Philosophers' Imprint

VOLUME 25, NO. 5
JULY 2025

Don't Burst My Blame Bubble

Hannah Tierney

University of California, Davis

DOI: 10.3998/phimp.3761

0. Introduction

Blame abounds in our everyday lives, perhaps no more so than on social media. With the rise of social networking platforms, we now have access to more information about others' blameworthy behaviour and larger audiences to whom we can express our blame. But these audiences, while large, are not typically diverse. Just as we tend to gather and share information within online social networks made up of like-minded individuals, much of the moral criticism found on the internet is expressed within groups of agents with similar values and worldviews. Like these epistemic practices, the blaming practices found on social media have also received criticism. Many argue that the blame expressed on the internet is unfitting, excessive, and counterproductive. What accounts for the perniciousness of online blame? And what should be done to address it?

To better understand what has gone wrong with blame on social media, we need to understand the structures in which it takes place. I develop an account of these structures, which I call 'blame bubbles', based on social epistemological accounts of echo chambers (Lackey 2021; Nguyen 2020). Given the similarities in how we blame and how we gather and share information on social media, one might argue that the structural features of these environments account for the toxicity of online blaming practices, just as many have argued that these structural features are responsible for the problematic nature of online epistemic practices (Sunstein 2017; Nguyen 2020). However, I argue that while blame bubbles can certainly be home to objectionable blaming practices, the structures of these environments are not in and of themselves morally objectionable. In fact, these structural features play an important, and potentially unique, role in holding perpetrators accountable and addressing wrongs done to victims. In my view, the problem with blame bubbles is not attributable to bad structures, or bad actors, but rather to bad communication. I conclude by reflecting on how to improve the communicative success of blame within blame bubbles without undermining these structures' important moral functions.

1. What is a blame bubble?

Before I explore the structures in which we express blame on social media, it will be helpful to first say a bit about blame. In this paper, I adopt a functionalist account of blame, according to which what blame is is determined by what blame does. Most theorists who defend functionalist views of blame argue that blame functions, at least paradigmatically, to communicate something to others (Kogelmann & Wallace 2018; Macnamara 2015; McGeer 2013; McKenna 2012; Fricker 2016; Shoemaker & Vargas 2021; Smith 2013). In the view of many of these theorists, blame communicates two kinds of information: information about the perpetrator and information about the victim. For example, Angela Smith argues that blame has two aims: 'to *register* the fact that the person wronged didn't deserve such treatment by *challenging* the moral claim implicit in the wrongdoer's action; second, to prompt moral recognition and acknowledgement of this fact on the part of the wrongdoer and/or others in the moral community' (2013: 43, emphasis in original). And Brian Kogelmann and Robert Wallace argue that expressions of blame can send an *accountability signal,* which communicates that the blamer intends to hold perpetrators accountable for their ill will, and a *quality signal*, which indicates that the blamer has goodwill towards the victims (as well as the wider moral community) (2018: 378). According to these views, when we blame, we not only communicate moral disapproval of the targets of our blame but also an acknowledgement of the moral status of their victims.

We can also express blame's messages to a range of individuals. If an agent wronged me in some way, and I blame them for doing so, I could directly express my blame to that agent. But I could also express my blame of the wrongdoer to other members of the moral community. Likewise, if we blame an agent for wronging someone else, we could express our blame directly to that agent as well as to the moral community at large. On social media, the intended audience of our blame is typically some subset of the moral community, perhaps in addition to the perpetrator. If an agent takes to social media to express their outrage about a particular politician's morally corrupt behaviour, they are not just, or even primarily, communicating this negative moral reaction to the politician in question. Rather, they are expressing their blame of the politician to other members of the moral community, specifically to those members who are in their social network. And, as is often noted by social scientists, our social networks are largely made up of individuals who are similar to us. Just as we tend to be exposed to theories and ideas with which we agree on social media (An, Quercia & Crowcroft 2014; Saez-Trumper, Catillo & Lalmas 2013), our own posts also tend to reach those who share many of our values and beliefs. The same is true for our expressions of blame.

Of course, there is nothing necessarily problematic about expressing blame within a group that shares a similar moral outlook. It can be morally neutral, and even beneficial, to share one's blame with those who agree with you, just as it can be epistemically permissible and perhaps fruitful to present a theory to a receptive audience. But our social networks on the internet are not just made up of people who happen to share our views. Rather, they are constructed, in part by these platforms' algorithms and those who can manipulate them, to amplify voices with which we agree and to dampen and filter out voices with which we don't. This has led many to argue that we gather and share most of our information on social media within *echo chambers*. According to Jennifer Lackey, echo chambers have three features:

> 1. There is an opinion that is repeated and reinforced, thereby amplifying it, often through re-sharing.
>
> 2. This occurs in an enclosed system or "chamber", such as a social network, allowing the opinion to "echo".
>
> 3. Dissenting voices are either absent or drowned out. (2021: 207)

Blame bubbles are a kind of echo chamber, but what is shared within them is not solely or primarily news stories or descriptive claims.

Rather, blame bubbles are filled with expressions of blame. While blame bubbles can form in many environments, they are particularly well suited to social media, where expressions of blame can echo through large and geographically dispersed social networks that are enclosed and resistant to encroachment from objecting voices.

Take, for example, the blame bubble that formed in response to an offensive tweet from Justine Sacco, which the journalist Jon Ronson documents in his book *So You've Been Publicly Shamed* (2015). In 2013, before boarding a flight to Cape Town, Sacco posted: 'Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!' to X (formerly Twitter). While Sacco was in the air, other users began to criticise her post, and a blame bubble formed. Eventually, over 100,000 posts were shared (Ronson 2015: 70), most of which expressed at least some degree of blame of Sacco. While the blame bubble originated on social media, it expanded to include other platforms and non-written expressions of blame.[1] Sacco was fired and her family members, whom she was visiting in Cape Town, blamed her in person.

If one followed the hashtag #hasjustinelandedyet, one would see thousands of tweets blaming Sacco, but very few dissenting voices. And those who did object to the severity of the blame being directed at Sacco were themselves blamed for doing so. The journalist Helen Lewis recounts her attempt to intervene on behalf of Sacco:

> Tentatively, at the time, I tried to suggest that perhaps the tweet wasn't that bad: certainly not bad enough to warrant the rape and death threats that were flooding in. Fellow tweeters began to argue that I was being a typical white, middle-class feminist, sticking up for a powerful PR executive and ignoring the voices of wronged people of color. So I did something I have been ashamed of ever since. I shut up and looked on as Justine Sacco's life got torn apart. (Lewis 2015)

In this case, dissenting voices were not simply left out, they were actively blamed. This is a common feature of blame bubbles. Often, those who object to the blame that is expressed in blame bubbles, and even those who simply fail to join, are morally criticised by those within the bubble. One might think that this sets blame bubbles apart from echo chambers. After all, on Lackey's account, those who disagree with the views shared within echo chambers are simply drowned out or ignored. But there are theorists who argue that echo chambers have a similar feature. For example, according to Thi Nguyen (2020), dissenting voices aren't just absent from echo chambers, they are systematically excluded and mistrusted:[2]

> I use "echo chamber" to mean *an epistemic community which creates a significant disparity in trust between members and non-members*. This disparity is created by *excluding non-members through epistemic discrediting*, while simultaneously *amplifying insider members' epistemic credential*. Finally, echo chambers are such that *in which general agreement with some core set of beliefs is a pre-requisite for membership*, where those core beliefs *include beliefs that support that disparity in trust*. (2020: 10, emphasis in original)

On this view, non-members of echo chambers are epistemically discredited. Similarly, non-members of blame bubbles are morally

1. One might worry that this blame bubble has more porous boundaries than that of an echo chamber, because anyone with an X account could witness the derision directed at Sacco and choose to join in. But this is true for many echo chambers as well. Classic examples of echo chambers include the followers of Rush Limbaugh and popular broadcasters on Fox News (Jamieson and Cappella 2008; Nguyen 2020), and anyone with access to a radio or cable television could follow these individuals. Of course, there also exist more exclusive echo chambers, but this is also true of blame bubbles. Take, for example, private Facebook groups, which require administrator approval, dedicated to criticising a particular individual.

2. Nguyen (2020) distinguishes between echo chambers and epistemic bubbles. According to Nguyen, relevant epistemic sources are merely left out of epistemic bubbles, but they are systematically excluded from echo chambers. Although I use the term 'blame bubble' in this paper, I take blame bubbles to be more similar to echo chambers than epistemic bubbles, since non-members are not simply left out, but morally discredited and excluded.

discredited, typically by being morally criticised for failing to join the bubble. Nguyen also argues that those within echo chambers amplify members' epistemic credentials and believe that members are more trustworthy than non-members. There is good reason to think that blame bubbles also possess an analogue of this feature. Blame bubble members (at least tacitly) believe that it is morally better to be a member of a blame bubble than to be a non-member. However, members need not conceive of themselves as members of an enclosed system, or think that such closed communities are morally good, to have this tacit belief. Rather, an agent possesses a tacit belief in the moral goodness of being in a blame bubble (and the badness of not being in one) when they (1) take it to be morally important not only to criticise the target of the blame bubble, but to *endorse* other members' blame, typically by repeating and reinforcing their criticisms; and (2) find fault in agents who fail to do these things.

The tacit belief in the moral goodness of blame bubble membership (and the badness of non-membership) can help explain several features of blame bubbles. First, it can explain why members of blame bubbles encourage others to share in their blame and criticise those who don't. Take, for example, familiar refrains like 'silence is violence' and 'complacency is complicity'. Blame bubble members will often make these remarks to encourage others to join in their blame and to criticise agents who fail to do so. This is because blame bubble members think it is morally important not only to recognise wrongs as such, or to privately condemn them, but to also *endorse* the blame of other members, typically by repeating and reinforcing their criticisms, i.e. to join the blame bubble.

The tacit belief in the goodness of blame bubble membership (and the badness of non-membership) can also explain how blame bubbles persist and grow over time.[3] If blame bubble members did not think that it was morally important to endorse other members' blame, then there would be little reason for members to continuously repeat and reinforce others' criticisms or to call on non-members to join in. This would cause blame bubbles to burst almost as soon as they form. But blame bubbles rarely behave this way. Those within blame bubbles will often repeatedly express blame of the third party in question and affirm others' expressions of blame as well. It is not unusual for agents to join blame bubbles long after they initially formed, causing the bubbles to not only persist but grow, sometimes quite substantially. This can be explained by the fact that agents are (at least partially) motivated to join blame bubbles because they take it to be morally good to do so and morally bad to remain silent.

I'm now in the position to offer an account of blame bubbles, synthesising elements from both Lackey's and Nguyen's accounts of echo chambers. In my view, there are four key features of blame bubbles:

1. Expressions of blame are repeated and reinforced, thereby amplifying them, often through re-sharing.

2. This occurs in an enclosed system or 'chamber', such as a social network, allowing the blame to 'echo'.

3. Dissenting voices are either absent, drowned out, or morally discredited.

3. There is much to say about the individuation and persistence conditions of blame bubbles. While an extended discussion of these matters is beyond the scope of this paper, it will be useful to say a bit about how blame bubbles come to exist and persist. There are many ways in which blame bubbles come into existence. They can arise within an already extant social structure, as is the case when a social group collectively turns their attention to a moral injustice. Blame bubbles can also develop in response to particular events. This best captures how the blame bubble that arose in response to Sacco came to be: The creation of #hasjustinelandedyet allowed strangers who blamed Sacco to join together, amplify one another's blame, and discredit their critics. Blame bubbles can also have a variety of targets, be they longstanding general focuses, like PETA's decades-long fight against all forms of animal exploitation, or relatively short-lived and narrowly focused topics, like the blame bubble directed at Sacco. In many ways, this mirrors the formation and shape of echo chambers, which include groups such as the long-time followers of Rush Limbaugh and his many causes (Jamieson and Cappella 2008; Nguyen 2020) and small groups that form in response to particular ballot measures and local campaigns.

4. Members (at least tacitly) believe that it is morally better to be a part of the blame bubble than to be a non-member.

Given these features, blame bubbles can occupy a variety of social structures. While social media platforms are well designed to give rise to the kinds of enclosed social networks in which blame bubbles thrive, blame bubbles can also form offline. Public demonstrations, including protests, picket lines, and boycotts, typically have all four blame bubble features, for example.

However, not *all* instances of public or group blame will possess the above features or count as blame bubbles. It is possible that large numbers of individuals could come to publicly share similar expressions of blame without being part of the same social network or intending to repeat and reinforce the blame of others. In such a case, these individuals would not be members of the same blame bubble, just as individuals who independently come to, and then express, similar beliefs are not members of the same echo chamber. There are also enclosed systems that feature repeated and reinforced expressions of blame that do not amount to blame bubbles. Take public debates, where each side amplifies and reinforces their own blame, but dissent is expected and even encouraged. Because dissenting voices are present and tolerated in these spaces, they do not constitute blame bubbles. Finally, there are also contexts in which blame echoes within an enclosed system that excludes dissenting voices but still do not constitute blame bubbles. Take, for example, friends who join together to admonish another friend for being inexcusably late. The friends might amplify one another and discourage dissent, since it is important to present a unified front in these circumstances, but it is unlikely that the friends would find fault in non-friends who failed to join in their blame. Indeed, they might find it objectionable, since their friend's tardiness, while wrong, is not anyone else's business.[4] In this case, the friends fail to have a tacit belief in the moral badness of non-membership, and so their blame does not take place within a blame bubble.

With an account of blame bubbles now in hand, I will next examine how their structural features can give rise to objectionable blaming practices.

2. What is so bad about blame bubbles?

Social media has spurred a rise in echo chambers, which many argue has damaged public discourse, led to the spread of fake news and conspiracy theories, and even degraded democracy (e.g. Sunstein 2017). Given the structural symmetries between echo chambers and blame bubbles, it isn't surprising that the blaming practices found on social media have also received criticism. Many of these criticisms focus on the ways in which online blame unjustly impacts its targets. In this section, I explore two ways in which blame bubbles go awry.

2.a. Unfitting blame

Humans are fallible, and sometimes we blame people who aren't in fact blameworthy. This is unjust — innocent people should not be blamed for things they did not do, and the harms associated with being wrongly blamed can be both serious and significant. While unfitting blame is not unique to blame bubbles, the risk of blaming innocent individuals and the harms associated with this are exacerbated within them. Recall that members of blame bubbles take it to be morally better to be a part of a blame bubble than to be a non-member. This tacit belief in the moral goodness of being within a blame bubble can bias members' judgements regarding the blameworthiness of others — they may be quicker to blame, less likely to seek out evidence of innocence, and more confident in their assessments of others' blameworthiness. This could increase the likelihood that members of blame bubbles will get it wrong — they may blame a person who isn't blameworthy. This risk is further exacerbated by the fact that blame bubbles exclude and discredit dissenting voices.

4. See Seim (2019) for discussion of how meddlesome blame can be objectionable.

Not only do blame bubbles pose a risk of generating unfitting blame, the harms associated with blaming the innocent could also be significant. Blame bubbles have the capacity to bring together large numbers of individuals, all of whom will express blame at a particular target. When that target is innocent, the consequences of this unfitting blame can be devastating. Take, for example, the blame bubble that targeted Tuhina Singh.[5] In May 2020, when Singapore was in lockdown to curb the spread of COVID-19, videos of a woman refusing to wear a mask at a market were posted online. The videos were shared widely on social media and generated a strong negative reaction. Within two days, users identified the woman as Singh and began posting her personal information online, including her phone number, e-mail address, and private pictures. The next day, it was revealed that the woman in the videos was not in fact Singh, and the actual woman in question was already in police custody.

2.b. Excessive blame

One might think that even if Singh had refused to wear a mask in the market, there would still be something inappropriate about people sharing her personal information online. This is another problem with blame bubbles — the blame within them risks becoming excessive. Just like unfitting blame, excessive blame is also unjust. It is wrong to excessively blame agents even if they are blameworthy, since such treatment is undeserved. There are two ways that excessive blame can arise within blame bubbles.

First, blame can ramp up. Because expressions of blame are repeated and reinforced in blame bubbles, these expressions become amplified over time. As members of blame bubbles are exposed to more expressions of blame directed at a third party, the more confident members will likely become in the blameworthiness of this individual, which can lead to harsher and at times excessive expressions of blame. The fact that those who disagree are silenced and seen as outsiders can exacerbate this effect. Notably, ramping up is also present in other groups in which contrary or dissenting viewpoints are absent, including echo chambers (Sunstein 2017).

Interestingly, blame within blame bubbles can also be excessive even if each member of the blame bubble expresses an appropriate degree of blame. This phenomenon, known as a 'pile-on', can occur when blame bubbles grow quite large. In large blame bubbles, it is possible that on their own, each expression of blame within the blame bubble is fitting, but taken together, they can have an outsized effect on their target. Recall the case of Justine Sacco. Thousands of people blamed Sacco for posting an offensive and racist message. Even if what Sacco posted was offensive and racist, it isn't clear that she deserved the suffering inflicted by a large-scale blame bubble. This could be so even if no individual member of the blame bubble blamed Sacco excessively.

3. Not all blame bubbles

Based on the above discussion, one might conclude that the structural features of blame bubbles are responsible for the ways in which blame can become unfitting and excessive within them. After all, it is because members (at least tacitly) believe it to be morally better to be a part of a blame bubble than to be a non-member. And the excessive blame that is characteristic of blame bubbles is attributable to the fact that expressions of blame are repeated and reinforced within these closed networks, allowing those expressions to amplify and echo. Given this, one could argue that blame bubbles are in and of themselves morally problematic.

This way of thinking about blame bubbles mirrors how some theorise about echo chambers. Many argue that the structures of echo chambers are responsible for their bad-making epistemic features (Sunstein 2017; Nguyen 2020). Because these theorists take echo chambers to be intrinsically epistemically problematic, they often conclude that it is

5. For coverage of these events, see www.newyorker.com/magazine/2020/09/28/the-public-shaming-pandemic, accessed 10 December 2020 and www.straitstimes.com/singapore/doxxed-ceo-wrongly-identified-as-sovereign-woman-thanks-supporters, accessed 10 December 2020.

epistemically bad for agents to be in them. For example, Alex Worsnip (2019) contends that we have an epistemic obligation to 'diversify our sources' and gather information from outlets that do not share our views and values, while Nguyen recommends that agents engage in Cartesian-style 'social-epistemic reboots', where we suspend all of our beliefs, particularly those about who to trust (2020).

But not everyone agrees that echo chambers are problematic in virtue of their very structure. Some, like Lackey (2021), argue that echo chambers are only problematic in so far as they contain unreliable information.[6] According to Lackey, it is possible to be within an echo chamber and not be led epistemically astray. In fact, she argues that echo chambers, particularly on social media, can be a powerful source of information (2021: 219). Can the same be said for blame bubbles? Are there blame bubbles that can be a source of moral good? In this section, I'll explore two ways in which blame bubbles do important moral work.

3.a. Accountability

Not all expressions of blame are successfully communicated to their targets. Sometimes, we brush off the blame that is directed at us, ignoring those that blame us and refusing to acknowledge or address the wrongs in question. This is particularly true of people in positions of power. There is a growing body of empirical work that suggests that power affects not only how agents express blaming emotions like anger, but also how they respond to these emotions. When powerful agents blame others by expressing anger, this tends to elicit fear in less powerful agents, which successfully alters their behaviour. In contrast, powerful agents do not typically alter their behaviour and even retaliate when others blame them by expressing anger (Lelieveld et al. 2012; Sinaceur & Tiedens 2006; Van Kleef & Côté 2007; Van Kleef et al. 2004). Powerful agents typically fail to uptake the blame that is directed at them and will often refuse to recognise or alter their wrongful behaviour.

While it can be relatively easy to deflect the blame of a single individual, it is much more difficult for perpetrators, even powerful ones, to ignore the blame of a collective made up of hundreds or thousands of individuals. First, blame bubbles amplify the blame within them — they are louder, both metaphorically and literally, than a single expression of blame could ever be. Blame bubbles can also help correct for the power differential between perpetrators and those who blame them. The larger the blame bubble, the fewer avenues a powerful perpetrator can take to retaliate against their blamers. Thus, when faced with a blame bubble, wrongdoers in positions of power are more likely to recognise their actions as wrongful, or at the very least alter their behaviour in light of being blamed. In this way, blame bubbles can play an important role in holding powerful perpetrators to account.

There are several recent cases that illustrate the important role blame bubbles play in holding wrongdoers accountable. Take, for example, the #FreeBritney movement. In 2008, Britney Spears was placed under a conservatorship, with her father Jamie Spears and lawyer Andrew Wallet as conservators. Under the conservatorship, designed for individuals who cannot manage their financial and/or daily lives, Spears was unable to retain her own attorney, access her funds, or make decisions about her healthcare, career, or personal life, without the approval of her conservators. Despite being declared mentally unfit, Spears maintained an extremely successful, though gruelling, career while under the 13-year conservatorship. She released four studio albums, went on several international tours, headlined a Las Vegas residency, and took on several television acting and hosting roles. Not only did Spears make a substantial amount of money during this time, all of which was under the control of her conservators, she was also responsible for paying these individuals for serving as her caretakers. Spears was required to cover her court-appointed lawyer Sam Ingham's $520,000 annual salary, and Jamie Spears reportedly

6. See also Regina Rini (2017), who argues that it can be epistemically virtuous to grant more credibility to those with whom one shares a partisan affiliation than to those with whom one doesn't.

collected an estimated $6,000,000 over the course of the conservatorship (Taylor 2021).

While the media was largely accepting of Spears's conservatorship, a small number of devoted fans began to form a blame bubble, challenging the conservatorship almost as soon as it was instituted.[7] These fans were convinced that Jamie Spears, and his associates, including Lou Taylor, were using the conservatorship to control, and profit off, Spears. For the next decade, the Free Britney blame bubble remained small and sequestered mostly to online spaces devoted to Spears. However, it grew in 2019 when the podcast *Britney's Gram*, hosted by Tess Barker and Babs Gray, aired a voicemail from an anonymous source stating that Britney was being held in a mental health facility against her will.[8] The episode was widely shared, the hashtag #FreeBritney began trending on social media, and a small protest was held in Los Angeles. At the time, those within the Free Britney blame bubble were labelled as conspiracy theorists, both by the Spears family and the media,[9] and several of its leaders were targeted with lawsuits.[10] Still, the bubble grew, drawing support from increasingly influential people, including other musicians and journalists. In addition to amplifying one another's blame of those responsible for the conservatorship within their social network, members also drowned out and morally discredited dissenting voices. For example, Britney Spears's sister, Jamie Lynn Spears, received harsh criticism whenever she defended the conservatorship or her decision not to speak up about it.[11] Members of the blame bubble also urged others to join them, particularly powerful media figures such as Kim Kardashian,[12] indicating that they (at least tacitly) believed that it was morally better to be a part of the blame bubble than to be a non-member.

Eventually, the movement appeared to affect even those closest to the conservatorship. In 2020, Ingham requested that future hearings be unsealed, citing the #FreeBritney movement: 'Far from being a conspiracy theory or a "joke"… this scrutiny is a reasonable and even predictable result of [Jamie's] aggressive use of the sealing procedure over the years to minimize the amount of meaningful information made available to the public'.[13] Things came to a head in June 2021, when Spears made a public statement in court accusing her conservators of abuse and making clear that she wanted the conservatorship to end.[14] From here, things moved relatively quickly. Within a month, Spears was able to hire her own attorney, Mathew Rosengart, who filed a petition to remove Jamie Spears from the conservatorship within a week of being hired, and the conservatorship was officially dissolved in November 2021. Spears's lawyers also entered the process of discovery and indicated that they may pursue charges of conservatorship abuse against Jamie Spears and others involved.[15]

7. Jordan Miller began posting critiques of the conservatorship on his Spears fan site BreatheHeavy early in 2009, signing his posts 'Free Britney'. See https://www.newyorker.com/news/american-chronicles/britney-spears-conservatorship-nightmare, accessed 19 April 2022.
8. https://britneysinstagram.libsyn.com/75-freebritney, accessed 19 April 2022.
9. https://i-d.vice.com/en_uk/article/8xznkz/how-the-freebritney-movement-took-stan-culture-too-far, accessed 19 April 2022; www.latimes.com/entertainment-arts/music/story/2020-08-01/britney-jamie-spears-dad-freebritney-conservatorship, accessed 19 April 2022.
10. www.hollywoodreporter.com/business/business-news/britney-spears-father-sues-freebritney-blogger-defamation-1221212/, accessed 19 April 2022; www.dailymail.co.uk/news/article-7228899/Britney-Spears-manager-Lou-Taylor-sues-FreeBritney-supporter-bought-domain-name.html, accessed 19 April 2022.
11. www.newsweek.com/jamie-lynn-spears-receives-death-threats-britney-spears-conservatorship-1606863, accessed 18 May 2023.
12. https://news.yahoo.com/fans-flooded-kim-kardashians-instagram-143856649.html?guccounter=1&guce_referrer=aHRocHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJkw8a6IatBAAEN68LPC_PgUD9xBFzsW7k56dk6DxK0Gd5YKG6s05y_IJgFbrDqAzyDMuBbIyhosD3gb3KArQYjJK5e_O-t6QIs6VDcQWI-ab6MGh52B9wO15yK6-oz-jbfaSGohDOMczbrijjqx7LRnDrYwDRNX_iJzWjMkXod, accessed 18 May 2023.
13. www.newyorker.com/news/american-chronicles/britney-spears-conservatorship-nightmare, accessed 19 April 2022.
14. https://237b6d3f-67e3-47e9-ab01-ef7c47479a89.usrfiles.com/ugd/237b6d_ac4d0cbfb9ce4b579093378dc7d9bc00.pdf, accessed 19 April 2022.
15. www.nytimes.com/2021/07/26/arts/music/britney-spears-conservatorship-father-jamie.html, accessed 12 May 2025.

According to those closest to the case, it is unlikely that Spears would have regained control over her person and estate if not for the blame bubble that formed in response to the conservatorship. Spears credits the movement with saving her life, while Rosengart stated: 'I think the support of the #FreeBritney movement has been instrumental…'[16] Of course, the existence of a blame bubble does not guarantee that justice will be served. Although Spears now has more control over her life than she has had for the past decade, no one responsible for the conservatorship has been charged with any crime relating to it, and it is possible that no one ever will be. Still, these recent events highlight how important blame bubbles can be in initiating the process of holding wrongdoers, particularly those in positions of power, accountable.

Examining how the blame bubble that formed in response to Spears's conservatorship and persisted in the months after it ended sheds even more light on the importance of blame bubbles in holding people in power accountable. While the blame bubble was initially focused on ending the Spears conservatorship, its members began to call out other incidences of conservatorship abuse. The last episode of *Toxic*, Barker and Gray's podcast dedicated to the Spears conservatorship, featured stories of others who have experienced conservatorship abuse and discussions of what kinds of legal interventions could prevent such injustices. And in an interview with *Rolling Stone*, Leanne Simmons, one of the leaders of #FreeBritney L. A., stated: 'We accomplished our initial goal, but as the movement acquired more knowledge about the whole system, we evolved. Now, there's a larger goal ahead of us: To change the system so that this doesn't happen to anyone else'.[17] This continued scrutiny from the Free Britney blame bubble has already yielded results. California Governor Gavin Newsom has signed off on a law, known as the #FreeBritney bill, designed to reform conservatorships in California, and other states are following suit.[18]

16. www.rollingstone.com/music/music-features/freebritney-whats-next-1291469/, accessed 19 April 2022.

17. www.rollingstone.com/music/music-features/freebritney-whats-next-1291469/, accessed 19 April 2022.

3.b. Address

Blame bubbles do not only play a role in holding perpetrators accountable, they can also address wrongs done to victims. Recall that perpetrators in positions of power do not typically uptake the blame that is directed at them by those in less powerful positions. Rather than acknowledge and attempt to repair their wrongdoing, many powerful perpetrators will either ignore the blame expressed towards them or retaliate. This places victims in a difficult position. If they express blame to those who have wronged them, they may suffer further harm and wrongdoing. But failing to blame perpetrators can also have negative consequences. For example, members of the moral community may come to think that the way the perpetrator acted is an acceptable form of treatment. There are also non-instrumental reasons to blame perpetrators as well. According to Amia Srinivasan (2018), getting angry about being treated unjustly is intrinsically good, because it is a valuable way of appreciating injustice in the world. She argues:

> Just as appreciating the beautiful or the sublime has a value distinct from the value of knowing that something is beautiful or sublime, there might well be a value to appreciating the injustice of the world through one's apt anger — a value that is distinct from that of simply *knowing* that the world is unjust. (Srinivasan 2018: 132)

According to Srinivasan, because there is intrinsic value in blaming those who have wronged you, being unable to express blame for fear of negative repercussions is unjust. She refers to this as 'affective injustice', which is the 'injustice of having to negotiate between one's apt emotional response to the injustice of one's situation and one's desire

18. www.politico.com/states/california/story/2021/09/30/newsom-signs-freebritney-bill-to-help-reform-conservatorship-laws-9427463, accessed 20 April 2022.

to better one's situation — a conflict between responsibilities that is "all but irreconcilable"' (2018: 135).

How do victims navigate this difficult and unjust landscape? Recent empirical work suggests that those who aren't in positions of power express indirect blame of perpetrators — specifically by sharing their anger at the perpetrator with others (Petkanopoulou et al. 2019). This allows agents to express their anger, and affectively appreciate the injustice they've been subjected to and communicate it to others, while also shielding themselves from retaliatory behaviour. Take, for example, the whisper networks that exist in many professions to warn women about the abusive men in their fields. Although many victims of workplace harassment and assault do not publicly accuse those who wrong them for fear of retaliation, this does not mean that they remain silent. Victims will often express their blame of perpetrators to their co-workers and colleagues, alerting them to the perpetrators' behaviour while also registering their own appreciation of the wrongs that have been done to them. Some criticise whisper networks, arguing that they render entire industries largely complicit in the perpetrators' wrongdoing. But given the lengths that some powerful perpetrators will go to silence accusers, and the generally dismal treatment of those who directly blame powerful perpetrators, expressions of indirect blame are sometimes the best available, though non-ideal, course of action.

Blame bubbles are excellent venues in which victims can express this kind of indirect blame. Because members of a blame bubble are united in their blame of a particular perpetrator, expressing blame towards that perpetrator within the bubble won't elicit the negative consequences that typically attend direct expressions of blame towards powerful wrongdoers. Victims can express blame, which has both instrumental and intrinsic value, while being shielded from retaliation from perpetrators.[19] Furthermore, as more people add their voices to a blame bubble, the larger it becomes, and the better it is at offsetting the power differential between the perpetrator and their victims. Eventually, this can allow victims to directly and publicly express blame to the perpetrator without fear of retaliation. This is likely one of the roles the #FreeBritney movement played in ending the Spears conservatorship. In her June 2021 testimony, Spears reflected on why she was reluctant to publicly accuse her father and others of conservatorship abuse:

> It's embarrassing and demoralizing what I've been through. And that's the main reason I've never said it openly. And mainly, I didn't want to say it openly, because I honestly don't think anyone would believe me… and that's why I didn't want to say any of this to anybody, to the public, because I thought people would make fun of me or laugh at me and say, "She's lying, she's got everything, she's Britney Spears".[20]

The prominence of the #FreeBritney blame bubble, and its increasingly large number of members who were prepared to believe Spears and trust her testimony, likely gave Spears the moral support necessary to openly testify about the abuse she suffered.

Blame bubbles' ability to provide moral support to victims is an important source of their value. Blame bubbles unite many agents who all, to varying degrees, blame a perpetrator or group of perpetrators for how they treated an individual or individuals. Recall that expressions of blame not only communicate information about the perpetrator, but also about the victim. When members of blame bubbles express blame, they are not only communicating that they take the perpetrator in question to be blameworthy; they are also communicating that they take the victim to be worthy of better treatment than they received. This is an important message to communicate, particularly when the victim is a member of an oppressed group or is vulnerable to further

19. This feature also allows blame bubbles to address other forms of injustice that victims of powerful perpetrators face, such as testimonial injustice (Fricker 2007). Thanks to David Bronstein for discussion on this point.

20. https://variety.com/2021/music/news/britney-spears-full-statement-conservatorship-1235003940/, accessed 4 July 2022.

mistreatment. In expressing blame to those who wrong these victims, members of blame bubbles affirm victims' status as members of the moral community and dissuade others from wronging them further. Thus, not only do blame bubbles address affective injustice by providing a space in which victims can express indirect blame of perpetrators, they also provide moral support to victims, laying the groundwork for these victims to one day be able to directly blame those who have wronged them.

3.c. Overview

While some blame bubbles are clearly pernicious, not all blame bubbles are guilty of these faults. Although the structural features of blame bubbles can exacerbate their negative qualities, they can also enhance their good-making features. Because expressions of blame are repeated and reinforced within blame bubbles, they become amplified, making it difficult for powerful perpetrators to ignore their message or retaliate against their victims. And because dissenting voices are excluded from blame bubbles, victims can express blame within them without facing affective injustice. Similarly, because blame bubbles are made up of many individuals who blame some perpetrator(s) for how they've treated their victim(s), they prove to be excellent sources of support for these victims. So, it is unlikely that the structures of blame bubbles are the complete source of their badness. Rather, when trying to determine what causes blame bubbles to go bad, we should follow Lackey's lead and look beyond structure.

4. Cheap talk

While there are likely several factors that contribute to the downfall of blame bubbles, in this section I present the hypothesis that one key contributing cause stems from failures of communication. Recall that blame possesses a communicative function, and expressions of blame can convey information about both the perpetrator and the victim. Roughly, expressions of blame communicate that the blamer takes the perpetrator to have acted in a morally objectionable way and that they take the victim of the perpetrator's actions to be deserving of better treatment. But expressions of blame do not always succeed at communicating this information. As we have seen, power is one factor that affects the communicative success of blame, but there are many others.

Recently, theorists have begun to draw on work in decision theory and evolutionary psychology to defend signalling theories of blame (Shoemaker & Vargas 2021; Kogelmann & Wallace 2018). According to these views, the communicative success of blame depends, in part, on how difficult it is to fake. This is because the information that blame conveys is highly valuable, difficult to independently discern, and agents can benefit from deceiving others that this information is true of them. When I express blame of a perpetrator, I'm communicating that I take this agent to have acted in a morally objectionable way and that their victim is deserving of better treatment. This is highly valuable information to my moral community, particularly those who also blame the perpetrator, for it indicates that I share (at least some) of their values and am willing to defend them when necessary. These agents can trust that I won't violate the norms in question and can rely on me to cooperate on moral projects. But if I am insincere in my blame and do not actually take the perpetrator to have done wrong or the victim to be worthy of better treatment, then my community members' trust will be ill-placed, and I could take advantage of them. This creates a coordination problem: If we can't trust others to follow the moral norms, then we ourselves have little reason to cooperate with them. Such a scenario would be a disaster, for we rely on the cooperation of others in most facets of our lives. In contexts where accurate and valuable information is difficult to ascertain, and deception could be beneficial, hard-to-fake signals often arise to ensure that coordination is possible among group members. If a signal reliably communicates valuable information and is difficult to fake, then others have good reason to think that the information is true, and they can reliably trust and cooperate with the signaller. Reflecting on the evolutionary and decision theoretic fitness of hard-to-fake signals has

led some philosophers to argue that *blame* can function as this kind of difficult-to-fake signal.

But what makes blame hard to fake? According to Neil Levy (2021), hard-to-fake signals are costly, involuntary, and/or self-validating. Typically, expressions of blame possess these features. One common way to express blame is through the expression of a reactive attitude like resentment, indignation, or guilt (McKenna 2012, 2022). These attitudes, like other emotions, are hard to fake (Frank 1988) because they are both costly and involuntary. The experience of resentment, for example, commits one to certain action tendencies that may not be in one's immediate short-term interests, like threatening and punishing behaviour (Shoemaker and Vargas 2021: 587). Emotions like resentment are also accompanied by changes in facial expression, posture, and tone (Ekman 1992), which are difficult to control but are visible to bystanders. Because resentment is hard to fake in virtue of being costly and involuntary, many argue that it conveys a degree of sincerity and seriousness that makes it particularly adept at serving the communicative function of blame (McGeer 2013). Expressions of blame can also be hard to fake when they take place in morally diverse environments (Kogelmann & Wallace 2018). Expressing blame among agents who do not agree is certainly costly. It opens one up to a range of negative reactions, including retaliation, punishment, and isolation. Expressions of blame in morally diverse contexts are also self-validating, since there is no other plausible explanation for why one would express blame other than the fact that one really does blame the agent in question. An agent who expresses blame within a group of opponents will not receive reputational gains, trust, or reliance from these agents — there is little reason to express blame other than the fact that one sincerely blames.

While in-person expressions of blame directed at perpetrators are typically hard to fake, this isn't so for the blame expressed within blame bubbles, particularly those that exist on social media. While emotions are hard to fake in real life, they are easy to fake online — one could easily create an outraged social media post without actually being outraged at all. When an individual expresses their moral outrage by posting on social media, no one can observe them engaging in the costly action tendencies or involuntary expressions that are distinctive of experiencing anger. And while our communities may be morally diverse, blame bubbles are not — they purposefully exclude dissenting voices. This means that members of blame bubbles do not bear the costs of facing negative reactions to their blame. In fact, expressions of blame within blame bubbles are typically met with approval and praise — members are rewarded for expressing blame. This makes the blame within blame bubbles not only relatively cheap but also less self-validating. There is an easily available explanation for why one would express blame in a blame bubble other than the fact that one really does blame the agent in question: one wants the positive benefits associated with being a blame bubble member. In this way, expressions of blame within blame bubbles are easy to fake — they aren't costly, involuntary, or self-validating.[21] But how does the fact that blame is easy to fake within blame bubbles contribute to their downfall? There are (at least) two explanations one might offer.

4.a. Bad actors

One could argue that the easier blame is to fake within blame bubbles, and the more benefits one can accrue from joining the blame bubble, the more likely it is that free riders will seek membership. Agents who don't care about the relevant perpetrator's behaviour or their victims might be motivated to express blame nonetheless, since doing so

21. This is not to say that blame within blame bubbles is *entirely* costless. As Levy argues, if agents want to gain reputational benefits, then they will need to use their real names and maintain a stable presence over time (2021: 9558). This comes at a cost, since it opens agents up to being charged with hypocrisy and other criticisms. But there are many benefits to being a member of a blame bubble that do not require an agent to provide their actual names or behave consistently over time. And, because blame bubbles are constantly growing and shifting focus, and members are geographically dispersed, it is difficult to determine whether members behave consistently over time and whether their real-world actions mesh with their online personas. Thus, blame within blame bubbles remains very, if not maximally, easy to fake and these costs do little to satisfy those who remain sceptical of the sincerity of online blame.

comes at little cost and can bring them goodwill from other blame bubble members. Not only could free riders take advantage of those who welcome them into their blame bubble, but they will also likely make the blame bubble itself worse, since free riders are more concerned with what the blame bubble can do for them than what they can do for victims and the moral community more generally.

This explanation coheres nicely with a series of arguments that Justin Tosi and Brandon Warmke defend in their book *Grandstanding: The Use and Abuse of Moral Talk* (2020). Tosi and Warmke argue that the unfitting and excessive behaviours associated with public moral discourse can be explained by the presence of moral grandstanders: agents who engage in moral discourse because they are overly concerned with impressing others with their moral qualities. According to Tosi and Warmke, grandstanders, because they are more concerned with receiving reputational gains through public blame than in the content of their blame, are likely to engage in behaviours like blaming the innocent, ramping up and piling on (2020: 44–62). This is because these agents are overly focused on impressing others and are thus more willing to engage in blaming behaviours even if they target innocents or cause excessive harm to their targets.

So, one could argue, because blame is easy to fake within blame bubbles, these structures attract free riders and grandstanders — agents who are more interested in boosting their reputations than holding wrongdoers accountable or supporting victims. Because free riders and grandstanders are not adequately interested in the relevant moral issues, they tend to express unfitting and excessive forms of blame, which cause blame bubbles to go bad. On this explanation, blame bubbles go bad not because they have inherently bad structures, but because they attract bad actors.

Tosi and Warmke are certainly right that there is something objectionable about moral grandstanding. They may also be right that moral grandstanders engage in the kinds of problematic behaviours associated with blame bubbles. However, it is less clear whether blame bubbles really do attract significant numbers of grandstanders and free riders. While there are surely some agents who join blame bubbles only or primarily to gain the moral approval of others, there is reason to think that most blame, even when it is easy to fake, is sincere. As both Evan Westra (2021) and Levy (2021) have noted, empirical work indicates that agents tend to genuinely feel moral outrage in low-cost environments even when their expressions of outrage are entirely anonymous and can't virtue signal successfully (Jordan & Rand 2020). It appears that most people, even if they blame problematically, do so sincerely and will continue to do so even if there is little in it for them. Thus, it is not clear that the presence of grandstanders and other kinds of free riders can fully explain how blame bubbles go bad.

4.b. Bad communication

Whether or not members of blame bubbles frequently engage in insincere blame, they are certainly frequently *accused* of doing so. Blame bubble members often face questions regarding their commitment to their cause as well as their motivations for engaging in public expressions of blame. The rise in criticism regarding blame on social media illustrates this point nicely. Members of blame bubbles are often accused of engaging in virtue signalling, grandstanding, and 'slacktivism'. This should not be surprising. It is very important to us that those with whom we engage in moral discourse do so sincerely. In environments when it is easy to false signal, and much stands to be gained by false signalling, we are particularly attuned to the possibility that those around us are acting selfishly and insincerely (Carlson & Zaki 2018).[22]

22. While one might expect outsiders to question the sincerity of blame bubble members, it is perhaps surprising that members of blame bubbles can become sceptical of one another's motivations. Even though members of groups tend to operate with a default presumption that other members are sincere, this default can be abandoned if certain triggering events occur (Levine 2014). When agents have an obvious motive for deception, or when they receive information from third parties about a potential deception, group members will abandon their presumption of trust and begin to scrutinise other group members (Levine 2014). The fact that it is easy to fake blame within blame bubbles generates a motive for deception: agents can gain significant benefits from

How do blame bubble members respond to challenges to their sincerity? I propose the following hypothesis: Because blamers have a significant moral interest in successfully communicating their blame, and remaining members of their blame bubbles, they will 'double down' on their blaming efforts to ensure that their blame is received as sincere. But focusing on being seen as sincere, even if one really is sincere, can distract agents from the content of their blame, which can have negative consequences. For example, agents may become quicker to blame, running the risk of blaming innocent people, to avoid doubts regarding their motivations. They could call for increasingly harsh punishments of the perpetrators in question to convince others of their commitment to the bubble. They may also repeat others' expressions of blame that were well received by those within the bubble to ensure that their own blame is also communicatively successful. But these are the very behaviours that are responsible for blame bubbles going bad, for they lead to ill-fitting and excessive blame.[23]

According to this explanation, because blame is easy to fake within blame bubbles, blame bubble members are subject to charges of insincerity. And, because blame bubble members are motivated to avoid and overcome these charges, they can become more concerned with successfully communicating their blame than its content, which can lead to ill-fitting and excessive expressions of blame. Notice that agents engage in these problematic blaming practices not because they are acting in bad faith but rather because blame bubbles create environments in which it becomes easy to doubt that others are acting in good faith. Thus, on this explanation, the badness of blame bubbles is attributable to bad communication, not inherently bad structures or bad actors.

joining a blame bubble even if they do not sincerely share in other members' blame. The larger the blame bubble, and the more support there is for members within the structure, the more salient this motive will likely become. This can explain why blame bubble members can start subjecting one another to the same scrutiny with which outsiders treat them.

23. These behaviours are also largely unsuccessful at demonstrating sincerity (Sawaoka & Monin 2018). This is due, in part, to the fact that these practices are easy to fake. It is no harder to fake the blaming of an innocent person than a guilty person, and it would be just as easy for someone to insincerely call for the firing and jailing of a perpetrator as it would be to call for their mild censure.

5. How to make better blame bubbles

In the previous section, I sketched two explanations of how blame bubbles can go bad because it is easy to fake blame within them. According to the bad actors explanation, easy-to-fake blame attracts free riders and grandstanders, who tend to engage in ill-fitting and excessive forms of blame. According to the bad communication explanation, easy-to-fake blame makes it difficult to successfully communicate, which tends to cause agents to engage in ill-fitting and excessive forms of blame. Much more work needs to be done to develop and assess these explanations, and there are likely many factors, in addition to easy-to-fake signals, that contribute to the badness of blame bubbles. But if easy-to-fake signals do play a role in causing blame bubbles to go bad, then on both the bad actors and bad communication explanation, we can improve these structures by implementing changes that make it costlier, more involuntary, and more self-validating to blame within them. In this section, I'll consider several ways of doing just that.

5.a. Take it offline

One way to make blame harder to fake within blame bubbles is to take blame bubbles offline. After all, it is much harder to express insincere blame in person than online, because in-person expressions of blame are both more costly and less voluntary. So, perhaps we should discourage agents from using social media platforms to express blame and attempt to cultivate in-person blame bubbles only.

While the practicality of such a plan is dubious, given the extent to which many of us live our lives online, it is also not clear we should want to pursue such a strategy even if it were possible. This is because taking blame bubbles offline undermines their ability to hold powerful perpetrators responsible. Social media allows us to reach a far greater number of people than we could communicate with in person, and to

do so more efficiently. As I argued above, large blame bubbles are key to holding powerful perpetrators accountable. It could take weeks and even years to form an in-person blame bubble that could rival the size of a blame bubble that formed online within mere minutes. If blame bubbles were taken offline, it is not clear that they would be able to grow quickly enough or to the size necessary to hold many powerful perpetrators accountable.

Furthermore, if we restrict blame bubbles to in-person environments, we will likely only be able to join blame bubbles with people that we know personally. While there is nothing objectionable about joining together with friends and family to blame a common enemy, it makes it more likely that our moral attention will be directed only towards wrongs that affect people who are very similar to us, i.e. people who live in our communities, work in our industries, and share our backgrounds. One of the great benefits of social media is that it connects people who would never meet one another in person. This allows us not only to learn about people who are very different from us, but also to morally support them. If blame bubbles were restricted to in-person environments, this important feature of blame bubbles would be hindered.

5.b. Diversify

Another way to make blame harder to fake within blame bubbles is to diversify the moral perspectives of their members. Expressing blame in morally diverse contexts is costly, since it is likely that such expressions will be met with disagreement. These expressions are also self-validating: Because expressing blame in morally diverse environments rarely serves agents' short-term interests, the most likely explanation for their behaviour is that they sincerely want the perpetrator to be held accountable and care about their victims. So, perhaps we should promote moral diversity in blame bubbles to make blame harder to fake within them.

Diversifying the perspectives of social networks, particularly echo chambers, is extremely difficult to do, and I suspect the same will be true for the moral perspectives within blame bubbles as well. Even if such interventions were to succeed, they would cause blame bubbles to cease to exist, since a defining feature of blame bubbles is that they exclude dissenting voices. This would likely do more harm than good, since blame bubbles' exclusion of dissenting voices serves an important moral function. Blame bubbles are well placed to address affective injustice because they provide spaces in which victims can blame their perpetrators and express anger without fear of retaliation. Blame bubbles are only able to provide such spaces because opposing voices are purposefully excluded from them. If agents can express scepticism and criticism of victims' blame without being ignored, shouted down, or morally discredited, then this space is no longer adept at addressing affective injustice, and one of the key moral functions of blame bubbles would be undermined.

5.c. Enhance the cost of address, not accountability

It shouldn't be surprising that attempts to take blame bubbles offline or diversify the moral perspectives within them will likely fail to improve blame bubbles. These changes alter the structure of blame bubbles, and these structural features are what allow blame bubbles to serve important moral functions. But if we cannot alter the structure of blame bubbles, how are we to make it harder to fake blame within them? Rather than alter the environment in which blame bubble members express blame, I suggest we alter our blaming expressions.

Recall that expressions of blame communicate two kinds of messages, one about wrongdoers and one about their victims. When we express blame, we communicate not only that the target of our blame has done something wrong but also that their victims were treated unjustly and deserve better treatment. While expressions of blame communicate information about both the perpetrator and victim, there are ways we can emphasise one message over the other. When we express blame by focusing on the culpability of the perpetrator and what they deserve in virtue of acting wrongly, we emphasise the perpetrator-directed message. When we attend to how the perpetrator's

actions affected their victims, and what victims are owed in virtue of being wronged, we emphasise the victim-directed message. And many of the problematic practices blame bubble members engage in heavily emphasise perpetrator-directed messages. For example, when we ramp up, we do so by calling for harsher and harsher treatment of the perpetrator, and piling on involves heaping moral criticism onto the target of blame. However, these practices are both morally objectionable and do little to make our expressions of blame more communicatively successful. So, it could be fruitful to examine ways in which we can enhance the communicative success of our blaming expressions by focusing on messages related to victims as opposed to perpetrators.

Importantly, simply altering which message we emphasise when we express blame will do little to make these expressions more communicatively successful. Within a blame bubble, it is just as easy to fake expressions of blame that emphasise victim-directed messages as perpetrator-directed messages. However, there are a variety of ways we can make blame harder to fake within blame bubbles by focusing on the victim-focused messages.[24] First, we can impose costs on blame that benefit victims. For example, fundraisers that benefit the victims of perpetrators' wrongful actions quite literally make expressions of blame more costly, and thus harder to fake. These fundraisers not only serve the moral function of addressing wrongs done to victims, they also enhance the communicative success of our blaming expressions. By making it easier for blame bubbles to host fundraising campaigns for victims, and cultivating norms that encourage blame bubble members to donate to these causes, we can make blame bubbles better. A growing body of empirical work on third-party compensation bears this out. Recent studies indicate that third parties who compensate victims are perceived as more trustworthy (Jordan et al. 2016; Dhaliwal et al. 2021) and are more frequently chosen to be partners in cooperative activities (Heffner & FeldmanHall 2019; Dhaliwal et al. 2021) than third parties who punish perpetrators. Dhaliwal et al. suggest that this is because third-party compensation is a costly, honest signal and it can more reliably signal cooperative intent than third-party punishment (2021: 48).

Another way to make blame harder to fake is by engaging in practices that amplify the voices of victims. Take the share the mic now movement, in which Black activists take over the social media profiles of famous white women to draw attention to the important work Black women are doing and magnify their stories.[25] We can cultivate similar practices within blame bubbles to address affective injustice. Powerful agents with large social media presences could share their platforms with victims from oppressed groups so that they may express their blame from a position of power. This practice could address affective injustice, since it provides victims with a way to express blame while decreasing the threat of retaliation or other negative consequences. Inviting victims to express blame from one's platform is less voluntary and more self-validating than expressing blame oneself. After all, one cannot control what others will say or do when they have access to your platform.[26] In fact, Myisha Cherry recommends that white allies

24. While I talk in terms of changes that 'we' can make in this section, I do not mean to be offering advice for how individuals should alter their blaming behaviour or police others' blaming behaviour. Often, the agents who are in the best position to directly impact these features of blame bubbles are the corporations that created these spaces and the governments that regulate them. This isn't to say that individuals cannot change norms, but how they do so is much more complex than simply identifying a problematic behaviour and abandoning and/or criticising it (Bicchieri 2016; Westra 2021). Thanks to Alejandro Naranjo Sandoval for discussion on this point.

25. www.instagram.com/sharethemicnow/?hl=en., accessed 14 May 2025.

26. One could attempt to invalidate such expressions by arguing that agents gain reputational benefits by amplifying victims' anger and sharing their platforms. But identifying a potential benefit for the platform-sharer doesn't mean that their action fails to be self-validating. Signals are self-validating when there is no plausible explanation for their action other than the fact that they possess the trait they are signalling that they have. Staking one's reputation on another agent's testimony would only make sense if one confidently believed their testimony to be true. Even if the agent is motivated to share her platform to improve her reputation, this would involve her sincerely believing that the target of blame is blameworthy and that the victim's voice should be heard, and these are precisely the beliefs that blame is meant to signal.

do exactly this when working to support people of colour who have been oppressed:

> A way not to give into this tendency to think or at least communicate that white feelings matter more than the feelings of other groups is to give people of color space to express their rage. To give space means to decenter or take the focus off oneself, if one is white, and allow the vulnerable (nonwhites) to be seen and heard, and to put their rage front and center. (Cherry 2021: 129)

By expressing blame in ways that emphasise victim-directed messages as opposed to perpetrator-directed messages, and doing so in hard-to-fake ways, we can make blame bubbles better. Not only do such practices make blame more communicatively successful and serve important moral functions, they also avoid many of the downfalls that perpetrator-focused blaming practices within blame bubbles have faced. Making blame harder to fake will reduce the risk that blame becomes unfitting and excessive. And when such forms of blame do arise, as is always the risk given blame bubbles' structure, they are much less worrying when the blame is victim-directed. Ramping up and piling on are not so morally problematic when they amount to allocating excessive financial, emotional, and moral support to victims, as opposed to excessive harm to perpetrators. While it would be unfortunate to direct one's moral attention to an agent who wasn't in fact wronged, doing so is much less bad than directing one's moral opprobrium to an agent who wasn't in fact a wrongdoer.

6. Conclusion: But is it blame?

One might argue that practices that support victims may be well and good, but they aren't *blaming* practices. What does compensating victims and providing them with platforms have to do with blaming their perpetrators? One could argue that blame, particularly the angry blame found in blame bubbles, involves a desire to punish or harm the perpetrator, and such motivations are absent from the victim-focused practices discussed above.

It is true that these practices have little to do with harming perpetrators, but I do not think that this precludes them from being meaningful expressions of blame. While some take motivations to harm to be essential features of blame and anger (Nussbaum 2016), many others disagree (Srinivasan 2018; Scanlon 2008; Fricker 2016). And the model of blame that I have adopted in this paper, according to which expressions of blame communicate that the blamer takes the perpetrator to have acted wrongly and takes the victim of the perpetrator's actions to be deserving of better treatment, certainly isn't committed to blame essentially involving a desire to harm wrongdoers.

One might press that these victim-directed practices do not count as expressions of blame even on communicative models. While these practices may be well placed to communicate the victim-directed message of blame, they cannot communicate blame's perpetrator-directed message. However, such an objection misunderstands the nature of blame's messages and the ways in which these victim-directed practices work. First, the victim-directed and perpetrator-directed messages of blame are inextricably linked. While it is possible to emphasise one message over the other, they do not come apart in the way this objection would require them to. The message that a victim has been wronged by another entails that there is a perpetrator who wronged them, just as the message that a perpetrator has wronged someone entails that there is a victim who has been wronged. Similarly, donating to a fundraiser for a victim of injustice is importantly different from donating to a fundraiser for a victim of a natural disaster. While the latter communicates that something *bad* has happened to the victim, the former communicates that the victim has been *wronged*. To communicate that an agent was wronged involves communicating that there was an agent who did the wronging, i.e. the perpetrator. Thus, engaging in the victim-directed practices of compensation and platforming can successfully communicate both the victim-directed and perpetrator-directed messages of blame.

Ultimately, though, not much hangs on whether these victim-directed practices are *blaming* practices or something else. So long as they get the job done, I'm happy to recommend that we adopt them within blame bubbles, regardless of whether we want to call them blame or not. But one might argue that these practices cannot get the job done. Blame bubbles perform two important moral functions: holding perpetrators accountable and addressing wrongs done to victims. While compensation and platforming might be able to address wrongs, one could argue that these practices will not be as successful at holding perpetrators accountable. And if blame bubble members focus all of their collective resources on supporting victims, then they will have no resources left to hold wrongdoers accountable, which would be a significant loss.

First, I do not propose that we replace all perpetrator-directed blaming practices with victim-directed practices within blame bubbles. Rather, my suggestion is that we institute practices that emphasise the victim-directed messages of blame, which have been under-emphasised in the past. It will be important for criticism of perpetrators, and calls to hold them accountable, to exist alongside efforts to support victims and provide them with platforms to express their blame. Second, victim-directed practices do not trade off with, but rather facilitate, efforts to hold perpetrators accountable. Consider how expensive it can be for a victim to try to hold a perpetrator, particularly a powerful one, accountable. It can involve job loss, legal fees, and impacts to one's mental health that require medical and therapeutic treatment. Fundraisers can be used to cover these costs, allowing victims to begin the legal and/or public process of holding perpetrators accountable. Furthermore, when members of blame bubbles engage in the victim-directed practice of platforming, they are validating and amplifying victims' expression of blame, which will include both victim and perpetrator-directed messages. For these reasons, instituting victim-directed practices will likely enhance, rather that stymie, blame bubbles' ability to hold wrongdoers accountable.

This is not to say that there will be no change in how blame bubble members treat perpetrators. By engaging in practices that emphasise the victim-directed message of blame, blame bubbles will likely produce less vitriol towards perpetrators, fewer pile-ons of insults and shame, and less ramping up of calls for punishment. But I do not think that this will impinge on blame bubbles' ability to hold perpetrators accountable. Justice rarely, if ever, calls for the death-by-a-thousand-paper-cuts style of punishment that pile-ons, ramp-ups, and public shamings generate. The #FreeBritney blame bubble wasn't successful because of the pain it caused Jamie Spears to read the thousands of tweets condemning him. It was successful because it made it impossible to ignore the fact that Jamie Spears committed a grave wrong and it forced those in power to begin the process of holding him, and others involved, accountable. Blame bubbles succeed in holding perpetrators accountable because they make it difficult for the perpetrator, and those in power, to ignore the fact that they have wronged the victim and to retaliate against them. Blame bubbles can pose this same challenge to perpetrators even when they are predominantly focused on attending to victims. When a large group of agents gathers to affirm that an agent has been wronged by a perpetrator, it will be just as difficult for the perpetrator to ignore their claims or take action to silence the victim as it would be if the group gathered to criticise and condemn the perpetrator. Thus, while victim-centric blame bubbles may treat perpetrators differently, this doesn't mean that they will be less effective at holding these agents responsible.

In this paper, I've argued that blame bubbles are morally important structures that we should seek to improve rather than eliminate. By focusing on victim-directed expressions of blame and finding ways to make these expressions of blame costlier, more involuntary, and more self-validating, we can make the blame within blame bubbles harder to fake. Such interventions will likely make the blame within blame bubbles more communicatively successful, less harmfully inapt and excessive, and better able to serve blame bubbles' morally important functions.

Works Cited

An, J., D. Quercia & J. Crowcroft (2014) "Partisan Sharing: Facebook Evidence and Societal Consequences," *Proceedings of the Second ACM Conference on Online Social Networks*: 13–24.

Bicchieri, C. (2016) *Norms in the Wild*. New York: Oxford University Press.

Carlson R. & J. Zaki (2018) "Good Deeds Gone Bad: Lay Theories of Altruism and Selfishness," *Journal of Experimental Social Psychology* 75: 36–40.

Cherry, M. (2021) *The Case for Rage: Why Anger is Essential to Anti-Racist Struggle*. New York: Oxford University Press.

Dhaliwal, N., I. Patil & F. Cushman (2021) "Reputational and Cooperative Benefits of Third-Party Compensation," *Organizational Behavior and Human Decision Processes* 164: 27–51.

Ekman, P. (1992) "An Argument for Basic Emotions," *Cognition & Emotion* 6: 169–200.

Frank, R. (1988) *Passions within Reasons*. New York: Norton.

Fricker, M. (2007) *Epistemic Injustice*. New York: Oxford University Press.

Fricker, M. (2016) "What's the Point of Blame? A Paradigm Based Explanation," *Nous* 50: 165–183.

Heffner, J. & O. FeldmanHall (2019) "Why We Don't Always Punish: Preferences for Non-Punitive Responses to Moral Violations," *Scientific Reports* 9: 13219.

Jamieson, K.H. and Cappella, J.N. (2008) *Echo Chamber: Rush Limbaugh and the Conservative*
Media Establishment. Oxford: Oxford University Press.

Jordan, J.J. & D.G. Rand (2020) "Signaling When No One is Watching: A Reputation Heuristics of Outrage and Punishment in One-Shot Anonymous Interactions," *Journal of Personality and Social Psychology* 118: 57–88.

Jordan, J., M. Hoffman, P. Bloom & D. Rand (2016) "Third-Party Punishment as a Costly Signal of Trustworthiness," *Nature* 530: 473–476.

Kogelmann, B. & R. Wallace (2018) "Moral Diversity and Moral Responsibility," *Journal of the American Philosophical Association* 4: 371–389.

Lackey, J. (2021) "Echo Chambers, Fake News, and Social Epistemology," in S. Bernecker, A.K. Flowerree and T. Grundmann (eds.) *The Epistemology of Fake News*. New York: Oxford University Press, 206–227.

Lelieveld, G., E. Van Dijk, I. Van Beest & G. Van Kleef (2012) "Why Anger and Disappointment Affect Other's Bargaining Behavior Differently: The Moderating Role of Power and the Mediating Role of Reciprocal and Complementary Emotions," *Personality and Social Psychology Bulletin* 38: 1209–1221.

Levine, T. (2014) "Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection," *Journal of Language and Social Psychology* 33: 378–392.

Levy, N. (2021) "Virtue Signalling is Virtuous," *Synthese* 198: 9545–9562.

Lewis, H. (2015) "The Digital Ducking Stool," *New Statesman* (London), 11 March. Available at: www.newstatesman.com/politics/2015/03/digital-ducking-stool (accessed 9 November 2020).

Macnamara, C. (2015) "Reactive Attitudes as Communicative Entities," *Philosophy and Phenomenological Research* 90: 546–569.

McGeer, V. (2013) "Civilizing Blame" in Coates & Tognazzini (eds.) *Blame: Its Nature and Norms*. New York: Oxford University Press: 162–188.

McKenna, M. (2012) *Conversation and Responsibility*. New York: Oxford University Press.

McKenna, M. (2022) "Guilt and Self-Blame within a Conversational Theory of Moral Responsibility" in A. Carlsson (ed.) *Self-Blame and Moral Responsibility*. Cambridge/New York: Cambridge University Press, 151–174.

Nguyen, C.T. (2020) "Echo Chambers and Epistemic Bubbles" *Episteme* 17(2): 141–161.

Nussbaum, M. (2016) *Anger and Forgiveness*. New York: Oxford University Press.

Petkanopoulou, K., R. Rodriguez-Bailón, G. B. Willis & G. A. van Kleef (2019) "Powerless People Don't Yell But Tell: The Effects of Social Power on Direct and Indirect Expression of Anger," *European Journal of Social Psychology* 49: 533–547.

Rini, R. (2017) "Fake News and Partisan Epistemology," *Kennedy Institute of Ethics Journal* 27: E-43–E-64.

Ronson, J. (2015) *So You've Been Publicly Shamed*. New York: Riverhead Books.

Saez-Trumper, D., C. Castillo & M. Lalmas (2013) "Social Media News Communities: Gatekeeping, Coverage and Statement Bias," *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*: 1679–1684.

Sawaoka, T. and B. Monin (2018) "The Paradox of Viral Outrage," *Psychological Science* 29: 1665–78.

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

Seim, M. (2019) "The Standing to Blame and Meddling," *Teorema Revista Internacional de Filosofía* 38: 7–26.

Shoemaker, D. & M. Vargas (2019) "Moral Torch Fishing: A Signaling Theory of Blame," *Nous* 55(3): 581–602.

Sinaceur, M. & Tiedens, L. Z. (2006) "Get Mad and Get More Than Even: When and Why Anger Expression is Effective in Negotiations," *Journal of Experimental Social Psychology* 42: 314–322.

Smith, A. (2013) "Moral Blame and Moral Protest" in Coates & Tognazzini (eds.) *Blame: Its Nature and Norms*. New York: Oxford University Press, 27–48.

Srinivasan, A. (2018) "The Aptness of Anger," *Journal of Political Philosophy* 26: 123–144.

Sunstein, C. (2017) *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ/Oxford: Princeton University Press.

Taylor, L. (2021) "Britney Spears Felt Trapped. Her Business Manager Benefited," *New York Times*, 19 December. Available at: www.nytimes.com/2021/12/19/business/britney-spears-conservatorship-tri-star.html (accessed 18 April 2022).

Tosi, J. & B. Warmke (2020) *Grandstanding: The Use and Abuse of Moral Talk*. New York: Oxford University Press.

Van Kleef, G. A. & Côté, S. (2007) "Expressing Anger in Social Conflict: When it Helps and When it Hurts," *Journal of Applied Psychology* 92: 1557–1569.

Van Kleef, G. A., C. K. W. De Dreu & A. S. R. Manstead (2004) "The Interpersonal Effects of Emotions in Negotiations: A Motivated Information Processing Approach," *Journal of Personality and Social Psychology* 87: 510–528.

Westra, E. (2021) "Virtue Signaling and Moral Progress," *Philosophy & Public Affairs* 49: 156–178.

Worsnip, A. (2019) "The Obligation to Diversify One's Sources: Against Epistemic Partisanship in the Consumption of News Media," in Fox & Saunders (eds.) *Media Ethics: Free Speech and the Requirements of Democracy*. New York/Abingdon: Routledge Press, 240–264.