Working out how blame works – Draft Paper

Anneli Jefferson, Cardiff University

Work in progress, please do not cite draft

Abstract

Recently, there have been two kinds of account of blame that aim to specify what blame is good for. One gives a functional explanation of blame, the other gives an instrumentalist moral justification of blame based on its effects. In the talk, I investigate the relationship between these two types of project. I argue that while in principle, an instrumentalist justification can draw on a functionalist explanation of blame to show that blame has morally desirable effects, there is no well-developed functionalist account of blame that can currently serve that purpose. Furthermore, a functionalist explanation need not paint blame in a morally favourable light. I then consider and reject an argument by Queloz which says that functionalist explanations of blame must necessarily come apart from consequentialist justification, and that the justification of blame needs to be non-consequentialist.

Intro and overview

When Marley blames Rhianna for cheating on her partner, telling her she should be ashamed of herself, he takes part in a very common human activity. We judge others to have done wrong and confront them angrily. But what is such blame good for? Philosophers have recently argued that blame fulfills the function of signaling commitment to moral norms (McGeer 2013, Shoemaker and Vargas 2021), or norm alignment and maintenance within a moral community. (Shoemaker 2024; Shoemaker and Vargas 2021; Fricker 2016).

Other philosophers give a justification of blame that appeals to its effects. They say that blame as a social practice of holding others to account in ways that are uncomfortable or painful to them serves the purpose and of making them adhere to and internalize moral reasons (Fricker 2016; Vargas 2013; Jefferson 2019; McGeer 2019). This kind of justification of blame sometimes also loosely talks about the function of blame.

Functional *explanations* of why people blame and *moral justifications* of blame that appeal to its effects seem to be natural bedfellows. However, this is a surprisingly controversial assumption. In this paper, I investigate the relationship between functional explanations of blame and consequentialist justifications of blame more closely.

One clarification is required before we embark on our exploration of functional accounts of blame. The term 'blame' covers a plethora of phenomena, and it is now commonplace to point out just how heterogenous the extension of 'blame' is. It can range from angry emotions, to dispassionate judgement, to expressing moral disapproval or disappointment to people's faces. We blame the living, the dead, we blame by confronting wrongdoers and in their absence. In this paper, I will be interested first and foremost in confrontational blame. That is because angry confrontational blame is the paradigmatic form of blame philosophers discuss, and it is also one that requires moral justification, as it is unpleasant for the person being blamed.

I should now lay my cards on the table: I am an instrumentalist about the justification of blame. Instrumentalism takes the consequences of blame on moral agency to be what justify blame, if anything does. What makes blame morally permissible or even required is that being subject to

it develops our sensitivity to moral reasons and increases the likelihood that we act on them. The instrumentalists holds that being confronted and blamed by others, seeing others being blamed or knowing that they will be blamed motivates people to act better. Blame also provokes negotiations of the norms we should adhere to and supports our sensitivity to moral norms and reasons. Instrumentalism can be read as a *justification* of blame in terms of its function, the practice of blame is justified in as far as it fulfils the function of building better beings, to borrow a phrase from Manuel Vargas (2013). So instrumentalism posits specific effects which then justify blame. It would therefore make sense that if blame has a function, this should align with its justification in some way. If the way blame works in the real world means that it is not conductive to building better beings, that endangers the instrumentalist justification.

Nevertheless, I have been somewhat skeptical about the recent proliferation of function talk in the literature on blame. A striking feature is that the notion of 'function' at work has remained somewhat underspecified. One goal of this paper is to establish whether talk about the function of blame can be made more precise. Only a more precise understanding of function will allow us to evaluate whether blame indeed has the function attributed and how its function relates to its justification.

A further motivation for digging deeper into this topic is that while functional explanation and an instrumentalist justification seem a good fit and some authors, most notably Miranda Fricker (2016), have endorsed it, some people reject this link. Some philosophers are functionalists about blame's explanation but not about its justification (Shoemaker 2024). Matthieu Queloz (2021) has gone one step further and argued that functional explanation and moral justification *have to* come apart. In this paper, I discuss what motivates his argument and why it is mistaken.

I will proceed as follows: First, I will give a very brief sketch of explanatory accounts of function and instrumental justification respectively and also outline a third kind of project, giving a functional *definition* of blame. I will then look at different ways in which 'function' might be understood in functional explanations of blame, either as an explicit aim, as an evolved function, or as a causal contributor to a large scale system. I will argue that the notion of function at play is frequently underdeveloped or only accounts for part of the phenomenon of blame, as in McGeer's evolutionary account.

I will defend the claim that instrumentalist accounts need blame to have the effect of norm-reinforcement, even if this effect need not be consciously aimed at or occur all the time. This means that accounts that see the role of blame as norm-enforcement are in principle well suited for a consequentialist justification. Sadly, none of the functional explanations discussed are currently sufficiently well developed to lend such support to instrumentalist justifications. I will then discuss the intriguing argument that in order for blame to have the role that it does, we need to justify it non-instrumentally. While unsuccessful, reflection on this argument shows something interesting about the relationship between the backward-looking phenomenology of individual acts of blame and the effects instrumentalists take to be crucial to justify blame. While an instrumentalist can (and should) allow that we don't always explicitly need to aim at moral improvement, this aim needs to be kept in mind as a constraint on appropriate blame.

Explanation, Definition and Justification

Accounts that talk about the function of blame can aim to explain or justify blame, but they might also want to provide a functional definition of blame. An explanatory account of blame

seeks to tell us why we blame, what blame is good for, what purpose it serves.[1] So, for example on Miranda Fricker's account, the function of blame is to reach a shared moral understanding between blamer and blamee. "Blame, then, functions as an invaluable interpersonal calibrator in moral agency." (Fricker 2016, 177) Shoemaker and Vargas claim that the function of blame is to provide costly signals of the norms we are committed to (Shoemaker and Vargas 2021; Shoemaker 2024). Broadly speaking, functional explanations concern the 'why' of blame in terms of effects. This glosses over significant differences between functional explanations that I will lay out in the next section, but it gives us an initial broad definition.

Separate but related to the explanatory project is the definitional project, according to which all and only those actions or emotions that have function x are instances of blame. Shawn Wang characterises the definitional project as follows: "Blame Functionalism. Blame is defined as whatever attitude or activity that serves a certain functional role (or certain functional roles)." (Wang 2024), p.608) The boundary between the definitional project and the explanatory one can be a bit fuzzy, especially as some functional accounts, for example Shoemaker and Vargas (2021), aim to identify a function that gets the pre-theoretic extension of our 'blame' concept right, trying to identify the unifying feature that is common to all types of blame. That sounds a lot like definition that aims to capture what falls under a concept.

However, a functional explanation need not also provide a definition that aims to capture the conditions of when x counts as an instance of blame. Some functional accounts explicitly focus only on paradigmatic instances of blame, such as communicative blame (Fricker 2016). Definitional projects will only concern us in as far as they overlap with explanatory ones.

Finally, a justification of blame can proceed in terms of its function, saying that what makes it morally justified to blame people is that blame normally has certain effects. At first glance, it would seem that the step from saying that blame has a certain function to arguing that that function is what justifies it is very small. The question 'What is blame good for?' and the question 'Why is it good to blame?' look as though they might well have the same answer.

Indeed, in some accounts, the two projects of explaining blame by its function and explaining why it is morally justified that we blame are very closely linked indeed. According to Fricker (2016), the function of blame is norm alignment between the person blaming and the person blamed. Provided that the norms that blamer and blamee converge on are good ones, this function looks like a promising candidate for justifying why it is a good thing that we blame. I take Fricker's justification to be broadly instrumentalist, we blame because it has desirable effects, i.e. agreement on moral norms. As the justifications of blame that are linked to its (explanatory) function I discuss are all instrumentalist and I myself am an instrumentalist about blame, I will henceforth talk about instrumentalist justifications for blame rather than functional justifications of blame. This has the added benefit of keeping explanatory and justificatory projects separate.

What notion of Function?

The notion of function at play in functional explanations of blame varies immensely, some authors will talk about blame functioning in a certain way without committing to a specific

---

[1] One exception is Randy Clarke's recent account of function, where 'functional' is to be read in analogy to functional role accounts of mental states in the philosophy of mind (Clarke MS).

account of function. While evolutionary, social or systems functions are mentioned, the notion of function at play seems fairly loose and combines several different theories of function. For example, Shoemaker's (2024) functional account of blame (and praise) subsumes a number of accounts normally seen as distinct in the philosophy of science, such as accounts of evolutionary function or causal role accounts of function.

I take it that whatever account of function is embraced, one common factor is that a practice, process, trait or artefact is normally explained by a subset of its effects or its intended effects. But this commonality does not go very deep. In evolutionary or selected effects accounts, the function of a trait or mechanism is ascribed as that effect or ability which was beneficial to survival and reproduction and was therefore passed on to further generations. So, it is an account that focuses on past effects, not on present ones. In the case of causal role or Cummins functions, the function of something is its causal contribution to a more large scale ability of the system it is part of (Cummins 1975). Famously, the heart contributes to blood circulation. Finally, artefact functions are those things an artefact does which that type of artefact was designed or created to do. For example, the function of a corkscrew is to open wine bottles. Opening bottles is what corkscrews are good for, that's what they were designed for and what we use them for.

It is important to note that even though effects are crucial to functional explanations, identifying a certain effect, such as norm reinforcement as the function of blame does not mean that blame always has to have that effect. It just needs to do so often enough for this to plausibly be the reason why we have such a practice.

Function and intended effect

One natural way of understanding function is in the way we do with artefact function, by locating the function in the intentional aims we pursue with a certain practice, in this case, blame. A natural aim for blame is norm enforcement or alignment, and this is indeed a commonly ascribed function. If the function was this kind of intended effect, this would make explanatory function and justification of blame coincide: The purpose of blame is norm enforcement, and enforcing norms is good, provided the norms are good. We can see that kind of function as intended effect in a practice that is closely related to blame, punishment, especially punishment of children. Schools and parents generally use punishment as a way of disincentivizing bad behaviour, and this is the explicit goal and justification of the practice.[2] However, explicitly and unabashedly consequentialist accounts of blame which take it to be consciously aiming to improve behaviour have long been rejected as being morally problematic (see (McGeer 2013) for a discussion of what she calls the anti-regulation concern) or as misdescribing our actual practice of blaming others. It was perhaps Strawson who most scathingly criticized the consequentialist for getting the phenomenology of blame completely wrong (Strawson 2008).

So, even accounts which take blame's function to explain *and* justify the practice often deny that we are consciously aiming at achieving a specific effect, such as moral improvement, in individual instances of blame. Even when philosophers concede that norm reinforcement may be the purpose of blame as a practice, it is not thought to be the main goal that individuals

---

[2] Things are more complicated in the case of criminal punishment, where there is normally both a retributive and a forward-looking element.

consciously and deliberately pursue. For example, Fricker claims that the aim of bringing the wrongdoer to acknowledge the moral significance of their action "need not of course be present as an intention in the psychology of the communicative blamer" (Fricker 2016, p. 173). Victoria McGeer, too, while instrumentalist in her justification of blame and functionalist in its explanation, does not think we need or should always be explicitly aiming at moral improvement (McGeer 2014, 2013). So, an effect can be the function, point or purpose of blame without the blamer consciously trying to bring about that effect.

This issue is further complicated by the fact that there are two ways in which the claim that the conscious purpose of blame is norm reinforcement might be questioned. We might question that people are consciously aiming at norm reinforcement in every token instance of blame. This seems undeniable, but it is compatible with the idea that in a reflective moment, we would assert that the reason we blame is to negotiate and reinforce the moral norms we are committed to. A different proposition would be to claim that the purpose of norm reinforcement is completely non-transparent to us. This would be the case if blamers explicitly disavowed any norm-reinforcement goals and for example said we should and do blame purely because it is fitting, deserved, or for some other reason unconnected to norm reinforcement. A functionalist story that still explains the function of blame as norm reinforcement would have to be very indirect indeed, but as we will see, some people do endorse this kind of theory.

While the function of blame need not be transparent to blamers in every instance of blame, what *is* required is that the effects of the functional account do in fact occur regularly, if not invariably. David Brink and Dana Nelkin (2022) criticise Fricker's functional account on the basis that blame does not in fact have the effect that her functional account ascribes to it. They cite empirical work on blaming children that purports to show that blame of children is counterproductive for norm-internalisation and fosters extrinsic, rather than moral motivation.[3] If blame does not engender remorse, moral alignment or better behaviour, this is clearly a problem for functional explanations of blame which state that this is the purpose or causal role of blame. Any account that says that we blame because it has the effect of norm reinforcement will be in trouble if it does not. Interestingly, this worry is more pronounced for explanatory accounts where norm reinforcement is not what we are consciously aiming at: If the function of blame is what we are consciously trying to do but it doesn't achieve what we want it to, then we have a problem as what we are trying to do is not working. However, an account that says we aren't consciously aiming at norm reinforcement but that it is its *de facto* social role will be completely baseless if it turns out that no norms are reinforced by blame.

To summarize, even philosophers who take norm alignment to be central both to the explanation and the justification of blame are not committed to this being what we consciously aim at when we blame, or at least not as their primary aim. In as far as we have an explanatory function at all here, it would be the typical effect. However, a typical effect on its own is insufficient for something being a function on most accounts. It needs to be an intended effect, or a contribution to a system's capacities. Shoemaker (2024, p. 24) gestures at such an account "functional explanations often articulate some causal contribution of a social system to the survival and thriving of society itself (sometimes referred to as a kind of organism)."

---

[3] See also (Moody and Nojoumian 2025) for a paper length defence of the view that blaming children does not have the effects that instrumentalist accounts of responsibility hope for.

While evocative, this is rather too underdeveloped to count as a causal role functional explanation. It requires us to say what counts as the survival of a society and to show the dependence on that practice for successful survival and or flourishing.

Evolutionary Accounts

We have seen that while a functional explanation trivially requires that blame has the effect the account ascribes to it at least some of the time, it does not require that this is the effect people are consciously pursuing when they blame. But if the function of blame is not determined by the explicit intention or goal we pursue when we blame, then what gives blame its function? One obvious place to look for functions are evolutionary or selected effects accounts, which provide a well-developed theory of function (Garson 2019). There is pretty wide spread agreement in the literature that punishment and the moral emotions have some kind of evolutionary function and that we penalize norm violations in order to police and reinforce moral and social norms (Martin and Cushman 2016; Levy 2016; McGeer 2013; Fischborn 2023; Mameli 2013).

Victoria McGeer sketches an evolutionary explanation of angry blame which starts out with the observation that 'many theorists argue that large-scale human cooperation could emerge only with the development of norms, supported and defended by cognitive/affective mechanisms specifically recruited to this task' (McGeer 2013, p.172) She then argues that moral anger and the tendency towards individually costly altruistic punishment -i.e. punishment of rule breaking which is costly to the person who punishes and has no immediate benefit for them – signals individuals commitment to moral norms and their enforcement. The presence of anger functions as a commitment device, something that lets the prospective wrongdoer know that their behaviour will be penalized and that they are unlikely to get away with it.[4]

The advantage of providing an evolutionary explanation of why we blame in the way we do is that it allows us to make anger part of the psychological machinery of blame even if it isn't clear that anger is always the best reaction in terms of getting a positive response. It also allows for blame to have a function that isn't necessarily transparent to those blaming. This is important because an account that closely links the function of blame to intended effects does not necessarily require anger. If we blame in order to improve others, it's not clear why we would also need to be angry. (And in fact, Smart (1961) advocates for a move away from condemnatory emotions.) However, if anger played a specific role in signalling commitment to the norms violated, it is clearer why angry blame might have been selected for and might be part of the blame package. More troublingly for some projects, it's also not clear that the evolutionary function of blame is moral improvement of those who have already done bad things, rather than credible threat to those tempted and exclusion of wrongdoers[5].

Relatedly, an evolutionary explanation of blame need not be one that lends itself to the moral justification of blame in terms of effects. The putative fact that angry blame may have been beneficial in terms of survival and reproduction and a precondition for developing more large scale societies tells us little about whether it is morally justified. The 'goals' of evolution and the

---

[4] Shoemaker and Vargas adopt this model in their costly signalling account of blame.

[5] As always with these evolutionary explanations, there is the worry about how to evidence them and that we might be dealing with a 'just so story'. I will shelve this concern, as I am primarily concerned with the structure of the explanation given here.

moral goals we pursue need not coincide. Compare for example the well-documented tendency to react very differently to in-group and out-group members. This may well have been evolutionarily advantageous in terms of the benefits to the survival of the in group. But on most theories of normative ethics, this does not make it morally justified.

McGeer's explanation of the function of blame is however not exclusively evolutionary. She argues that angry blame functions as a commitment signal, and that this elemental angry blame was subsequently recalibrated into more sophisticated blame reactions and emotions. These more sophisticated forms of blame attend to other people's mental states, specifically whether they intended harm and what their behaviour reveals about their character, and blamers also care about reactions going forward, demanding recognition of wrongdoing and remorse or restitution. These more sophisticated forms of blame are justified by the goals they (implicitly) pursue. "On the account I offer, the point and power of blame is to draw wrongdoers into a kind of exchange where they are perforce challenged to exercise their capacities as responsible agents, to reflect on what they have done, whether or not it is legitimate, and if it is not, to take responsibility for what they have done and for what they will do in future." (McGeer 2013, p. 180) Here, the gap between the functional explanation (how blame works) and its justification vanishes.

This means that in the end, McGeer's proposed evolutionary function of anger against norm-violations is just a stepping stone to a more encompassing account of how blame works. Evolutionary framing not-withstanding, the description of how blame works posits mechanisms in the here and now that are not given an evolutionary gloss. The one thing that the evolutionary account of blame does is to explain why anger is such a core element of blame and why it plays an important role in how blame functions even now. It is a signal that is uniquely suited to being taken seriously because we humans are attuned to paying attention to angry reactions (McGeer 2013, p.182).[6]

As in the case of Fricker's account, McGeer's is not clear on what notion of function it is employing when it talks about the function of blame as we practice it now, rather than the evolved function of anger. But, also like Fricker's, it suggests that when things go well, blame has the effect of leading to norm alignment and proposes a way in which this can happen. It is thus vulnerable to the same empirical concerns about whether blame actually has the effects on norm alignment that she proposes. And indeed the idea that angry blame is a good way instigating moral communication has recently been criticized on empirical grounds (Brandenburg 2025). Nevertheless, it does provide an interesting, evolutionary function for anger, a key reactive attitude in blame. This is useful for two reasons: first, it explains why anger or the threat that someone may react angrily to norm infringements can be useful in norm alignment. Second, it also explains why we might get anger even in cases where it is counterproductive.

Interim summary

While Fricker and McGeer have posited effects that help to explain why it makes sense for people to blame each other they haven't given a functional account that involves a worked out notion of function. McGeer's comes close by proposing an evolutionary function for anger as part of blame, but this is only part of a larger project that aims to explain the desirable effects of

---

[6] This idea is then taken up and further developed by Shoemaker and Vargas (2021) and Shoemaker (2024).

interpersonal blame. Maybe this merely reflects the fact that both the accounts I have discussed are consequentialist about the *justification* of blame. We might therefore think that functional accounts that are less concerned with justification might give a different functional explanation that doesn't also look like a candidate for an instrumentalist justification. As it happens, Shoemaker (2024) provides a functional account that is supposed to be purely explanatory, but also identifies norm maintenance as the core function of blame. I have not discussed his account in detail as it isn't a clear representative of any of the notions of function I have outlined.

Interestingly, all prominent functional accounts see some form or norm-reinforcement or alignment as the purpose of blame, whether or not they are instrumentalists in the justification of blame or not.

The relationship between functional explanation and justification of blame

Not everyone who gives a functional account of blame also intends this account to provide a functionalist moral justification of blame. For example, David Shoemaker (2024) identifies the function of blame and praise as norm maintenance. However, he also protests "Nothing about NMT (Norm Maintenance Theory) is, or is meant to be, normative. It is certainly not a justification for that system, nor is it a justification for any particular deployment of blame within that system." (Shoemaker 2024, p. 38)

To some extent, this is fair enough. We've already seen above that functional explanation and instrumental justification are in principle two different things. And yet, it's an odd move, especially if the norms in question are moral norms. In essence, it says that the purpose of doing x and what makes it permissible or morally right to do x are two completely separate matters.

Note that the gap between justification and function cannot merely consist in the fact that some norms are objectively not worth enforcing and that therefore, norm maintenance is sometimes morally undesirable. Every theory of blame comes with the inherent risk of getting blame wrong by getting things wrong morally. Assume Miss Watson blames Huckleberry Finn for helping her slave Jim escape because he has helped deprive her of her property. This is, to be fair to Miss Watson, not what happens in the book, but it is the reaction Huck fears and expects. In this scenario, her blame would not be morally justified on an instrumentalist account, because she gets the norms that are worth reinforcing wrong. If she insisted that what happened to her was theft, this would be a case of reinforcing existing moral norms, but at the same time it is also a case where she would be incorrectly categorising Huck's action as wrongful. So, blame would be still having the role of maintaining and reinforcing moral norms, it would fulfil its purpose on many functional explanations and doing what consequentialist justifications say blame should do, reinforce moral norms. But it would still not be morally justified. Moral mistakes pose a problem for any justification of individual instances of blame for specific actions, be it forward looking or backward looking, because people can simply be wrong in their moral norms and consequently in their moral assessment of specific actions.

Fricker, who gives both a functional explanation and justification of blame readily concedes that blame can lead to the acceptance of morally undesirable norms. Similarly, Jules Holroyd (2024) provides a functional account of praise according to which praise performs the social function of entrenching values and putting pressure on others to entrench these values. As she points

out, this means that praise can contribute to oppression by reinforcing problematic values. However, it also provides a potential justification for praise where the *correct* values are reinforced.

So why separate function and justification? One reason might be that one simply has pre-existing commitments when it comes to what justifies blame. Retributivists think blame is not justified by its effects at all. Let's say one believes that expressing moral condemnation is a good in and of itself, no matter its effects. If one also believes that as a matter of fact, blame has the function of norm enforcement, then one has to have a separate account of blame's explanatory function and of its justification.

Does a functional explanation preclude an instrumentalist justification?

We've seen that if you don't like instrumentalist justifications of blame, you can in principle still have a functional explanation of blame. But what if, like me, you do like instrumentalist justifications of blame? Instrumentalist justifications of blame require blame to work in certain predictable ways and have normatively desirable effects. It is hard to see how a practice of blaming could be justified by its effects without reliably generating desirable effects. This means that a functional explanation of blame in terms of norm reinforcement is grist to the instrumentalist's mill. However, it also means that instrumentalists will likely be somewhat biased when providing a functionalist explanation, as they will be inclined to look for what they want to find. In the interest of intellectual and research integrity, the instrumentalist is therefore better advised to proceed as follows: Say what kind of effects of blame would morally justify it, then see whether these effects can actually be observed. An alternative would be to appropriate an independent functional account as a building block for a consequentialist justification of blame.

I now want to turn to an account that explicitly says that functional explanation and the justification of blame must come apart. Matthieu Queloz (2021) argues that a functional explanation of blame must come apart from an instrumentalist justification of blame. Queloz sees norm alignment between the blamed and the blamer or the moral community more broadly as the role or function of blame. However, he argues that blame cannot fulfil this role unless it is self-effacing and that blame's justification *cannot* be instrumentalist. He claims that in order for blame to fulfil its function of improving the person blamed morally, that function cannot be transparent to the person blamed: 'the functionality of blame itself demands that it be justified by reasons other than the instrumental reasons that figure in its explanation'(Queloz 2021, 1362).

Why would he say something like that? According to Queloz, as we are all both blamers and blamed in the blame game, the blamer and the wrongdoer need to have the same understanding of what justifies blame to avoid incoherence. I cannot believe that you blame me purely because I deserve it if the only reason I myself blame others is to improve them. So the function should be opaque to those blaming, and they should instead blame for purely backward-looking reasons.

But why think that blame would be ineffective if the blamer thought they were being blamed in order to improve their behaviour or to make them adhere to certain norms? Queloz says that in order for blame to be effective, it needs to be perceived as morally justified (p. 1364). Let's concede this for the sake of argument. So the line of argument in Queloz is roughly this – the

reason why we blame, the explanation for blame is that it leads to norm alignment. However, it can only fulfil this function if norm alignment isn't what we use to justify blame. The justification of blame, Queloz says, needs to be in terms of the moral wrongness of the act that the person is blamed for.

As an objection to instrumentalist accounts of blame, this is rather odd, because no consequentialist says that the wrongness of the action can or should be irrelevant. These accounts explicitly take blame to be a *reaction* to morally bad or wrongful behaviour and ask what the morally correct reaction to such behaviour is. This means that the assessment of wrongfulness is already baked into the theory. I have made this point in past work (Jefferson 2019), as has Per Milam who notes that "blame is necessarily a response to a person's behavior" (Milam 2021, 446).

Here is another way of seeing this: the most common way of becoming a consequentialist about blame is by considering the appropriate reaction to wrongdoing and rejecting retributivism as morally unjustified. This could either be because one does not believe we have the free will necessary for retributivist desert (Pereboom 1995), or because one has the strong moral intuition that harming others can never be non-instrumentally justified. Blame that is painful for its recipient can therefore can only be justified by its effects. But this does not require blame to be anything other than a reaction to wrongdoing. There is therefore no basis for claiming consequentialist blame must appear morally unjustified to the one blamed.

It's also not clear why anyone should see the fact that they are being blamed because the blamer wants them to behave better as morally problematic. So, Queloz has not given us a reason to think functional explanation and instrumentalist justification must come apart.

All this does not mean that there aren't numerous reasons why blame should not be self-effacing in many instances. One is descriptive adequacy: as a matter of fact, we tend not to calculate the consequences every time we blame someone and our reactive attitudes are backward looking. Another is the usual problem consequentialism faces in calculating the consequences of an action (Lenman 2000) applied to the case of blame. In many cases, we simply won't know whether a specific blaming act will have good consequences. Furthermore, any kind of interpersonal practice that aims at stable results will do well to be dependable, as parents of young children know well. So good consequences are more likely when what triggers blame is an assessment of culpable wrongdoing, rather than an assessment of likely consequences of the blaming act. As we have seen already, most consequentialists explicitly make space for this in their theory by going more indirect. Vargas (2013), for example, argues that the practice as a whole is justified by its consequences, not individual tokens of blame. Both Fricker and McGeer, but also Barrett (2020) and others stress that one need not and should not be calculating individual consequences in every instance of blame.

Maybe a worry in the background here is that an instrumentalist justification of blame at the level of the practice generally but not for individual instances of blame is inherently unstable. It is indeed correct that even if individual instances of blame are normally backward looking, it would be wrong to think that the way we blame each other day to day should be completely insulated from the general justification of blame as a practice.

If, like me, you think that the practice of blame is justified by its effects, you should also think that we should stop blaming people if it did more harm than good. This means that our everyday

practices of angry blame are subject to revision if it turns out that they don't work as the instrumentalist hopes.

So, the claim that a functional explanation and an instrumentalist justification of blame are incompatible turns out to be mistaken. Furthermore, theories about the effects of blame but also the etiology of our blaming psychology matter to consequentialist justification. The very fact that instrumentalists justify blame by its role in norm alignment or because it builds better beings means that they have an obligation to make sure that blame works, and if it doesn't work, instrumentalists have a problem.

If there is good evidence that angry blame is counterproductive, this is a fact that instrumentalists need to take seriously. If blame does not work the way they think or hope it works, then they should become revisionist and do away with blame in as far as that is psychologically possible. This is also why McGeer's psychological hypothesis that getting rid of the angry parts of blame is not feasible and her assertions about the unique communicative role of anger matter. If anger as part of blame does something morally valuable, then it can stay because it is justified. If we can't get rid of angry blame because it is an immutable part of our moral psycholopgy, it will have to stay even if it is not morally desirable. However, if we would be better off without it and we *can* do away with it, then this is what we should do.

Conclusion

Where have we got to? An account of how blame works and an account of what justifies blame are two separate things. Functionalist explanations of blame pursue different projects, some try to explain its social role, some the evolutionary function of blame or aspects of blame, some posit effects of blame. Interestingly, nobody seems to be pursuing the most obvious functional explanation, whereby the function of blame is what people explicitly use it for. Given that functional explanations may be pursuing quite different projects, they would benefit from being precise in their use of the term 'function'.

Explanatory projects that try to identify the function of blame are important for instrumentalist justifications. Even if the functional explanation of blame ends up having a social role that is orthogonal to its moral justification, the effects of blame matter to its moral justification. So people who want to pursue a consequentialist justification of blame should pay attention to any account that aims to identify blame's effects. Furthermore, instrumentalists need to see whether blame has the effect of reinforcing, negotiating and aligning moral norms for an instrumentalist theory to get off the ground. Instrumentalists about blame need to find out how blame works, because if it doesn't work, we need to find some other way of holding each other to account.

References

Barrett, Jacob. 2020. "Optimism about Moral Responsibility." *Philosophers' Imprint* 20 (33):1-17.
Brandenburg, Daphne. 2025. "A Philosopher goes to the Therapist." *Journal of the American Philosophical Association* 11 (2):269-286. doi: 10.1017/apa.2024.18.
Brink, David, and Dana K. Nelkin. 2022. "The Nature and Significance of Blame." In *The Oxford Handbook of Moral Psychology*, edited by Manuel Vargas and John M. Doris, 177-196.

Cummins, Robert C. 1975. "Functional analysis." *Journal of Philosophy* 72 (November):741-64. doi: Doi 10.1086/289488.

.

Fischborn, Marcelo. 2023. "The Evolutionary Roots of Moral Responsibility." *Philosophy of Science* 90 (4):817-835. doi: 10.1017/psa.2023.60.

Fricker, Miranda. 2016. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50 (1):165-183. doi: https://doi.org/10.1111/nous.12067.

Garson, Justin. 2019. *What Biological Functions Are and Why They Matter*. Cambridge: Cambridge University Press.

Holroyd, Jules. 2024. "Proleptic praise: A social function analysis." *Noûs* 58 (4):905-926. doi: https://doi.org/10.1111/nous.12482.

Jefferson, Anneli. 2019. "Instrumentalism about Moral Responsibility Revisited." *The Philosophical Quarterly* 69 (276):555-573. doi: 10.1093/pq/pqy062.

Lenman, James. 2000. "Consequentialism and Cluelessness." *Philosophy & Public Affairs* 29 (4):342-370.

Levy, Neil. 2016. "Dissolving the Puzzle of Resultant Moral Luck." *Review of Philosophy and Psychology* 7 (1):127-139.

Mameli, Matteo. 2013. "Meat made us moral: a hypothesis on the nature and evolution of moral judgment." *Biology & Philosophy* 28 (6):903-931. doi: 10.1007/s10539-013-9401-3.

Martin, Justin W., and Fiery Cushman. 2016. "The Adaptive Logic of Moral Luck." In *A Companion to Experimental Philosophy*, 190-202. John Wiley & Sons, Ltd.

McGeer, Victoria. 2013. "Civilizing Blame." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal Tognazzini, 162-188. Oxford University Press.

McGeer, Victoria. 2014. "P.F.Strawson's Consequentialism." In *Oxford Studies in Agency and Responsibility: 'Freedom and Resentment' at 50*, edited by David Shoemaker and Neal Tognazzini. Oxford: Oxford University Press.

McGeer, Victoria. 2019. "Scaffolding agency: A proleptic account of the reactive attitudes." *European Journal of Philosophy* 27 (2):301-323. doi: https://doi.org/10.1111/ejop.12408.

Milam, Per-Erik. 2021. "Get Smart: Outcomes, Influence, and Responsibility." *The Monist* 104 (4):443-457. doi: 10.1093/monist/onab011.

Moody, Kristoffer, and Makan Nojoumian. 2025. "Blame: What is it good for?" *Philosophical Explorations* 28 (1):32-50. doi: 10.1080/13869795.2024.2405523.

Pereboom, Derk. 1995. "Determinism al Dente." *Noûs* 29 (1):21-45. doi: 10.2307/2215725.

Queloz, Matthieu. 2021. "The self-effacing functionality of blame." *Philosophical Studies* 178 (4):1361-1379. doi: 10.1007/s11098-020-01479-y.

Shoemaker, David. 2024. *The Architecture of Blame and Praise - An Interdisciplinary Investigation*: Oxford University Press.

Shoemaker, David, and Manuel Vargas. 2021. "Moral torch fishing: A signaling theory of blame." *Noûs* 55 (3):581-602. doi: https://doi.org/10.1111/nous.12316.

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

Strawson, Peter F. 2008. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1-28. Abingdon: Routledge.

Vargas, Manuel. 2013. *Building Better Beings: A Theory of Moral Responsibility*: Oxford University Press.

Wang, Shawn Tinghao. 2024. "Rethinking Functionalist Accounts of Blame." *The Journal of Ethics* 28 (4):607-623. doi: 10.1007/s10892-023-09468-z.