4</sup> Note that this route introduces a substantial and somewhat surprising complication, in so far as, it renders standing to blame *mode-of-presentation* sensitive. After all, the messy shopper certainly knows that he is himself: that is, Anand is an  $x$  and the messy shopper is a  $y$  such that  $x$  knows that  $x$  is  $y$ . So merely appending the disjunct “unless  $x$  knows that  $x = y$ ” does not *by itself* avoid the problem. To get off the ground, the suggestion must involve the commitment that, while it is true that the messy shopper knows that he is himself, he only knows this under a first-person mode of presentation; he doesn’t know it under a “messy shopper” mode of presentation. And standing to blame is mode-of-presentation sensitive, in the sense that  $x$  may have standing to blame  $y$  when  $x$  thinks of  $y$  under description D, but also lack standing to blame  $y$  when  $x$  thinks of  $y$  under description D’.

But our contention is that this addition to the IHB principle, like the former one, makes the principle unduly *ad hoc*. The non-hypocrisy norm is a norm on the appropriateness of *blame*. But the modified IHB principle builds in a restriction, so that it has the effect of bracketing-off cases of self-blame. That is, the principle essentially states that when a case of blame is *not* a case of (known) self-blame, then the non-hypocrisy condition applies. This renders the relevant norm highly *disjunctive*. And our contention is that, when we grasp the relevant norm, the norm we grasp is not any kind of *disjunctive* norm – a norm that says something like, “Blame the guilty party only if you aren’t guilty of the relevant transgression yourself, unless you know that you *are* the guilty party.” The norm is just: “Blame the guilty party only if you aren’t guilty of the relevant transgression yourself.” And our point is that *this* would be seem to be the relevant “norm of blame” – and that *this very norm* is a norm we cannot meet with respect to ourselves.

## 2. *Bifurcate blame?*

Instead of pursuing the (we think implausible) route of saying that the non-hypocrisy norm is disjunctive, one might instead *bifurcate blame*. On this response, there are two distinct attitudes, subject to different sets of norms: there is *other-blame*, and then there is *self-blame*. (It isn’t that the norm is disjunctive; it is that there are two distinct attitudes subject to different norms.) Now, *other-blame* is subject to some relevant non-hypocrisy norm such as IHB. *Self-blame*, however, is not subject to any such norm. Since self-blame is not subject to the relevant norm, self-blame remains appropriate, even if the self-blamer is guilty of the very offence for which he is self-blaming – but not so, of course, for other-blame. The paradox, then, involves *equivocation* on “blame”. And once we see that IHB concerns other-blame and that ASB concerns self-blame, we can accept both principles without inconsistency.

On this response, it would indeed always be inappropriate for any given subject to other-blame himself – no one could have the standing to do so! – but typically, what we call “self-blame” is not other-blame directed towards the self, but, well, a distinct attitude altogether, viz., self-blame.

Our response to this suggestion is brief. In short: Self-blame is a form of blame. So, self-blame is subject to the norms of blame. So, self-blame is subject to the non-hypocrisy norm. That is, we find it implausible that *self-blame* could be anything other than, well, *blame* directed towards the self. But then “self-blame” and “other-blame” are indeed instances of one and the same attitude – the difference simply being in the respective *objects* of that attitude. (In the one case, that attitude is direct towards others. In the other case, *that very attitude* is directed towards the self.)

But then it is deeply plausible that the non-hypocrisy norm is a norm exactly on the attitude of blame itself; and if this is so, it remains a norm on that attitude, whatever its object.

Note. Our contention here is *not* that all forms of blame must be subject to all of the same norms. (Perhaps blaming one's benefactors, say, is subject to certain norms to which blaming one's peers is not.) Our contention is that all forms of blame must be subject to all of the general norms *governing blame*. And our contention, once more, is that self-blame is indeed a form of blame. Thus: since self-blame is a form of blame, and since the non-hypocrisy norm is a norm *on blame*, self-blame is subject to the non-hypocrisy norm.

### 3. *The nature of hypocrisy*

And so perhaps we should try a new approach. Perhaps the paradox points to an important fact about the non-hypocrisy condition on standing. On this approach, for someone to fail to meet the non-hypocrisy condition, it is not enough, say, that that person is an adulterer blaming another adulterer. It must be added that this is an adulterer *who fails to blame himself* for his adultery, blaming an adulterer. The thought here appears simple. If I blame *myself*, say, for violating the given norm, then it doesn't follow that I lack standing to blame with respect to that norm. On this approach to the non-hypocrisy condition, it is the *inconsistency* of the relevant blamer that, in some sense, removes that person's standing: it is the fact that this person is prepared to blame *others*, but not himself. If I blame *myself* for not replying to your emails on time, I am *therefore* not hypocritical with respect to blaming *you* for *your* failure to reply to my emails on time: I blame the both of us, and therefore my blaming you is not without standing.<sup>5</sup> The suggestion, then, is that IHB must be amended as follows:

(IHB') For all subjects  $x$  and  $y$  and norms  $N$ , if  $x$  is a violator of  $N$  **and  $x$  doesn't blame  $x$  for violating  $N$** , then it is inappropriate for  $x$  to blame  $y$  for violating  $N$ .

---

<sup>5</sup> Cf. Wallace 2010, Fritz and Miller 2018. However, it is worth noting that this intuition about standing is contested; see, e.g., Rivera-López (2017: 343), who (in response to Wallace) writes, "Acknowledging (or criticizing myself for) my own faults does not restore my moral standing to blame others, regardless of how sincere such self-blame might be."

IHB' is equivalent to saying that in order to meet the non-hypocrisy condition, then either (a) you have to not be a violator of  $N$ , or (b) you have to self-blame for violating  $N$ . Thus, self-blaming is sufficient to meet the non-hypocrisy condition.<sup>6</sup>

Far from resolving the paradox we have presented, however, this suggestion simply deepens it. But here we must back up. In this paper, we are primarily concerned with one *necessary condition* on standing: the non-hypocrisy condition. More generally, however, we assume that the non-hypocrisy condition is one of a certain set of necessary *and sufficient* conditions on standing. In particular, the assumption we wish to make is that if a given agent meets the non-hypocrisy condition, as well as any other conditions on standing (such as, perhaps, the “business condition”, or the “non-complicity condition”, or the “warrant condition”<sup>7</sup>) then that is *sufficient* for the agent’s standing. Letting  $C_1, \dots, C_n$  be any other conditions on standing, then we suggest:

(Standing) For all subjects  $x$  and  $y$  and norms  $N$ , it is appropriate for  $x$  to blame  $y$  for violating  $N$  iff [( $x$  meets the non-hypocrisy condition) &  $C_1$  &  $C_2$  & ... &  $C_n$ ]

In the current dialectical context, we can simply assume that the relevant agent has met all of the other conditions on standing. Thus, *if* that agent meets the non-hypocrisy condition, that will *suffice* for that agent’s standing.<sup>8</sup>

Now the point. Recall that, according to IHB', self-blaming is sufficient to meet the non-hypocrisy condition. And now suppose that, in the relevant context, all the other conditions  $C_1, \dots, C_n$  are met, so that an agent’s meeting the non-hypocrisy condition is sufficient for that agent’s having standing. The result is simple. On this approach, since self-blame can be sufficient for meeting the non-hypocrisy condition, in the relevant context, one can render it the case that one has standing to blame oneself ... precisely by blaming oneself. However, it is senseless to suggest that *if one blames oneself* for violating the given norm, then ... one has standing to blame oneself for violating that norm. Having the standing to blame a given wrongdoer is a conceptually prior necessary condition on the appropriateness of one’s blaming that wrongdoer. Thus: one cannot *render* one’s blame appropriate simply and precisely *by blaming*. Said differently: one cannot

---

<sup>6</sup> Note: IHB' is of course consistent with ASB, which can be seen as follows: Let subject  $s$  be a witness for a true instance of ASB. Then  $s$  is a blameworthy violator of some norm and it is appropriate for  $s$  to blame  $s$  for the offence. By IHB' it follows that  $s$  indeed blames  $s$  for the offence – and so indeed has standing to blame  $s$ .

<sup>7</sup> For more on the business condition, see Radzik 2011 and Seim 2019, and for more on the non-complicity condition, see Cohen 2006, and for more on the warrant condition, see Friedman 2013 and Coates 2016; for discussion, see Coates and Tognazzini 2018.

<sup>8</sup> Notably, however, Todd (2019) contends that the “business” and “warrant” conditions are not genuine conditions on *standing*, and the non-complicity condition reduces to the non-hypocrisy condition. Thus, meeting the non-hypocrisy condition is necessary *and sufficient* for standing. Here, however, we remain neutral about whether there are any other conditions on standing than the non-hypocrisy condition.

bootstrap the appropriateness of one's blaming a given wrongdoer precisely by blaming that wrongdoer. In the present context: the appropriateness of blaming oneself is exactly what is at issue; this appropriateness cannot simply be achieved by doing the blaming.

One might suppose that the problem here can be solved with a dispositional fix. Perhaps when I myself am a violator of  $N$ , it isn't my *actually blaming* myself for violating  $N$  that suffices for my standing to blame with respect to  $N$ , but instead my being *disposed* to blame myself that suffices for my standing. Thus, suppose we tried:

(IHB'') For all subjects  $x$  and  $y$  and norms  $N$ , if  $x$  is a violator of  $N$  **and  $x$  isn't disposed to blame  $x$  for violating  $N$** , then it is inappropriate for  $x$  to blame  $y$  for violating  $N$ .

But this second suggestion seems little better than the first. After all, having the right to  $\varphi$  is a conceptually prior necessary condition on its being appropriate for one to be *disposed* to  $\varphi$ . In other words, just as one cannot render it appropriate for one to  $\varphi$  precisely by  $\varphi$ -ing, so similarly one cannot render it appropriate to  $\varphi$  by becoming *disposed* to  $\varphi$ . It *first* must be appropriate for one to  $\varphi$ ; *then* it may be appropriate for one to be *disposed* to  $\varphi$ .

There is, however, another way of bringing out the severity of the problems facing IHB' and IHB''. The problem is that these principles do not seem apt to give to their proponents what they want. In short, from the perspective of a rival theory to our own – one on which self-blame is sometimes appropriate – then these principles will face an *undergeneration* problem: they will undergenerate cases of what *should* be appropriate cases of self-blame. After all, notice that both IHB' and IHB'' in fact *vindicate* the position of Jones the Adulterer. In particular, notice that Jones the Adulterer neither blames himself, nor is disposed to blame himself. The result: according to IHB' and IHB'', Jones lacks the standing to blame himself! (Just let  $x = y$ , and note that, in this case, we do *not* have an  $x$  who is disposed to blame  $x$ .) Of course, we are content to accept the result that Jones the Adulterer lacks standing to blame himself – but the point is that, if, as our opponents will contend, self-blame is ever in fact appropriate, it seems wholly unacceptable that Jones should be *prohibited* from blaming himself *just because* he isn't in fact disposed to do so.

The point that, according to IHB'/IHB'', Jones lacks the standing to blame himself doesn't in itself entail that, given these principles, *no one ever* has standing to blame himself. However, we do think these principles would, if true, go some ways towards supporting that conclusion. Consider the following argument:



1. If self-blame is sometimes appropriate, then it is appropriate for Jones the Adulterer to blame himself [even if he doesn't].
2. But given IHB'/IHB", it is not appropriate for Jones the Adulterer to blame himself [because he doesn't blame himself, and isn't disposed to]
3. So, given IHB'/IHB", self-blame is never appropriate.

We have defended (1) and (2), and (3) follows.

#### 4. *Bifurcate the blamer*

The responses to the paradox so far have all somehow attempted to modify the non-hypocrisy principle so that it is consistent with there being appropriate cases of self-blame. Another type of response insists that the cases we ordinarily call "self-blame" are not really cases where  $x$  blames  $x$ . Strictly speaking, there are never appropriate cases of self-blame. ASB is false. But, according to this response, the ordinary cases we call "self-blame" are nevertheless often appropriate because they really involve non-hypocritical other-blame. The strategy here is to *bifurcate the blamer*. There are two ways one might wish to carry out this project: bifurcate the blamer *at a time*, and bifurcate the blamer *over time*.

The first strategy begins with a familiar – and certainly ancient – model of the self. On this model, the self is composed of "parts" – for instance, as in Aristotle, the rational part, and the appetitive part. And the central suggestion of the current strategy is the following. In cases of what we might ordinarily call "self-blame" – cases we might ordinarily describe as someone blaming herself for what she has done – what is *really* going on in these cases is that one *part* of the relevant "self" is blaming a *distinct* part of the given "self". Thus, the thing that is doing the blaming (say, the "rational part" of the person) is in fact not identical to the thing that is being blamed (say, the "appetitive part" of the person). And thus, this response contends, no violation of the requisite norm on blame is implied. For the given norm on blame prohibits  $x$  from blaming  $y$  when  $x$  is guilty of the relevant transgression. But even in cases of "self-blame" (where we might have thought  $x = y$ ), we do not *really* have any such  $x$  blaming  $x$  – instead, we have some relevant  $x$  blaming some *distinct*  $y$  (of course, some  $y$  to which it bears a particularly intimate relation). In short, the suggestion is that the relevant *subjects* and *objects* of the attitude of *blame* are not what we might ordinarily consider "persons" (single human organisms, say) – rather, the subjects and objects of these attitudes are *parts* or *aspects* of persons. (Alternatively, one might suggest that any

given human organism is constituted by several distinct persons [a human Trinity, perhaps]; but we set this even more radical suggestion aside.)

We do not hereby wish to take issue with the theory that the self has “parts”. We instead wish to take issue with the given *application* of that theory – which is subject to several serious problems. We mention only the following. Consider blame directed towards *others*. The problem is that if this strategy works to show that the relevant instances of what we ordinarily call “self-blame” imply no violation of the standing norm, it would also show that paradigm cases of violation of that norm are in fact no such cases at all. Suppose that I blame you for failing to reply to my emails on time. You then point out, correctly, that I regularly fail to reply to *your* emails on time. I then say: well, it is my *appetitive* part that fails to reply to those emails on time, but it is my *rational* part that is blaming you for failing to reply on time – and so I am no hypocritical blamer after all. You will, of course, reject this explanation. You will say that if *I* fail to manage my “appetitive part” in the requisite way, then I am in no position to blame *you* when *you* fail to manage *your* “appetitive part” in an exactly analogous way. And so the current suggestion about the subjects and objects of the attitude of blame – even if it were plausible in itself (which it isn’t) – would not undergird the desired distinction, according to which self-blame can never violate the non-hypocrisy norm on blame, whereas other-blame can (and often does).

The second way of carrying out the project of bifurcating the blamer is to bifurcate the blamer *over time*. The thought here is again familiar. On this approach, in cases we might ordinarily describe as “self-blame”, the  $x$  doing the blaming is never in fact identical to the  $x$  being blamed. On this approach, for instance, Jones-at-t1 is a distinct entity as compared to Jones-at-t2; and when Jones-at-t2 blames Jones-at-t1, we do not have any entity blaming itself, for Jones-at-t2 is not identical to Jones-at-t1. Besides the obvious intrinsic metaphysical costs associated with this theory, however, the problem is that this response is subject to the same problems associated with the response above. Suppose that I blame you for failing to reply to my emails on time. You then point out, correctly, that I regularly fail to reply to *your* emails on time. I then say: well, that was *me-at-t1*, and it is *me-at-t2* who is now blaming you for your failure. You will, of course, reject this explanation – but any rate, we can once again note that even if this position were plausible in itself, it would not undergird the desired distinction, according to which self-blame can never violate the non-hypocrisy norm, whereas other-blame can (and often does).

## 5. Some accounts in the literature – and literature

At this stage, one might worry that the paradox arises, perhaps, only given some relatively idiosyncratic understanding of the “non-hypocrisy” condition on standing to blame – an understanding that is in conflict with recent statements of that condition in the literature. Needless to say, we can’t consider every statement of the non-hypocrisy condition that exists in the literature; we will content ourselves to observe that a conflict between the non-hypocrisy condition and the appropriateness of self-blame seems to arise given two prominent recent accounts of that condition offered by Todd (2019) and Fritz and Miller (2018 and 2019).

Todd’s account is simple and straightforward: “one has moral standing to blame [a given responsible wrongdoer] if and only if one is morally committed to the values that condemn the wrongdoer’s actions.” (2019: 357) Todd’s account is thus a *moral commitment* account of standing. Here we must be brief. Plausibly, in order to match the intuitive data concerning who “lacks standing”, the “commitment” at issue in Todd’s account needs to be very strong. However, if the “commitment” at issue is strong enough to match the data, it will also be strong enough to imply that violators of norms, insofar as they are, are never in fact committed to the values undergirding those norms – in which case those violators will, given Todd’s account, lack standing to blame those who violate them, *including themselves*. Thus, given Todd’s account, we could always, for example, reason impeccably as follows:

1. Only committed vegetarians can blame people for ordering steak. (Application of standing norm)
2. I am not a committed vegetarian. (See: I just ordered steak.) So,
3. I cannot blame myself for ordering steak.

Thus, Todd’s account generates a conflict with the appropriateness of self-blame. Of course, our claim is that this is not, in itself, a *defect* in Todd’s account, but a result with which we can learn to live.<sup>9</sup>

---

<sup>9</sup> More recently, Todd (ms) defends what he calls the *be better* norm of criticism, according to which one must criticize  $x$  with respect to norm  $N$  only if one is *better than*  $x$  with respect to  $N$ . (For a similar proposal, see Rivera-López 2017: 344.) What motivates Todd’s analysis, it seems, is the idea (to which we ultimately appeal below) that to blame is (inter alia) to *stand in judgment* – and the idea that the imagery of *standing* in judgment is to be taken seriously. To *stand* in judgment is to *look down on* the one judged. But when  $x$  is no better than  $y$ ,  $x$  cannot *look down on*  $y$ . In any case, the important point in the present context is the following: Todd’s *be better* norm immediately rules out the appropriateness of self-blame. Plainly, if in order to appropriately blame, the blamer must be better (in the relevant respect) than the blamee, then since for no  $x$  is  $x$  better than  $x$ , and necessarily so, no one appropriately self-blames. In a similar vein, Rivera-López contends that standing requires moral superiority with respect to the blamee. Interestingly, Rivera-López seems to sense a tension between this view and the appropriateness of self-blame; he writes: “In my view, what we call ‘self-blame’ is rather the expression of other kinds of attitudes: repentance, guilt, remorse, among others. I only blame myself (for example, by expressing indignation) for my own act metaphorically.

Fritz and Miller's (2018) account is also simple and straightforward. For Fritz and Miller, lacking standing is a matter of being unfairly *differentially disposed to blame*. Simply put: if you are unfairly differentially disposed to blame violators of some norm *N*, then you lack standing to blame with respect to norm *N*. Given Fritz and Miller's account of standing, we have strong reason to conclude that no one has standing to blame himself. How so? The problem, we hope to show, parallels closely the problem of *undergeneration* mentioned at the end of Section 3.

To begin, note that it is a central component of Fritz and Miller's view that even the *inconsistent* blamer lacks standing to blame. For instance, consider this case developed by Todd (2019: 370). Suppose Michael has two colleagues, Ellie and Suilin, both of whom regularly fail to reply to his emails on time. However, Michael is only really disposed to blame *Ellie* for this failure, but not Suilin – and there is no adequate moral justification for this differential disposition. Fritz and Miller insist that it therefore follows that since (i) Michael rejects the equality of persons with respect to norm *N* (replying on time), he (ii) therefore forfeits the right to blame with respect to norm *N*, which amounts to saying that he lacks the standing to blame both Ellie *and* Suilin for their failure to reply to his emails on time.<sup>10</sup> But now consider. Suppose *Michael* in fact himself often fails to reply to his colleagues' emails on time. And suppose that he knows that he does so, knows that this is wrong, and blames himself for this failure. Other things being equal, is Michael's blame appropriate, in the sense that Michael has the standing to blame himself for this failure?

Fritz and Miller's account straightforwardly tells us that it isn't. After all, from what we've been told above, Michael rejects the equality of persons with respect to *N*: he is (unfairly) disposed to blame Ellie, but not Suilin – and that, Fritz and Miller insist, is enough to count as rejecting the equality of persons with respect to the given norm, which equality grounds one's right to blame. By a straightforward application of Fritz and Miller's theory, it therefore follows that Michael lacks the standing to blame *with respect to violations of N*. But then: Michael therefore lacks the standing to blame *his own violations of N*.

From the perspective of a moral theory on which self-blame is sometimes appropriate, it is wholly unacceptable that Michael could be prohibited from blaming himself for violating *N*, simply and precisely because he rejects the equality of persons with respect to *N*! In other words, if self-blame is ever appropriate, then Michael's blaming himself should be an appropriate case of

---

This would require splitting myself into two selves: one superior self who admonishes the other, inferior self. This can only be metaphorical." (2017: 360) Rivera-López suggests that what we normally call "self-blame" is not literally blame directed at oneself; instead it is really just guilt or remorse. We agree that colloquially "self-blame" may often just mean "guilt". But we think Rivera-López goes too far. We agree that self-blame would only be *appropriate* if there were two selves – one who didn't violate the norm and the other who did. But insisting that *appropriate* self-blame is impossible is very different from insisting that self-blame is impossible. (More on these themes below.)

<sup>10</sup> Todd (2019) is highly critical of this implication of Fritz and Miller's theory, regarding it (roughly) as a *reductio* of that theory. Fritz and Miller (2019), however, further embrace and defend this result.

self-blame. The mere fact that he is inclined to blame Ellie but not Suilin shouldn't imply that he is *not* permitted to blame himself. But on Fritz and Miller's account, Michael's blaming himself is in fact not an appropriate case of self-blame, since it is an instance of a blamer blaming a blamee for violating a norm, when the blamer rejects the equality of persons with respect to that norm.

Slightly more carefully, we can put our argument here as follows:

1. If self-blame is sometimes appropriate, Michael's blaming himself is appropriate.
2. If Fritz and Miller's account is correct, Michael's blaming himself is not appropriate. [He is unfairly differentially disposed to blame with respect to the violation in question.]
3. So, if Fritz and Miller's account is correct, then self-blame is never appropriate.

Thus, both the account from Todd and the account from Fritz and Miller give us the resources, in different ways, to reach some version of our startling conclusion: we never have the standing to blame ourselves.

Before we turn to our own preferred solution to the paradox, it is worth appreciating that the paradox arises, not merely given recent statement of the non-hypocrisy norm in the literature, but also given what would seem to be foundational expressions of the non-hypocrisy norm familiar from non-philosophical contexts. Consider, for instance, Jesus' famous intervention on behalf of the woman caught in adultery: "Let he who is without sin cast the first stone." (John 8:7) As various commentators have noticed, this standard is, of course, considerably more stringent than the one at issue in this paper – which is a standard more like "Let he who is at least not an *adulterer* cast the first stone." But the point here is simple. According to the moral lesson we are (presumably) meant to learn from this episode, to blame, of course, is akin to casting a stone. But if only *he who is without sin* may cast a stone, and only *he who is with sin* is a candidate for stoning, then of course no one should be in the business of casting any stones on *himself*. To a defense of that thought we now turn.

## 6. *Self-blame and guilt*

We turn now to one final response to the paradox we have presented we wish to consider – and this is the response that we recommend. This response is simply to accept the surprising conclusion. We accept the non-hypocrisy principle, and accept that it renders self-blame

inappropriate: insofar as we recognize that we are *candidates for blame* with respect to a given norm, we should therefore accept that we are not *candidates for blaming* with respect to that norm – and we should therefore simply accept that, insofar as we are blameworthy for violating a norm, we should not blame even ourselves for violating that norm. Precisely the condition of the guilty self that makes the guilty self ill-suited to blame others likewise makes the guilty self ill-suited to blame *itself*. Now, above we maintained that it is implausible to *bifurcate blame* – to maintain that we have two fundamentally distinct attitudes involved in *self-blame* and *other-blame*. However, it is not, we contend, similarly implausible to contend that the following involve two fundamentally distinct psychological attitudes: blame, and *guilt*. Whereas it is implausible to suggest that *self-blame* and *other-blame* are distinct attitudes, it is not similarly implausible to suggest that *blame* is one attitude – subject to a singular, non-disjunctive non-hypocrisy norm, whereas *guilt* is another attitude altogether, subject to no such norm at all.

Revisit the case of Anand. What is it that becomes appropriate once Anand discovers that *he* is the messy shopper? What becomes appropriate is *not*, we contend, one and the same attitude that Anand exhibited as he strode around the supermarket, looking for the messy shopper – except that it now becomes appropriate that Anand should *knowingly* exhibit that attitude towards himself. For *that* attitude was the attitude of *blame* – of, we might say, *standing in judgment* with respect to the wrongdoer. And when Anand discovers that *he* is the messy shopper, it is strained and implausible to suggest that it now becomes appropriate that Anand should stand in judgment with respect to Anand. What becomes appropriate, perhaps, is Anand’s recognition that, in principle, it *would* be appropriate for some *conscientious* shopper *who isn’t himself* to exhibit the relevant attitude towards himself. And it is appropriate, perhaps, all else equal, for Anand to be pained by that fact: it is fitting that others should blame me.<sup>11</sup> But recognizing that it would be appropriate for someone else to exhibit an attitude towards oneself, and being pained by that fact, does not imply the appropriateness of one’s exhibiting that attitude towards oneself. In short: what becomes appropriate is not Anand’s *blaming* himself for his messy shopping; that was not appropriate when he was in ignorance that he was the messy shopper, and it remains inappropriate once he realizes that he is. What becomes appropriate, if anything, is Anand’s pained recognition *of his guilt* – in short, of his feeling guilty. But this attitude must be distinguished from the attitude of blame directed towards the self.<sup>12</sup>

---

<sup>11</sup> Cf. Sartre’s position that “[shame] is the recognition of the fact that I am indeed that object which the Other is looking at and judging.” (Sartre 1943: 261).

<sup>12</sup> Very often, it is not; see, e.g., Carlsson 2017 and Carlsson forthcoming, who identifies guilt with self-blame. However, see, e.g., Gibbard 1992: 139: “Guilt and shame, then, are first-person counterparts to anger and disdain. Not that guilt is self-directed anger; feeling guilty is different from feeling you could kick yourself.” Cf. also Duggan 2018: 299, who, in distinguishing between self-blame and guilt, writes: “Guilt doesn’t aim to hurt. It just hurts.” Note:

But how, precisely, should these attitudes be distinguished? We suggest the following account.

**Blame:**  $x$  blames  $y$  iff  $x$  stands in judgment with respect to  $y$  and condemns  $y$ , accompanied by some sort of negative affect (or “hostile attitude”<sup>13</sup>).

Of course, given this construal of blame, we can thus distinguish:

*Other-blame:* when an agent stands in judgment with respect to an agent and condemns that agent, accompanied by some sort of negative affect, where the agent who stands in judgment is distinct from the agent judged.

*Self-blame:* when an agent stands in judgment with respect to an agent and condemns that agent, accompanied by some sort of negative affect, where the agent who stands in judgment is identical to the agent judged.

But we contend that self-blame must be distinguished from *guilt*:

**Guilt:**  $x$  feels guilt/feels guilty iff  $x$  judges (*de se*) that  $x$  is blameworthy, and is pained (in the right way) by that judgment.<sup>14</sup>

Having thus distinguished these attitudes<sup>15</sup>, our core contention is that one may – and one should – feel *guilt* without also *self-blaming*.

---

Duggan distinguishes between blame (and therefore self-blame) and guilt at least partially in terms of their respective social roles. However, we are skeptical of the claim that blame “aims to hurt” in the relevant sense; for instance, it seems strained to suggest that the angry resentment I feel towards the actions of distant agents I hear about on TV *aims at* their suffering.

<sup>13</sup> Bell (2013: 265).

<sup>14</sup> We will not try to specify fully what it takes to be pained “in the right way” relevant to guilt. Presumably, for instance, the pain needs to be emotional in some way (and not merely “physical” – as in a pain in one’s elbow) – and the pain needs to be non-deviantly caused by the judgment. One further issue: there may be a subtle “one thought too many” issue here. Does one feel guilty because one is pained by the judgment that one is blameworthy? Or does one feel guilty because one is pained by the judgment that one has performed the act in virtue of which one judges oneself to be blameworthy? If the reader thinks the latter is more to the point, we may modify our characterization of guilt appropriately.

<sup>15</sup> Linguistically, it is worth noting that “blame” is a *transitive* verb relating a subject to a direct object (as in “give” or “poke”), whereas “(feeling) guilt” is an *intransitive* verb (as in “sneezed” or “cries”). In other words, blame involves essentially a relation between a blamer and a blamee. Guilt, however, is no sort of relation.

## 7. Advice

Perhaps it will help if we further develop the sort of *recommendation* that (we contend) follows from the arguments developed in this paper. Our contention is that, when we recognize that we are blameworthy, we should (*ceteris paribus*) feel *guilt*, but we should not self-*blame*. More particularly, consider the following sort of advice.

You have been blaming yourself, poor soul? And for what? For being a scoundrel, you say. But consider he who it is that condemns you. Consider that he who condemns you is too a scoundrel. And consider that – even if you *are* a scoundrel – you would never concern yourself with the condemnation of a fellow scoundrel. Thus: silence the voice inside you who condemns you. The voice of the one who condemns is the voice of one no better than the one condemned. Or do you look down on yourself? But the one who looks down is no higher than the one looked down upon.

Nevertheless: Feel badly that you are blameworthy. But do not self-blame.

We admit that this is advice that is not immediately clear we know how to follow. How can we go about feeling badly that we are blameworthy, *without* blaming ourselves? Our contention, however, is that this is advice it is *possible* to follow. One can feel guilt, without self-blaming.

Consider the following analogy. It certainly seems possible for Jones to feel praiseworthy – to judge that he is praiseworthy, and to feel good about his being praiseworthy – without its being the case that Jones *praises himself*. Or consider that it certainly seems possible that Sarah should feel admirable, without its being the case that Sarah *admires herself*. Indeed, it seems evident that it is sound advice that though we should (in the right way, and to the right degree) sometimes *feel admirable*, we should, in general, refrain from *admiring ourselves*. To feel good about being admirable is, well, just that. To *admire oneself* is something else entirely: it is, so to speak, to “go before the mirror” and admire what one sees. It is sound advice that we should, at least sometimes, feel an appropriate sort of pride (in the right way, and to the right degree). But that advice is certainly entirely different in kind from the advice that we should, on those occasions, *praise ourselves*



to the relevant degree. One can feel praiseworthy, without ever praising oneself – and without anyone else ever praising one either.<sup>16</sup>

Our advice now comes into focus. Just as it may be sound advice that we should (sometimes) feel praiseworthy, but not praise ourselves, so it is sound advice, we contend, that we should (when we are blameworthy) feel blameworthy, but not blame ourselves. Setting aside cases like that of Anand, to *blame* oneself is, so to speak, to go before the mirror and become hostile with what one sees. To feel blameworthy, however, is to feel badly on account of one's blameworthiness. And these are different things. One may do one without doing the other. When one is blameworthy, one can sit with one's guilt. One can condemn the *qualities* of one's self that make one blameworthy – or one can condemn the *acts* one performed that make one blameworthy. One can, on that accord, *feel badly* about having those qualities, or about having performed those acts. But one can nevertheless refrain from being angry with oneself for having those qualities or for having performed those acts.

Phenomenologically, we suggest, the attitude of self-blame can be associated with what we might call *self-othering* – with *addressing* oneself in hostile tones second-personally:

*You fool! Yet again, you didn't email her back on time. This is just like you. Dropping the ball. Letting everyone down.*

Let us be clear. Our point here is not the familiar one that we shouldn't *go too hard on ourselves* or *self-flagellate* – that is, we shouldn't punish ourselves *disproportionately* (though of course we shouldn't). Our point is that precisely insofar as these attitudes are blame *at all* then they are inappropriate. You may be right: you didn't email her back on time. And yes, that is just like you. *We* blame you for that. That is *our* job. Feel badly about what you did. But leave the hostile

---

<sup>16</sup> Consider one further comparison (which we lack the space to fully articulate). Consider the phenomenology of finding someone attractive. There is a sort of sexual desire for the object of one's attraction. But now consider the phenomenology of *feeling attractive*. This is (in the right context) certainly a positive feeling. One puts on the dress, or suit (or...) and looks in mirror – and one feels attractive (feels "sexy"). But this is, of course, entirely different from feeling sexual desire for the person one judges to be sexy, viz., oneself. That is, when one puts on a special dress or suit and feels sexy, this is *not* to be conflated with feeling attraction to (or desire for) oneself. Importantly, this same phenomenon seems to hold when the valence is changed. Consider the phenomenology of finding someone *unattractive* (or sexually repellent). There is a sort of revulsion here, a sort of desire to *not* become intimate with the object that repels you. But now consider the phenomenology of *feeling unattractive* or sexually unappealing. One puts on one's drab, tired clothes and looks in the mirror, and feels down about how unappealing (or perhaps even repellent) one takes oneself to be. But this is entirely different from feeling sexual aversion for the person one judges to be sexually repellent, viz., oneself. That is, when one feels sexually unappealing, this is *not* to be conflated with feeling sexual aversion for oneself – a desire to not be intimate with oneself. Now our core claim. Just as one may feel unattractive or sexually unappealing, but not be sexually repelled by oneself, so one may feel guilt or shame, but not be angry with or hostile with oneself.

feelings to us.<sup>17</sup> Is that unintelligible? Well, why should it be? We can understand the following. You may be right: you worked so hard to raise all of that money for charity. We praise you for that. Feel good about what you did. But leave the praising to us. Don't go (even internally) praising yourself for being so good. Just feel good for being so good. Don't do the *additional* thing; don't praise yourself.<sup>18</sup>

## 8. Conclusion

In this paper, we have argued that it is always inappropriate to blame ourselves. We regard the arguments of this paper for this conclusion to be compelling – and certainly they point to difficult questions about the nature of blame, hypocrisy, and moral standing.<sup>19</sup> If we are right, however, then you shouldn't blame yourself – even if you don't accept them.

## References

- Bell, Macalester. 2013. "The Standing to Blame: A Critique," in Coates & Tognazzini 2013, pp. 263–281.
- Carlsson, Andreas. 2017. "Blameworthiness as Deserved Guilt," *The Journal of Ethics* 21: 89 – 115.
- Carlsson, Andreas. forthcoming. "Shame and Attributability," in Shoemaker, ed., *Oxford Studies in Agency and Responsibility* vol. 6.
- Clarke, Randolph. 2016. "Moral Responsibility, Guilt, and Retributivism," *The Journal of Ethics* 20: 121 – 137.
- Coates, D. Justin. 2016. "The Epistemic Norm of Blame," *Ethical Theory and Moral Practice* 19:

---

<sup>17</sup> How does the thesis of this paper bear on the question of how standing might be *regained*? We are officially neutral on this question, but one natural suggestion would be this: one regains standing to blame when one feels the appropriate amount of guilt for the appropriate amount of time, and has resolved to do better. In this case, one is – perhaps – no longer best described as a *violatee* of the relevant norm. One is instead best described as a *former violatee* of that norm. Again, however, a full discussion of this complex issue must lay beyond the scope of this paper. (We thank an anonymous referee for encouraging us to address this issue.)

<sup>18</sup> To be clear: we are not insisting that self-praise is indeed inappropriate; we are simply insisting that it is possible to feel praiseworthy without self-praising, and that it is similarly possible to feel guilty without self-blaming. It is only self-blame that we are arguing is inappropriate.

<sup>19</sup> For more on these questions, see Shoemaker (ms), which independently raises a paradox of the kind we develop in this paper; we learned of Shoemaker's draft at a late stage, and so we can't engage with it here – although we hope to do so in further work.

457–473.

Coates, D. Justin and Neal A. Tognazzini. 2013. *Blame: Its Nature and Norms*. New York: Oxford University Press.

Coates, Justin and Neal Tognazzini. 2018. “Blame,” in the Stanford Encyclopedia of Philosophy.

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

Dover, Daniela. 2019. “The Walk and the Talk,” *Philosophical Review* 128: 387 – 422.

Duggan, A.P. 2018. “Moral Responsibility as Guiltworthiness,” *Ethical Theory and Moral Practice* 21: 291 – 309.

Friedman, Marilyn. 2013. “How to blame people responsibly,” *Journal of Value Inquiry* 47: 271 – 284.

Fritz, Kyle G. and Daniel Miller. 2018. “Hypocrisy and the Standing to Blame,” *Pacific Philosophical Quarterly* 97: 1– 22.

Fritz, Kyle G. and Daniel Miller. 2019. “The Unique Badness of Hypocritical Blame,” *Ergo* 6: 545 – 569.

Gibbard, A. 1990. *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.

Herstein, Ori. 2016. “Understanding standing: permission to deflect reasons,” *Philosophical Studies*. Online First. <https://doi-org.ezproxy.is.ed.ac.uk/10.1007/s11098-016-0849-2>

Isserow, Jessica and Colin Klein. 2017. “Hypocrisy and Moral Authority,” *Journal of Ethics and Social Philosophy* 12: 191 – 222.

Lippert-Rasmussen, Kasper. 2021. “Why the moral equality account of the hypocrite’s lack of standing fails,” *Analysis* 80: 666 – 674.

- Perry, John. 1979. "The Problem of the Essential Indexical," *Noûs* 13: 3 – 21.
- Radzik, Linda. 2011. "On Minding Your Own Business: Differentiating Accountability Relations within the Moral Community," *Social Theory and Practice* 37: 574–598.
- Rivera-López, Eduardo. 2017. "The Fragility of our Moral Standing to Blame," *Ethical Perspectives* 24: 333 – 361.
- Rudy-Hiller, Fernando. 2018. "The Epistemic Condition for Moral Responsibility," *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/moral-responsibility-epistemic/>.
- Sartre, Jean-Paul. 1943/1993. *Being And Nothingness*. Hazel Barnes (trans.), New York: Washington Square Press.
- Seim, Maria. 2019. "The Standing to Blame and Meddling," *Teorema* 38: 7 – 26.
- Sher, George. 2009. *Who Knew? Responsibility without Awareness*. New York: Oxford University Press.
- Shoemaker, David. ms. "The Trials and Tribulations of Tom Brady: Self-Blame, Self-Talk, Self-Flagellation." [https://www.dropbox.com/s/hyxg98cqyofy5j/Self-Blame\(Jan2020Revision\).pdf](https://www.dropbox.com/s/hyxg98cqyofy5j/Self-Blame(Jan2020Revision).pdf)
- Tierney, Hannah. Forthcoming. "Hypercrisis and Standing to Self-Blame," *Analysis*.
- Todd, Patrick. 2019. "A Unified Account of the Moral Standing to Blame," *Noûs* 53: 347 – 374.
- Todd, Patrick. ms. "Let's see you do better," Philpapers. <https://philpapers.org/rec/TODLSY>
- Wallace, R. Jay. 2010. "Hypocrisy, Moral Address, and the Equal Standing of Persons," *Philosophy & Public Affairs* 38: 307– 341.