**ORIGINAL ARTICLE**

*European Journal of Philosophy* **WILEY**

# Enforcing social norms: The morality of public shaming

## Paul Billingham[1] | Tom Parr[2]

[1]Department of Politics and International Relations, University of Oxford, Oxford, UK

[2]Department of Politics and International Studies, University of Warwick, Coventry, UK

**Correspondence**

Tom Parr, Department of Politics and International Studies, University of Warwick, Coventry CV4 7AL, UK.
Email: t.parr@warwick.ac.uk

**Abstract**

Public shaming plays an important role in upholding valuable social norms. But, under what conditions, if any, is it morally justifiable? Our aim in this paper is systemically to investigate the morality of public shaming, so as to provide an answer to this neglected question. We develop an overarching framework for assessing the justifiability of this practice, which shows that, while shaming can sometimes be morally justifiable, it very often is not. In turn, our framework highlights several reasons to be concerned about the increasingly widespread phenomenon of *online* public shaming.

## 1 | INTRODUCTION

We are witnessing a re-emergence of the practice of public shaming, especially shaming carried out with the use of the Internet. The following two cases are typical of the phenomenon.

In October 2012, Lindsey Stone was on a trip to Washington DC as a caregiver for a group of adults with learning difficulties. Stone had a running joke with a colleague, Jamie Schuh, where they took humorous photographs, such as them smoking in front of a "No Smoking" sign. While at Arlington National Cemetery, Schuh photographed Stone raising her middle finger and pretending to be shouting in front of a sign reading "Silence and Respect". Thinking it hilarious, Schuh posted the photo on Facebook, with Stone's consent. Four weeks' later, Twitter and Facebook were abuzz with outrage at the photo. Messages ranged from "Lindsey Stone hates the military and hates soldiers who have died in foreign wars" to "Send the dumb feminist to prison" to "Hope this cunt gets raped and stabbed to death". A "Fire Lindsey Stone" Facebook page was created, and attracted 12,000 likes overnight. The next day, Stone lost her job. As a result, "she fell into depression, became an insomniac, and barely left home for a year".[1] Stone applied for many other jobs as a caregiver during this time, but never heard back. Eventually she did manage to get a new job, but lived in constant fear that her new employers would discover the photo and fire her.

In June 2015, biochemist Tim Hunt gave a toast at a lunch sponsored by the Korea Federation of Women's Science and Technology Associations, during the World Conference of Science Journalists in Seoul. According to one account, in the speech, Hunt said:

"It's strange that a chauvinist monster like me has been asked to speak to women scientists. Let me tell you about my trouble with girls. Three things happen when they are in the lab: You fall in love with them, they fall in love with you, and when you criticize them, they cry. Perhaps we should make separate labs for boys and girls. Now seriously, I'm impressed by the economic development of Korea. And women scientists played, without doubt, an important role in it. Science needs women, and you should do science despite the obstacles and despite monsters like me!"[2]

Connie St Louis, a science journalist who was present, tweeted Hunt's comments regarding crying female scientists and single-sex labs, stating that the lunch had been "utterly ruined by sexist speaker Tim Hunt".[3] The comments quickly attracted a great deal of criticism on social media, with many arguing that they reflected broader misogynistic attitudes within the scientific community. The mainstream media picked up on the story, and reported it widely.[4] As a result, Hunt was forced to resign from his positions at University College London, The Royal Society, and The European Research Council. A few days later, Hunt apologized for his comments, which he said were "inexcusable" but were made "in a totally jocular, ironic way".[5] He later said that the barrage of criticism had led him to consider taking his own life.[6]

What should we make of such cases? The most obvious point is that many of those who took part in this public shaming acted reprehensibly by subjecting the shamed to threats, insults, and abuse. This is clearly morally objectionable. In addition to this, the consequences for the shamed ended up being severely disproportionate. Neither Stone nor Hunt did anything to warrant losing their jobs or suffering such deep distress.

Beyond this, matters are less clear. We might think that Stone and Hunt acted wrongly. Their actions arguably violated justified social norms, which play an important role in enabling us to live together civilly. Upholding social norms might well involve criticizing those violators. Making this criticism *public* can help to reinforce our collective commitment to the norms, and can deter prospective future violators. Public shaming may thus be one way in which ordinary people can shape our shared social environment for the better. Perhaps we should laude it as a democratized way in which the public can mold civil society.

This conclusion strikes many of us as too quick, however. One might be troubled by our opening cases, even if one accepts that Stone and Hunt did something wrong and that those who criticized them were seeking to uphold justified social norms. What, then, if anything, is wrong in these kinds of cases? This paper answers this question by conceptualizing public shaming as a sanction imposed on norm violators, and then developing a framework for assessing the justifiability of this sanction. Specifically, we identify five constraints on public shaming, and contend that a particular instance of shaming is justifiable if and only if each of these constraints is met. We then consider *online* public shaming in particular, and highlight several reasons to be concerned about this practice. Online public shaming frequently violates the constraints we identify, and the nature of online interaction makes it difficult, although not impossible, for those constraints to be fulfilled. Public shaming is justifiable only as a morally acceptable sanction, and many instances of online public shaming will not qualify as such.

We conceptualize public shaming as a practice of public moral criticism in response to violations of social norms. This practice is our focus throughout, and we take our opening cases to be paradigmatic examples. We use the term "public shaming" to refer to this practice for two reasons. First, "public shaming" has become the common term for this phenomenon. It has been popularized by Jennifer Jacquet, who also highlights the connection between public shaming and social norms, and Jon Ronson, who describes many of the cases that we discuss.[7] Of course, there are many things going on in these cases besides norm-reinforcing public moral criticism. For example, Jacquet also discusses public moral criticism intended to create new social norms, which falls outside the scope of our inquiry. Nonetheless, norm-reinforcing criticism is undoubtedly one practice commonly picked out by the term "public shaming". Second, "public shaming" is a philosophically appropriate label in so far as public moral criticism is likely to cause shame, the presence of shame may play a significant causal role in explaining its efficacy as a sanction, and these facts are relevant to our moral evaluation of the practice.

We recognize that this conceptualization differs from two alternative definitions of shaming that might seem natural. The first identifies shaming with actions aimed at inducing shame. We choose not to adopt this focus, since those who participate in the practice in which we are interested are not necessarily aiming to induce shame in their targets. Those who engage in online public shaming are seeking to draw attention to a social norm violation, and to rally others to their cause.[8] They may or may not also be aiming to induce shame. Furthermore, the phenomenon in which we are interested—that is, public moral criticism in response to violations of social norms—is a widespread phenomenon whose moral dimensions are underexplored, yet merit sustained analysis.[9] The moral status of actions aimed at inducing shame, and of the resulting experience of shame, in contrast, has received greater attention.[10]

The second definition identifies shaming with especially nasty forms of public criticism, such that only highly personalized, abusive, and vitriolic attacks that demean their victims count as "public shaming". In this case, public shaming would presumably be impermissible in (nearly) all cases. An advantage of our definition is that it corresponds to a widespread and philosophically interesting phenomenon, yet does not settle important moral disputes in advance. In contrast, a moralized definition identifying public shaming with demeaning and humiliating conduct makes it true by definition that there is something objectionable about the practice. All of the action would therefore be in determining what forms of criticism count as shaming.

Those who favor one of these alternative definitions—or indeed some other one—might consider our usage of "public shaming" to be a misnomer. Ultimately, we need not take a stand on this terminological dispute. The reader should feel free to adopt a different name for the practice on which we focus, such as "public blaming", if she thinks it important to do so. The key is that we are interested in a particular kind of response to social norm violations, namely public moral criticism of the violation. This is what we mean by the term "public shaming".

The paper is structured as follows. We first outline some of the pertinent features of social norms, and discuss the role of public criticism in upholding those norms. We then identify five constraints on public shaming. Next, we respond to an objection that claims that public shaming is always unjustifiable. We then apply our framework, focusing on *online* public shaming, and highlighting the features of online interaction that often lead to the violation of the constraints. We conclude with some reflections on what might be done about the widespread phenomenon of unjustified online public shaming.

## 2 | SOCIAL NORMS AND PUBLIC CRITICISM

Social norms are rules that govern the attitudes and behavior of members of a group, specifying how those individuals ought (not) to act. Four features of social norms are relevant for our discussion. First, norms specify standards to which members expect one another to adhere.[11] In order to be effective regulators of behavior, they typically must satisfy conditions for publicity. They must be (i) a matter of common knowledge, or at least widely known; (ii) sufficiently determinate, such that members know how they are required to act; and (iii) publicly verifiable, such that there can be widespread knowledge of the extent to which others comply with their demands.

Second, social norms concern a wider range of conduct than that regulated by law. These norms govern conduct that ought to be insulated from legal regulation, for principled or practical reasons, as well as conduct that should be subject to legal regulation, but in fact is not. Some theorists define norms as rules that lack an official source, such as government, and are not enforced through official sanction.[12] On this view, laws are distinct from norms. Others see norms and law as overlapping, with crimes being violations of social norms that society enforces through formal punishment.[13] Either way, all agree that norms encompass a litany of rules that go beyond the law—rules of etiquette, manners, fair play, respect, and so on.

Third, social norms are enforced through both internal and external sanctions. Members of the group generally take the norms to be authoritative for them. They accept or internalize these norms, such that they consider themselves duty-bound to comply with them and are likely to feel guilt or shame if they violate them. Sometimes, these internal sanctions may be sufficient to ensure that the norms are upheld. Often, however, this depends also on the

possibility that others will find out about norm violations and subject norm violators to criticism, as well as negative reactive attitudes. These external sanctions can morally fortify individuals who cannot reliably depend on internal sanctions doing the job.[14] Ideally, these external sanctions should not crowd out the internal sanctions. That is, it is preferable for individuals to uphold social norms because they recognize the moral reasons that support those norms, and not wholly because they fear losing social standing.

Finally, the combination of internal and external sanctions enforcing social norms makes them effective regulators of individuals' behavior.[15] As Emily McTernan notes, empirical research shows that "social norms are powerful determinants of behavior, and secure stable patterns of behavior from the majority of those who internalize the norm".[16] Further, they are essential for ensuring peaceful and cooperative coexistence. As Daniel Solove puts it, "Norms bind societies together; they regulate everyday conduct; they foster civility. They are the oil that reduces the friction of human interaction… In short, norms are a central mechanism through which a society exercises social control".[17]

As we have already intimated, public criticism is an important way in which social norms are upheld and reinforced.[18] It helps to distinguish three roles that such criticism can play. First, public criticism can play a _communicative role_. One aspect of this role is making the violator aware that she has violated a morally authoritative social norm.[19] Ideally, this will lead her to recognize that she has acted wrongly, feel remorse for this misconduct, and commit to complying with the norm in future.[20] These responses might be appropriately reflected in her apologizing for her action. A second aspect of the communicative role relates to the victims of wrongdoing, in cases where norm violations wrong specific individuals. Those individuals can recognize that others are unwilling to let the wrongdoing go uncriticized, and thus have a strengthened sense of solidarity with the group. In some cases, others might have _duties_ to criticize norm violators publicly, in order to stand with the wronged and avoid conniving with wrongdoing.

Second, public criticism can play a _deterrence-based role_ in reducing future norm violations. Individuals are less likely to disobey social norms if they believe that they will be externally sanctioned for doing so. In this way, public criticism can help to protect potential victims against future violations.

Third, public criticism provides a way in which to affirm morally authoritative social norms publicly. We can call this the _expressive role_. This role has instrumental value, in that it can strengthen our shared sense of commitment to the norm. This relates to our earlier remarks about publicity: public criticism can increase our common knowledge of (a) what the norm is; (b) what it demands in particular situations; and (c) when it has been violated. In these ways, public criticism enables us collectively to reaffirm our endorsement of the norm, and of the values it promotes or respects.

When social norms directly instantiate norm-independent moral requirements, public affirmation may also be intrinsically valuable. For example, we might think that it is intrinsically valuable publicly to condemn racism and other offensive views, irrespective of whether this serves any communicative or deterrence-based purpose. Such condemnation is in itself a fitting response to serious moral violations.

For criticism to play all three of these roles, or at least to play all three well, it must be _public_. While privately chiding the norm violator might successfully cause her to recognize her wrongdoing (one aspect of the communicative role), it is likely to be less effective as a deterrent. This is both because private criticism typically imposes fewer burdens than public criticism, and because others might not be aware of private criticism and so will not be deterred from similar norm violations. Moreover, it is essential for the criticism to be public if it is to reinforce our shared commitment to the norm (the expressive role) and to communicate solidarity with the wronged (the other aspect of the communicative role). Public moral criticism thus involves a practice of public accountability, in which individuals publicly hold one another responsible for norm violations in a way that serves several morally valuable purposes.

The public nature of such criticism also brings distinct complications, however, since it increases the risk that criticism will violate moral constraints. This is because, when criticism is conducted in front of larger audiences, or when more people participate, it becomes more difficult to control the effects of criticism. In turn, this fuels a number of concerns that jeopardize the justifiability of the practice, including worries relating to proportionality, respect

for privacy, and non-abusiveness—all of which we discuss below. For these reasons, criticism being public is a double-edged sword that can be a force for both good and bad. It is the public nature of public criticism that makes it particularly morally interesting and that calls for the attention that we give the phenomenon in this paper.

# 3 | CONSTRAINTS ON PUBLIC SHAMING

In this section, we identify five constraints on public shaming.[21] There is a justification for public shaming if and only if each of these constraints is met. This is *not* to say that any shaming act that meets these conditions is in fact justified. In order for that to be the case, the individuals who participate in public shaming must also have the right to perform this act. The fact that there is a justification for shaming, by virtue of the constraints being met, does not necessarily mean that everyone has the right to shame. The possession of this right might further require a certain kind of moral standing and/or acting with certain intentions. We lack space to explore these issues, and limit our attention to the prior question of when public shaming has a justification. To put this another way, our focus is on when public shaming is *justifiable* rather than when it is *justified*.

Two further clarifications are in order. First, our analysis identifies the moral principles that govern public shaming. There is a distinct question about which *laws and conventions* should govern this practice. We turn to this more downstream issue in the final section. Second, we provide an account of only the prima facie justifiability of public shaming. An instance of public shaming that is prima facie unjustifiable according to our framework may turn out to be justifiable all things considered if, say, it is necessary for avoiding catastrophe. We suspect there are few exceptions of this kind.

## 3.1 | Proportionality

An instance of public shaming is proportionate when its negative consequences are not excessive in comparison with its positive consequences.[22] Sometimes, the negative consequences are not excessive because they fall exclusively on an individual who has made herself *liable* to this treatment by culpably violating a morally authoritative social norm. The positive consequences morally outweigh the negative, and the negative are within the bounds of that to which the norm violator has made herself liable. When this is the case, public shaming is *narrowly proportionate*. In other cases, the negative consequences exceed that to which the norm violator is liable and/or they fall on third parties. In these cases, in which public shaming is narrowly disproportionate, the negative consequences count as excessive *unless* the positive consequences of public shaming are much more valuable than the negative consequences it imposes. When this is the case, public shaming is *widely proportionate*.

The standard for what counts as excessive is more demanding in calculations of wide proportionality than in calculations of narrow proportionality. This is because negative consequences are more difficult to justify when they fall on those who are not liable. This distinction between narrow and wide proportionality enables us to make sense of the judgment that costs that it would be disproportionate to impose on third parties may be proportionate when imposed on those who have acted wrongly.

The rest of this subsection further clarifies these ideas, to show their bearing on the case of public shaming. First, for public shaming to be proportionate—narrowly or widely—the social norm that it enforces must be *morally authoritative*, in the sense that individuals are under a duty to comply with it. This is not to say that the norm itself is morally required, or even morally optimal. Suppose we were to discover a new method of allocating seats on buses that was fairer and more efficient than queuing. This would call into question whether the current social norm is morally optimal. But it would not—immediately, at least—call into question whether compliance with the existing social norm is morally required. Given that queuing is acceptably fair and efficient, and creates goods of peaceful coordination, individuals still have duties to comply with the existing norm.[23]

The reason that shaming targeted at violations of norms with which individuals lack a duty to comply cannot be proportionate is that the norm violator will not have acted wrongly, and so will not have made herself liable to this treatment.[24] This is the case for both social norms that lack moral content (such as a social norm within a workplace that women wear skirts) and social norms whose moral content is not sufficiently weighty to generate duties (arguably, such as a social norm to take out one's rubbish on the morning of collection rather than the night before). It is consistent with this conclusion that other kinds of responses to these norm violations may be justifiable, such as private criticism or public persuasion.

Any account of when an individual is duty-bound to comply with a given norm will depend on one's broader moral and political commitments. However, it is essential not to limit this to only those norms that instantiate norm-independent moral duties. We should allow that individuals are under a duty to comply with some norms of politeness and etiquette, even though they do not directly instantiate norm-independent moral requirements. Adopting legal terminology, we might say that violations of such norms are *mala prohibita* rather than *mala in se*.

With regard to narrow proportionality, several factors affect the extent of an individual's liability, and thus the negative consequences they can justifiably be made to bear. Perhaps the most important of these is the seriousness of the norm violation. Other things equal, those who violate more important social norms—that is, norms with which we have stronger duties to comply—such as norms against racist speech, can be more liable than those who violate less serious social norms, such as norms of etiquette. Similarly, those who violate norms to a greater degree can be more liable than those who violate to a lesser degree.

A second factor that can affect liability is culpability. The norm violator's culpability increases with the extent to which she is responsible for the violation and not excused. Precisely which conditions must be satisfied here is a matter of intense dispute. But almost all accounts agree that, in order to be culpable, the violator must have been able to gain knowledge of the relevant norm and her conduct must have breached the norm in a way that she could have reasonably foreseen. This fits with our earlier discussion of publicity.

At this stage, it bears emphasizing that we restrict our attention to public shaming conceived of as a practice of public moral criticism in response to violations of social norms. That is, the cases in which we are interested are ones in which public moral criticism sanctions an individual who violates a prevailing social norm. It is in these cases that the violator's liability hinges on her being able to foresee that her action violates the norm. We set aside two sets of closely-related but distinct cases in which the relationship between liability and culpability is more complex. First, there are cases in which public moral criticism is used in response to an individual violating a pre-existing moral duty that is not embodied in any prevailing social norm. Second, there are cases in which public moral criticism is used to establish a new social norm. These cases are like ours in so far as they involve public moral criticism, but they are unlike ours in so far as this criticism does not arise in response to the violation of an extant social norm. For this reason, these cases fall outside the scope of our investigation, and thus our view implies nothing about the relationship between liability and culpability in these tricky cases. We limit the scope of our inquiry in this way precisely because these related cases bring with them further complications that we lack space to explore. Those complications warrant sustained analysis, but our aim here is to establish a framework for assessing the justifiability of the simpler cases that fall within our conceptualization of public shaming.

A third factor affecting liability is the amount of harm caused by the norm violation. There are difficult questions about how to characterize this condition. For example, should we take into account only the specific harms directly caused by this particular norm violation, or does the wider social context also affect the individual's liability, such as when this norm violation contributes to a broader harmful practice? Hunt's case is instructive here, assuming that he culpably violated a norm against sexism. If we take Hunt's remark as an isolated incident, then his liability will be fairly restricted, given the fairly trivial effects that his words were likely to have. However, if we see his comments as contributing to a wider harmful practice of sexism within the scientific community, as his critics did, then we might see his liability as much greater. This will seem objectionable to some, since it means that Hunt's personal liability is increased by others' wrongdoing. But harmful practices are often constituted by many small infractions, each of which does little harm when considered in isolation, but that on aggregate cause great harm.[25] If our assessment of

liability does not take this wider context into account, then we are likely to find that no one is liable to face public shaming. This seems unsatisfactory in the light of the overall harm, suggesting that the broader assessment is appropriate. Individuals can be highly liable due to contributing to a practice that causes great harm on aggregate.

In order to satisfy narrow proportionality, public shaming must produce some positive consequences. However, the size of the positive consequences necessary decreases as the seriousness of the norm violation, culpability, and harm caused increase. Negative consequences that would otherwise be excessive in comparison with the positive consequences may turn out not to be excessive once we are appropriately sensitive to the degree to which the norm violator has made herself liable.

When performing the narrow proportionality calculation, it is important to take into account the full array of positive and negative consequences. On the positive side, we have the goods that public shaming brings about—the fulfillment of the communicative, deterrence-based, and expressive roles. On the negative side, we must include any material costs that follow, such as the loss of one's job, as well as reputational, and psychological harms, such as feelings of distress, humiliation, embarrassment, and shame. One factor that is likely to affect the magnitude of these reputational and psychological harms is the size of the audience before whom an individual is shamed. Shaming carried out in front of a larger audience, or in which more people participate, might well better fulfill the three roles of public criticism, but it will also impose greater harms. The size of the audience and the number of participants is thus one thing that is governed by the proportionality constraint. Even if unrestrained, mob-like attacks on norm violators carried out before the general public are in some respects especially effective, they will often violate narrow proportionality.

## 3.2 | Necessity

In order to be justifiable, public shaming must effectively serve a communicative, deterrence-based, and/or expressive role and, crucially, it must achieve this outcome in a way that is no more harmful than any alternative mechanism that is not ruled out on other grounds. This constraint explains why it would be unjustified to shame an individual publicly if we could take an alternative course of action that served these roles equally well while imposing fewer burdens, providing that this alternative did not violate other constraints.

When judging whether a given instance of public shaming meets this constraint, we must consult all three of the roles that we have distinguished. Even if an act of public shaming lacked deterrent effects, for example, it might remain justifiable because of its communicative and/or expressive effects. Nonetheless, there certainly can be cases where public shaming is unnecessary to convey the wrongness of the norm violation to the violator, deter future violations, or to sustain the norm itself, such that this constraint is violated. Much public shaming might well take the form of "moral grandstanding" or "virtue-signaling", whereby participants indicate their own moral virtue without any of the roles of public shaming being served.[26]

The necessity constraint is importantly distinct from the proportionality constraint. Whereas the former requires comparing an instance of public shaming with alternative means that achieve the same aims, the latter requires a comparison between the positive and negative consequences of an instance of public shaming.[27]

## 3.3 | Respect for privacy

Public shaming involves drawing attention to the conduct of norm violators, but it is essential that it is carried out in a way that respects rights to privacy. This generates two requirements. First, public shaming should not involve dredging up irrelevant information or irrelevant facts about the norm violator's past. Criticism should focus on the relevant attitudes and behavior of the norm violator, not on wider facts about her life. It can be wrong to publicize such information, even when doing so is not disproportionate or abusive. This is because there is a right to privacy.[28]

Second, it may sometimes be wrong to publicize information that is relevant but *highly sensitive*. For example, perhaps we ought not publicly to shame someone for having an extramarital affair, on the grounds that this might be an entirely private matter. Norm violators may have rights against others publicizing certain sensitive information. The plausibility of this idea depends partly on which publicizing activities we have in mind. The right to privacy might condemn public shaming that involves making the information *fully* public by publishing it online or in newspapers, for example. But it is less plausible to hold that it also condemns shaming that reveals information only to a select group of individuals.

Violations of privacy will likely affect proportionality, but also provide independent grounds to consider instances of public shaming unjustified. Public shaming can be impermissible *because it violates privacy*, as well as because its invasion of privacy causes a violation of the proportionality constraint. To see this, note that the privacy constraint operates in a quite different way to the proportionality constraint. Whereas the latter asks us to consult the distribution of positive and negative consequences of public shaming, the former tells us that certain kinds of public shaming are ruled out from the start. Public shaming that violates privacy is off the table, so to speak, even if it is more effective than alternative forms of public criticism, and thus meets the proportionality constraint.

This constraint is interesting partly because of further questions it raises. For example, is public shaming more likely to be justifiable when targeted at politicians and other public figures? Those who respond positively must supply us with an explanation for this verdict. One possibility is that public figures have forfeited certain rights to privacy. An alternative is that information that is usually irrelevant is relevant with respect to public figures. For example, the private life of a politician may be relevant to our assessment of her vote-worthiness, since this may reveal something significant about her character. We lack the space to pursue these possibilities any further here.

## 3.4 | Non-abusiveness

This constraint rules out a range of abusive, insulting, and demeaning conduct from playing a role within public shaming.[29] First, threats of murder, rape, or grievous violence are clearly impermissible, even if they are empty threats and not perceived as credible. Second, this constraint prohibits speech that disparages or attacks the norm violator on the basis of their race, sex, or other socially salient characteristics. Third, it also rules out insults, mockery, and shaming that amounts to mere character critique, rather than focusing on the norm violation.

There are two justifications for this constraint. First, the forms of treatment that it prohibits are generally inconsistent with respecting the norm violator's moral status as a free and equal person. The inherently disrespectful nature of abusiveness means that it is never (or at least only exceptionally) appropriate to treat other people in this way.[30] Second, the conduct prohibited by this constraint is ill-suited to contribute to a practice of public accountability that sanctions norm violations and that strengthens our collective commitment to morally authoritative social norms. This practice requires criticism that is focused on norm violating acts, and that encourages the violator to recognize their wrongdoing and to express remorse. Threats of violence, insults, and racist speech do not appeal to the norm violator as a moral agent. Instead, they demean, belittle, and intimidate.

Both of these justifications are important, since in some cases one or the other might not apply. In some cases, abusive shaming might be highly effective as a deterrent, or vividly communicate to the norm violator the extent to which others disapprove of her conduct, such that one might argue that it fulfills some of the functions of public shaming (so the second justification does not apply). In other cases, insults or jibes might seem too mild to be disrespectful of the targets' status as free and equal (so the first justification does not apply). Together, however, the two arguments justify a constraint against a wide range of abusive conduct. Such conduct is prohibited when either of the two justifications applies. Most kinds of abusiveness will be ruled out by both arguments.

The range of conduct ruled out by this constraint is thus identified by referring to its two justifications. The precise extent of this range, and thus where we draw the line for what counts as abusive, is a difficult question. We surely should not rule out all criticism that in any way strays from the specific norm-violating act. Comments on norm

violators' more general character or vices can respect violators' moral status and play a productive role within public shaming. But such comments can also fall into disrespectful character critique, which should be prohibited. Similar vagueness surrounds what counts as an insult, or whether certain statements are sexist. We cannot draw precise boundaries here, but the fact that there are difficult cases does not show that this is not a genuine constraint.

Abusiveness also affects proportionality, but like respect for privacy it supplies independent grounds for objecting to instances of public shaming. Public shaming can be impermissible *because it is abusive*, as well as because its abusiveness leads to a violation of the proportionality constraint. Again, abusive forms of public shaming are off the table, so to speak, even if they are proportionate.

In reply, a critic might maintain that it is possible to conceive of cases in which abusive public shaming is justifiable. This might be the case if the norm violation is very serious and being abusive serves as a much more effective deterrent than any other means. We acknowledge the possibility of such cases, though we doubt that they are common. Moreover, even if abusive public shaming can in principle be justifiable all things considered, it remains prima facie unjustifiable by virtue of its abusiveness. Abusive public shaming always infringes rights, even if it does not always violate them.[31] By characterizing non-abusiveness as a constraint on the prima facie justifiability of public shaming, we maintain this attractive conclusion.[32]

## 3.5 | Reintegration

Public shaming must aim at, and make possible, the reintegration of the norm violator back into the community, rather than permanently stigmatizing them. Shaming appeals to the norm violator as a moral agent, so it must not exclude them from the moral community or permanently assign them an inferior status.

Toni Massaro draws a distinction between shaming and contempt: "shame-humiliation does not renounce the object permanently, whereas contempt-disgust does".[33] Within communities with healthy practices of shaming, it is "not understood... as an expression of contempt that permanently renounces or expels the shame[d] member",[34] but instead shows that norm violations will be criticized and punished, with the aim of reconciling the violator and the community. A central feature of reintegrative shaming is that it seeks, and is receptive to, repentance on the part of the shamed. It features as a means of moral education, seeking to bring the offender to an awareness of the wrongness of their conduct and to produce reformation.[35] In other words, public shaming need not, and should not, threaten an individual's membership in the community or their dignity.[36]

Non-reintegrative shaming is prohibited precisely because it constitutes a threat to the norm violator's dignity. It is therefore ruled out even if it could function as an especially effective deterrent, or was seen as strongly communicating solidarity with those wronged by norm violations. While these are important functions of public shaming, they should not be achieved at the cost of the norm violator's basic social standing.[37] This also explains why reintegration is a distinct constraint. A lack of reintegration will factor into proportionality, and non-reintegrative shaming will often be narrowly disproportionate, due to the severe negative consequences it imposes on the norm violator. But there is also a distinctive wrong in excluding someone from the moral community, rather than this simply being another negative consequence to take into account within proportionality analysis. Some public shaming might be proportionate, for example on account its deterrence effects, but nonetheless be impermissible due to failing to make reintegration possible.

A corollary of this constraint, and of our remarks about both non-abusiveness and proportionality, is that shamers should retain some control over the nature and extent of public shaming.[38] At the very least, shamers must ensure that they are aware of the risks of escalation, leading to violations of these constraints, and adjust their behavior accordingly. A lack of this kind of control is a central concern with respect to online public shaming, as we show below.

This completes our account of the constraints on justifiable public shaming. At this point, a critic might ask how we know that this framework is complete. Might there not be other constraints that we have overlooked? It is

impossible in principle to rule this out; clearly, we cannot consider every possible constraint one might suggest. However, the burden is on the critic to justify further constraints that are distinct from those we have outlined. In particular, they would have to show that the proposed constraint is not subsumed by existing constraints. Our suspicion is that many of the normative concerns underlying putative further constraints would be adequately captured by being factored into the analysis of proportionality. Further, below we directly apply our framework to online public shaming, and the plausibility of our analysis counts in favor of the completeness of the framework that we have developed. In other words, that the five constraints we have identified supply plausible results in a range of cases supports the verdict that they do not need to be supplemented. Before turning to this, however, we first consider and reply to a pressing objection.

## 4 | THE PROHIBITION OBJECTION

A critic might object that our analysis thus far ignores the most normatively salient element of public shaming: the fact that it causes shame. Shaming should not be a feature of our interpersonal relationships, because, given the nature of shame, shaming always involves demeaning another, suggesting that they are inferior or a worse kind of person.[39] Inducing shame in others is in itself wrongful, and ought to be prohibited. In other words, there is a constraint on shaming itself, just as we have claimed that there is such a constraint against, say, abusiveness. This is the *prohibition objection*.

This objection has particular force because shame is widely considered to be a particularly unwelcome, even morally defective, emotion. It is the emotion that one feels when one regards oneself as being negatively appraised by an audience that one respects, due to falling short of some standard of value that is accepted by that audience.[40] It thus threatens one's social standing and "is linked with the awareness of inadequacy, strangeness, limitation, or defeat".[41] Whereas guilt focuses our attention on the victims of wrongful acts, shame focuses on the "nature of the self that has done these acts".[42] Shame is "less directed at the wrong done than at how we appear, or how others will receive us, or what good or bad opinion we are entitled to have of ourselves".[43] For proponents of the prohibition objection, public shaming wrongs the shamed individual by condemning them to a lowered social standing and a sense of being a "bad person".

While this is an important objection, it can be defused. Three responses are in order. First, and most importantly, recall the nature of the shaming that is deemed justifiable by our framework: proportionate, privacy-respecting, and non-abusive public moral criticism, in response to violations of social norms with which individuals are under a duty to comply, that does not permanently stigmatize the shamed individual. While wrongs such as abuse, violations of privacy, or suggestions of permanent moral taintedness are often associated with shaming in practice (as we highlight below), they are distinct from it, and are ruled out by our constraints. The fact that shame is caused by shaming that meets our constraints is insufficient reason to prohibit it, even if such shame is an unwelcome or morally unattractive emotion. These harms are insufficient to justify the constraint on which the prohibition objection relies.

Second, shame is a likely consequence of the most effective forms of public criticism—indeed, this is why the term "public shaming" is appropriate. Public criticism will most effectively serve its various purposes when it is in response to a widely-endorsed norm that the norm violator accepts and internalizes. An individual who is rebuked by others in response to her violation of such a norm is likely to feel shame.[44] Such public shaming is likely to cause the norm violator to recognize her wrongdoing, and perhaps to make amends, and to be especially effective as a deterrent and as a means to strengthen our shared commitment to the norm. In other words, the moral goods that can be achieved through public criticism are most likely to be realized when that criticism causes shame. Further, the shame might sometimes not merely be a side-effect, but itself be causally efficacious. The fact that an act causes shame might sometimes be what *makes* it an effective form of public criticism, and so one that has a strong justification.

The point here is not simply that public shaming should not be prohibited because it can have good consequences. Rather, the point is that, in ruling out all acts that (foreseeably) induce shame, the prohibition objection is implausibly over-inclusive. The constraint would effectively rule out the whole practice of public moral accountability. The objection is more plausible if it is narrowed such that it applies only to particular kinds of public shaming, such as those that involve highly personalized attacks, character assassinations, and so on. But our framework already rules these out, since they violate the constraints we have identified.[45]

Third, the presence of shame will itself factor into the proportionality analysis, and thus affect whether the proportionality constraint is fulfilled. We have already seen that it can weigh in favor of public shaming, by increasing its positive consequences. But it also weighs against, since shame is itself a painful emotion, and thus a negative consequence. The harmfulness of shame is thus accounted for within calculations of proportionality.

## 5 | APPLYING THE FRAMEWORK

Public shaming can be justifiable. Consider a simple case in which Jane queue-jumps, skipping ahead of hundreds of people who are waiting patiently in line for tickets for a once-in-a-life-time gig. Some in the queue publicly criticize Jane for this, expressing their disapproval of Jane's conduct. This is a case of public moral criticism in response to the violation of a morally authoritative social norm,[46] and so is an example of public shaming. Moreover, it can certainly serve the three roles of public shaming: communicating to Jane her wrongdoing in a way that elicits remorse, deterring Jane and others from violating the norm in future, and strengthening the group's collective commitment to the norm.

Our five constraints on public shaming can be fulfilled in this case. First, those in the queue can make sure that Jane is aware of her wrongdoing, and induce appropriate levels of remorse, without disproportionately harming her. Second, there is no less harmful way to achieve the relevant goods, and thus public shaming is also necessary. Additionally, the comments can be made in ways that do not make use of private information, are non-abusive, and do not permanently ostracize. In these ways, the public shaming can satisfy the third, fourth, and fifth constraints.

Other cases of justifiable public shaming occur within particular associations and professions. Physicians have social norms regarding patient confidentiality and when disclosure of information is (im)permissible. Some of the duties that arise from these norms are legally enforced, but others are upheld through professional codes of conduct. If a doctor believes that another has acted in ways that violate confidentiality norms, by inappropriately revealing details regarding a patient's condition or carelessly leaving documents containing patient information within the view of others, for example, then it may be justifiable to shame the norm violator publicly, in front of their colleagues. The public nature of the shaming makes it possible for the norm violator to recognize that he has violated a social norm shared by his professional community, enables others to affirm the standards that he must live up to, and provides an opportunity for the resolution of any interpretative disputes regarding those standards. This criticism can also meet the constraints we identified.

What about online public shaming? We opened this paper with two stories in which individuals were shamed online for their words and actions. Can this kind of public shaming, carried out using social media, be justifiable?

There is reason to think that it can be. The online world features communities of individuals interacting and participating in a shared practice of communication and public expression, which can be used effectively to uphold morally authoritative social norms. The large-scale, democratic, and self-regulating character of these communities may even give us reason to be more hopeful about the role of public shaming in this sphere than in others. This kind of decentralized regulation (arguably) reflects the values shared by ordinary users of the Internet, and so is not the product of the decisions of a powerful elite.

Online public shaming can be used against violations of social norms that occur both in the offline world, such as in the cases of Stone and Hunt, and online. If an individual on Twitter violates a social norm, perhaps by tweeting something racist, then as fellow members of that community we might want to criticize this tweet and call attention to the fact that it fails to meet the standards to which Twitter users should adhere. Publicly responding to the offending tweet by expressing condemnation of it can encourage others to recognize the relevant norm and affirm their own commitment to it, and might cause the norm violator to acknowledge her wrongdoing. It may also have a deterrent effect. We might even believe that we have a duty publicly to condemn certain wrongful tweets in this way.

However, the features that make online public shaming attractive also generate risks. In particular, the highly decentralized nature of this practice makes it very difficult to control. This counts against the justifiability of online public shaming. Several of the constraints that we have outlined are hard to fulfill when public shaming happens online. All of them *can* be fulfilled, but they often are not in real-life cases, and the conditions of online interaction mean that there is always a high chance that at least some of the constraints will be violated. We expect that most readers will believe that there is something wrong with the kind of shaming involved in the cases with which we began. Applying our framework helps us to identify precisely why this is so.

## 5.1 | Proportionality

As we noted in relation to our opening cases, online public shaming often seems grossly disproportionate.[47] There are at least four important concerns regarding online public shaming's capacity to meet this constraint.

First, with respect to violations of social norms that occur online, we should note that online norms are at an early stage in their development. There are active debates about what standards of behavior we ought to uphold and whether certain forms of speech or expression, such as offensive jokes, are permissible. This means that an individual might not know the relevant norm, or might not be able to foresee that her action violates this norm. This lowers her culpability, which in turn reduces the level of negative consequences that she can be liable to bear.

Second, cases of online shaming often involve actions that are seemingly misunderstood or misinterpreted by those engaged in the shaming. The Stone case is of this kind, as is the (in)famous case of Justine Sacco, who tweeted "Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!"[48] Sacco meant this as an ironic comment about the attitudes of middle-class white Americans toward AIDS, but many considered it racist, leading her to face a barrage of online criticism and abuse. Sacco and Stone should have recognized that their actions could easily be understood as breaching social norms. Sacco's tweet appears racist and Stone's photo appears to mock the dead. On the other hand, both were acting ironically in order to cause amusement.

We might think that one can be culpable for foreseeably causing offence by *seeming* to breach a norm, even if one has not in fact breached that norm. Even so, there remains a duty to interpret others' actions in a charitable way. Stone's and Sacco's actions were certainly misjudged, but they should not be considered to have culpably violated norms in a way that makes them liable to public shaming. More generally, the risks of misunderstanding and misinterpretation within online public shaming are high. Justifiable criticism for norm violations must be based on a reasonable judgment that someone has culpably violated the norm, and this requires charitable interpretation of her words and actions. This kind of interpretative charity is often absent in online interactions.

Third, online public shaming is typically carried out by a large number of disparate individuals, none of whom has much control over the final outcome for the shamed individual. This can lead to forms of public shaming that are both narrowly and widely grossly disproportionate. We often observe "piling on", where an increasing number of individuals express their condemnation, leading the norm violator to experience deep distress and humiliation.[49] Even if most participants do not want their targets to bear extreme burdens, the nature of online public shaming means that such outcomes can easily occur. Moreover, the burdens typically fall on both norm violators and their associates, such as friends, family, and employers.

Finally, with respect to wide proportionality, there is an important worry that the practice of online public shaming will have negative systemic effects. Even if some individual instances of online shaming appear proportionate, the fear of being shamed might lead many—and especially those who are the most sensitive to others' views—to engage less in online interactions, or to refrain from posting anything that could be controversial. This could lead to an online environment that is less active and discursive. Ronson closes his book on public shaming by expressing this concern, quoting from a friend who no longer dares to post many of his jokes and observations online anymore. The friend states: "I suddenly feel with social media like I'm tiptoeing around an unpredictable, angry, unbalanced parent who might strike out at any moment".[50] The idea that the over-zealous enforcement of social norms might have a chilling effect that prevents free expression and limits individuality is of course familiar from Mill.[51] These dynamics pose a grave threat in the online world. Indeed, a recent report on social media trends found that "the concern around digital footprints, and what you publish or 'like' on social networks coming back to haunt you, is making young people increasingly self-censoring, risk averse and socially rigid".[52] This is an important, yet neglected, reason for why so much online public shaming is unjustifiable.

## 5.2 | Necessity

Regarding the necessity constraint, online public shaming is likely to be highly effective in playing the communicative, deterrence-based, and expressive roles. It is doubtful that there are alternative courses of action that can serve these roles just as well or better, in a less harmful way. Indeed, public shaming is one of the most effective tools we have for playing the communicative and expressive roles—arguably sometimes more effective than state-based punishment. Again, this is due to its large-scale, democratic, and self-regulating character, which allows it directly to reflect the values of participants. However, the decentralization that helps online public shaming fulfill the necessity constraint is also what causes it frequently to violate other constraints.

Further, as we noted above, some public shaming amounts to little more than "moral grandstanding", which will often violate the necessity constraint. To the extent that grandstanding serves the expressive role, it could do so just as effectively without being critically targeted at a specific norm violation, and thus without being shaming.

## 5.3 | Respect for privacy

This constraint can be fulfilled in principle, but in practice is often violated. This is illustrated by a very early case of online shaming.[53] In June 2005, a young woman's dog defecated on the subway in South Korea, and she refused to clean up the mess when urged to do so by other passengers. Someone took photos of her and posted them on a popular blog. Within hours, she had been labeled "dog poop girl" and the picture, along with various parodies, were widely circulated online. People requested information about the woman and her family, and her name, age, and school were soon revealed. In the words of one journalist, "online discussion groups crackled with chatter about every shred of the woman's life that could be found".[54]

The decentralized nature of online public shaming increases the likelihood of the violation of this constraint. It is difficult to prevent people from finding and spreading private information about shamed individuals. Invasions of privacy of this kind are such a common feature of online public shaming that a new term has been coined to describe them—doxing. Doxing is defined as searching for and publishing private or personally-identifying information about a person on the Internet, usually with malicious intent. The popular link-sharing and discussion website Reddit has introduced a strict rule against doxing, and threatens to ban violators.[55] This illustrates how members of online communities are aware of the risks of privacy violations, and are seeking to develop social norms that encourage respect for privacy. Online public shaming would be justifiable more often if such norms became well-established and respected.

## 5.4 | Non-abusiveness

This is another constraint that can be fulfilled, but often is not—as we saw in Stone's case. Again, the risk of this constraint being violated in any particular case is high, due to the large-scale and decentralized nature of online public shaming. The former makes it more likely that those with a propensity to abuse will become aware of the case and participate, while the latter makes such participation very difficult to prevent or control. Even if an individual intends to shame in a non-abusive way, the likelihood that others will be abusive is something that she should take into account. Even if her own contribution is non-abusive, her participation raises awareness of the case and makes it more likely that the criticized individual will face abuse.[56]

The extent to which an individual's participation increases the chance of abuse depends on her own prominence within the relevant online community—for example, the number of followers she has on Twitter. More prominent individuals should be particularly aware of the risks of exposing people to abusive attacks. Of course, prominent individuals also have a greater ability to bolster the relevant norm, precisely because they will draw more attention to the violation and thus cause more people to consider their own commitment to the norm. An individual's prominence thus gives her both more and less reason to engage in online public shaming: the putative benefits are greater, but so are the risks of the non-abusiveness constraint being violated. This point also applies to several of the other constraints we identify. For example, shaming from a more influential individual is more likely to be an effective deterrent, but also raises greater concerns of disproportionality.

A further grave concern is that online abuse is disproportionately directed at women and members of ethnic minorities.[57] Women are much more likely than men to receive threats of physical violence, rape, and murder.[58] This pattern is reflected in the abuse directed at Stone. It is striking and deeply troubling that those who are more vulnerable more regularly suffer such treatment.

## 5.5 | Reintegration

Public shaming can occur in a way that makes clear that it does not constitute a permanent rejection of the individual. However, there are limits to how clearly this can be done when shaming occurs on the Internet. Online public shaming often occurs through media, such as Twitter, that severely limit the space for explanation and are not well-suited to nuance. Furthermore, the mere fact that many people participate in the criticism and express disapproval makes it harder for the reintegrative aspect to be properly emphasized.

Reintegration also sometimes requires that misdeeds are forgotten, or can no longer be brought up. This raises a further concern regarding online public shaming because, in principle, online material can remain forever, and can be dredged up at any time. It thus threatens to be a permanent stain on an individual's record, "an indelible blemish on a person's identity".[59] We must overcome these hurdles to reintegration, perhaps through tighter regulation, if online public shaming is to be justifiable more often.[60]

In the light of the numerous concerns regarding online public shaming that we have identified, can it ever be justifiable in real life? We think that it can, and that a case from within academia illustrates this point. In January 2016, a furore broke out regarding a paper by Jean-Yves Béziau, published in a special issue of the journal *Synthese*. The paper contained a passage that many readers considered homophobic and sexist. The paper was discussed on social media and on several prominent philosophy blogs, including Daily Nous and Feminist Philosophers.[61] Many participants in these discussions were highly critical of both Béziau and the editors of *Synthese*. Many considered Béziau and the editors to have fallen below the standards of good scholarship and breached the norms of the academic community, lowering their standing within that community.

Plausibly, all of the constraints for justified public shaming were met in this case. A morally authoritative norm had been breached, and the shaming was proportionate.[62] The public shaming was also necessary, in the sense that there was no less harmful mechanism that would have been as effective in reinforcing social norms against

homophobia and sexism, as well as in encouraging journal editors to ensure that articles published with their journals do not violate these norms. Indeed, the editors of *Synthese* put a moratorium on special issues, in order to reexamine their policies.[63] The criticism respected the privacy of the relevant parties and was non-abusive. Moreover, there was no suggestion that Béziau was to be shunned from the academic community, or that *Synthese* was to be completely mistrusted as a journal. Overall, this seems to be a case where online public shaming was carried out in a way that met the constraints we identified, and so was justifiable.

## 6 | LOOKING FORWARD: ACCOUNTABILITY AND ANONYMITY

Our arguments raise the question of what we ought to do in response to the more general phenomenon of online public shaming, much of which is unjustifiable. A central solution is for us to develop clear and better social norms about the use of the Internet. We need norms that encourage restraint in criticism, sympathy in one's interpretation of others' words and actions, and a willingness to overlook past infractions so as to allow people to be restored into the community in good standing. We also need strong norms against doxing, abusive conduct, and seeking disproportionate punishment. In other words, while the unjustified enforcement of social norms is the central problem we have discussed, social norms also provide one of the most promising solutions.[64]

Central to achieving this is greater *accountability*. It is desirable that those who shame others can themselves be held to account, taking responsibility for their criticism and its effects, and being open to being criticized themselves. This requires that the norm violator, and perhaps others, enjoy a right of reply. This is desirable for two reasons. First, like due process laws, greater accountability acts as a safeguard against inaccurate and unwarranted criticism. It both deters potential shamers from engaging in such criticism and makes it easier for those wrongly accused of norm violations to correct these errors.[65] Second, the shamer's vulnerability to criticism displays her sincerity, which enables public shaming to play its various roles better. Others may be more inclined to take the shamer's claims seriously given her willingness to grant the right of reply.

Several of the concerns regarding online public shaming that we have already discussed recur here. Currently, online shaming is often questionably accurate, and there are no clear mechanisms for correcting unwarranted criticism. Moreover, its disparate nature means that no particular participating individual can be held fully to account for the consequence of online public shaming. This is a familiar problem with so-called "collective harms".[66]

Nonetheless, there might be ways in which we can increase accountability. Those who engage in online public shaming could come to recognize the benefits of accountability and endeavor to make themselves open to it, by being ready to listen to those they criticize, inviting open discussion of the norm violation, and admitting any errors they make. This kind of openness to accountability can be seen in the Béziau case: the shamers discussed what level of criticism was appropriate and were ready to listen to responses from Béziau and *Synthese*'s editors, as well as from third parties.[67]

It is important to highlight that those who respond to public shaming sometimes themselves engage in public shaming of the initial shamers, who they accuse of acting unjustifiably. This kind of exercise of the right of reply must itself abide by the constraints that we have identified. In practice, shamers often face a "shame backlash" that violates these constraints.[68] For example, Connie St Louis, who shamed Tim Hunt, suffered online abuse from those who considered her to have uncharitably misrepresented Hunt's words. This is clearly unjustifiable, whether or not one believes the shaming of Hunt to have itself been justifiable.

Importantly, if a shame backlash is severe then it might seem to jeopardize the justification for the initial shaming, by undermining its efficacy at fulfilling the communicative and deterrence-based roles. This is because public shaming that is rejected by its target and broader audience is less likely to elicit remorse or to act as a deterrent. This would be a troubling result. As Karen Adkins shows, this kind of unjustified retaliation is especially likely to be directed at members of minority groups, who often seek to enforce social norms that the majority express allegiance to but are inconsistent in following.[69] Perversely, it thus might seem that the risk of shame backlashes undermines

minorities' moral justification for public shaming, making it a tool that can be justifiably used by the powerful only. However, even shaming that provokes a backlash expresses solidarity with victims and plays an important expressive role, so its justification is not in fact extinguished, even if it is weakened. The risk of shame backlashes means that members of minority groups unfortunately have stronger prudential reasons to be cautious about engaging in public shaming,[70] but such shaming can still be justifiable. Of course, ideally, the right of reply should itself be exercised in justifiable ways, thus removing these concerns.

An issue closely related to online accountability is anonymity. Anonymity might well be considered a central reason that online public shaming so regularly fails to meet the constraints we have discussed, due to the way it reduces the accountability of shamers. Anonymity makes it more difficult for the shamed, and others, to exercise their right of reply, which makes it less likely that inaccurate criticism will be corrected. Also, anonymous shamers make themselves less vulnerable to criticism, which may make others less likely to take their claims seriously. In turn, this means that their actions serve the communicative and expressive roles less effectively, and so reduce the value of their public shaming.

Empirical evidence links anonymity to abusiveness. Summarizing psychological research in this area, Arthur Santana notes that "anonymity can foster a sense of impunity, loss of self-awareness, and a likelihood of acting on normally inhibited impulses in a way that is markedly inconsistent with a person's offline self".[71] Santana provides empirical evidence that anonymity increases the chance of abuse. He studied the comments on articles on the websites of 11 newspapers and found a markedly higher level of "incivility"—a category that included personal attacks, threats, abusive or hateful language, epithets, and racist sentiments—among anonymous comments. Specifically, 53% of anonymous comments were uncivil, compared to 29% of nonanonymous comments.[72] It is likely that anonymity also increases the chance of invasions of privacy.

These considerations might lead some to conclude that those engaging in online public shaming—and indeed online discussion more generally—ought not to post anonymously. Some online fora do not permit anonymity. Some newspapers require online commenters to use their real names, and Facebook requires that users use their authentic name and identity. These fora often appeal to accountability to justify these policies.

There are also considerations in favor of anonymity, however. The ability to write anonymously allows individuals to express unpopular or marginal views with less fear of being personally attacked or criticized. Anonymity increases the number of participants and the range of views articulated within online discussions.[73] Vulnerable individuals might be more likely to interact online if they can do so anonymously. Perhaps ironically, this argument in favor of anonymity is persuasive precisely because online discussion and criticism is so often conducted in an objectionable way. There would be less need for anonymity if abuse, invasions of privacy, and so on, were not so rife. While anonymity has been shown to *contribute* to these problems, it might also be *justified* in the light of those same problems. More specifically, even engaging in anonymous online public shaming might be justifiable, especially for members of marginalized groups, in the light of the risks of shame backlashes.[74] At the very least, anonymity has certain benefits in the world in which we live.

It is also worth noting that banning anonymity would not be a panacea. Santana found a troublingly high level of incivility even among nonanonymous comments—29% were uncivil. Again, this shows the broader pressing need for the kinds of better social norms we listed above.

Developing these kinds of norms is no small task, especially given the huge numbers of people who participate online. However, it is only through their emergence that the Internet can become a place where public criticism and shaming occurs in a justified way. In his 2007 book, Solove notes that the Internet was "now maturing into its second decade in mainstream culture—its teenage years".[75] The implication of this was that norms around its appropriate use, and the way online interactions ought to occur, were still emerging and developing. Ten years on, the need for greater maturity remains.[76]

## ORCID

*Paul Billingham* 🄳 https://orcid.org/0000-0003-2262-9064
*Tom Parr* 🄳 https://orcid.org/0000-0002-5873-6382

## ENDNOTES

[1] Ronson (2015, p. 202).

[2] Quoted in Foreman (2015). Some of the details of this quote are disputed. See Waddell and Higgins (2015).

[3] See https://twitter.com/connie_stlouis/status/607813783075954688?ref_src=twsrc%5Etfw.

[4] It bears emphasizing that Connie St Louis was herself subject to a backlash from defenders of Hunt. We address this feature of the case in the final section of the paper.

[5] Quoted in McKie (2015).

[6] As reported by Jon Ronson. See http://www.lbc.co.uk/sexism-shame-scientist-considered-suicide-124010.

[7] Jacquet (2016) and Ronson (2015).

[8] Norlock (2017).

[9] For example, Jacquet defends public shaming (conceptualized in a broadly similar way to us), but her normative claims are somewhat underdeveloped, since she is primarily concerned with the mechanics and efficacy of shaming. Our project is thus broadly compatible with Jacquet's, but our focus is different.

[10] Calhoun (2004) and Nussbaum (2004).

[11] Brennan et al. (2013, pp. 15–39). See also Bicchieri (2017, pp. 1–49).

[12] Klonick (2016, pp. 1035–1037).

[13] Solove (2007, p. 84).

[14] Howard (2017).

[15] This is not to imply that social practices are necessary for norms to exist. See Brennan et al. (2013, pp. 20–21).

[16] McTernan (2014, p. 95).

[17] Solove (2007, p. 6).

[18] This is also a central claim in Jacquet (2016), although she does not distinguish the three roles that we outline here.

[19] We define this term in Section 3.1.

[20] Duff (2001).

[21] Attentive readers will have noticed that we have switched to the language of "public shaming", which was absent from the previous section. The previous section analyzed the value of a practice of public moral criticism in response to violations of social norms. As explained above, we use the term "public shaming" to refer to this practice.

[22] We are indebted to Tadros (2011, chap. 15) and McMahan (2016).

[23] They may also have reason to seek to change the prevailing norm in favor of the fairer system, of course.

[24] This parallels the claim that all unjust wars are disproportionate. See McMahan, "Proportionality and Necessity in Jus in Bello".

[25] This is the structure of Parfit's famous "Harmless Torturers" example. See Parfit (1984, pp. 79–82).

[26] Tosi and Warmke (2016) and Norlock (2017) raise related concerns.

[27] McMahan (2016).

[28] Marmor (2015).

[29] Certain kinds of abusive language are also *legally* prohibited in some jurisdictions, such as the United Kingdom.

[30] Archard (2013).

[31] For this distinction, see Thomson (1990, pp. 82–104).

[32] Parallel comments apply to the case of privacy-invading public shaming.

[33] Massaro (1997, p. 653), quoting Silvan Tomkins.

[34] Massaro (1997, p. 682). See also Braithwaite (2000).

[35] Garvey (1998).

[36] Arneson (2007, p. 37).

[37] This does not mean that norm violators must always be welcomed back into every community of which they were a member. An individual who egregiously violates important professional norms might justifiably be permanently professionally ostracized. This individual should not be excluded from the moral community more generally, however.

[38] Whitman (1998, pp. 1088–1091).

[39] Nussbaum (2004, pp. 172–221).

[40] Kahan (1996, p. 636).

[41] Massaro (1997, p. 656).

[42] Arneson (2007, p. 38).

[43] Calhoun (2004, p. 128).

[44] This is not necessarily to imply that shame can be felt *only* when one personally endorses the norms or ideals against which one is judged. See Rawls (1999, pp. 390–391); Massaro (1997, p. 660); Calhoun (2004, p. 129).

[45] Proponent of the prohibition objection might hold that the label "public shaming" should be reserved for these kinds of acts. In that case, however, the disagreement is purely terminological. See our discussion in Section 1.

[46] We assume that Jane's society has a queuing norm that is morally authoritative, as a fair and efficient way to allocate scarce resources.

[47] Jacquet (2016, pp. 127–129) notes this concern, although her conception of proportionality is different to ours.

[48] Ronson (2015, pp. 63–77).

[49] Solove (2007, p. 95) and Klonick (2016, p. 1054).

[50] Ronson (2015, p. 268).

[51] Mill (1859, esp. chap. III).

[52] The Kantar Media report on Social Media Trends 2018, available at https://uk.kantar.com/tech/social/2018/gen-z-is-the-generation-taking-a-stand-for-privacy-on-social-media/.

[53] Solove (2007, pp. 1–3).

[54] Krim (2005).

[55] Reddit's Content Policy at https://www.reddit.com/help/contentpolicy.

[56] There is disagreement about how others' wrongdoing should feature in an individual's calculations about how to act. See Tadros (2016).

[57] Gardiner et al. (2016).

[58] Ronson (2015, p. 121).

[59] Solove (2007, p. 94).

[60] The so-called "right to be forgotten", which was recently endorsed by the European Court of Justice, is relevant here. See Warman and Barratt (2014).

[61] Justin W., "Hey did you know logical pluralism is 'connected to homosexuality'", *Daily Nous*, 20th January 2016, available at http://dailynous.com/2016/01/20/hey-did-you-know-logical-pluralism-is-connected-to-homosexuality; and "Homophobic and Sexist Rant in Synthese", *Feminist Philosophers*, 20th January 2016, available at https://feministphilosophers.wordpress.com/2016/01/20/homophobic-and-sexist-rant-in-synthese/.

[62] There is room for debate about precisely how culpable the editors were, given that the article was published in a special issue and that *Synthese* publishes large numbers of articles, and so some mistakes are inevitable. See Dutilh Novaes (2016).

[63] This moratorium was lifted 4 months later. See Justin W., "Synthese Editors Issue Letter on Special Issues", *Daily Nous*, 24th May 2016, available at http://dailynous.com/2016/05/24/synthese-editors-letter-on-special-issues/.

[64] This is not to say that the law has no role to play here. One possibility is to grant employees stronger legal protections against being fired by their employer on the grounds that the employee's actions might damage the firm's reputation and profitability.

[65] One might think that some form of accountability is more fundamental, perhaps serving as an additional constraint on justifiable public shaming. We tentatively reject this possibility, holding instead that public shaming can in principle be justified in the absence of accountability, given the instrumental nature of the concerns that motivate our focus on accountability.

[66] These problems mirror the complications that affect the proportionality constraint in cases such as Parfit's "Harmless Torturers".

[67] Justin W., "Statement from Synthese Editors / Moratorium on Special Issues", *Daily Nous*, 27th January 2016, available at http://dailynous.com/2016/01/27/statement-from-synthese-editors-moratorium-on-special-issues/; and Justin W., "Reply from Jean-Yves Beziau re: Synthese Controversy", *Daily Nous*, 29th January 2016, available at http://dailynous.com/2016/01/29/reply-from-jean-yves-beziau-re-synthese-controversy/.

[68] We take the term "shame backlash" from Adkins (2019), although our use of the term is slightly broader than hers.

[69] Adkins (2019).

[70] Of course, members of majority groups can also face shame backlashes, so have these prudential reasons. The risk is simply lower.

[71] Santana (2014, p. 23). See also Jacquet (2016, pp. 129–131).

[72] Santana (2014, p. 27).

[73] McCluskey and Hmielowski (2011).

[74] This is a further reason that accountability is not a *constraint* (see fn 65).

[75] Solove (2007, p. 4).

[76] Previous versions of this paper were presented at the May Workshop at the LSE, the seventh Braga Meeting on Ethics and Political Philosophy, a workshop on "The Political Theory of Social Norms" at the 2016 MANCEPT Workshops, a CELPA seminar at the University of Warwick, and a CSSJ seminar at the University of Oxford. Thanks to all who attended on those occasions, and especially to Marta Wojciechowska, who served as the paper's discussant at the LSE. We are grateful to many others for helpful discussions and/or written comments, including Paul Bou-Habib, Clare Burgum, Kimberley Brownlee, Simon Caney, James Christensen, Matthew Clayton, Gideon Elford, Cécile Fabre, Mollie Gerver, Sally Haslanger, Chris Mills, Jonathan Parry, Felix Pinkert, Jonathan Seglow, Tom Simpson, Adam Slavny, Ashley Taylor, Laura Valentini, Andrew Williams, and Caleb Yong.

## REFERENCES

Adkins, K. (2019). When shaming is shameful: Double standards in online shame backlashes. *Hypatia*, *34*, 76–97.

Archard, D. (2013). Insults, free speech and offensiveness. *Journal of Applied Philosophy*, *31*, 127–141.

Arneson, R. (2007). Shame, stigma, and disgust in the decent society. *The Journal of Ethics*, *11*, 31–63.

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. New York, NY: Oxford University Press.

Braithwaite, J. (2000). Shame and criminal justice. *Canadian Journal of Criminology*, *42*, 281–298.

Brennan, G., Eriksson, L., Goodin, R., & Southwood, N. (2013). *Explaining norms*. Oxford, England: Oxford University Press.

Calhoun, C. (2004). An apology for moral shame. *Journal of Political Philosophy*, *12*, 127–146.

Duff, A. (2001). *Punishment, communication, and community*. Oxford, England: Oxford University Press.

Dutilh Novaes, C. (2016, January 22). *In defence of journal editors who make mistakes. NEW APPS: Art, politics, philosophy, science*. Retrieved from http://www.newappsblog.com/2016/01/in-defense-of-journal-editors-who-make-mistakes.html

Foreman, J. (2015, September 1). The Timothy hunt witch hunt. *Commentary Magazine*. Retrieved from http://jonathanforeman.info/the-timothy-hunt-witch-hunt-commentary-sept-2015/

Gardiner, B., Mansfield, M., Anderson, I., Holder, J., Louterand, D., & Ulmanu, M. (2016, April 12). The dark side of Guardian comments. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments

Garvey, S. (1998). Can shaming punishments educate? *The University of Chicago Law Review*, *65*, 733–794.

Howard, J. (2017). Punishment as moral fortification. *Law and Philosophy*, *36*, 47–75.

Jacquet, J. (2016). *Is shame necessary: New uses for an old tool*. London, England: Penguin Books.

Kahan, D. (1996). What do alternative sanctions mean? *University of Chicago Law Review*, *63*, 591–694.

Klonick, K. (2016). Re-shaming the debate: Social norms, shame, and regulation in an internet age. *Maryland Law Review*, *75*, 1029–1065.

Krim, J. (2005, July 7). Subway fracas escalates into test of the internet's power to shame. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2005/07/06/AR2005070601953.html

Marmor, A. (2015). What is the right to privacy? *Philosophy & Public Affairs*, *43*, 3–26.

Massaro, T. (1997). The meanings of shame: Implications for legal reform. *Psychology, Public Policy, and Law*, *3*, 645–704.

McCluskey, M., & Hmielowski, J. (2011). Opinion expression during social conflict: Comparing online reader comments and letters to the editor. *Journalism*, *13*, 303–319.

McKie, R. (2015, June 13). Tim Hunt: 'I've been hung out to dry. They haven't even bothered to ask for my side of affairs'. *The Guardian*. Retrieved from https://www.theguardian.com/science/2015/jun/13/tim-hunt-hung-out-to-dry-interview-mary-collins

McMahan, J. (2016). Proportionality and necessity in Jus in Bello. In H. Frowe & S. Lazar (Eds.), *The Oxford handbook of ethics of war*. Oxford, England: Oxford University Press.

McTernan, E. (2014). How to make citizens behave: Social psychology, liberal virtues, and social norms. *Journal of Political Philosophy*, *2014*, 84–104.

Mill, J. S. 1859. *On liberty*.

Norlock, K. (2017). Online shaming. *Social Philosophy Today*, *33*, 187–197.

Nussbaum, M. (2004). *Humanity: Disgust, shame and the law*. Princeton, NJ: Princeton University Press.

Parfit, D. (1984). *Reasons and persons*. Oxford, England: Oxford University Press.

Rawls, J. (1999). *A theory of justice*(Revised ed.). Cambridge, MA: Harvard University Press.

Ronson, J. (2015). *So you've been publicly shamed*. London, England: Picador.

Santana, A. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, *8*, 18–33.

Solove, D. (2007). *The future of reputation: Gossip, rumor, and privacy on the internet*. New Haven, CT: Yale University Press.

Tadros, V. (2011). *The ends of harm: The moral foundations of the criminal law*. Oxford, England: Oxford University Press.

Tadros, V. (2016). Permissibility in a world of wrongdoing. *Philosophy & Public Affairs*, *44*, 101–132.

Thomson, J. J. (1990). *The realm of rights*. Cambridge, MA: Harvard University Press.

Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs*, *44*, 197–217.

Waddell, D., & Higgins, P. (2015, November 9). Saving Tim Hunt. *Medium*. Retrieved from https://medium.com/@danwaddell/saving-tim-hunt-97db23c6ee93

Warman, M., & Barratt, D. (2014, May 13). Google must delete your data if you ask, EU Rules. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/technology/google/10827005/Google-must-delete-your-data-if-you-ask-EU-rules.html

Whitman, J. (1998). What is wrong with inflicting shame sanctions? *The Yale Law Journal*, *107*, 1055–1092.