The Moral Risks of Online Shaming

Krista K. Thomason

The Oxford Handbook of Digital Ethics
Edited by Carissa Véliz

Subject: Philosophy, Moral Philosophy Online Publication Date: Nov 2021
DOI: 10.1093/oxfordhb/9780198857815.013.8

Abstract and Keywords

Shaming behaviour on social media has been the cause of concern in recent public discourse. Supporters of online shaming argue that it is an important tool in helping to make social media and online communities safer and more welcoming to traditionally marginalized groups. Objections to shaming often sound like high-minded calls for civility, but this chapter argues that shaming behaviour poses serious risks. It identifies the moral and political risks of online shaming. In particular, shaming threatens to undermine our commitment to the co-deliberative practices of morality. As a result, online shaming can undermine the very goals it is supposed to accomplish.

Keywords: online shaming, social media, shame, co-deliberation, racism, sexism

Introduction

Shaming practices are not new. Evidence of shaming practices can be found in ancient and pre-modern societies in many parts of the globe (Stearns 2017: ch. 2). Given shaming's long history, the prevalence of online shaming should not be surprising. The technology involved has shifted quickly and sometimes dramatically, but online shaming seems to follow it wherever it goes. Chatrooms, fan forums, blogs, and social media platforms can all provide examples of shaming. It seems that as long as we have been online, we have been finding ways to digitally shame each other.

Let me begin with a brief survey of the current literature about online shaming. Philosophical work on online shaming is not as extensive as one might think, given how the topic captures public attention. Perhaps fittingly, philosophers who have written about online shaming often do so for online media outlets. For example, both Beard (2016) and Raicu (2016) have discussed the moral problems with shaming in brief public pieces. Beard presents four common arguments in favour of shaming and then raises possible problems with them. His main goal is to examine 'the internal logic of mass online shaming' (2016). Raicu warns against online shaming because it resembles vigilantism and because it quickly becomes disproportionate to the original wrong (2016). Although these pieces are helpful and informative, since they appear in short-form essays, the arguments presented are brief. Two books written for popular audiences have thorough discussions of online shaming: Jon Ronson's *So You've Been Publicly Shamed* (2015) and Jennifer Jacquet's *Is Shame Necessary? New Uses for an Old Tool* (2015). Ronson is a journalist and a documentary filmmaker, and his book explores the downsides of online shaming by interviewing people who have been its targets. Jacquet's view is more optimistic; she describes shaming as a 'delicate and sometimes dangerous' tool that can help us solve problems (2015: 26). As an environmental studies scholar, many of Jacquet's arguments involve the shaming of corporations or organizations, so it is unclear how well her account applies to individual cases (2015: 174). Philosophers have written about the role of the emotion of shame in democratic politics and their arguments can be instructive in thinking about online shaming. Nussbaum (2004) and Locke (2007), for example, are sceptical of the value of shame in political life, while Tarnopolsky (2004) draws on Plato to defend it.

More recently, Norlock (2017), Billingham and Parr (2019), and Adkins (2019) have tackled online shaming directly. Norlock situates online shaming in the context of imaginal relationships (2017: 188). Our presence on social media proliferates online relationships. Because we are at a distance from the people with whom we interact on these platforms, we imaginatively fill in the context of those relationships (2017: 189–190). For example, people who shame online often construct the object of their shaming to be a person in a position of power who needs to be 'taken down a peg', even if that turns out not to be true (2017: 192). Norlock argues that we should be more cognizant of the ways that imaginal online relationships can affect us in real life in positive and negative ways (2017: 194). Billingham and Parr draw on John Locke's work to argue that online shaming can be an important informal social sanction, especially for racist and sexist behaviour (2019: 5–7). Online shaming can play the 'reparative' and 'restraint' roles that Locke assigns to punishment in the state of nature: it can lead the shamed person to make amends and it can deter others from engaging in the same behaviour (2019: 6). By contrast, Adkins calls into question the assumption that online shaming for sexism is as effective as it appears. She argues that some incidents of online shaming for sexist behaviour result in a 'shame backlash' against the shamers (2019: 77). On Adkins' view, successful shaming requires that the shamer has epistemic and social authority in the community, which most members of marginalized groups do not possess (2019: 82–83).

I will draw on this work as well as some related literature for the purposes of this chapter. Here, I wish to identify and articulate the risks of shaming. The risks that I have in mind are moral and political risks: shaming threatens to undermine some of our moral commitments or political goals. Given the risks and given that there are alternative and comparable strategies, even if shaming can be effective, we have good reason to avoid it.

Let me set some initial parameters for this chapter. My primary focus is online shaming that occurs between individuals. As such, I will not directly address cases in which individuals try to shame corporations or organizations. I also will not directly address shaming legal punishments because these cases introduce questions about the justification of punishment and the relationship between citizens and the state. My primary examples will involve the kind of shaming that occurs online, which means that in these cases most people are shamed for types of speech. I will sometimes refer to these remarks as 'actions' or 'misdeeds'. I do not make a sharp distinction between speech and action, although I recognize some may argue that shaming is justified for one and not the other.

What is shaming?

To explore questions about online shaming, we have to first understand what shaming is. 'Shaming' is a broad term that often encompasses (at least) the following things: *feelings of shame, shaming practices*, and *stigmatizing* (Thomason 2018: ch, 5). Philosophers draw distinctions between these concepts differently. For example, Tarnopolsky distinguishes between the 'occurrent experience of shame' and 'acts of shaming' (2004: 475). Roughly, Tarnopolsky's distinction tracks the difference between feelings of shame and shaming practices. Adkins (2019) and Nussbaum (2004) draw a close connection between shaming practices and stigmatizing. As Adkins writes, shaming is a 'stigmatizing judgment, where an actor or group condemns another actor or group for failing to adhere to a shared ideal or norm' (2019: 77). Shaming practices, feelings of shame, and stigmatizing are, of course, related concepts. Yet because they can all occur independently from each other, it is a mistake to theorize them together.

Although it is reasonable to think that an act of shaming is meant to induce feelings of shame, there is no necessary connection between the two. Shaming someone might not make that person feel shame. He or she might think that someone's attempt to shame them is laughable and dismiss it as ridiculous. Members of stigmatized groups might embrace that membership and celebrate it rather than feel shame about the stigma. The movement to fight 'slut-shaming'—the stigmatizing of women for having multiple sex partners—is an example of an unashamed response to both shaming and stigmatization (Poole 2013). Feelings of shame can also arise when we are neither shamed nor stigmatized. Philosophers are divided on the question of whether an actual or imagined audience is required to experience feelings of shame (Taylor 1985; Velleman 2001; Calhoun 2004; Deonna et. al. 2012; Thomason 2015, 2018). Shaming and stigmatizing, by contrast, seem to require an audience. Shaming also should not be confused with cases in which we use the phrase 'Shame on you' in conversation. Often, 'Shame on you' is a stand-in for 'You have done a bad thing' or 'You should feel bad about yourself for what you have done' (Thomason 2015: 18). We can, of course, try to induce feelings of shame in others, but shaming practices are only sometimes undertaken for this purpose.

The kind of shaming that occurs online is usually an act of shaming or an instance of a shaming practice. To get a better sense of what shaming is, consider one of the most famous cases of shaming: Hester Prynne from *The Scarlet Letter* (Hawthorne 1850/2003). Hester is a member of a Puritan community during the 1600s and is forced to wear the scarlet letter on her clothing because she has an affair that produces a child. Hester's shaming is complicated: it is simultaneously a legal penalty, a tool to get her to confess the name of her lover, and the public display of her sin, both for her own sake and for the sake of the community. It is this final aspect of her shaming that illustrates the key features of the practice. Philosophers have noted that a certain kind of publicity is central to shaming (Jacquet 2015; Thomason 2018; Adkins 2019; Billingham and Parr 2019). Shaming practices require what I have called the ‘marshalling of communal attention’ (Thomason 2018: 181). In order to shame someone, her flaw or offense must be pointed out to others. Of course, not all shaming happens in large public forums. We can shame others in front of a handful of people, but the practice assumes an audience who is there to see the shamed person’s flaw or misdeed (Adkins 2019: 77). In wearing the scarlet letter, Hester is meant to be a ‘living sermon against sin’ (Hawthorne 1850/2003: 58); that is, the scarlet letter is meant to draw the community’s attention to her.

What is the purpose of marshalling communal attention towards the shamed person? There are generally two motivations for shaming. First, we might say that shaming is meant *to inspire self-consciousness or self-awareness* (Tarnopolsky 2004; Thomason 2018; Billingham and Parr 2019). We can see this in Hester’s shaming: it is meant to make her appreciate the seriousness of her sin. As such, her shaming is meant to inspire a change of heart or moral self-reflection. Second, shaming is meant to *send a message of condemnation on behalf of and to the community* (Nussbaum 2004; Tarnopolsky 2004; Norlock 2017; Thomason 2018; Adkins 2019; Billingham and Parr 2019). I am using the term ‘community’ in a loose sense here. ‘Community’ can refer to a small group, a group of peers, or, in the case of social media, members of the online community. In Hester’s case, shaming is a form of censure that tells her she has done something that her community finds unacceptable; she has failed to live up to or take seriously the values of that her community holds dear. Additionally, Hester serves as an example of what not to do for the benefit of onlookers. Her shaming is not just directed at her, but it is also directed at her peers. The rest of the village receives the message that this behaviour is unwelcome or prohibited. Shaming is thus an expression of condemnation both from and to the rest of the group.

The question that is the focus of most of the literature on online shaming is whether this practice is morally good or morally justifiable. We should be careful here to distinguish the efficacy of shaming from the ethics of shaming. For example, Jacquet’s (2015) arguments in favour of shaming are mostly arguments claiming that it is a powerful tool in shaping the behaviour of powerful corporations. We might grant her conclusion and yet nonetheless argue that online shaming is unjustified for other reasons. Compare such arguments to arguments that punishing the innocent effectively deters crime: this may be true, but it would not automatically license the conclusion that such measures are morally justified. Philosophers are divided on the issue of effectiveness. Adkins argues that online shaming can produce shame backlashes, which may diminish its effectiveness (2019: 76). By contrast, Billingham and Parr cite cases of shaming that led offenders to apologize, make amends, and change their future behaviour (2019: 1–2). Questions of shaming’s effectiveness are certainly relevant to its justifiability. If some argue that we should use shaming because it is an effective way to police community norms, and yet it turns out not to be effective, we ought to rethink its use. Nevertheless, I will focus most of this chapter on questions of justifiability apart from effectiveness; that is, even if we assume that shaming is effective, the controversy over its justifiability is not settled.

The case in favour of shaming

Let us first construct the case in favour of online shaming. When and why might shaming be a good thing to do? Shaming online is pervasive, and people are shamed for a wide range of alleged offences. Optimists about online shaming do not defend every instance of it. Typically, they argue that the most compelling candidate for good shaming is shaming directed towards racist and sexist behaviour (Jacquet 2015; Thomason 2018; Adkins 2019; Billingham and Parr 2019). Since this is taken to be a strong candidate for justified shaming, I will focus on these types of cases. Supporters of shaming argue that it is an important tool in helping to uphold and enforce community norms and values. As Billingham and Parr put it, ‘public shaming provides a way that we can express our endorsement of valuable social norms, thus strengthening our shared sense of commitment to those norms, and the values that they promote or respect’ (2019: 7). Social media platforms are most often the places where we see the ‘calling out’ of racist and sexist behaviour, though shaming practices need not be confined to social media. Any online forum where there is a recognizable community and where communication is visible to its users is a place where shaming can occur.

To get a sense of how an episode of online shaming works, consider the case of Hank and Adria from Jon Ronson’s book, *So You’ve Been Publicly Shamed* (2015). At a technology conference, Hank made a joke (containing a sexual innuendo) to another male friend during one of the presentations (Ronson 2015: 114–115). Adria was sitting one row in front of the two men and heard the joke. She turned around, took their picture and posted it to her Twitter account with a message indicating what Hank had joked: ‘Not cool. Jokes about forking repo’s in a sexual way and ‘big’ dongles right behind me’ (2015: 114). Adria was retweeted thousands of times. She then wrote a blog post explaining why she had done it: ‘Yesterday I publicly called out a group of guys at the PyCon conference who were not being respectful to the community …Yesterday the future of programming was on the line and I made myself heard’ (2015: 115–116).

Notice that Adria articulates two common motivations for shaming identified in the previous section (‘What is shaming?’). Adria wants Hank to realize that his joke was offensive (self-awareness), and she wants the rest of the tech community to realize that jokes like his do not reflect the values that the community should stand for (sending a message of community condemnation). She marshals communal attention by posting Hank’s joke to her Twitter account. In doing so, she invites the community to likewise judge the joke as offensive. She intends for them to join her in condemning it and to refrain from making similar jokes.

Online shaming in cases like this seems positive for a number of reasons. Consider first how it inspires self-awareness in the offender. Racism and sexism are morally offensive and the people who express racist and sexist remarks are doing something morally disrespectful. Shaming is therefore a way of showing the people making these comments that they have done or said something disrespectful (Billingham and Parr 2019: 6). If people make such remarks online, the fact that other members of the community can ‘call them out’ for it might inspire the self-reflection they need to realize what they said was wrong and make appropriate amends. Being subject to public criticism for racism or sexism should, at the very least, make someone think twice about doing it again. Also, shaming people for these sorts of remarks can allow others to stand up for people who have been subject to this kind of invective and allow the targets of the invective to communicate their offence. As Billingham and Parr put it, ‘shaming helps to protect potential victims against future violations’ (2019: 7). According to arguments like these, shaming functions as another kind of moral criticism or condemnation. In the same way that I would object to a sexist joke made in my presence, online shaming is a way of objecting to offensive online communication.

Additionally, shaming can act as a form of condemnation when other kinds of consequences are unavailable or not forthcoming (Jacquet 2015: 106; Billingham and Parr 2019: 5–6). Sexism and racism expressed online are often unpunished and unacknowledged. Women of all races are routinely told that they are overreacting or imagining things when they report misogyny online. Similarly, complaints of racism made by people of colour are either treated as not serious or as imagined. Philosopher George Yancy, for example, was targeted with racist abuse after he wrote an opinion piece on racism called ‘Dear White America’ for the *New York Times* (Evans and Yancy 2016). Shaming provides the targets of abuse with a way of showing that these things in fact occur and that they are not just ‘making it up’.

Finally, shaming for racism and sexism allows the community to uphold the values it cares about. In this way, it functions as a sanction that can empower vulnerable members of the community. Shaming is one of the few strategies that the community can use to regulate the behaviour of its members—social media is free and open to everyone, and platforms like Twitter cannot (or perhaps will not) shut down profiles of people who say offensive things that stop short of threats. Because any member of the online community can call the community’s attention, each person can be a standard bearer for community values (Etzioni 1999: 47; Solove 2007: 92; Jacquet 2015: 26). At its best, then, shaming is a tool that can help to shape communities according to morally desirable norms. As Arneson puts it, ‘A society that strives to be just cannot dispense with tools that help get the job done’ (2007: 32).

The moral risks of shaming

Many philosophers have expressed scepticism about the promise of online shaming, and objections to the practice take a variety of forms. Here I have divided them into three broad risks: the risk of disproportionality, the risk to co-deliberation, and the risk to creating safe communities.

Risk of disproportionality

One of the problems with online shaming is that the negative attention the shamed person receives is often far worse or more sustained than the harm or offense caused by the original misdeed. A commonly cited example is the case of Justine Sacco. Sacco was a public relations manager for a magazine. In December of 2013, she was flying from the United States to South Africa to visit family and tweeted the following joke: 'Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!' (Ronson 2015: 68). Although Sacco meant for the joke to be an ironic commentary on the obliviousness of Americans, Twitter users did not read it that way (2015: 78). Her joke was retweeted thousands of times and she became a trending topic. When she landed in South Africa, people followed her through the airport, snapped her photo, and posted it to Twitter (2015: 71–72). She eventually lost her job, suffered from insomnia, and had trouble meeting new people because Google search results continued to link her to the Twitter incident (2015: 79–80).

Shaming sceptics point to cases like this one to show that online shaming can have much larger effects than it appears (Solove 2007: 95–96; Norlock 2017: 188–189; Thomason 2018: 192; Billingham and Parr 2019: 2). There are two senses of disproportionality to which critics object. First, the shamed person receives far more negative attention than he or she deserves. Second, online shaming can have lasting or permanent effects that it should not have. To use Sacco as an example, even if we grant that her joke was offensive, having thousands of strangers condemn her for it is excessive. Moreover, in person, one offensive joke would not be sufficient grounds to fire someone from a job, nor would it license being followed and having one's photo taken by strangers. The consequences she suffered were too harsh given the nature of her offence. Additionally, sceptics of online shaming often follow the objections presented by Nussbaum, who argues that shaming stigmatizes another person as having a 'spoiled identity' (2004: 230). On her view, shaming sends the message to the offender and to the community that the offender is irredeemable and is no longer welcome as part of the community. This kind of permanent stain is incompatible with the belief that the shamed person is of equal worth and dignity (2004: 239).

Optimists about online shaming have responded to worries about proportionality. First, they argue that online shaming is only justified when the shamed person has culpably violated a serious moral norm (Billingham and Parr 2019: 8–9). This provision would prohibit online shaming that is mere bullying or revenge (Solove 2007: 98). Second, they argue that online shaming must be reintegrative: there must be a way for the shamed person to make amends or return to the community (Billingham and Parr 2019: 10–11; Solove 2007: 95). Although instituting reintegrative shaming online is difficult, philosophers have argued that greater awareness of the dangers and effects of shaming might help to dampen its negative effects (Norlock 2017: 193–194; Billingham and Parr 2019: 13). Additionally, philosophers have argued that shamers bear responsibility for ensuring that they are shaming in morally justifiable ways. Shamers ought to better consider the well-being of the target of their shame (Norlock 2017: 194). As Billingham and Parr put it, 'Those who accuse others of violating social norms should be willing to listen to the other side of the story and consider whether their criticisms might be misplaced' (2019: 14). Sceptics argue that such responses paint an overly rosy picture of online communication. We might hope that people will behave better on social media, but experience tells us not to expect it.

Some have advocated for a technical solution to the problems with online shaming. For example, Basak et. al. (2019) designed an application called BlockShame, which would be used on Twitter to limit online shaming. BlockShame uses an algorithm to detect shaming tweets and automatically block the shamer who sends them (2019: 217–218). Since one of the problems with online shaming is its context sensitivity, it is unclear whether an algorithm would be sophisticated enough to detect shaming reliably. Although the ethics and technology of online communication can surely present novel solutions to online shaming, it is unclear whether those solutions will satisfy sceptics. Optimists think that online shaming can be justified, provided it is done well, and mitigating concerns about disproportionality may help to show that it can be done well. Yet there are sceptics who will reject online shaming no matter how well it is done. Those rejections likely stem from the concerns that extend beyond disproportionality.

Risk to co-deliberation

The second risk is what I will call the risk to co-deliberation. There are two types of co-deliberation I will examine here: epistemic co-deliberation and moral co-deliberation.

Levy's chapter in this volume, 'Fake News: Rebuilding the Epistemic Landscape', discusses the negative effect that fake news has on our epistemic landscape. Although the objections to shaming are usually moral, some sceptics are concerned that shaming might have similar negative effects. These objections are broadly Millian in nature. In chapter 2 of *On Liberty*, John Stuart Mill makes his famous case that suppressing views—even those we believe to be false and offensive—is illegitimate on two grounds: 'We can never be sure the opinion we are stifling is a false opinion, and if we are sure, stifling it would be an evil still' (1859/2003: 88). Proper pursuit of truth, in Mill's view, is a collective endeavour that requires free and open conversation. It is only 'by discussion and experience' that we are able to correct our mistakes in reasoning and our false beliefs (1859/2003: 90). Sceptics will argue that shaming can be a form of suppressing free and open conversation. The target of online shaming is the object of communal condemnation and is typically prevented from further participating in the conversation. Further, when other people witness instances of online shaming, they may be less willing to speak their minds. Online shaming may therefore have a 'chilling effect' on speech (Billingham and Parr 2019: 12). According to Mill, the chilling of speech makes us worse off epistemically: not only will we lose the opportunity to learn from others and test our own opinions against theirs, but we also risk holding our own views as 'dead dogma, not a living truth' (1859/2003: 103).

We can see a possible example of the negative effects of shaming in online echo chambers. As Nguyen argues, echo chambers are epistemic communities that create a disparity of trust between members and non-members: members of echo chambers portray non-members as unreliable or dishonest (2020: 146). Once non-members are the objects of distrust, the members of the echo chamber become more impervious to any claims or evidence that contradict their beliefs (2020: 147). Shaming might be used as a way to discredit certain voices within a community and create echo chambers. In drawing communal attention to the shamed person's flaw or misdeed, others tend to see the shamed person as defined by that flaw or misdeed (Thomason 2018: 205–206). If the shamed person is seen as flawed on the whole, he or she is more likely to be dismissed or discredited. Sceptics may claim that widespread use of shaming can create echo chambers and damage our abilities to hear opposing views.

Optimists will respond that free expression does not entail freedom from criticism, but we must ask whether shaming is simply another form of criticism. To explore this question, let us now turn to the second risk: the risk to moral co-deliberation. On the surface, it appears that online shaming is simply another form of moral criticism or moral communication. Tarnopolsky argues in favour of shaming on these grounds because it causes a 'potentially salutatory discomfort' in which someone comes to a newfound self-awareness (2004: 479). This kind of moral communication is most familiar in interpersonal cases. Suppose I have said something hurtful to a friend and others have witnessed the incident. Someone might say something to me along the lines of, 'How would you like it if she said that to you?' This person is attempting to get me to realize or appreciate that what I said was hurtful by getting me to see my words differently than I currently see them. She is hoping that, if I come to see my words differently, this realization will lead me to perhaps apologize or at least lead me to stop saying similar hurtful things (Thomason 2018: 179). If sexist and racist speech—whether in person or online—is wrong or hurtful, online shaming seems to be no more controversial than other kinds of negative attention we direct towards anyone who commits a moral wrong (Billingham and Parr 2019: 8–9).

Yet there is a difference between shaming and other kinds of moral communication: shaming requires the marshalling of communal attention (Thomason 2018; Adkins 2019; Billingham and Parr 2019). We can—and often do—communicate disapproval to those who are disrespectful without drawing the attention of others. In interpersonal communication, we direct our disapproval directly to the offending party. If, for example, someone makes an offensive joke in my presence, I might object and say, 'That is not funny.' By contrast, shaming publicizes the person's wrongdoing and calls on other people to direct disapproval to the offending party. In the case of the offensive joke, I would shame the offending party by calling out to a group of people, 'This person just made an offensive joke.' In typical interpersonal communication, we do not put the offending party on display merely by protesting her actions or comments. The publicity or drawing of communal attention distinguishes shaming from regular moral communication. In light of this, sceptics worry that shaming communicates with others in ways that are morally problematic.

Some philosophers phrase this concern as a lack of due process. For example, Nussbaum argues that shaming is akin to 'justice by the mob', which is not 'deliberative, impartial or neutral' (Nussbaum 2004: 234). Solove objects that 'There are no rules and procedures to ensure that the internet norm police are accurate in their assessments of who should be deemed blameworthy' (2007: 97). Others talk about this risk in terms of the context-sensitivity of norms. Billingham and Parr warn that what will count as a norm violation is dependent on a particular social context (2019: 8–9). Norlock makes a similar point about the nature of jokes online—a joke is a 'social thing' that has to be taken up as such in order to count as a joke (2017: 192). In offline communication, norms vary with our social context and it is not always clear when someone violates those norms. Online communication exacerbates this problem. For example, social media posts are generally short and do not invite nuanced distinctions. As Roache shows in this volume (What's Wrong with Trolling?), activities like trolling rely on sarcasm, intentional insincerity, and bad faith arguments. We often struggle to discern the basic meaning and intent of online communication. Sceptics will argue that, since it is hard to tell what norms apply to online communication and when those norms have been violated, we are rarely (if ever) in a position to know whether someone deserves to be shamed.

Underlying both of these concerns is a certain conception of moral communication: a moral community must discover together which moral norms are authoritative through a process of co-deliberation (Scanlon 1998; Calhoun 2000, 2004; Walker 2007). As Walker argues, moral life is 'a continuing negotiation *among* people' (2007: 67, emphasis original). We do not, in other words, come to have moral knowledge of either our own or others' behaviour without 'collaboration and communication in identifying moral problems and resolving them' (Walker 2007: 61). As moral co-deliberators, we figure out together where moral obligations, responsibilities, and violations lie. To be able to receive and offer moral evaluations, we operate with a 'presumption in favour of accounting to others and trying to go on in shared terms' (Walker 2007: 69). In other words, we see each other as part of a moral conversation the terms of which are not unilaterally dictated and the point of which is to come to collective decisions about how to live together.

Shaming is in tension with this picture of moral co-deliberation. To see this, let us look more closely at the way shaming is supposed to communicate. Shaming is supposed to inspire moral self-reflection because it is unpleasant and uncomfortable for the shamed person to be the target of people's anger. As such, shaming relies on what Scanlon calls a 'sanction' model of moral criticism (1998: 268). The painful weight of public disapproval is a negative consequence for bad behaviour. Notice, however, that the sanction model of moral criticism does not directly aim to get the shamed person to appreciate what was hurtful or disrespectful about his or her remarks. The weight of disapproval sends the message that these remarks were out of bounds, but the disapproval does not by itself communicate why they were out of bounds. Also, the shamed person is offered no opportunities to respond to the negative attention. Shaming does not presume *mutual* accountability; it presumes that the shamed person is accountable to the group and not the other way around (Billingham and Parr 2019: 14). Particularly on social media, the shamed person is expected to simply accept the disapproval and preferably apologize rather than try to respond to the charges. The shamed person is offered no say in the process of hammering out whether or not his or her remarks were offensive or out of bounds. If we are truly committed to the idea that moral life consists in ongoing negotiation, we are not licensed to unilaterally exclude others from participating in the negotiation process, even when they have said or done something wrong—as Walker puts it, this negotiation 'occurs in real time' (2007: 71). Similar to the risk of epistemic co-deliberation, the marshalling of negative attention that is central to shaming typically has the effect of simply silencing the shamed person. Shaming is risky because it closes off the negotiation process that is essential for working out our moral understandings. It is, of course, possible for shaming in some ideal situations to be compatible with co-deliberation, but sceptics will say that in practice this compatibility is unlikely.

One could object that the targets of shaming do not deserve to be treated as co-deliberators because they refuse to recognize the standing of others as co-deliberators. As Billingham and Parr argue, the strongest cases for online shaming are directed towards people who express racist and sexist views (2019: 8). Is there no place for forms of social sanction within the co-deliberative model, especially when the violations are serious? Why not sanction people who flagrantly refuse to treat marginalized groups as co-participants? As Arneson argues, shaming is a response to a 'failure to comply' with norms that we think are valuable and important (2007: 38). According to this position, shaming is a way of enforcing anti-racist and anti-sexist norms that the community has already decided to uphold.

Although this argument is compelling, it relies on the claim that our moral norms have already been decided and that violations of them are easy to identify. If we are committed to moral co-deliberation, exactly when norm violations occur will be a matter of debate and negotiation. Since morality is an ongoing practice of negotiation, the boundaries of moral violations are contestable—as Walker puts it, there is no 'pure core of moral knowledge' (2007: 73). This is not to say that there are no moral violations, but a commitment to morality as co-deliberative requires that we are ready to articulate our reasons for thinking that a violation has taken place and ready to respond to those who disagree. Only through moral protest and renegotiation is such behaviour finally seen as disrespectful. It is tempting to think that if an online user makes an offensive remark, we are licensed to shame him or her rather than draw them into conversation because in making an offensive remark they have demonstrated that they are 'on the wrong side'. Yet if morality is co-deliberative and there is no pure core of moral knowledge, we cannot determine who has the wrong views without the process of negotiation. Being a co-deliberator is not conditional on having the right views. Determining which views are right and which are wrong is something we must figure out together. As such, being open to challenge is something all members of the moral community must accept.

Moral conversation is—and often should be—uncomfortable and difficult. Of course, the targets of abusive rhetoric do not have to listen to it and engage with the person hurling it (Hill 2000; Bell 2013). If, for example, someone defends themselves against sexist comments and the person making those comments refuses to stop or to listen, he or she is well within their rights to withdraw from the conversation. Being open to challenge and negotiation does not amount to an obligation to talk to someone who mistreats you. Yet we can maintain these positions while at the same time claiming that when we do engage in moral dialogue we should do so in a way that does not presume we are right. Since shaming operates as a sanction, we are not merely objecting to or challenging someone else's behaviour; we are punishing it. The problem is not that we are never in a position to morally judge others because we never have enough information. The problem is with shaming and not with moral criticism.

Even if we accept that ruling out shaming need not rule out all forms of moral criticism, online moral communication is difficult to do well. The public nature of social media platforms makes online moral criticism without shaming difficult. An exchange that starts out a moral criticism can become shaming if many users retweet or share the exchange. All conversations have the potential to turn into public spectacles. We are left with a considerable challenge. Online communication may not lend itself to healthy, respectful moral conversation, but more and more conversations are happening online. If moving all moral criticism offline is not a viable option, philosophers need to do more work in the ethics of online communication.

The risks to creating safe communities

One of the powerful arguments in favour of shaming is that it helps shape our communities to be more welcoming to members of marginalized groups. On this view, we need to make public the behaviour that violates the values of the community to make clear what the community's values are. The public element of shaming is essential to this task. It is only by holding up racist or sexist behaviour that we are able to show everyone that such behaviour is out of bounds. If we do not 'call out' this behaviour, we are condoning it or we are complicit in allowing it to continue. Bell, for example, states the position this way: 'To respond civilly to [a racist person] is to risk condoning the [vices] they express, thereby further damaging moral relations' (2013: 219). If we allow people to say racist or sexist things online, we are allowing them to violate the values of the community and, in turn, we are failing in our own obligations to uphold those values. According to arguments like these, shaming is an important tool for creating communities that are properly anti-racist and anti-sexist and therefore safer for members of marginalized groups (Arneson 2007: 51; Adkins 2019: 78; Billingham and Parr 2019: 7).

The first question to ask is whether shaming is necessary to accomplish the task of creating safer communities. In the argument above, there is a false equivalence made between condemning behaviour and shaming a person who engages in that behaviour (Thomason 2018: 202–203). If I fail to shame my fellow community members when I discover they have not upheld some value, it does not follow that I have no other options. I can speak out about the importance of the value in general without shaming individual people who do not uphold it. In order to ensure that our communities embody the values we want them to, it is important to make clear which behaviour is acceptable and which is unacceptable, but we do so by focusing on the behaviour rather than the people who engage in it. To see how this might work, consider the following: some women have started posting misogynistic messages they receive from online dating sites, but they exclude the names and photos of the men who send them (Dewey 2015; Levy 2015). In these cases, women are trying to draw people’s attention to the types of harassment they receive rather than to the harassers. We may, of course, think that harassers and abusers do not deserve to be shielded in this way. Giving harassers a taste of their own medicine is no doubt ‘pleasurable and therapeutic’ for those who have encountered them (Locke 2007: 156). Yet this conclusion does not count in favour of shaming them unless giving harassers and abusers a taste of their own medicine is helping to improve women’s online experience. We have to ask whether engaging in this kind of shaming is creating the kind of community we want: is shaming misogynists actually making online platforms safer and more welcoming to women? Here is one of the cases in which the effectiveness of shaming and its justifiability overlap. If shaming does not make the online community safer, then it may not be justified on these grounds. Adkins provides reasons to think that shaming does not accomplish this task. She points to at least two high-profile examples of a shame backlash (2019: 83). Two different women took to an online platform to shame sexist behaviour within their professional communities (2019: 81). As a result, they ‘became the objects of sustained, public, and shaming scrutiny’ (2019: 83). According to Adkins, shaming requires that one has credibility and competence to speak on behalf of one’s community (2019: 86). Since women are often not seen as legitimate members of their community, they are seen as lacking this credibility and competence. Women who ‘risk speaking up and challenging bad behaviour are themselves the focus of sustained shaming, and those who wish to speak up in the future are implicitly cautioned against doing so’ (2019: 86). If members of marginalized groups are likely to face shame backlash, it is unclear whether shaming will work to make online communities safer for them.

Even if shame backlashes were not a problem, we may still wonder whether shaming creates a welcoming environment. Etzioni argues that one of the upsides to shame is that it is ‘democratic’ (1999: 47); that is, anyone in the community can shame anyone else, so all members of the community can enforce its values. Sceptics of shaming will argue that what Etzioni identifies as an advantage can quickly become a disadvantage (Thomason 2018: 203–204). One of the risks is that the community can quickly become one that is moralizing (Driver 2005: 138). Moralizing takes several forms, but here moralizing refers to an objectionable form of perfectionism. As Driver writes, ‘When one accepts the values [of the community], there is a sense that one may be slacking off in not supporting those values whenever and wherever one can do so’ (2005: 141). The trouble is that enforcing community values becomes a way of publicly expressing one’s commitment to those values. The more I am ready to shame the behaviour of other online patrons, the more committed to our values I appear. Especially online, there is an expectation that people must ‘weigh in’ or ‘make a statement’ about any troubling incident within the community. Failure to do so is often read as a half-hearted commitment to the community’s values. When we require people to call out violations of community values to prove that they adhere to those values, we engender an atmosphere of objectionable moral perfectionism. Additionally, if publicly ‘calling out’ bad behaviour is seen as a sign of one’s commitment to the values, there is the risk of moral grandstanding (Tosi and Warmke: 2016). Moral grandstanding occurs when ‘one makes a contribution to moral discourse that aims to convince others that one is “morally respectable”’ (2016: 199). The public element of shaming makes this risk particularly salient. First, shamers might be more apt to shame so that they can be recognized by others as pillars of the community for enforcing values. Second, as Tosi and Warmke point out, grandstanding often involves the phenomenon of 'ramping up', in which people make increasingly strong moral claims in an exchange (2016: 205). If enforcing the community's values showcases one's good character, then the harshness of one's condemnation will further reinforce the appearance of that good character. Sceptics will argue that the 'democratic' nature of shaming turns us all into moral police.

Shaming is often touted as a way of protecting marginalized groups, but shaming may turn out to create an environment that is more precarious for them. Adkins points out that marginalized people may face a shaming backlash from the dominant group (2019). Tessman argues that similar hierarchies can exist within marginalized groups, which can give rise to what she calls dangerous loyalty (2005: 133). Although loyalty is often treated as a virtue, especially in feminist and anti-racist movements, Tessman argues that dangerous loyalty can 'present an impediment to a political resister's own flourishing' and fail to bring about 'the goals of the very community that serves as an object of loyalty' (2005: 134–135). Tessman provides the example of Chicana feminists who are seen as betraying racial liberation struggles when they are critical of Chicano masculinity (2005: 138). Individuals who are committed to liberation struggles are often labelled as traitors when they choose to reject 'hegemonic beliefs and practices' within the marginalized group that are damaging either to them as individuals or to the goals of liberation (2005: 135). In Tessman's view, dangerous loyalty often arises in communities in which 'identity and politics are closely tied' (2005: 136).

Using shaming as a way to uphold community values creates the conditions for dangerous loyalty. Since shaming requires collective negative attention, initiating and participating in shaming becomes one of the ways that community members display to others their own commitment to the values. As such, relying on this kind of enforcement both requires and encourages public displays of loyalty. These displays of loyalty are similar to the problems with moral grandstanding: members of the liberatory group will face increased pressure to appear the most committed to the cause. The demand for public displays of loyalty, however, frequently undermines internal criticism of the values of the community. As Tessman points out, some 'feminist communities and racialized communities of colour have fallen into the mistake of silencing internal dissent, for they have tended to portray departures from the communities' hegemonic beliefs and practices … as reprehensible acts of treason' (2005: 144). Any community that identifies itself with certain values, but does not also allow for critique of those values or different ways of expressing those values runs the risk of inculcating dangerous loyalty. If upholding values is too strongly equated with enforcing values, the space of critique and dissent might become so narrow that critique and dissent start to look like violations of the values and thus fair targets for shaming. People who are already vulnerable may be put in a position of further vulnerability if they decide to be critical of the values of the community. Even if shaming effectively changes racist and sexist behaviour, it may do so by enforcing particular views about what racism and sexism are, which can be harmful to the vulnerable populations it is meant to protect. Additionally, the risk to epistemic co-deliberation re-emerges here. If certain views about racism and sexism become too dominant in public discourse, the conversation surrounding them may become impoverished. Dogmatic or rigid views about racism and sexism may prevent communities from addressing behaviour that does not immediately fit the categories on which they rely and from coming to new understandings about the categories.

Alternatives to shaming

In this final section, I want to identify some of the goods that shaming is meant to accomplish and suggest that we can come about them by other means. Shaming seems attractive because it appears to be a tactic that (a) is a way of defending people who are vulnerable to online abuse; and (b) calls attention to actions that violate community values.

As I have argued above, shaming may not improve the situation of marginalized people as much as we think. Still, some might argue that shaming is sometimes our only defence. When members of marginalized groups are abused, often no one does anything about it. Shaming seems like the only way to get people to pay attention to what is happening. Clementine Ford, for example, a weekly columnist for *Daily Life*, posted screen shots of abusive comments from men on her Facebook page. She explained her actions in an interview by saying: ‘These men don’t get to just go around leaving these kinds of comments and attempting to degrade women just for the hell of it. Why should they get away with it? Why should there be no consequences at all for them?’ (Levy 2015). Ford’s claims clearly reflect her frustration that no one takes misogynistic abuse seriously, and it is only once she starts shaming that her abusers face any sort of consequences for their actions. In this case, however, we only shame because nothing else works. We shame because we think that abusers ‘should not get away with it’. The underlying problem is that there are often no mechanisms in place on online platforms for marginalized people to stop abuse. As such, reforming reporting policies and terms-of-use policies is key to protecting vulnerable members of the community. In addition, we can ensure that vulnerable groups have supportive enclaves to which they can turn for help. We can take abuse seriously and protect people from it without shaming.

The other main draw of shaming is that it allows members of marginalized groups opportunities to make visible the online abuse that they face. Just as we sometimes conflate shaming and moral criticism, here I think we conflate visibility and shaming. There are ways to shine a light on this kind of abuse without shaming people who engage in it. For example, in 2016 a sports podcast posted a video in which male sports writers read examples of abusive misogynistic comments that their female colleagues received online. Even though the tweets did not name the abusers, the video was picked up by major news organizations and was viewed over three million times on YouTube (Willingham 2016; Dunlap 2016; Curtis 2016). This strategy both makes visible the kind of misogyny that women face online and serves as a critique of the online sports community. The video asks viewers to imagine what it might be like to say such abusive things to someone’s face rather than firing off a hateful social media post from a distance. Projects like this inspire a community to reflect on its own practices, while simultaneously revealing that the community is not living up to the values that it could or should have. Campaigns to increase the visibility of misogyny and racism help show that our communities often are not what we take them to be and can start conversations about what our values are, but we do not have to engage in shaming to accomplish these important tasks.

Conclusion

The strategies I have suggested in this chapter are part of a forward-looking, rather than a backward-looking, approach to positively shaping our communities. Since shaming is a reactive strategy, it takes place largely as a response to occurrences of wrongdoing. It is therefore less likely to address or change the background conditions that lead to wrongdoing in the first place. Additionally, shaming may not be a response to wrongdoing that will create the sort of communities we want to live in. If we want people to see the importance of certain values, will punishing them for not expressing those values accomplish this task? Instead of relying on social sanctions we might be better off trying to model the values that we think are important. As Locke puts it, rather than 'focusing on shaming those who shame us, let's make films, tell stories, tend parks, paint murals, open farmers' markets, build schools and universities, support clinics, and foster misfit salons' (2007: 159). In order to uphold the value of anti-sexism, for example, I can promote positive signs of respect for women, I can praise people who treat women with respect, and I can work for policy changes that help women combat the sexism they face. Arguments against shaming often sound like high-minded calls for politeness, but sceptics argue that the concerns run much deeper. Even if shaming seems like a good way to enforce anti-racist and anti-sexist values, it may put at risk the very kinds of communities we want to create.

References

Adkins, K. (2019). 'When Shaming is Shameful: Double Standards in Online Shame Backlashes', *Hypatia* 34, 76–97.

Arneson, R. (2007), 'Shame, Stigma, and Disgust in the Decent Society', *Journal of Ethics* 11, 31–63.

Basak, R, Sural, S., Ganguly, N., and Ghosh, S. K. (2019), 'Online Public Shaming on Twitter: Detection, Analysis, and Mitigation', *IEEE Transactions on Computational Social Systems* 6(2), 208–220.

Beard, M. (2016), '4 Arguments for Ethical Online Shaming (and 4 Problems with Them)', *The Conversation*, 18 May, **https://theconversation.com/4-arguments-for-ethical-online-shaming-and-4-problems-with-them-59662**., accessed 11 August 2021.

Bell, M. (2013), *Hard Feelings* (New York: Oxford University Press).

Billingham, P., and Parr, T. (2019), ‘Online Public Shaming: Virtues and Vices’, *Journal of Social Philosophy* (Early view December) 1, 1–20, doi: **https://doi.org/10.1111/josp.12308**.

Calhoun, C. (2000), ‘The Virtue of Civility’, *Philosophy and Public Affairs* 29, 251–275.

Calhoun, C. (2004), ‘An Apology for Moral Shame’, *Journal of Political Philosophy* 12, 127–146.

Curtis, C. (2016), ‘Men Read Terrible Tweets to Female Sportswriters in Eye-Opening PSA,’ *USA Today*, 26 April 26, **http://ftw.usatoday.com/2016/04/sarah-spain-julie-di-caro-harassing-tweets-video**, accessed 11 August 2021.

Deonna, J., Rodogno, R., and Teroni, F. (2012), *In Defence of Shame: The Faces of an Emotion* (New York: Oxford University Press).

Dewey, C. (2015), ‘Can Online Shaming Shut Down the Internet’s Most Skin-Crawly Creeps?’, *Washington Post*, 16 September, **https://www.washingtonpost.com/news/theintersect/wp/2015/09/16/can-online-shaming-shut-down-the-internets-most-skin-crawly-creeps/**, accessed 11 August 2021.

Driver, J. (2005), ‘Moralism’, *Journal of Applied Philosophy* 22, 137–151.

Dunlap, T. (2016), ‘Men Read Hate Tweets Sent to Female Sports Journalists’, *People Magazine*, 28 April, **http://www.people.com/article/men-read-mean-tweets-female-sports-anchors**, accessed 11 August 2021.

Etzioni, A. (1999), ‘Back to the Pillory?’, *The American Scholar* 6(3), 43–50.

Evans, B., and Yancy, G. (2016), ‘The Perils of Being a Black Philosopher’, *New York Times*, 18 April, **http://opinionator.blogs.nytimes.com/author/george-yancy/?_r=0**, accessed 11 August 2021.

Hawthorne, N. (1850/2003), *The Scarlett Letter* (London: Penguin Books).

Hill, T. E., Jr (2000), ‘Must Respect Be Earned?’, in *Respect, Pluralism, and Justice: Kantian Perspectives* (New York: Oxford University Press), 87–118.

Jacquet, J. (2015), *Is Shame Necessary? New Uses for an Old Tool* (New York: Pantheon).

Levy, M. (2015), ‘Hotel Worker Michael Nolan Sacked over Facebook Post to Clementine Ford’, *The Sydney Morning Herald*, 1 December, **http://www.smh.com.au/national/hotel-worker-michael-nolan-sacked-over-facebook-post-to-clementine-ford-20151130-glc1y4.html**., accessed 11 August 2021.

Locke, J. (2007), ‘Shame and the Future of Feminism’, *Hypatia* 22, 146–162.

Mill, J. S. (1859/2003), *On Liberty* (New Haven, CT: Yale University Press).

Nguyen, C. T. (2020), ‘Echo Chambers and Epistemic Bubbles’, *Episteme* 17, 141–161.

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

Nussbaum, M. (2004), Hiding from Humanity: Disgust, Shame, and the Law (Princeton, NJ: Princeton University Press).

Poole, E. (2013), ‘Hey Girls, Did You Know? Slut-Shaming on the Internet Needs to Stop’, *University of San Francisco Law Review* 48, 221–260.

Raicu, I. (2016), ‘On the Ethics of Online Shaming’, *ABC*, 25 February, **http://www.abc.net.au/religion/articles/2016/02/25/4413372.htm, accessed 11 August 2021**.

Ronson, J. (2015), *So You’ve Been Publicly Shamed* (New York: Riverhead Books of the Penguin Books Group).

Scanlon, T. M. (1998), *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Solove, D. (2007), *The Future of Reputation: Gossip, Rumour, and Privacy on the Internet* (New Haven, CT: Yale University Press).

Stearns, P. N. (2017), *Shame: A Brief History* (Urbana, IL; Chicago, IL; Springfield, OR: University of Illinois Press).

Tarnopolsky, C. (2004), ‘Prudes, Perverts, and Tyrants: Plato and the Contemporary Politics of Shame’ *Political Theory* 30, 468–494.

Taylor, G. (1985), *Pride, Shame, and Guilt* (Oxford: Oxford University Press

Tessman, L. (2005), *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (New York: Oxford University Press).

Thomason, K. K. (2015), ‘Shame, Violence, and Morality’, *Philosophy and Phenomenological Research* 91, 1–24.

Thomason, K. K. (2018), *Naked: The Dark Side of Shame and Moral Life* (New York: Oxford University Press).

Tosi, J., and Warmke, B. (2016), ‘Moral Grandstanding’, *Philosophy & Public Affairs* 44, 197–217.

Velleman, J. D. (2001), ‘The Genesis of Shame’, *Philosophy and Public Affairs* 30, 27–52.

Walker, M. U. (2007), *Moral Understandings: A Feminist Study in Ethics* (New York: Oxford University Press).

Willingham, A. J. (2016), ‘Brutal Video is Leaving Sport Twitter Speechless’, *CNN*, 26 April, **http://www.cnn.com/2016/04/26/us/women-journalists-target-of-offensive-tweets/**, accessed 11 August 2021.

Krista K. Thomason

Swarthmore College, Philosophy