

# Blame and Fault: Toward a New Conative Theory of Blame

**László Bernáth**

Eötvös Loránd Research Network, Budapest  
Eötvös Loránd University, Budapest

DOI: 10.2478/disp-2020-0018

BIBLID [0873-626X (2020) 59; pp.371–94]

## **Abstract**

This paper outlines a new conative theory of blame. I argue that the best-known conative approaches to blame (Scanlon 1998, 2008, Sher 2006a) misrepresent the cognitive and dispositional components of blame. Section 1 argues, against Scanlon and Sher, that blaming involves the judgment that an act or state is the *fault of* the blamed. I also propose an alternative dispositional condition on which blaming only occurs if it matters to the blamer whether the blamed gets the punishment that she deserves. In Section 2, I discuss objections to judgment-based accounts of blame (that they cannot tell the difference between blaming and judging to be blameworthy, that they cannot explain why blame is often accompanied by emotion, and that they cannot make sense of irrational blame), and I argue that my proposal can handle all of them.

## **Keywords**

blame, conative theory, fault-judgments, irrational blame, punishment

## **Introduction**

If blaming someone means having some sort of belief or judgment about them, (Zimmerman 1988, Hieronymi 2004) then it is hard to explain why blaming someone is different from judging her blameworthy. (Coates and Tognazzini 2012). On the other hand, emotive theories of blame, (Wallace 1994, Pickard 2013) while not subject to this worry, have a difficult time explaining why blaming can occur without any emotional response. (Sher 2006a) Conative approaches

of blame were developed to address these challenges by retaining the best of both approaches. Roughly, conative theories say that blaming involves having certain beliefs as well as having certain dispositions or attitudes, the latter of which explain why blame is so often colored by emotion.

Although conative theories can make sense of the connection between doxastic and dispositional aspects of blame, I will argue that existing conative theories fail to do justice either to the judgmental or to dispositional components of blame. Before I propose an alternative, I will raise objections to Georg Sher's and Thomas Scanlon's celebrated conative theories of blame. Specifically, I will argue that these theories fail because blaming involves judging that something was the *fault* of the blamed. I am not going to give a full analysis of this concept; all I claim is that  $x$ 's blaming  $y$  for  $E$  involves one of the following three judgments:

- (i)  $y$  (non-causally or causally) produced  $E$  (at least, in part)
- (ii)  $y$  could have prevented  $E$  but failed to do so
- (iii)  $y$  could have made a difference to  $E$  but failed to do so.

In other words, if we blame someone for an event  $E$ , we think that  $E$  was (partly) her fault. Blaming involves *fault judgments* which, in turn, involve (minimally) the attribution of some sort of difference-making role along the lines of (i), (ii), or (iii). I will argue for this claim in detail, which will lead to a new conative account of blame. But let me turn to existing conative theories first.

## 1 Against existing conative theories of blame

### 1.1 The main problem with Scanlon's and Sher's accounts

Both Scanlon (1998, 2008) and Sher (2006a) maintain that one can blame someone even if one does not hold that the blamed one has difference-making role in the sense of (i)–(iii). My complaint against Scanlon and Sher, to be elaborated below, is that this is false, for the following reason:

- (1) If agent  $x$  blames agent  $y$  for  $E$ , then  $x$  judges that  $E$  was (partly)  $y$ 's fault.
- (2) If  $x$  judges that  $E$  was (partly)  $y$ 's fault, then  $x$  believes at least one of the following:
  - (i)  $y$  (non-causally or causally) produced  $E$  (at least, in part)
  - (ii)  $y$  could have prevented  $E$  but failed to do so
  - (iii)  $y$  could have made a difference to  $E$  but failed to do so
- (3) Therefore, blaming involves attributing difference-making role in the sense of (i)–(iii).

As far as I can tell, (3) contradicts most theories of blame on the market today, as most of them fail to take fault judgments—as defined by (2)—to be essential components of blame.<sup>1</sup> According to Miranda Fricker, the reason why theories of blame tend to miss the importance of fault judgments is that they are looking for the most comprehensive *analysis* of blame. (Fricker 2014) There are attitudes that are called “blame” in everyday moral practice but clearly do not involve a robust attribution of difference-making (for example, in

<sup>1</sup> The only exceptions I know of are Fricker (2014), Björnsson and Persson (2012), Björnsson (2017) and Coates (2016). However, the present paper differs from these theories in important ways. Coates claims (among other things) that blaming involves fault judgments in the sense that it implies the attribution of moral responsibility. Since the latter does not necessarily involve causal responsibility, Coates and I argue for different claims. Fricker analyses fault-judgments only, neglecting the other components of blame, and she does not analyse fault-judgments in terms of difference-making. Björnsson and Persson do offer an interpretation of “fault”, but I think their interpretation is wrong. The root of the problem is that, on Björnsson and Persson’s view,  $x$  can be judged blameworthy for an event (or, in the present terminology, event  $E$  can be judged to be  $x$ 's fault) only if one of  $x$ 's substandard attitudes can *explain* the event. I deny that the focus of blame judgments are attitudes (or other dispositions and states), as opposed to the agent herself. The attitudes of an agent and the agent herself are different entities (even if agents are bundles of states or dispositions), and I argue that fault-judgments attribute the fault *to the agent* and not to her attitudes. Hence, even if blameworthiness should be spelled out in terms of explanation, the focal point in the explanation should ultimately be the agent herself.

cases of pure causal blame), so a definition which fits all forms of blame must omit mentioning this difference-making role.

I disagree with Fricker's diagnosis. Most theories of blame focus on a particular type of blaming—in fact, most of them aim to justify “hostile” or “deep” blame. (Wolf 1994) This type of blame makes the blamer disposed to resent, condemn, avoid, reproach, scold, denounce, or punish the person blamed. The most well-known proponents of conative theories of blame, Sher and Scanlon, themselves focus on hostile blame. (Whenever I use “blame” from now on, it refers to hostile blame, unless otherwise indicated.)

As I see, fault judgments tend to be left out of the picture—even if the theory in question focuses on hostile blame—because it is hard to see how someone could have the relevant difference-making role with regard to their *emotions, beliefs, attitudes, character traits, or omissions*, which can very well be targets of blame. Thus, for many (including Scanlon and Sher), it seems plausible that fault judgments are not necessarily implicated in blaming.

I will return to this problem in section 1.3. Before that, I'd like to put some flesh on (1)–(3) by presenting counter-examples to Scanlon and Sher. My goal is to present cases where nobody is blamed (because no fault is attributed) but, under Scanlon's and Sher's accounts, we are forced to say that blaming occurs.

## 1.2 Counter-examples to Sher

Here is Sher's summary of his own view:

To blame someone, I have argued, is to have certain affective and behavioral dispositions, each of which can be traced to the combination of a belief that that person has acted badly or has a bad character and a desire that this not be the case. (Sher 2006a: 115)

More formally, Sher is committed to the following theory of blame (SH):

$x$  blames  $y$  for action  $A$  or character trait  $C$  iff

(SH1)  $x$  believes that  $A$  or  $C$  is morally bad [*cognitive condition*], and

(SH2)  $x$  desires that  $x$  refrain from performing  $A$  or that she lack

$C$  [*dispositional condition*], and

(SH3)  $x$  has the relevant affective and behavioural dispositions (dispositions that result in expressing resentment, hostile behavior, reproach, etc.) due to some combination of the beliefs and desires mentioned in (SH1) and (SH2).

It is worth noting that Sher is not a Kantian and so he denies that events or states are morally bad only if they are the agent's fault in the sense of (SH). (Sher 2006a, 61–70) According to Sher, what is morally relevant when it comes to the justification of hostile blame is that morality aims to guide actions and it applies to everyone who finds herself in the same morally relevant circumstances at any point in time.

I believe that (SH1) is too weak. And because of (SH1), (SH) classifies cases without blame as cases of blaming. Consider the following scenario:

**PERFECT HARD-DETERMINIST**

Dirk believes that free will is impossible and nothing is up to anyone. As a result, he trains himself to be a person who does not resent anyone. But this does not stop him from pointing out wrongdoers' faults to them. On such occasions, he expresses sadness, shock, and hurt because of the vices and wrongdoings he perceives. He has mastered his emotions to the extent that if someone does something morally wrong, Dirk is not angry with *her*, he is only angry at *the fact* that something morally bad happened. Even though he believes that wrongdoers typically grasp the moral reasons that tell against their conduct and vices, Dirk is not angry with them because he thinks it is not up to them in the relevant metaphysical sense whether they act in accordance with moral reasons. On his view, this is ultimately determined by external factors and is never the wrongdoer's fault. Dirk even trained himself to feel pity for moral transgressors on the grounds that they are simply unlucky. He wishes we could transform every morally imperfect person into a morally perfect saint without inflicting punishment.

It is not important whether we agree with Dirk about the impossibility of free will and the possibility of faultless wrongdoing (but see

arguments for such a view in Pereboom 2001 and Pereboom 2014). The point is, rather, that Dirk *blames* wrongdoers for their morally wrong actions according to (SH), even though Dirk attempts to do his best not to blame anyone—and he seems to be good at it. Dirk agrees with Sher that morality aims to guide actions and applies to all persons who find themselves in the same morally relevant circumstance at any point of time. Dirk's desire for a morally perfect world, combined with his moral beliefs, even *causes* him to criticize people for their vices and wrongdoings. But it would be bizarre to say that he is in any way hostile to them and *blames* them. Indeed, Dirk is angry *at* morally bad things in the same way as he is angry at natural disasters that harm people. But Dirk is not angry *with* anybody because he believes that moral vices and morally wrong actions are not the agents' fault but—so to speak—the world's. If he took wrongdoers to be faulty, he wouldn't just issue a negative assessment of their behavior, he would *blame* them without pity.

One could argue that the counter-example fails because Dirk does not blame anyone, according to (SH). Dirk should have affective and behavioral dispositions which typically result in expressing resentment, hostile behavior, etc., in order for (SH) to imply that he blames anyone.

I think this objection fails because Dirk has the requisite dispositions. His belief that the character traits and actions of a given agent are morally wrong, combined with his desire that these wrong actions and character traits be absent, gives rise to dispositions that *typically* cause blame-behavior (such as expressing moral disgust). They do not cause blame-behavior in Dirk's case only because Dirk has extreme self-control and holds a hard-determinist worldview.

One could respond by simply biting the bullet and claiming that Dirk, in fact, does blame wrongdoers. He suppresses his blame because of his peculiar views and extreme self-control. But I regard this move as dialectically problematic. This strategy can be used to 'defend' all kinds of theories that impose extremely weak conditions on blaming (for example, theories that conflate blaming with judging to be blameworthy). One just claims that the protagonist of the counter-example suppresses blame or hides it. We should only tolerate such defenses if we have a strong independent argument for the theory in question. I will now attempt to show that Sher does not

have such arguments.

In my view, Sher's main arguments for (SH) are based on thought experiments in which the blamed person (allegedly) had no causal role in acquiring the morally wrong characteristic trait. As a result judgements about fault are not necessary for blame. Consider the following scenario from Sher:

**BAD JOKE**

Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. Consequently, she is bewildered and a bit hurt when her rambling anecdote about a childless couple, a handicapped person, and a financial failure is not well received by an audience that includes a childless couple, a handicapped person, and a financial failure. (Sher 2006b: 290)

Sher thinks that we blame Ryland not only for offending her audience but also for being self-centered. If we think of the case more carefully, Sher claims, we will conclude that Ryland is probably not causally responsible for being self-centered and had no real chance to choose or change her character traits. We are not inclined to speculate about past or present opportunities to change, we simply blame Ryland for the bad joke and for being insensitive. According to Sher, this shows that blame does not necessarily involve robust judgments about the difference-making role of the blamed agent. That is, blaming someone does not entail believing that what they are blamed for was their fault.

I agree with Sher on two things: (i) Ryland can be blamed for both her action and her character and (ii) those who would blame her are typically not inclined to inquire into her causal or difference-making role. However, it does not follow from (i) and (ii) that blame does not involve judgments about the causal or difference-making role of the agent. It is possible that blame just involves by a defeasible intuitive assumption that the event or trait in question was the agent's fault.

In fact, this second interpretation seems more plausible than Sher's. In most cases when we consider an agent's character, we have little or no information about her past, and we automatically and often unconsciously tend to assume that her life history was relatively

normal and ordinary. And since such a past cannot by itself explain a morally significant character trait, we tend to assume that the agent was responsible for having acquired such traits. Thus, in most of the cases when we unreflectively focus our attention on negative character traits, we tend to assume that it is the agent's fault that she has that trait. We blame people for their character traits not because we conclude upon reflection that having that trait was probably the agent's fault but because we tend to assume (defeasibly) that agents have some difference-making roles with regard to the formation and presence of their morally significant character traits.

To support this hypothesis, let me modify *BAD JOKE* to bring out the defeasible assumption that I believe is in play. By assuming that Ryland's life history is unusual and her insensitivity is not her fault, the intuition that she is blameworthy becomes much less stable. This, in turn, suggests that we had the intuition in the case of *BAD JOKE* because we took Ryland to be responsible for her character traits.

### **Bad Joke 2**

Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. She is an only child who was excessively pampered. Her parents taught her that her own feelings and achievements are the only things that matter. Her unfortunate upbringing shaped Ryland's beliefs, desires, and attitudes, through no fault of her own, in such a way that it rarely if ever occurs to her that she might offend other people. Consequently, she is bewildered and a bit hurt when her rambling anecdote about a childless couple, a handicapped person, and a financial failure is not well received by an audience that includes a childless couple, a handicapped person, and a financial failure.

I claim that this modified story, which stipulates that it wasn't Ryland's fault that she became insensitive, does not trigger the intuition that Ryland is blameworthy. If this claim is correct, then Sher jumped to the wrong conclusion. His example does not show that we can blame agents for things that aren't their fault. It shows that we make defeasible assumptions about the difference-making role of agents in terms of their character traits, beliefs, emotions, attitudes and omissions.



Sher has a second argument for the claim that judgments about the agent's difference-making role are not necessary components of blaming. This argument appears in a discussion about control and blame, but it is relevant here as well:

The basic source of the difficulty is that many of the acts over which agents exercise control are themselves manifestations of traits over which they lack control. When someone performs such an act, his control over it may not go very deep. If it does not, then we may wonder whether it is any fairer to blame him for manifesting his bad trait than merely for having it. If there is no difference in fairness, then anyone who insists that it is not unfair to blame an agent for such an act will indeed have to reject F [which says that it is unfair to blame someone for something that was not up to her]. (Sher 2006a: 61)

Fault-judgments clearly qualify as judgments about control in Sher's sense, because being in control requires having the kind of influence that can result in something's being the agent's fault. So Sher's second argument can be reformulated in the following way: if blame involves fault-judgments, then blame could not be a fair response to the manifestations of those character traits the having of which is not the agent's fault. But we do blame agents for the manifestations of such traits. Therefore blame does not involve fault judgments.

There are two problems with this reasoning. First, it does not establish that there is no relevant difference between events caused by the agent and events not caused by her. Sher just asserts that causing *E* does not provide sufficiently 'deep' control over *E* if the agent does not have direct or indirect control over exercising her causal powers in the way she does. I am sympathetic to this idea but many philosophers are not. For instance, Björnsson and Persson (2012) believe that a person can be blameworthy for an event even if she is not blameworthy for the attitude that is the causal origin of the event.

Second, denying that blame involves fault-judgements solves the problem of blameworthiness for mental states and character traits not by discovering some feature of blame but by inventing one. Beyond the (in my view, solid) intuition that no-one can be blameworthy for something that is not her fault, there is considerable empirical evidence that blamers solve the problem raised by mental states and character traits in a different way. Based on the findings of Cusimano

and Goodwin (2019), it is reasonable to conclude that blamers do assume that the blamed person has control over her mental states and character traits, regardless of the fact that most philosophers regard this kind of control as implausible. This is why laypeople do not find it problematic to blame agents for attitudes, desires, beliefs, and so on. Cusimano and Goodwin put the moral of their research as follows:

The present results show that people do tend to see most mental states as moderately controllable, believing that others choose the beliefs, desires, emotions, and attitudes they have to a considerable extent, and could opt not to have such mental states if they chose to. [...] [J]udgments of control predict judgments of blame for mental states, just as they do for behaviors. (Cusimano & Goodwin 2019: 24.)

### 1.3 Counter-examples to Scanlon

Here is a brief summary of Scanlon's account of blame:

Briefly put, my proposal is this: to claim that a person is blameworthy for an action is to claim that the action shows something about the agent's attitudes toward others that impairs the relations that others can have with him or her. To blame a person is to judge him or her to be blameworthy and to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate. (Scanlon 2008, 128–129)

More formally, we have the following theory (SC):

If  $x$  blames  $y$  for  $E$ , then

(SC1)  $x$  takes  $E$  to show that  $y$  has attitudes that are harmful to personal relationships [*cognitive condition*],

and

(SC2)  $x$  modifies her relationship with  $y$  in a way appropriate in light of (SC1) [*dispositional condition*].

In Scanlon's view, it does not matter whether it is the agent's fault that she has the attitude that results in a blameworthy action. Rather, it is the *quality* of the attitude that matters. Further, the agent is no

less responsible if she does not identify with the attitude in question. Even if she would like to change, the attitude can still be attributed to her (Scanlon 2008: 194–5). For the same reason, the agent is not rendered blameless if the genesis of the problematic attitude is not due to her (Scanlon 2008: 197–8).

As earlier, one can construct counterexamples to this account. Let me outline one:

#### THE TEACHER

Frances is watching the execution of Robert Harris. Harris murdered two of Frances' students to take their car. But Frances knows that Harris had a deeply troubled, tragic life. His parents abused him, then he was taken to a foster home where he was also abused, for years and years. Frances thinks that Harris' attitude impaired his social skills to the fullest possible extent, leading inevitably to conflict and tragedy. Before the murders, Frances wanted Harris to change and she wished his life would take a turn for the better. In order to facilitate this change, Frances attempted to help him find motivation to participate in continuing education. Since the trial, however, she stopped caring whether Harris reforms himself on death row—she thinks he is too depraved to change in a morally significant way. And of course, she sees no point in helping his education. On the other hand, Francis continues to believe that Harris is the victim of his parents and the institutions that raised him. Such horrible communities, Frances thinks, are guaranteed to foster bad character traits.

Frances does not blame Harris for his attitudes and actions. She takes Harris to be a terrible person, but she does not blame him. One may wonder whether Frances is right to think that Harris was guaranteed to become a bad person and that he had no real chance of changing his attitudes. But even if Frances is wrong, it is still a fact that, because of her pessimistic view about the power of environment on character, she does not blame Harris in a hostile way but takes him to be a victim. Even though Frances has changed her attitude toward Harris (she is not interested in his moral and intellectual development any more), this is not due to the fact that she blames him. It is because she thinks that Harris is not capable of the relevant kind of

change any more.

One could point out that, according to standard English usage, Frances *can* be said to blame Harris in a sense. I agree, but I would add that she definitely does not blame him in a hostile way. She feels pity for him and she changes her attitude toward him in a calm, not to say calculated, manner. Her pity and her calm deliberation are not compatible with hostile blame. If someone blames others in a hostile manner, she is unable to feel pity for them at the same time; moreover, changes in her attitude toward them are not brought about by calm deliberation about potential consequences of the attitude change.

It is plausible to add that, in Frances' case, the assumption that the character trait in question is Harris' fault would lead to the judgment that Harris deserves blame. If Frances thought that Harris was causally responsible for acquiring and sustaining his sociopathic attitudes, she would blame him for having them. In this alternative scenario, the source of hostility against Harris would not be—even in part—a *practical* consideration regarding how one should handle such a terrible person; hostility would arise from the blame itself.

Note that Frances' case is especially challenging for Scanlon since she changes her attitude toward Harris in the very way that would be appropriate (according to Scanlon) if Harris were blameworthy, yet the attitude change happens *precisely because* Frances does not consider Harris the proper object of moral blame. After the murders, the most important change in Frances' attitude is coming to see Harris as an agent who falls outside the boundaries of the moral community. She stops caring about his fate and moral development not merely because he violated her trust and values but because she has stopped thinking of him as a moral being. Before the murders, she saw Harris as someone who is fighting against bad inclinations and often fails. After the murders, she regards him as a beast created by the moral failures of his environment. Confronted with his incomprehensible cruelty, she stops blaming him and puts the blame on society instead, even though Scanlon's view implies, in my view mistakenly, that she starts blaming Harris. The upshot is that failing to see a morally wrong action (no matter how terrible) as the fault of the wrongdoer entails failing to blame the wrongdoer, even if one's attitudes change in the appropriate way.

In response to this, Scanlon could agree that having the offending character trait or attitude must be up to the agent in some sense. He does think that some sort of control is necessary for blameworthiness—if we blame somebody for an action or state, we believe that it was up to her. The relevant sort of control is control over attitudes. Agents are *in principle* able to control their attitudes. As Scanlon explains:

If a person holds a certain judgment-sensitive attitude, then, because this attitude is in principle sensitive to and dependent upon his judgment, it is appropriate in a general sense to ask him to defend it or to disown it. By contrast, he could not properly be asked to defend (that is to say, give his reasons for) his height or eye color. If he is unable to see the force of some reason that counts against this attitude, this does not alter the fact that the attitude and the judgment that it is warranted are properly attributable to him. Any errors involved in these attitudes are also attributable to him, and he is therefore properly criticized for holding them. (Scanlon 1998: 289)

Scanlon talks about a type of control that comes from the dependence of rational attitudes on rational judgments. Our judgments about actions or character traits influence our attitudes toward them. In this sense, attitudes of an agent are up to the agent because they depend on rational judgments, which are attributable to the agent.

That said, agents are responsible for their attitudes only to very limited and restricted sense, on this account. The dependence of attitudes on rational judgments does not involve much more than counterfactual dependence: if judgments changed, the relevant attitudes would change. It is undeniable that such dependence is crucial for blaming, but this sort of dependence cannot by itself guarantee that the agent herself created the attitude and had the power to change it. And so the claim that blaming involves the judgment that something was the agent's fault goes far beyond Scanlon's contention that we take the agent responsible for her attitudes as a result of a counterfactual dependence between judgments and attitudes.

Furthermore, one can, I believe, argue that (SC) is too weak. Attributing an attitude, character trait, or action to an agent is not sufficient for blaming her for it. We only blame someone if we judge that her action, or her having the relevant attitude or trait, was her

fault in the sense that she was in some sense its primary source. In my view, *THE TEACHER* already shows that (SC) does not provide a correct analysis of blame as a result of this. Frances attributes socially harmful attitudes to Harris but she does not blame him for them. Let me introduce another thought experiment to drive this point home.

#### AMERICAN HISTORY

Ellie is a historian. Her research field is the microhistory of 19th century American rural communities. She acquires the diary of a 19th century man, John, who lived in a small village in the South. The diary clearly shows that John was heavily influenced by contemporary racist propaganda. However, John never actually witnessed the plight of slaves: he died at 20 and never travelled outside his village, which had no African American inhabitants. Based on the diary and other information, Ellie believes that John had no reliable evidence that racism was socially harmful. Moreover, Ellie thinks that it was ultimately not John's fault that he had such depraved views about African Americans. Everyone in his village was racist, along with all the literature that John had access to. Ellie also believes that if John had read *Uncle Tom's Cabin*, his attitudes toward African-Americans would have changed, because John was otherwise a sensible person who could have had a reasonable attitude to African-Americans but lacked access to appropriate evidence.

I claim that Ellie does not blame John for being a racist, even though she attributes a racist attitude to him. Moreover, Ellie thinks that John's attitude was judgement-sensitive and that he had in principle control over it: he would have changed his attitude if he had been confronted with relevant evidence. The main reason why Ellie does not blame John but rather feels some kind of pity toward him is that she believes he had no real chance to reconsider the awful views that had been inculcated in him by his environment. If one thinks that John was defenseless against indoctrination, then one cannot blame him for being a racist throughout his relatively short life but she will feel pity for John epistemic deprivation (especially if John has some importance for her). And so if one thinks, like Ellie, that it was not John's fault that he was a racist, one will not blame him even though one will attribute racist attitudes to him in the Scanlonian sense.

#### 1.4 A modified conative theory

Against Scanlon and Sher, I claim that blaming involves the judgment that the act or state that the agent is blamed for was her fault, specifically, that she had a significant role in bringing about or sustaining the state or event in question. I think we should acknowledge that blaming includes fault-judgments in this sense. Such judgments are not simply the preconditions of blame but are parts of the mental state of blaming. More formally, I'm proposing the following characterization of blame:

(B) If  $x$  blames  $y$  for  $E$ , then

(B1)  $x$  judges that  $E$  is the fault of  $y$ . [*cognitive condition*]

(B2)  $x$  judges that  $y$  violated some moral norm by causing  $E$ . [*cognitive condition*]

(B3)  $x$  has a specific type of attitude or disposition with regard to  $y$ 's having caused  $E$ . [*dispositional condition*]

(B1) and (B3) must be further specified. In the case of (B1), it is important that  $x$  does not simply believe that  $y$  had some role in producing (or failing to prevent)  $E$  but that  $y$  as an agent had some difference-making role in producing  $E$ . Since I think that having a difference-making role as an agent has to do with intentional action in one way or another, and the supposed epistemic states of the agent have a crucial role in attributing normative values to intentional actions, I believe fault judgments take into account the supposed epistemic states of the agent. Specifically, if  $x$  thinks that  $E$  is the fault of  $y$ , then  $x$  believes that  $y$  had difference-making role in producing or sustaining  $E$  and  $y$  was or should have been aware of her potential for doing some relevant action  $A$ , and of the normative value of action  $A$ . The details of this epistemic component of fault judgments go beyond the scope of this paper.

(B3) must be amended by specifying the attitude or disposition that is the characteristic component of blame. As indicated earlier, both Scanlon and Sher have a proposal about this component. Sher thinks it is the desire that  $E$  not be the case. For Scanlon, it is related to  $x$ 's changing her attitude to  $y$  in one of the following ways: (1)

having negative emotions toward *y* because of *E*, (2) withdrawing trust from *y*, (3) decreasing readiness to enter into special relations such as friendship with *y*, (4) decreasing willingness to help *y* with her projects, (5) being less likely to take pleasure in things going well for *y* person and to feel sad or regretful when they do not (Scanlon 2013: 105–6).

The following thought experiment is meant to show that something is wrong with Scanlon's and Sher's account of the dispositional component of blame. I claim that the protagonist of this story has the relevant dispositions but does not blame anyone.

#### THIEVES

Anna is an ingenious car thief. She thinks she does not have any other talent that would allow her to earn a decent living. One day, she sees George steal a car. She would prefer this event not to occur, because she does not like competition and she thinks stealing is wrong. Before the incident, she took George to be an innocuous, disabled homeless person. Now she realizes that it was a trick and George is a resourceful thief like herself. As a result, Anna does not trust George anymore and she feels no sympathy for him. However, she does not blame him. She thinks that George is just another thief. She agrees that most people would blame him for such actions, but she cannot, because she would do the same.

This story is coherent; it could actually be true. But it is at odds with Sher's and Scanlon's account of the dispositional component of blame. Anna wishes that George had not stolen any cars, and she changes her attitude toward him in a way that fits Scanlon's conditions. Yet Anna does not blame George. Or, at any rate, she does not believe she does. And I think she is right. I claim that Anna does not blame George because she does not care whether George gets what he deserves for stealing. Anna is a reflective person who recognizes the similarities between George and herself. She thinks it is OK for her to get away with stealing, and she understands that her belief rationally compels her to adopt a similar position with regard to George. As a result, she does not blame George, precisely because she does not care whether George gets the punishment he deserves. Note, further, that even if Anna cared whether George gets what he deserves, as long as she cared only because his being punished would



be advantageous for her, she still would not blame George. In contrast, if she cared because of the inherent moral worth of deserved punishment, then she would blame him. In the thought experiment, she is indifferent to the moral good of deserved punishment because she learned all too well to disregard this aspect of theft.

I would like to use this thought experiment to propose a new dispositional condition in place of (B3):

(B\*) If  $x$  blames  $y$  for  $E$ , then

(B1)  $x$  judges that  $E$  is the fault of  $y$ . [Cognitive condition]

(B2)  $x$  judges that  $y$  violated some moral norm by producing/sustaining/not preventing  $E$ . [Cognitive condition]

(B3\*) It matters to  $x$  for its own sake whether  $y$  gets the punishment that  $y$  deserves for  $E$ . [Dispositional condition]

This account of blame still a conative theory thanks to (B3\*), but it prominently features cognitive components too in the form of fault judgments. The dispositional condition is indispensable for at least two reasons. First, (B3\*) makes it easy to distinguish blaming from judging blameworthy, and it helps explain both detached and affective blame. I will turn to these issues in the final section of the paper. Second, (B3\*) allows one to mount an additional argument for (B\*) based on general considerations about the proper justification of hostile blame and its role in the system of morality.

As mentioned before, theories of blame typically seek to justify hostile blame. This is not surprising, since even the mildest expressions of blame put a serious burden on the blamed person, namely the burden to justify her deeds and redress her mistakes. Moreover, hostile blame arguably aims at some type of suffering as punishment.

This last trait of hostile blame may seem quite atavistic; hence the frequent attempts to explain it away. Sher, for instance, attempts to characterize the hostile aspect of blame without reference to the desire to cause some type of suffering. Instead, he argues that the hostile dispositions connected to blame come from a commitment to morality. I have argued that this commitment does not explain why the hostile attitude that is constitutive of blame is incompatible with coolness, or even pity, toward the blamed one. Scanlon, similarly to

Sher, seeks to deny that blame aims at some type of suffering. Rather, he proposes that suffering, if present, is only a side-effect of blame and the main point of blame is to change the relationship between the blamer and the blamed one. I argued that this approach fails to explain why blame is incompatible with coolness and pity even if such an attitude changes the relationship between the blamer and the blamed one in an appropriate way.

That said, connecting blame to the desire for proper punishment seems to clash with the fact that we often blame those we love. The reason why the tension is only apparent is that blame aims at a *specific kind* of suffering, namely, suffering that results from recognizing one's fault and its normative relevance. (Fricker 2014) In other words, the primary aim of hostile blame is to generate guilt. (Carlsson 2017) This is why fault judgments are necessary components of hostile blame: hostile blame seeks to make the blamed person (or, in some cases, people with specific connections to the blamed one) recognize their fault.

Even though I hold, similarly to Carlsson (2017), that the primary aim of blame is to cause guilt, the dispositional component of blame often goes beyond this. Blame arises from the desire that the blamed person get the punishment she deserves. If the fault is serious, the blamer will be not satisfied by mere guilt if the latter is not deemed sufficiently painful relative to the disvalue that the fault represents. So, in the case of serious faults, blame typically involves the desire for additional punishment because the blamer feels that the punishment should be both deserved and in some sense *proportional* to the disvalue of the fault. This explains why the desire for additional suffering emerges so naturally when we blame someone.<sup>2</sup>

In the next section, I turn to the objections to judgment-based theories of blame. My goal is to argue that (B\*) can handle all of them.

<sup>2</sup> I do not claim that the desire for pain over and above guilt is *justified*. I'm trying to analyze hostile blame. Its detailed justification is beyond the scope of this paper.

## 2 Problems for judgment-based theories of blame

### 2.1 The difference between blaming and judging blameworthy

The classic argument against judgment-based theories of blame is that there is a difference between blaming someone and judging that she is blameworthy. The conative theories outlined in §1 can all evade this objection. Although they take judgments to be essential for blaming, they are able to make a distinction between blaming and judging blameworthy (Coates & Tognazzini 2012, 200.). In a nutshell: If the cognitive conditions of blame are met but the dispositional one isn't not, then  $x$  judges  $y$  blameworthy without blaming her. If all conditions are met, then  $x$  blames  $y$  for  $E$ . For instance, in THIEVES, Anna judges George blameworthy but does not blame him.

### 2.2 Detached blame vs affective blame

In the case of detached (or cold) blame, blaming occurs without being accompanied by emotions such as anger, indignation, or resentment. In contrast, affective blame involves such emotional response. (Picard 2013) My proposed analysis of blame, (B\*), explains why these distinct forms of blame are both possible. The key is the dispositional condition, (B3\*). Whether blame is cold or affective depends on the sense in which it seems to the blamer that the blamed person gets what she deserves, more specifically, it depends on the way the blamer represents the blamed agent and their acts. If the agent or the act is represented as 'distant', then the blamer does not feel anger, indignation, or resentment. But if the agent or the act is represented as something 'close', we have a case of affective blame. Let me try to illuminate this distinction with the following thought experiment:

#### **Descendants**

Maddy is an American. She is interested in the history of Belgium because she is of Belgian descent. She reads about Leopold II, a Belgian monarch who perpetrated horrible crimes on the natives of the Congo Free State and robbed their land of its natural resources. According to some estimates, he was responsible for the

death of fifteen million people. Maddy tries to find out whether Leopold was ever punished for his crimes. She is disappointed to learn that he became immensely rich and died in peace and prosperity. Even though she is disappointed, Maddy does not feel indignation, resentment or anger. Uma feels differently. Living in the Democratic Republic of the Congo, she is curious what happened to her grandparents and great-grandparents. Her parents tell her that they all died as a result of forced labour and human rights abuses, thanks to Leopold II. While listening to her forebears' tragic story, Uma is overcome with anger. She is shocked that Leopold II received no punishment and lived a happy life.

Maddy only reads books about Leopold II's crimes, while Uma knows about them from the testimony of her parents. Books represent the events in question in a somewhat sanitized way, while first-hand (or, in this case, second-hand) testimony tends to be much more direct and upsetting. Representations of the former kind show us events from a distance, while representations of the latter kind show them as if they happened yesterday. This is the reason why Maddy feels no anger, indignation or resentment even though it matters to her a great deal whether Leopold II got what he deserved. If Leopold II's crimes were represented to Maddy the same way they are to Uma, Maddy would also feel anger. Emotions are integral parts of blame when we encounter 'direct' representations like Uma's parents'. Emotions are explained by the fact, in accordance with (B3\*), that the object of blame matters a great deal to the blamer, and hence the blamer is likely to feel strong emotions when the basis of blame is represented in a direct and evocative way.

### 2.3 Irrational blame

Hanna Pickard (2013) has argued that most theories of blame are wrong because they cannot explain irrational blame. For instance, if one believes that judgments are necessary parts of blame, then it seems impossible to blame someone without judging her to be blameworthy. Pickard presents the following example to show that this is, in fact, possible:

A couple come home from work exhausted and stressed, after a long,

hard day. One has failed to see to a minor household duty, which annoys and inconveniences the other, but for which, let us suppose, they have a reasonable excuse. Nonetheless, a critical remark is uttered, and a blazing row ensues, in which each takes their exhaustion and stress out on the other. The inconvenienced party blames their partner for the inconvenience, even though they know at heart they have a reasonable excuse. The blamed party blames their partner for the unfair blame, even though they too know that the background of stress and exhaustion might excuse their partner for failing to better manage their annoyance at the inconvenience. As the fight develops, both parties become increasingly incensed, self-righteous and adamant that they occupy the moral high-ground. But even in the heat of the argument, both know that in truth the blame they feel towards the other is undeserved, and not good for their relationship. The point is that this knowledge does not, in the moment, moderate their emotions and their behaviour, blame included. (Pickard 2013: 615)

In this story, both parties believe that the inconvenience is not the other person's fault, but they allegedly blame each other nonetheless. This contradicts my claim that fault judgments are necessary components of blame. Pickard solves the dilemma by arguing that blame is a complex emotion that consists of a first-order emotion (anger, indignation or resentment) and a second-order feeling that the blamer is entitled to have the first-order emotion in the situation at hand.

Although I agree that judgment-based theories of blame have not provided a plausible account of irrational blame so far, I do not accept Pickard's solution, because taking blame to be some kind of emotion (complex or otherwise) makes it impossible to explain detached (cold) blame. Emotion-based accounts of blame must—as Picard does—deny that detached blame exists or they must take it to be the same as judging someone blameworthy. But both options are implausible. In *DESCENDANTS*, Maddy blames Leopold II for his crimes. On the other hand, it is possible that someone who also knows history agrees with Maddy that Leopold II is blameworthy yet does not blame him, because she is indifferent to the characters in the story.

Still, I would like to borrow Pickard's strategy but deploy it in a reversed manner. Instead of suggesting that calling detached blame "blame" is just a manner of speaking (Pickard 2013: 614–5, 621), I claim that the couple in Pickard's story do not *blame* each

other (irrationally or otherwise), and the contrary is just a manner of speaking.<sup>3</sup>

My first point is this: it is rather weird to say things like “I know that *S* is not at fault/blameworthy but I cannot help blaming her”. It is a lot more natural to say: “I know that *S* is not at fault/blameworthy but I cannot help being angry with her”. If both of Pickard’s characters know that the other is not blameworthy, it would be more plausible to claim that they are angry with each other (in spite of the fact that neither blames the other) because of the stress and inconvenience that they caused each other. In other words, it seems plausible that their mental state is not a state of blaming but the state of being angry.

My second point is that it is not implausible at all that the couple make speech acts characteristic of blame in order to express their annoyance and anger while failing to blame each other. Uncontrolled anger results in uncontrolled behaviour, it tends to make people say the first thing that comes into their minds. We often make speech acts characteristic of blame to express our anger at inanimate objects, but it would be strange to say that we typically blame those objects in such cases. When I accuse my computer of sabotaging my work (“You do this on purpose, don’t you?!”), I’m not really trying to assert that it is consciously plotting against me. I just want to vent. If someone agreed with my accusation, I would be puzzled and I would explain that I didn’t mean it.

I certainly do not deny that irrational blame exists. But the irrationality of blame does not consist in *actual* inconsistency between blaming and judgments or beliefs. Irrational blame is irrational either because it is based on wrong reasons or because it is non-occurrently inconsistent with the agent’s doxastic attitudes.

## Conclusion

I have argued that blaming an agent for an event or state involves judging that the state or event in question is the agent’s fault. I first presented a few thought experiments to show that the best-known conative theories of blame (Scanlon 1998, 2008 and Sher 2006a) fail

<sup>3</sup> I owe this idea to Elia Zardini.

because they find blame in cases where there is none. I also claimed that the mistake results from denying that fault judgments (judgments that attribute a primary causal role to the blamed person) are necessary parts of blame. I then argued that we should modify the dispositional condition of blame as well, taking it to require that it matter to the blamer that the blamed person gets what she deserves.

In section 2, I addressed three important objections to judgment-based accounts of blame, and I tried to show that my proposal can meet all of them. In 2.1, I claimed that blaming and judging blameworthy come apart on the present view, thanks to the former's dispositional component. In 2.2, I explained why affective blame is so common despite the (alleged) fact that blame necessarily involves judgments but does not necessarily involve emotion. Again, one must take into account the dispositional element of blame: since it matters to  $x$  whether  $y$  gets the punishment she deserves, it is not surprising that  $x$  is likely to have strong emotions toward  $y$  or  $y$ 's acts. Whether blame is affective depends on the way  $x$  represents  $y$ . Finally, in 2.3, I argued that fault-judgments are indispensable for genuine blame and that in cases in where people seem to irrationally blame someone while consciously acknowledging that the person in question is not at fault, there is no genuine blame at all. Rather, in these cases, agents make speech acts characteristic of blaming to express and vent their anger.<sup>4</sup>

László Bernáth  
Institute of Philosophy  
Research Centre for the Humanities of the Eötvös Loránd Research  
Network  
Eötvös Loránd University  
Bernath.Laszlo@btk.mta.hu

<sup>4</sup> I would like to thank Dániel Kodaj and Judit Szalai for their suggestions and the stimulating exchanges we had on many occasions. I am also grateful to Elia Zardini and an anonymous referee for their comments on a previous version of this paper. The research was supported by János Bolyai Research Scholarship of the Hungarian Academy of Sciences (code: BO/00432/18/2), the OTKA (Hungarian Scientific Research Fund by the National Research Development and Innovation Office) Postdoctoral Excellence Programme (grant no. PD131998), the Higher Education Institutional Excellence Grant (Autonomous Vehicles, Automation, Normativity: Logical and Ethical Issues) at the Eötvös Loránd University, and two other OTKA research grants (grant no. K132911, K123839).

### References

- Björnsson, G. 2017. Explaining away epistemic skepticism about culpability. In *Oxford Studies in Agency and Responsibility*, ed. By D. Shoemaker. Oxford: Oxford University Press.
- Björnsson, G. and Persson, K. 2012. The explanatory component of moral responsibility. *Noûs* 46: 326–54.
- Carlsson, A.B. 2017. Blameworthiness as deserved guilt. *J Ethics* 21: 89–115. <https://doi.org/10.1007/s10892-016-9241-x>
- Coates, D. J. 2016. The epistemic norm of blame. *Ethical Theory and Moral Practice* 19: 457–73.
- Coates, D. J. and Tognazzini N. A. 2012. The nature and ethics of blame. *Philosophy Compass* 7: 197–207.
- Cusimano, C. & Goodwin, G. P. 2019. Lay beliefs about the controllability of everyday mental states. *Journal of Experimental Psychology: General* 148(10): 1701–32.
- Fricker, M. 2014. What's the point of blame: a paradigm based explanation. *Nous* 50: 1–19.
- Hieronymi, P. 2004. The force and fairness of blame. *Philosophical Perspectives* 18: 115–48.
- Pereboom, D. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Pereboom, D. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Pickard, H. 2013. Irrational blame. *Analysis* 73: 613–26.
- Scanlon, T. M. 2013. Giving desert its due. *Philosophical Explorations* 16: 101–16.
- Scanlon, T. M. 2008. *Moral Dimensions*. Cambridge: Harvard University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Sher, G. 2006a. *In Praise of Blame*. Oxford: Oxford University Press.
- Sher, G. 2006b. Out of control. *Ethics* 116: 285–301.
- Wallace, R. J. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Wolf, S. 1994. *Freedom Within Reason*. Oxford: Oxford University Press.
- Zimmerman, M. 1988. *An Essay on Moral Responsibility*. Totowa NJ: Rowman & Littlefield.