University of Wisconsin Milwaukee
UWM Digital Commons

Theses and Dissertations

May 2015

The Function of Blame

Nihar Nandan Nilekani
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Philosophy Commons

THE FUNCTION OF BLAME

by

Nihar Nilekani

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

in Philosophy

at

The University of Wisconsin-Milwaukee

May 2015

ABSTRACT

THE FUNCTION OF BLAME

by

Nihar Nilekani

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Professor Husi

This paper sets forth a theory of blame. Many currently proposed theories of blame fail to

capture all instances of blame. This motivates a pluralism about blame, suggesting that

there are many kinds of blame. These varieties of blame are nonetheless united in serving

a particular function in our interpersonal decision making. This function is to flag the

blamed behavior in such a way as to unsure that we factor it into future interactions with

the blamed person. Thus any feature of our psychology that generally fulfills this function

is a kind of blame. Since on this theory an instance of blame can count as such even if it

does not fulfill the function of flagging in that particular instance, as long as it is still is a

kind of blame, this theory can account for virtually all cases of blame.

TABLE OF CONTENTS

1. Introduction

We have some contradictory intuitions about blame. For instance, we think that blame and moral judgments are intimately linked, yet we recognize irrational or unjustified cases of blame that clash with our judgments. We think that blame plays an important role in our interpersonal relationships, yet we recognize cases of blame that are private or unexpressed. Many theories can account for some paradigm cases of blame, but not for all. In this paper I will present a theory of blame that I believe can capture all paradigm cases of blame, while excluding all that is clearly not blame. This theory is pluralistic about what blame is in psychological terms, but identifies a commonality between the various kinds of blame. It posits that all the different types of blame share a certain role in our decision making – that of 'flagging' or marking out blameworthy behavior to ensure that we will take it into account in our future decisions. This includes inspiring us to condemn people or to change our behavior towards them. It is critical to understand, however, that this theory identifies blame as having that function in the general case rather than identifying blame as being that which actually fulfills that function, thus allowing for cases where blame does *not* in fact fulfill that function. Sometimes there will be overriding reasons not to change our decision, or the blame is actually flagging something that we have no need to take into account in our future decisions.

I will first outline the aim of my theory of blame. Then, I discuss the difference between constitutive and functional theories of blame. I offer a brief overview of many of the competing theories of blame, and show how they all manage to cover some but not all

cases of blame. I shall argue that this should motivate a kind of pluralism about the features of our psychology that make up blame, and thus that any theory of blame that identifies it with a single specific feature will fail. I will then discuss Thomas Scanlon's view, in which blame is identified with an impairment of a relationship, in more detail, and show both how his theory can be modified to accommodate such pluralism and why it would still be unable to account for many cases of blame. I will then show how to create a functional theory of blame that can avoid this deficiency, by distinguishing between blame's general role and its role in a given instance. I will argue for my candidate for the function of blame, that of a 'flag' that ensures we take the blamed behavior into account in our decision making. Finally, I will explore the implications my theory has on certain longstanding issues in the literature about blame.

2.   Making Sense of Blame

The dialectic of blame is often confused. For instance, some discuss blame in the context of free will, and how we can justifiably blame others if determinism is true (Stawson, 1962), while others seem to be laying out the proper role of blame in our interpersonal relationships (Scanlon, 2013) (Wolf, 2011), yet others try to find what it is we are renouncing when we forgive someone (Hieronymi P. , 2001). In other words, blame is often discussed in the context of a larger agenda, and the simple question of what precise elements constitute blame is often confused with other questions such as when it is appropriate to blame, or whether blame is justified. In the interest of clarity, I shall try to focus only on the question of what blame actually is, sidelining other issues as much as possible.

Therefore, my theory of blame is descriptive, rather than prescriptive or normative. That is, I am interested in the question of what blame *is*, not what it should be. I will not distinguish between justified and unjustified blame, but regard both as instances of the phenomenon I am trying to understand. In other words, I am not trying to answer questions like "which instances of blame are proper and justified and which are not?" but questions such as "what is it that we pick out by the term 'blame'?" I believe that we need to know what blame is before we can discuss what it should be. Since the dialectic often rotates around putatively paradigm examples of blame, I take that the goal of such a theory of blame is to account for all such examples. Thus, I shall here try to outline a set of criteria that identifies as blame all paradigm examples of blame, while excluding all that is clearly not blame.

Over the course of this paper, I will divide the various theories of blame into two main kinds: what I call *constitutive theories* and *functional theories*. A constitutive theory of blame is a theory that identifies blame – that is, the commonality between all instances of blame – with some feature of our psychology. An emotion, a belief, a desire, a plan, or some combination of the above or of similar things. I call them constitutive theories because they hold that what is common to all instances of blame is what constitutes blame; that some psychological feature is present in every instance of blame. A functional theory of blame, by contrast, is one that holds that what is common to all instances of blame is the role that blame is serving in those instances – in our psychology, in our decision-making, in our social interactions, or, again, in some combination of the above. Such a theory might identify many different features of our psychology as being kinds of blame, as long as they fulfill that role.

I defend in this paper a functional theory of blame: I consider the most important aspect of blame to be the role it plays in our decision making, and crucially think that the function that blame serves is the only thing that unites different cases of blame. I still identify blame with certain features of our psychology, but I think that no single such feature is common to all instances of blame, and that what *makes* those features of our psychology count as kinds of blame is their function. Since many different features of our psychology may serve a role in our decision making, my theory can make sense of many different kinds of blame. Accommodating such pluralism at the level of psychological realization is one of the main motivations for my theory. The next few sections will try to motivate said pluralism, by showing why any theory that identifies blame with just one feature of our psychology will not cover all cases of blame.

3. Pluralism

Whenever we attempt to analyze a complex phenomenon, we are confronted by the possibility that we are actually talking about several related but distinct phenomena. Blame is certainly complex and multifaceted enough that pluralism appears to be a live possibility even before we consider any theory of blame in detail. Further, there are some good reasons for thinking that any theory of blame should be pluralist in some sense.

The first is that blame appears in several different contexts. The type of blame that occurs when we make an impersonal accounting of responsibility – when we *assign* blame – does not seem to be quite the same thing as what is happening when we blame someone in an interpersonal context. This is often said to be the difference between

judging someone to be blameworthy and *blaming* them (Coates & Tognazzini, 2012). To a certain extent this is a settled issue, and most of the theorists discussed in this essay are interested in interpersonal blame first and foremost, and so implicitly or explicitly agree that the word blame comes with many senses.

Even in an interpersonal context, however, there seem to be two major kinds of blame. The intuitive way to divide these up would be as 'cold' blame and 'hot' blame – the judgement that someone has wronged and the anger one feels at that knowledge. Right now I do nothing more than note the prima facie plausibility that these are separable. As we will see below, different theories of blame are better at accounting for one than the other, which supports the idea that this is a real distinction. But even before we get into those difficulties, I think that there is good reason to suspect that these are two different mental states.

A further, tentative reason to think that there are many kinds of blame is the evolutionary one. One must be careful when doing evolutionary psychology, as evidence is hard to come by and it is very easy to construct a just so story – a plausible explanation that fits the available data but which is only one of several possible explanations. Still, it seems likely that we developed some capacity to feel resentment *before* we developed either language or the capacity to make sophisticated moral judgements. Supporting this is that monkeys and apes seem to be capable of feeling resentment. One interesting example comes from *The Mental Life of Monkeys and Apes:* the researchers have recently changed the layout of the experiment used to test the problem solving abilities of a rhesus macaque. Yerkes describes:

But when on July 21 he was brought into the apparatus for the second series, he exhibited a wholly new form of behavior, for instead of attending diligently to the open doors and devoting his energies to trying to find the right box, he instead, after gazing at them for a few seconds, turned toward the experimenter and jumped for him savagely, throwing himself against the wire netting with great force. This was repeated a number of times during the first two or three trials, and it occurred less frequently later in the series. Since nothing unusual had happened outside of the experiment room, the suggested explanation of this sudden change in attitude and behavior is that the monkey resented and blamed on the experimenter the difficulty which he was having in obtaining food. From this time on until the end of my work, Sobke was always savage and both in and out of the apparatus he was constantly on the watch for an opportunity to spring upon me. […] Doctor Hamilton when told of this behavior, reported that several times monkeys have shown resentment toward him when they were having trouble in the experiment. (Yerkes, 1916)

More recent research has shown that resentment and punishing behavior is widespread in the animal kingdom, particularly in primates (de Waal F. , 2006).[1] This is a fascinating area of research, but the details are not germane to this current project. The takeaway is that there is good reason to think that much of what we think of as being related to blame – resentment, punishment, perhaps even forgiveness – predated certainly language and probably any notion of moral norms.

However, it is hard to shake the notion that blame *does* involve, in some way and at least sometimes, the transgression of moral norms. If so, this is a form of blame that developed much later, and we therefore have reason to think that it is a different thing than resentment. Exploring what these differences are, and what these different kinds of blame may be, are the subjects of the next few sections.

---

[1] Interestingly, there is also research into 'reconciliation', which appears to be related to forgiveness, if not quite the same thing. Chimpanzees, for instance, kiss and make up after fights. (de Waal & van Roosmalen, 1979)

4. Constitutive Theories of Blame: Cognitive

Different constitutive theories of blame identify it with different parts of our psychology. These theories can be further subdivided into some broad types.[2] The most straightforward theories of blame might be called cognitivist: theories that make sense of blame as a kind of judgment or evaluation of a person or an action. Others identify blame with our emotions and our responses to people. Still others might be called conative: they focus on our attitudes and expectations of other people, and identify blame with a change in those attitudes. I will discuss each of these in turn: In this section I will focus on cognitive theories of blame, and in the next I will discuss emotional ones. After that, I will examine one prominent theory that identifies blame with a change in our attitudes.

Cognitivist theories are a natural way of thinking about blame, as most of us think of blaming others as finding them morally culpable for some kind of wrongdoing. Some philosophers argue that blame is a negative evaluation that implies moral responsibility (Smart, 1961) or diminishes moral standing (Zimmerman, 1998), while others claim that it is a judgment that the blamed has failed to show proper regard for others (Hieronymi, 2004). What they have in common is that they identify blame as being a belief or judgement, or at least having a belief-like character (e.g. a mind-to-world direction of fit). But these theories face many challenges. Even though moral judgments do seem to be an important part of blame, we have reason to think that they are neither sufficient nor necessary for blame.

---

[2] My thanks to (Tognazzini & Coates, 2014) for this terminology.

For the case for insufficiency, consider the case of a criminal who admires a co-conspirator for skillful crime. The recognition of the act's wrongness may in fact be *part* of the admiration, and they are certainly holding the other morally responsible for the crime, but it would be strange to say that one is *blaming* the other. (Tognazzini & Coates, 2014) There seems to be more to blaming someone then simply judging that they are culpable for an action that the judger recognizes violates some moral norm.

Whether this example is a problem for cognitive theories seems to me to depend on the specifics of the theory in question. For instance, Hieronymi might say that the admiring criminal may recognize that his partner's moral culpability, but does not judge him to show ill will to anyone the criminal cares about. This reply, though, opens up the objection that it is the *caring* that matters, rather than the judgement. It is not whether or not the person judges another to have shown ill will that matter for determining that one person is blaming the other, but whether or not that first person cares about that fact.

This more general argument for insufficiency is similar to the moral motivation problem: cases of blame that consist merely of judgements that do not motivate us or cause us to change our behavior or stance towards the blamed person don't seem to count as being real instances of blame (this is often described as being the characteristic 'force' of blame. (Tognazzini & Coates, 2014) (Scanlon, 1986)). This is what the criminal example is meant to highlight: a case where judgement is present but blame is not. It is also why many current cognitive theories[3] identify blame as a judgement on how the action in question reflects on our moral or practical selves. That is, we do not simply hold

---

[3] Most of the ones mentioned above, particularly (Hieronymi, 2004) (Scanlon, 1986) (Zimmerman, 1998)

the person to be responsible for an action, but hold that their moral standing has diminished as a result, or that their action represents an ill will. Since people care about these things, they care about their judgements of blame. In short, to be a cognitivist about blame is to be an externalist about blame's characteristic force.

Is this a problem for cognitive theories? At the very least it means that blame is not *merely* making a judgement but requires a certain background interest in the terms of the judgement. If you do not possess the requisite care for other's moral standing or for whether one has been shown ill will (or however you wish to slice it), then you can make the relevant judgement but you are not thereby blaming. Thus the criminal example. But with that clarification, I do not think this presents a very great problem for cognitive theories. True, blame is not merely a judgement, but the core of what blame is on such theories is still a judgement. Being largely sympathetic to externalism in general, I think that the case of insufficiency can be addressed by such theories.

What seems to me a more pressing worry about cognitive theories of blame is that judgments of blameworthiness seem also un*necessary* for blaming. This is because we are capable of unjustified or irrational blame. I might blame the person who bought the last item in stock for depriving me of the same, but I don't hold them morally culpable for some sort of transgression. I don't judge that they have shown me ill will, or that their moral standing is diminished. Yet I blame them, just the same. I feel resentment towards them, and if I were to interact with them I might have to put some effort in to not have that blame alter my behavior. And certainly I could resist that irrational resentment from spilling over into actual action, but the very fact that I *do* have to put effort into resisting

it shows that it exists, and the effects it has on my thought processes seem to me to have all the hallmarks of blame.

We are capable of blaming people for what are clearly accidents and, even more so, what we *acknowledge* to be accidents (that a reasonable person who was not being careless might have made). A mother might blame the man who hit her child with his car for the injury even while acknowledging that he took all reasonable precautions. We are even capable of feeling resentment towards people against our better judgment – the recognition that blame is irrational is not sufficient to dissolve it. That last point – that blame is not always responsive to rationality – is to me the strongest one. If blame were simply a judgement, then the disavowal of that judgement would lead to there being no longer blame, but this does not always happen. All these examples point to there being more to blame than just our judgments. And if there are instances of blame where there are no judgements of any sort involved, then blame must at least sometimes consist of something else.

5. Constitutive Theories of Blame: Emotional

Other prominent theories of blame identify blame with our reactive emotions: with emotions such as resentment, indignation, guilt and anger.[4] Peter Strawson's

---

[4] There exist cognitivist theories of *emotions*, theories that identify our emotions with a suite of judgments. Were they correct, there would be no fundamental difference between cognitive and emotional theories of blame. It seems to me, however, that those who reject cognitivist theories of blame in favor of emotional theories of blame would reject cognitivist theories of emotions for similar reasons – they would think that such theories miss something very important about how our emotions are non-optional

"Freedom and Resentment" (Stawson, 1962) is among the most influential theories of blame, and is generally considered the foremost emotional theory of blame. As the title suggests, he identifies blame primarily with the emotion of resentment. There are other emotional theories of blame (Wolf, 2011) (Wallace, 2011); what they share is a commitment to thinking that to blame someone is to experience a certain kind or kinds of emotion. And certainly, it is very plausible that we do feel reactive emotions of this sort when we blame someone. But just as with cognitive theories, there are arguments that emotions are neither sufficient nor necessary.

For the case of insufficiency, it is hard to identify blame with *solely* an emotion when not every instance of that emotion is a case of blame. Not every instance of anger is an instance of blame, nor resentment, nor indignation. A beggar may resent the fortunes of another without blaming them for being responsible for their state of poverty. A woman may be indignant at being doubted, but is she blaming the questioner for their doubt? It seems odd to me to say so. Even if we *were* to grant that every instance of resentment or indignation is a case of blame (perhaps in the first instance the beggar is blaming the society for its inequality, and perhaps in the second the woman *is* blaming the doubter) not every case of blame is accompanied by either of those emotions. The emotional components of blame seem to have widely varying contents: besides those mentioned, there is sometimes grief, resignation, even occasionally admiration (at a well-executed prank, for instance). For many of these it is easy to think of instances of these emotions that are clearly not cases of blame, but what separates the two kinds of

---

(Stawson, 1962), for instance. In any case, such theories are controversial, and I will set them aside for the rest of this paper.

instances into blame and not-blame? If it is not the emotional content, then the theory must demarcate blame by more than just emotion, which would be an argument that emotions alone are insufficient for blame.

Further, if more than one emotion is associated with blame – and this does seem to be the case – then this demands an explanation for what is the unifying factor of these emotions. The answer is unlikely to be another emotion, but more plausibly is something else that produces the emotions. A theory that accounted for all the variant emotions we might feel when we blame someone by explaining the cause of them would be a theory of blame; but not, I think, an emotional theory of blame, since it would identify blame with that common cause more readily than with the emotions produced.

The argument that emotions are unnecessary for blame flows from instances of blame that do not involve emotional responses of the kind these theories envision. We seem to be capable of emotionless blame: that is, there seem to be paradigm cases of blame that consist solely of holding someone to be responsible in a disinterested manner, with no emotions involved. Further, we can blame friends and family without thereby holding negative emotions towards them (Sher G. , 2006). Emotional theorists have, broadly, two ways to deal with these cases.

The first is to argue that these are, despite appearances, not cases of blame but only of judging blameworthy (Wallace, 2011). And yet these cases still seem to possess blame's characteristic force. We might judge someone to have wronged us, and still *care* about that wrong without necessarily feeling strong emotions from it. These instances of seeming blame motivate us to action, cause us to change our evaluation of the blamed person and our behavior towards them, and otherwise seem more than merely judgments

of blameworthiness. At this point it becomes implausible to say that these are not cases of blame. And if there are cases of blame that are largely or entirely emotionless, then blame consists of more than just emotion.

Dismissing these as being cases of true blame seems unsatisfactory: while our theory may be forced to claim that some cases of seeming blame are not true instances of blame, it is still preferable to capture all such cases in the theory if it is possible to do so. Further, while this response might deal with the first case, that of disinterested blame, I think it is a weaker response for cases where we blame while still feeling positive emotions about the person blamed, as in some cases of us blaming friends and family.

The other tack is to claim that in these cases of blame we *are* in the corresponding emotional state but are not experiencing the affect usually associated with that emotion (Hurley & Macnamara, 2010). This is not wholly implausible: it is likely that emotions consist of more than just affect, and might still be present when the affect is not. However, as with the previous reply, this response struggles to deal with cases of friends and family where we might blame not while experiencing no affect but while experiencing *positive* affect towards the person blamed. Even if emotions are not solely affect, they *are* tied up with affect in so substantial a way that arguing that they are present while the felt affect is the entire opposite of what you would expect seems implausible. And there do seem to several cases of blame where the emotions associated with blame are not present (even if other emotions might be). These serve as examples that blame can exist without those emotions, and so that those emotions are unnecessary.

A brief overview cannot fully discuss every possible counter objection that the defenders of cognitive or emotional theories might present. However, while there is much

more that can be said about these various theories considered in isolation I think there is a striking fact about them considered together: though none seem to cover every instance, they each best account for the cases of blame the others fail to account for. Unjustified blame, or blame that is not responsive to rationality, is easy to understand in an emotional theory, since our emotions are often similarly irrational and not easily dismissed by the recognition that they are so. Similarly, disinterested or emotionless blame is easy to understand in terms of moral judgments. Further, they each seem to be getting at something important about blame. Strawson argues that reactive emotions play an important part, and are in fact partly constitutive of, our interpersonal relations, and that to lose them would be in a clear sense a diminishment; I think this is an important point. But the intuition that to blame someone is to hold them morally responsible is equally pressing. If we wish to account for both intuitions, we might do well to embrace both kinds of theories.

Thus, the problems both cognitivist and emotional theories of blame have with accounting for all cases motivates a pluralism about blame's constitutive elements. It is my contention that there are many different features of our psychology that are different kinds of blame, and that reactive emotions and the judgment of responsibility are among these features. In many instances of blame all of these features of blame are present, but in some cases only one might be, and we still count those cases as instances of blame. To embrace this pluralism, however, does not mean we have to give up the idea that there is something common to all instances of blame. It does mean that we either need to find that commonality somewhere other than blame constitutive features, or else find the common element *of* those features.

6. Constitutive theories of blame: Scanlon

In addition to theories that identify blame with judgments or emotions, some identify blame with our attitudes or intentions. In this section I will discuss a particular theory of this kind in some detail, and take away some lessons from its failures. The theory is that of Thomas M. Scanlon (Scanlon, Moral Dimensions, 2008) (Scanlon, Interpreting Blame, 2013), one of the most developed theories of blame. Scanlon argues that to blame someone requires not just that we judge someone blameworthy but that we modify our behavior with regards to that person in an appropriate way. Blame is the impairment of a relationship.

Scanlon is using the word 'relationship', however, in a somewhat non-standard sense: "a set of intentions and expectations about our actions and attitudes towards one another that are justified by certain facts about us" (2013). In other words, Scanlon's understanding of relationships can be said to be the sets of rules that govern our interpersonal interactions. It seems to me that such rules include conditional rules – 'if someone helps me, then thank them', for instance – and that thus this account of relationships is *dispositional.* In particular, you can have such an attitude without actually interacting with the person you have that attitude towards, which is important when considering the blame of strangers. An *impairment* of a relationship is an appropriate modification of our behavior with regard to the other person – i.e. appropriate with respect to the judgment of blameworthiness.

This is not a functional theory of blame. Scanlon identifies blame with a feature of our psychology – with a change in our conative attitudes, a modification of our intentions. The first worry with this is that though the modification of our dispositional relationships with others is associated with blame it is *caused* by blame rather than being what blame is. There is an important sense in which blaming others seems to be prior to any impairment to our relationships. If one asked me why I treat a former friend harshly, I might tell you it is *because* I blame them for some particular wrongdoing. The fact that this is an intelligible thing to say seems to imply that blame is at least separable from a change in our attitudes, since there is something causing such a change that is being recognized as blame.

There is, however, a functional theory of blame that can be constructed along similar lines as Scanlon's: a theory that identifies blame not with the change in attitude itself, but with having the function of changing our attitudes. One can see how such a theory is able to accommodate both this worry and the above mentioned pluralism about blame: resentment and other reactive emotions, moral judgments[5] might all be admitted as instances of blame, since it is plausible that they all cause us to modify our attitudes towards others in the appropriate way. A full theory would need to say more about what the appropriate way *is*, but we might be able to take cues from Scanlon in this regard. However both Scanlon's theory and my proposed functionalist modification have some significant flaws. In particular, as with many theories of blame there are many instances of what seem to be blame that it fails to cover.

---

[5]and various other claims about what blame is, such as a desire that the act had not been committed (Sher G. , 2006)

For instance, there seem to be cases of us blaming others without necessarily modifying our relationships at all. Susan Wolf (2011) gives the example of a short-tempered but nonetheless loving family that regularly blame each other for all sorts of things but never modify their relationships with each other thereby. She is especially critical of the idea that there is anything that might be characterized as an *impairment* to the relationship. Similarly, we might think of cases of private or unexpressed blame, where we still feel the blame without it modifying our behaviors or our dispositional relationship.

The other major worry about Scanlon's theory is that we seem to be capable of blaming those with whom we do not have *any* sort of relationship, even in his dispositional sense of the word. We blame total strangers, or politicians on the other side of the planet who we have no expectation of being in the same country as, let alone interacting with (Sher, 2013). Scanlon attempts to make sense of these cases by insisting that we all are in the "moral relationship" with each other (Scanlon, 2013). This is a relationship that holds between any two rational agents, involving the set of intentions and expectations they have for dealing with each other – which, remember, is Scanlon's idea of a relationship – which are governed by moral norms. In other words, it is our set of moral rules for interpersonal interactions. Scanlon uses this idea both to explain cases of blaming strangers as well as cases of the kind of emotionless blame discussed above. In the latter case, he argues that we might modify our moral relationships.[6]

---

[6] But he also thinks that we can modify our other relationships *without* impairing the moral relationship. For instance, "a Mafioso can be said to blame an associate for violating the code of *omertà*" (Scanlon, Interpreting Blame, 2013) despite violating no

While it is possible to understand what a judgment of blameworthiness is in the context of such a relationship (i.e. the judgment that someone has violated a moral norm), it is not entirely clear what an impairment of such a relationship constitutes – and it is with such impairments that Scanlon identifies blame. When I blame a stranger, have the moral norms I operate under when interacting with him changed in any way? It does not seem likely, and it is those norms which constitute the relationship: so it seems that it has not even been modified, let alone impaired.

Should one be convinced by this move, there are still cases of blame that even the idea of the moral relationship has trouble covering. For instance, we are capable of blaming historical figures. I think it is safe to say that most of us blame Hitler or Pol Pot or any other dictator for many of the crimes of their regimes – and rightly so! But we do not hold intentions or expectations towards people we know are dead[7], not even ones based solely on moral norms. We don't have intentions towards people who no longer exist.[8] Further, we don't just blame distant historical figures but more recent ones, and we blame dead people for present ills. On a personal level, for instance, a child might blame their deceased parents upon learning of their disinheritance even if they first learned of it when the will was read, long past the time for any sort of intentions or expectations.

---

moral norm. For him, any impairment of any relationship is a kind of blame. This can allow him to explain certain kinds of unjustified blame, though not all.

[7] Some people who believe in an afterlife might, but the point is that you don't *need* to believe in an afterlife to blame dead people

[8] Certainly one could tell a story about time travel, but I am quite certain any intentions I may have about Hitler once time travel is brought into the picture did not exist prior to the possibility being brought up, whereas I am equally certain that my blame *did* exist prior to that.

For a final and perhaps more controversial case, consider the issue of fictional characters. Prima facie, it seems that we do blame fictional characters. Many of the kinds of blame discussed above can be felt in relation to them. We feel both resentment and moral outrage towards fictional characters, for instance. At the very least, I think that a theory of blame that can account for the blame of fictional characters has an advantage over ones that do not, even if it is a small one, and Scanlon's theory cannot handle this case at all. We, of course, do not have intentional attitudes towards fictional characters.

It is my belief that a proper functional theory of blame can solve all of these problems of Scanlon's theory, while retaining many of its advantages. The key lies in the realization that the fact that a feature of our psychology serves a particular function in the general case does not mean that it will serve that function in every instance it is present.

7. Functional Theories of Blame: General Functions vs. Specific Cases

A functional theory of blame identifies blame not with any particular feature of our psychology but by its role. Since such theories claim that what makes something blame is the function it serves, not what it is constituted by, they can afford to be silent or expansive about the attitudes or combination of attitudes that make up blame. Thus, all functional theories of blame share a certain advantage: they can be very flexible about what blame is. For those of us motivated towards a pluralism about blame's constitutive features, this is very attractive. But if we are motivated towards such pluralism from a desire for our theory to accommodate all instances of blame, then there is still a worry. Functional theories can have the same problem as other theories: that of there being cases

that seem to be blame but which the theories do not identify as blame. The modified version of Scanlon's theory discussed above shares all the shortcomings of its parent theory when it comes to strangers or historical figures, for instance.

For another example, some philosophers identify the function of blame as a kind of moral protest (Hieronymi P. , 2001) (Smith, 2013). On this view, what we protest is the implicit moral claim of the one being blamed – namely, that it is acceptable to act in that way. In judging them blameworthy, we reject that implicit moral claim. Thus we change our attitudes in order to repudiate said claim (according to Smith), or feel resentment as a way of challenging it (according to Hieronymi). In seeing how these different authors account for different constitutive theories and unite them by their function, we can see the strengths of functional theories in action. These moral protest theories, however, have much of the same difficulties as cognitive theories of explaining unjustified and irrational blame, for in such cases it is unclear how there can be a moral claim one is protesting.

So, functional theories all struggle with cases of blame where their candidate function is clearly not being fulfilled. But in fact all functional theories have an answer to objections of this sort. The general function of blame need not be one that blame fulfills in every instance. Instead, one can argue that there are various features of our psychology that are kinds of blame. Each of these features shares a particular function, which is what makes it blame. But because they are features of our psychology, they will not always fulfill their function when they are present. Sometimes they will 'misfire', responding to perceived stimuli even though there is no actual reason for them to do so.

Consider our emotions, as an example. Fear is an emotion that pretty clearly serves a function in our decision making and reactions. Its purpose is to ensure that we are cautious in those situations that we should be cautious in. Yet there are times when we feel afraid when fear is unwarranted. We have all experienced irrational fear or jumped at shadows. But that sometimes we feel fear in circumstances where that fear serves no purpose or is even counterproductive does not invalidate the fact that the *general* function of fear is to promote appropriate caution.

When it comes to our judgments and evaluations, we usually operate under imperfect information or with limited time to examine all the available information. This means that our judgments must sometimes go wrong and fail to serve any function they might be serving. For a slightly more concrete example, consider our tendency to categorize things. Categorization is a kind of judgment and an evaluation in the light of that judgment, and it serves the function of streamlining our decision making. Yet, it does not always succeed in doing this, and sometimes we actually need to slow down our thought processes to deal with the inherent biases of reflexive categorization. Yet, even though it sometimes works in cross purposes to its function, it's still true in general that we classify things into categories because doing so lets us make decisions more quickly.

In addition, the features of our psychology, be they emotions or judgements or expectations, have a notable tendency to overcompensate. The paradigmatic example is pattern recognition: Humans are notoriously quick to find patterns where there are none, mistaking fuzzy pictures for meaningful images, seeing a face on the moon and shapes in clouds, and finding trends in random data. We developed in this way because having an overactive tendency to recognize patterns is better than having an underactive one; as it is

sometimes said, it is better to jump at a hundred shadows than to miss the one time it actually *was* a tiger. Similarly, many of our emotional responses to events occur even when they seem unwarranted from a more rational perspective, because the stimuli to which the emotional response is designed to react to was apparently present. Arguably it is the point and purpose of fiction to produce exactly those stimuli as would produce false positives in our emotional responses, and so move us to happiness, sadness, anger or, yes, blame, even though those responses do not, in such a context, serve their normal purpose. This, then, is an explanation of why a functional role theory of blame can succeed as a general explanation of blame even if not every instance of blame succeeds in fulfilling the function ascribed by that theory. If we can recognize the general function of something even though it does not always fulfill said function, it follows that we can recognize something *by* its general function as well, and so use the function of blame to demarcate between blame and not-blame.

This distinction is crucial for functional theories to make sense of blame. I believe that any functional theory should possess the following claim: what makes a feature of our psychology a kind of blame is the fact that it fulfills whatever role the theory claims is the function of blame, but what makes something an *instance* of blame is the presence of one of more of these features. In other words, it is two-tier theory of blame which first identifies instances of blame with several features of our psychology, and then identifies a commonality to all those features which makes them count as kinds of blame. For anything except the most restrictive (and thus implausible) candidate functions of blame, this leads us to a pluralism about blame's constitutive features. Since I think this is in

accord with our intuitions, I consider this a strength of such theories[9].

From all this, however, it does not follow that we can expect any function at all to be a candidate function of blame, waving aside any evidence otherwise by claiming that those instances don't count. It is highly implausible, for instance, to claim that the function of blame is to engender good feelings towards the person blamed – this is clearly false. For something to plausibly be the function of blame, it still must be the case that blame fulfills that function most of the time, or under usual conditions. Ideally, our theory can explain the occurrences of blame that do not fulfill their function. Thus a functional theory is not removed of the burden of arguing for the plausibility of its candidate function, nor the burden of explaining instances that its theory does not identify as blame. However, such a theory has at least some resources for making sense of cases where blame is not fulfilling the function, and it is these resources that this section was dedicated to showing.

8. My Theory of Blame

Thus far I have discussed functional theories in the abstract. The following section shall discuss what I think is the correct function of blame. I take cues from Scanlon, since I think the idea of blame modifying our dispositional relationships to one another to be on the right track. Rather than talking about relationships, however, I would

---

[9] I also think evolutionary theory motivates pluralism about blame. For instance, it is plausible that our reactive emotions evolved long before we had the capability to judge others according to moral norms, and when the latter evolved it was merely overlaid on the first rather than replacing it. This is speculative, however, and in any case there are other reasons to be pluralist about blame.

prefer to talk in terms of heuristics. We all possess a set of general rules or norms for interacting with others, both in the general case and in the case of specific persons, which govern our behavior with them. I believe that we engage in blaming behavior because doing so is useful to us in our decision making. Once we begin to examine things from this perspective, it is easy to imagine how this could be so: blame serves as a sort of warning or flag, a mental note identifying the person being blamed as one not to be trusted, and also identifying the behavior the blamed is engaged in as worthy of condemnation. Even if we do not communicate our blame to other people, we still take it into account for all future interactions with that person, and modify our attitudes and expectations towards them accordingly.

Of course, this alone is not sufficient to distinguish blame from other modifications to our dispositional attitudes towards other people. After all, as a matter of course we take previous interactions with and past actions of a person into account when deciding how to interact with them in the future. Blaming is merely an example of taking into account past behaviors, and what distinguishes it is its role as a *warning*, a marker of our disapproval. Such a marker often manifests psychologically as a reactive emotion such as resentment, but it doesn't necessarily have to. It also generally causes us to change our outward behavior towards the person being blamed, or to express our blame in some way, but again it doesn't have to if there is overriding reason for us not to do so. The important function of blame is to serve as a warning mechanism for future interactions.

While one can imagine a modified version of Scanlon's theory of blame that proceeds along similar lines, I prefer thinking about blame in these terms than in

Scanlonian terms because it is easier to make sense of private or unexpressed blame. After all, one can be said to legitimately take something into account in their decision making even if, ultimately, it does not actually cause them to *change* their decision – as long as it is a *factor* in their decision making. I think Scanlon errs in identifying blame with an *actual* modification to our relationship, instead of as a feature of our psychology that might *cause* us to modify our relationships. In my theory, unexpressed blame is simply when we blame others – i.e. take their actions into account in the appropriate way when we make further decisions about dealing with them – but that blame does not cause us to actually change our behavior, for whatever reason: because we have overriding reason not to, or for simple lack of opportunity if they are strangers or we otherwise do not interact with them.

One can see how many of the kinds of blame discussed above serve this function. When we hold others morally responsible for their actions we will of course take that into account when dealing further with them. Resentment and other reactive emotions similarly have a clear effect on our decision making. In the case of unjustified blame that I recognize as irrational, it seems to me that I need to exert a certain force of will in order for it *not* to affect my decisions – the fact that I need to do this I view as proof that blame of that kind does indeed affect our decision making and is therefore, on my theory, blame.

Further, by thinking of blame in terms of taking things into account in our decision making, we can see why we blame people we are never likely to interact with. The heuristics by which we operate are, by practical necessity, general rules rather than specific ones. Thus, our psychology tends to operate according to norms such as "blame

someone if they do X" even if that someone in question is on the other side of the planet, dead, or fictional. Many of the blaming features of our psychology, particularly our reactive emotions, are not sufficiently *discriminatory* to care that the person we are blaming is not someone we are likely to interact with.[10] And because I recognize this, and because my theory is that what makes something an instance of blame is the mere presence of one or more of these features, I am inclined to say that these *are* cases of blame: that we do blame strangers, historical figures and fictional characters, even if that blame will never actually affect our decision making in the appropriate way.[11]

To sum up my theory: there are many features of our psychology that we call 'blame' – these include our ascriptions of moral responsibility, our reactive emotions and changes in our intentions regarding others. What all these kinds of blame have in common is that they serve to flag the blamed behavior and/or person, ensuring that we will take it into account in our future decisions, whether the ones involving that person or persons behaving similarly. We say that someone is blaming another, however, based solely on the presence of these features regardless of whether the blamer will actually change their behavior or their decisions. In this way, the theory accounts for at least all

---

[10] Presumably, they evolved under conditions where us hearing about someone we are unlikely to interact with was very rare.

[11] It is worth noting through all of this that I am not claiming that this is the 'proper' or 'correct' role of blame. What role blame *should* be playing in our decision making, or in our interpersonal interactions, or our society's views about justice and punishment etc. etc. are all very interesting questions, but not ones I am trying to answer here. I am making an argument about the definition of 'blame', claiming that we recognize a particular feature of our psychology as a kind of blame if it fulfills the (general) role I have outlined.

the putative examples of blame given above. There may be other paradigm cases of blame that this theory cannot cover, but I cannot think of any.

That is not to say that my theory is unfalsifiable. If it could be argued that there are some kinds of blame that do not have the role in our decision making that I have outlined, it would be a good reason to think that my candidate function is not the correct one. If it can be shown that the various kinds of blame all serve different functions, it would be good reason to think that that functional theories are not the way to go if we wish for a unified accounting of blame. I see, however, no evidence that either of these is the case.

9. Implications

If my theory is correct, it has several implications about the way we think about blame and the questions of what situations it is appropriate to blame. I largely avoided these questions for this paper in order to focus on the question of what blame *is*. Having laid out my theory of blame, I devote this section to discussing those implications.

The first question is an old one: is blame justified? A lot of the initial impetus for emotional theories of blame was the worry that judgments of responsibility would be unjustifiable were determinism to be true. Addressing that worry is what *Freedom and Resentment* (Stawson, 1962) was about, hence the title. The idea was that if determinism were true, it would not be possible to assign just moral responsibility to individuals – and thus, to blame them – as they are not truly free. Strawson's answer was that blame consists of reactive attitudes, not assignments of moral responsibility, and so it was still

appropriate to blame people, as those reactive attitudes don't care about the truth of determinism. Rather, those reactive attitudes are partly constitutive of our personal relationships, and so need no further justification.

Like all emotional theories, this does little to deal with cases of unemotional blame. Since I do believe in cases of blame where reactive emotions are not present, Strawsonian explanations for how it is appropriate to blame others given the truth of determinism do not solve the problem for me. But if my theory of blame is true, and the function of blame is to mark the blamed as untrustworthy for future interaction, then even disinterested blame serves that function. And that function is a useful one even if determinism is true. If anything, noting someone's behavior in order to take it into account in future interactions is useful only if there is a certain level of reliability to the way people behave, and that reliability is one of the things compatibilist accounts of determinism emphasize when outlining why our freedom of action does not conflict with determinism.

This is not a new argument, in its essence. Compatibilists (or 'optimists' as Strawson refers to them in his paper) often justify our moral practices in general and blame in particular by pointing out that they regulate our behavior in desirable ways. However, these accounts traditionally focus on social sanctions, and how praise or blame affects other people's behavior, and struggle to account for private or unexpressed blame. What my theory offers is a function for blame that even private blame can fulfill, and a function which is useful even if determinism is true. But the idea is the same as most compatibilist accounts of blame – a justification for blame that is forward looking, where we blame people because doing so is useful for our future decision making.

Another idea that many theories of blame aim to capture is the feeling that blame plays an important role in our interpersonal relationships (though, of course, it is far from the only thing that does). My theory places as central the role of blame in our decision making when it comes to interacting with others, and so is no different in this regard. Further, I think that this account generalizes to blame's converse, praise, more easily than some other accounts. The idea is simple enough if you have grasped my theory thus far: praise refers to those facets of our psychology that have the general function of updating our decision making heuristics for interacting with the person praised (or for other people when it comes to things related to the person praised). The difference is that praise's function is not as warning but as commendation. Scanlon's theory might hash out praise as the appropriate modification of a relationship, but I think would have a harder time that with blame, for while it is plausible that an impairment of a relationship is blame, surely not every positive modification of one is praise.

Finally, I wish to talk a little about forgiveness. This is a large topic I can only brush on in this paper, but it is worth asking how my theory handles forgiveness. Straightforwardly, you forgive someone on my theory when that instance of blame no longer serves as a flag or warning in your future interactions. When stated like that, it doesn't quite sound right. Forgiveness is surely not just a matter of not letting the blame change your behavior, but something *weightier* than that: something more significant, and harder.

But that is to misunderstand my theory. Remember that one of the points I made earlier was that blame can serve as a marker even if it doesn't actually cause us to change our behavior, as long as we *take it into account* in our decision making. It may be that

overriding reasons, or through effort, it will not actually affect our final decision. But that's not forgiveness. It is forgiveness when that blame no longer factors into our decision making at all (or at least, no longer factors in as a warning or flag): when we don't even have the flag to remind us that the blamed incident occurs. And especially for that blame that manifests as resentment, this is not an easy thing. It is one thing to prevent blame from influencing ones decision making, but quite another to actually set it aside and truly forgive. And that has the ring of truth about it.

10. Conclusion

To summarize, I put forth in this paper a functional theory of blame that identifies many different kinds of blame by the role that they play in our decision making. The theory of blame presented in this paper can be thought of as belonging to a family of functional theories about blame. Even if one does not find my candidate function for blame convincing in particular, I believe there are very good reasons to think that a proper theory of blame must come from this family of views. The fact that there are various different kinds of blame means that a constitutive theory cannot make sense of all cases of blame. A functional theory can accommodate pluralism without giving up a common feature to all cases of blame. Such a theory can also make sense of instances of blame that do not fulfill the function by recognizing that a general function need not always be fulfilled. Although I cannot claim that my account is anything more than a step in the right direction, I am reasonably confident than any successful description of blame must share at least some features of this account.

Works Cited

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

de Waal, F. (2006). *Primates and Philosophers: How Morality Evolved.* Princeton: Princeton University Press.

de Waal, F. B., & van Roosmalen, A. (1979). Reconciliation and Consolation among Chimpanzees. *Behavioral Ecology and Sociobiology*, Vol. 5, No. 1 (1979), pp. 55-66.

Hieronymi. (2004). The Force and Fairness of Blame. *Philosophical Perspectives*, 18(1): 115–148.

Hieronymi, P. (2001). Articulating an Uncompromising Forgiveness. *Philosophy and Phenomenological Research*, 62: 529–555.

Scanlon, T. (1986). The Significance of Choice. *The Tanner Lectures on Human Values VIII.* Salt Lake City: University of Utah Press.

Scanlon, T. (2008). *Moral Dimensions.* Cambridge, MA: Harvard University Press.

Scanlon, T. (2013). Interpreting Blame. In D. J. ed. Coates, & N. A. Tognazzini, *Blame: Its Nature and Norms* (pp. 84-99). New York: Oxford University Press.

Sher. (2013). Wrongdoing and Relationships: The Problem of the Stranger. In D. J. Coates, & N. A. Tognazzini, *Blame: Its Nature and Norms* (pp. 49–65). New York: Oxford University Press.

Sher, G. (2006). *In Praise of Blame.* New York: Oxford University Press.

Smart, J. (1961). Free Will, Praise and Blame. *Mind*, 70(279): 291-306.

Smith, A. (2013). Moral Blame and Moral Protest. In D. J. Coates, & N. A. Tognazzini, *Blame: Its Nature and Norms* (pp. 27-48). New York: Oxford University Press.

Stawson, P. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48 1-25.

Tognazzini, N., & Coates, D. J. (2014, June 21). *Blame*. Retrieved from The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/sum2014/entries/blame/

Wallace, R. J. (2011). Dispassionate Opprobrium: On Blame and the Reactive Sentiments. In K. a. Wallace, *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon* (pp. 348–372). New York: Oxford University Press.

Wolf, S. (2011). Blame, Italian Style. In K. a. Wallace, *Reasons and Recognition: Essay on the Philosophy of T. M. Scanlon,* (pp. 332–347). New York: Oxford University Press.

Yerkes, R. M. (1916). *The Mental Life of Monkeys and Apes: A Study of Ideational Behavior.* Retreived from http://www.gutenberg.org/cache/epub/10843/pg10843.html.

Zimmerman, M. (1998). *An Essay on Moral Responsibility.* Totowa, NJ: Rowman & Littlefield.