THE TECHNOLOGY OF PUBLIC SHAMING

By Harrison Frye*

Abstract: *This essay argues that online public shaming can be productively understood as a problem of technology. In particular, the technology of public shaming is ambiguous between two senses. On the one hand, public shaming depends on various technologies, such as social media posts or, more historically, pillories. These are the artifacts of shame. On the other hand, public shaming itself is a social technology. In particular, public shaming is a way for communities to promote cooperation. Ultimately, I claim there is a mismatch between the artifacts of shame and this important social technology of shame. Social media drifts toward disintegrative shame, which tends to corrode cooperation. This suggests that we must either realign the technology of public shame or reject shame as a legitimate option.*

KEY WORDS: cooperation, public shaming, social media, social norms, technology

I. Introduction

A few weeks before writing this introduction, I saw a video and message shared by a friend on Facebook. In this video, a white woman accuses a black teenager of stealing her phone. The father of the teenager was the one recording the video, and responds off-camera to the woman that her accusations are false (which they are). She does not cease harassing the teenager and father by demanding to see the phone to verify that it is her stolen phone (which it isn't). Eventually she grabs at the teenager before the video stops. The father posted the video on Facebook along with a message that the woman is a "lunatic," and points out that her own phone is returned to her by an Uber driver shortly after the video was filmed and that she did not sufficiently apologize for wrongdoing. There is an obviously implied racial element to the condemnation of her behavior (ungrounded suspicion of young black males is, I hope, an uncontroversial example of racism). Shortly after seeing the Facebook post, I noticed that the story was picked up by news outlets, both major and minor.[1] In the meantime, the woman in question has been arrested for some unknown reason, suffered countless

* Department of Political Science, University of Georgia, hpfrye@uga.edu. Competing Interests: The author declares none. I would like to thank Matthew Adams, Nora Benedict, Kirun Sankaran, and an anonymous referee for their helpful comments on earlier versions of this essay.

[1] For example, https://www.nbcnews.com/news/us-news/woman-who-falsely-accused-black-teen-phone-theft-returns-new-n1253629. A brief update in December 2021: The incident (and subsequent court case based on hate crime charges) continues to attract major media attention: https://nypost.com/2021/11/08/soho-karen-miya-ponsetto-wishes-she-apologized-differently/.

doi:10.1017/S0265052522000085

public comments condemning her and her behavior, and now bears stigma by search engine—putting her name in a search engine inevitably brings up this event and her part in it.

When I reflect on this case, I find myself of two minds. On the one hand, it is clear that the woman in question acted wrongfully. To the extent that acting wrongfully renders one liable to response, I can understand the retributive impulse behind those who take to the Internet in such situations. On the other hand, I find myself deeply unnerved and unsettled by the reaction on the Internet. There is something disturbing to me about thousands and thousands (potentially millions and even billions) of people who do not know the individuals involved suddenly mobilized through the Internet to express certain attitudes and beliefs about what ought to be done to this woman. Further adding to my concern is that I suspect this person will never be free of her association with this particular act. This ambivalence reflects the fact that this case is just the most recent episode in the ongoing series of public shaming on the Internet.[2] The script at this point is a somewhat tired one: A person does something wrong. The wrongdoing gets picked up and posted throughout social media alongside messages of condemnation (it goes viral). Most often, there is a call for the wrongdoer to be fired, and the employer usually obliges. The stigma by search engine remains as a kind of modern-day scarlet letter.

The debate over public shaming in both its academic and popular guises tends to subsume the morality of public shaming under the more general heading of the morality of punishment. For example, Paul Billingham and Tom Parr focus on the question of proportionality. In their words, "[C]oncerns [surrounding public shaming] arise mainly because there is currently such a high volume of online public shaming that is disproportionate."[3] Similarly, Guy Aitchison and Saladin Meckled-Garcia focus on the question of how online public shaming is an impermissible form of punishment. In their words, "Our central argument is that the practice of OPS [online public shaming] is an attempt to incite a public, collective punishment of people for the kind of person they are (their moral personality) and therefore mistreats them."[4] These approaches seek to evaluate public shaming by the lights of the same moral principles according to which we evaluate punishments. While this is a fine way of going about the problem, I think it is limited. Because these approaches focus on the morality of public shaming, they give us no insight into public shaming as a social phenomenon. I believe it is better to have a grasp on the underlying

[2] See Jon Ronson, *So You've Been Publicly Shamed* (New York: Riverhead Books, 2015).

[3] Paul Billingham and Tom Parr, "Online Public Shaming: Virtues and Vices," *Journal of Social Philosophy* 51, no. 3 (2020): 383.

[4] Guy Aitchison and Saladin Meckled-Garcia, "Against Online Public Shaming: Ethical Problems with Mass Social Media," *Social Theory and Practice* 47, no. 1 (2021): 3.

structure of public shaming—we need to understand what explains the behavior and how the behavior tends toward wrongdoing.[5]

My aim in this essay is to describe public shaming as a problem of technology. Ultimately, I argue that the technology of public shaming involves two different dimensions. Recognizing these two different dimensions of technology renders clear a tension within the practice of public shaming. The most obvious dimension is that public shaming depends on technology understood in its more traditional sense. Like the case I introduced the essay with, most public shaming these days occurs through social media and the Internet: we publicly shame people by posting videos recorded on our phone, sharing their Tweets with condemning messages, creating hostile petitions on websites for all to observe and sign, and so forth. Earlier, cruder, forms of public shaming also depended on technology to operate. Consider that the sandwich board, the stockade, the pillory, the whipping post, and so on are all pieces of technology that make public shaming possible. I will call this aspect of the technology of public shaming the *artifacts of shame.* The less obvious dimension is that public shaming is itself a form of technology. More specifically, public shaming is a piece of social technology that helps groups achieve particular ends. Public shaming is a way that groups express their value judgments, ostracize individuals, develop solidarity, deter would-be wrongdoers, among other things.[6] Of most importance, I will suggest, is the role of public shaming in promoting cooperation. I will call this aspect the *social technology of shame.* The technology of public shaming, then, refers to two different ideas: the artifacts of shame and the social technology of shame.

The upshot is that to understand public shaming as a social problem requires attention to *both* dimensions. First, what is public shaming useful for? Second, what artifacts of shaming help us best achieve this end? To show the value of this approach, I will focus on what I take to be a fundamentally important function of public shaming: promoting cooperative behavior among individuals for the good of the community. Drawing on the criminologist John Braithwaite's distinction between reintegrative and disintegrative shaming, I describe how shaming, when reintegrative, can promote pro-social behavior under the right conditions.[7] However, if we use social media and the Internet to this end, this will be in fact self-defeating

[5] In this way, I'm interested in providing something like what Joseph Heath calls a "normative model" of public shaming. Joseph Heath, "Three Normative Models of the Welfare State," *Public Reason* 3, no. 3 (2011): 13–14.

[6] On various uses of public shaming, see Krista K. Thomason, *Naked: The Dark Side of Shame and Moral Life* (New York: Oxford University Press, 2018), 181–82; Billingham and Parr, "Online Public Shaming," 1-2, 6–7.

[7] John Braithwaite, *Crime, Shame and Reintegration* (New York: Cambridge University Press, 1989). Billingham and Parr also discuss the reintegrative aspects of shaming, however they focus on reintegration in distinctly moral terms (e.g., what is wrong with disintegrative shaming is that it constitutes "a threat to the norm violator's dignity"). Paul Billingham and Tom Parr, "Enforcing Social Norms: The Morality of Public Shaming," *European Journal of Philosophy* 28, no. 4 (2020): 1005.

to the end of promoting cooperation. I will argue that social media is as a practical necessity prone to disintegrative shaming, which is largely counterproductive toward the end of promoting cooperative behavior among individuals. There is a misfit between the artifact of shame and the social technology of shame, suggesting that we either must consider how to realign these dimensions of shame or must reject shame as a legitimate option.

II. Public Shaming: A Brief Account

To begin, it is important to get a handle on what exactly public shaming *is*. I will not provide a full account of public shaming here. Instead, I will describe some core features to undergird a working account of public shaming for the purposes of this essay. I will expand on some of these ideas as they become relevant in later sections.

At a most basic level, public shaming involves public expressions of moral disapproval of some wrongdoer that marks that wrongdoer as having a defective status.[8] Let me draw attention to five features of this working account.

The first feature is that public shaming tends to be grounded in social norms within a community. This is often noted in the extant academic literature on public shaming.[9] By social norm, I mean, roughly, a general, informal rule accepted by a given population.[10] Following Geoffery Brennan et al., let us say that social norms have two features.[11] First, they are normative in nature. Social norms purport to provide a reason for action. Second, they are social facts. Social norms exist to the extent that enough people within a group accept the norm or believe that others accept the norm. Public shaming relates to such norms as public shaming is often mobilized in response to perceived violations of norms. An important implication of this feature of public shaming is that it can operate independently of the law—just because a legal norm has been violated does not necessarily implicate public shaming as I understand it. That depends on whether or not that legal norm is connected with a social norm within the community. Thus, when the law attempts to enlist public shaming in its service, as it has historically, it can only do so when the law itself reflects the more informal norms of a community.

[8] See also Aitchison and Meckled-Garcia, "Against Online Public Shaming: Ethical Problems with Mass Social Media," 9-10.

[9] For example, Parr and Billingham, "Enforcing Social Norms."

[10] I am using social norm in a fairly broad and inclusive sense. In this way, I am setting aside the controversy about how social norms relate to moral norms. See Philippa Foot, "Morality as a System of Hypothetical Imperatives," *The Philosophical Review* 81, no. 3 (1972): 305-16; Nicholas Southwood, "The Moral/Conventional Distinction," *Mind* 120, no. 479 (2011): 761–802; cf. Joseph Heath, "Morality, Convention and Conventional Morality," *Philosophical Explorations* 20, no. 3 (2017): 276–93.

[11] Geoffrey Brennan, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood, *Explaining Norms* (Oxford: Oxford University Press, 2013), 3–4.

The second feature is that public shaming is public—there needs to be an audience in order for an act to count as public shaming. If I take my friend aside and chide the friend for improper table etiquette, this act is not public shaming. In contrast, it is important in the introductory case that the father in question took to social media and allowed his post to be shared "globally," viz. with people who have accounts on Facebook, but no formal friend connection to the sharer. It would be very different if his outrage were limited to the event as it occurred, or if he even simply just recounted the events in private conversations with friends in passing. There needs to be an audience for public shaming to count as such. The audience need not be known to the shamed (a key point I will return to later), but neither must it be unknown.

The third feature is that public shaming involves *expressions* of moral disapproval. This need not be limited to verbal expressions. Perhaps the most famous of all examples of public shaming is Nathaniel Hawthorne's scarlet A. Without question, the A stands in for adultery, but the shaming consists in the mark that Hester Prynne must wear for the rest of her life. In the introductory case, it is not just the words and description that feature into the shaming act, but also the video of the wrongdoing for all to watch and rewatch in near perpetuity.

The fourth feature is that public shaming is not the same as public moral disapproval. When I announce in public that I think it is wrong for a particular person to try to get someone fired on the basis of protected speech, I am not shaming that person, even though I am saying they did something wrong. Shaming is not equivalent to criticism. What distinguishes shaming from criticism is an attempt to go beyond the wrongdoing and attack the character of the person in question. This attack on the status of the individual is an attempt to put social distance between the wrongdoer and other people. Notice in the introductory case that the father did not just say that the actions of the woman in question were wrong, but also described the woman as "lunatic" pejoratively. In this way, shaming is linked to stigmatization. To stigmatize is to mark someone as having a diminished status. As Erving Goffman puts it, a stigmatized individual is reduced "from a whole and usual person to a tainted, discounted person."[12] Stigmatization is often understood in particularly dramatic terms, as it involves marking someone off as a part of a class of deviants.[13] Not all forms of shaming amount to stigmatization, if we understand stigmatization in this dramatic sense.[14] The call for distance in shaming can come in degrees (seeking to fire someone, seeking to get people to slightly distance

[12] Erving Goffman, *Stigma: Notes on the Management of Spoiled Identity* (New York: Simon and Schuster, 1963), 3.

[13] For example, Braithwaite, *Crime, Shame and Reintegration*, 12–13.

[14] For a milder account of stigmatization, see Joseph Heath, "A Defense of Stigmatization" (unpublished manuscript), available at https://www.academia.edu/31792827/A_Defense_of_Stigmatization. On the complexities of the concept of stigma, see Bruce G. Link and Jo C. Phelan, "Conceptualizing Stigma," *Annual Review of Sociology* 27 (2001): 363–85.

themselves from the person in question, and so on). As will become important in later sections, this call for distance is not necessarily permanent: some shaming offers an opportunity for redemption.

The fifth, and final, preliminary feature is that public shaming, because it involves stigmatization, involves a success-condition. If I try to publicly shame you and no one pays me any mind, I have attempted and failed to publicly shame you. You are not stigmatized if everyone treats you the same as they treated you before my (and perhaps others') actions. Public shaming requires uptake by an audience to count as such, in my view.[15]

III. The Technology of Public Shaming, Two Ways

Previous approaches to the problem of public shaming tend to apply the principles supplied by the moral theory of punishment: due process, proportionality, and so forth. Getting past this moral theory of punishment does not mean throwing away punishment as a way of understanding public shaming. It is obviously true that public shaming can act as a form of punishment. There is a seeming retributive impulse behind much participation in public shaming. But we want to get a clearer understanding of public shaming, even if it ends up being best understood as a form of punishment. To get past the moral theory of punishment, let us take a step back from the idea that public shaming is a problem of punishment. Instead, let me suggest that public shaming can be fruitfully seen as a problem of technology.

I previously stated that the technology of public shaming is ambiguous and can mean two things. On the one hand, there is the technology used in public shaming. These are the artifacts of shame. On the other, there is the idea that public shaming is itself a kind of technology. This is the social technology of public shame. Let me now explain these ideas with a little more detail.

[15] Some minor points worth noting: First, this account says nothing about the motivations of people who engage in public shaming. People can engage in public shaming for all sorts of reasons. People may publicly shame as a form of moral grandstanding, or as a way of trying to look morally good in the eyes of others. (On moral grandstanding, see Justin Tosi and Brandon Warmke, "Moral Grandstanding," *Philosophy and Public Affairs* 44, no. 3 [2016]: 197–217.) People may publicly shame due to a retributive impulse and see punishing a rule violator as sufficient reason to engage in the act. (On the idea that human beings are "rule-following punishers," see Gerald Gaus, *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* [New York: Cambridge University Press, 2011], 103–122.) And so on. The second caveat is that public shaming as a social phenomenon comes apart from shame as an emotion (or guilt for that matter) in the wrongdoer. Shame tends to refer to the feeling we get when we have failed in the eyes of others. (In contrast, guilt focuses on the feeling we get when we fail by our own lights). You can publicly shame a person without the person feeling any shame (or guilt for that matter). (On the distinction between shame and guilt, see Joseph Henrich, *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* [New York: Farrar, Straus and Giroux, 2020], 34.)

It is no shock that a synonym for public shaming is "pillorying." While pillory as a verb acts as a substitute for public shaming, the verb appears to have its etymological root in the nominal form of the word, which refers to a wooden contraption that locks a person's head and hands between two pieces of wood.[16] The two pieces of wood are elevated by a post, making it so that the person is standing with his or her head on display. Pillories are set up in public places on platforms, so that the person locked in a pillory is presented for all to see. Sometimes a placard is placed on the platform, to describe what the person has done to end up in this situation. Pillorying tends to occur for hours.

The pillory makes vivid the idea that public shaming occurs through artifacts and invented objects. I could have also used the stocks as an example of this. Stocks lock in the person's feet or wrists, leaving the head free from bondage (this is less uncomfortable than the pillory). The pillory and stocks were a widely used form of punishment from the Middle Ages in Western Europe.[17] It is only in the nineteenth century that such forms of public shaming became widely outlawed, a point I will return to in later sections. Less brutal, sandwich boards also show up as a form of public shaming. In this case, a person stands on a roadway, between two large pieces of cardboard that describe in writing what the person did ("I stole money from my employer," and so on).[18] The most fitting artifact of shame for the concerns of this essay is the social media post, which often involve videos and pictures along with words against the target of public shaming.

It is important to notice two features of artifacts of shame. The first feature is that artifacts of shame create *publicity*: artifacts of shame help demarcate the audience that is relevant for the act of public shaming. Publicity can vary by degree. For example, pillories limit the public to those in the town square. A billboard on a highway reaches the larger audience of all who drive by. A social media post reaches a potential audience of all those who have access to the technology to view that post. The second feature of artifacts of shame is that they act as a mark for the relevant stigma in a similar way to how a traffic light acts as a signal to coordinate the behavior of others. Standing on a platform in the middle of town square is a public act but not necessarily an act of public shame; wearing a sandwich board on a street corner could be advertisement; and so forth. The artifact of shame includes with it the relevant mark: it is the *pillory* that marks the person as degraded in status; the sandwich board *with* the morally thick description of the action; the video or picture of the person alongside the description of wrongdoing; and

[16] A good example of the image of the pillory is in Robert Chambers, *Book of Days, Volume 1* (1879), 832, "The Pillory," available at https://digicoll.library.wisc.edu/cgi-bin/History/History-idx?type=turn&id=History.BookofDaysv1&entity=History.BookofDaysv1.p0862

[17] Peter N. Stearns, *Shame: A Brief History* (Urbana: University of Illinois Press, 2017), 27–28.

[18] Ted Poe, a judge in Houston from 1981–2003, would (in)famously inflict these kinds of punishments on convicts. Ronson, *So You've Been Publicly Shamed?* 82–86.

so on. Thus, artifacts of public shame raise two questions: How public does this particular artifact make a given act of public shaming, and what sort of stigma does the artifact produce?

The second sense of the technology of public shame is the idea that public shaming is itself a form of social technology. The term "social technology" is a bit ambiguous, and you can find the term used in a variety of ways. By "social technology," I mean (i) a social practice (ii) that advances the ends of a particular group. Take, for example, cultural rituals. A broad class of rituals involves movements and dances coordinated with other individuals along with music and simultaneously chanted words to help heighten the sense of oneness between participants. A hypothesized function of such rituals is that they help produce a sense of solidarity among a group, and thus lubricate social relations within the group.[19] This example suggests that ends advanced by social technology may not be consciously accessible to individual participants. For example, when I attended my graduation ceremony from college, I did not think I was contributing to a practice that helped develop a sense of solidarity with my fellow students and alumni. But surely that is part of the point of commencement as a ceremony.[20] Thus, it makes sense to abstract away from the motivations of participants in public shaming if we are interested in trying to get at the social technology of public shaming.

Describing public shaming as a social technology draws our attention to its function in group life. Yes, this function may be punishment. But public shaming encompasses a wide range of goals: it can itself generate a sense of solidarity (imagine the feeling among the crowd in the town square), it can signal what is and is not appropriate (the placard and message along with the pillory or shared video), and it can isolate intolerable individuals (severe stigma that marks someone as an outsider). Importantly for my purposes, not all of these functions are valuable in a universal, objective sense (however you might understand that idea). Nor is this intended to be an exhaustive list of functions, but rather to give a sense of the variety of ways through which public shaming can advance the ends of the group.

We have two dimensions of the technology of public shaming: the artifacts of shame and the social technology of shaming. I think we can already begin to see how these two dimensions can come apart from one another. For example, if the social technology of public shaming is being used to generate cohesion among the shamers, the sandwich board on the side of the street is unlikely to be successful. Why? There is no observed coordination of shaming—the sandwich board operates on its target as disparate individuals pass by. Contrast this with the pillory and a public announcement of

[19] Henrich, *The WEIRDest People in the World*, 76-78.

[20] I suspect universities utilize the ritual of commencement in part to give a sense of obligation to the group and help promote alumni donations.

shaming. People gather in the town square for these events, and yell almost as one against the wrongdoer.

This is an initial brief gesture toward how the two dimensions of the technology of public shaming may come apart and work against each other.[21] The rest of the essay will develop an argument for a specific way that artifacts of shame can work against what I take to be the most fundamental social technology of public shaming: the promotion of cooperation of an individual for the good of the community.[22] I will develop the case in the next section for how public shaming can promote cooperation. Following that, I will argue that social media is poorly equipped to advance this end, and public shaming ends up operating in a pathological function in that context.

IV. Public Shaming, Social Norms, and Cooperation

As noted in my description of public shaming, much of this behavior is grounded in social norms. People engage in public shaming largely in response to perceived norm violations. If we are interested in public shaming as a social technology, I believe it is helpful to begin with how norms more abstractly operate as a social technology.

A growing consensus among a wide range of disciplines suggests that social norms play a key role in promoting cooperation among individuals.[23] I do not mean cooperation in the broad sense of whenever people get together to do something. Rather, I mean cooperation in the game theoretic sense of getting people to act in a way that promotes socially optimal outcomes, even if rational self-interest provides temptations to act in ways that produce socially suboptimal outcomes. The exemplary game here is the Prisoner's Dilemma. In this way, cooperation is distinguished from coordination. Coordination focuses on how two agents must act in concert to achieve socially optimal ends. However, unlike cooperation, coordination does not require restraining self-interest.[24] Think here of what side to drive on. Drivers have a strong interest to make sure they pick the same side to

[21] There is some similarity between the approach taken here and John Thrasher's call for "comparative-functional analysis" in evaluating norms. See John Thrasher, "Evaluating Bad Norms," *Social Philosophy and Policy* 35, no. 1 (2018): 210–15.

[22] To say that promoting cooperation in this way is the fundamentally important function of public shaming is not to say that it is the only important function. Further, I cannot fully argue for the claim that promoting cooperation is the most fundamental function of shaming relative to others in this essay. Nonetheless, the next section makes the case for why promoting cooperation is a deeply important function of public shaming, and this is enough for my purposes here: to show how public shaming on the Internet works against a particularly important function of such shaming.

[23] Cristina Bicchieri, *The Grammar of Society* (Cambridge: Cambridge University Press, 2006), 8–28. Joseph Heath, *Following the Rules* (Oxford: Oxford University Press, 2008), 66–71; Joshua Greene, *Moral Tribes* (New York: Penguin Press, 2013), 22–25; Joseph Henrich, *The Secret of Our Success* (Princeton, NJ: Princeton University Press, 2016), 143-45; Michael Tomasello, *A Natural History of Human Morality* (Cambridge, MA: Harvard University Press, 2016), 98–103.

[24] Bicchieri, *The Grammar of Society*, 38–39.

avoid collisions. It takes no restraint on self-interest to avoid head-on collisions! In contrast, a norm against, for example, stealing can help produce socially optimal outcomes (a world where people do not have to excessively fear stealing makes it pay to produce, invest, and save), but it is not difficult to imagine situations in which an individual could advance their individual self-interest by stealing. To summarize, social norms help promote cooperation in the sense that they coax individuals within a community to engage in cooperative behavior for the sake of that community.

When linked to social norms, public shaming can promote cooperation in a variety of ways. There is the obvious one: public shame acts as a deterrent to wrongdoers. Being publicly shamed is not a desirable outcome. The possibility of facing public shaming for violating a norm might alter the deliberative calculus such as to render wrongdoing unprofitable. For example, if a potential trouser thief knew that he would be forced to wear a pair of pants on his head in the town square and be ridiculed by his fellow townspeople, I think he would at least think twice before stealing those pants.[25] While clearly a part of the story, this obvious approach to thinking about public shaming resembles too closely the analysis of public shaming as a form of punishment, as punishment is commonly justified by its deterrent effects. This would depict public shaming as of a kind with imprisonment or fines: an abstract cost we attach to some wrongdoing to get less of that wrongdoing. But this does not help us understand the potential value of public shaming for cooperation as opposed to these alternative deterrents. Can we say something specific about public shaming that helps link it to cooperation and social norms but is distinct from its role as a deterrent?

Let me suggest here that public shaming can advance cooperation through at least two avenues: through reintegration and through ostracism or disintegrative shaming. Ultimately, I will argue that reintegrative shaming is what we should aim for if we are interested in promoting cooperative behavior among individuals. However, I also acknowledge that disintegrative shaming has a place in protecting a community in extreme situations even if it does not induce individuals to behave in cooperative ways. Prior to this, however, I must explain these two forms of shaming.

John Braithwaite describes reintegrative shaming as an important form of shaming. In his words, "[r]eintegrative shaming means that expressions of disapproval, which may range from mild rebuke to degradation ceremonies, are followed by gestures of reacceptance into the community of law-abiding citizens."[26] For shaming to be reintegrative, it is important that it is temporary on some level. Consider some of the artifacts of shame noted above: The pillory and the stocks. These devices are temporary in their results. You only spend hours in these devices. Similarly, consider another

[25] This is an example of an actual act of public shaming. See Stearns, *Shame: A Brief History*, 44.
[26] Braithwaite, *Crime, Shame and Reintegration*, 55.

artifact of shame surrounding adultery present in Classical Greece.[27] Once caught, adulterers would be required to shave half of their head. The half-shaven head acts as the public mark and, in turn, humiliates the adulterer. (In the relevant community, everyone knows what the half-shaven head means). However, the adulterer only bears the shame until the hair grows in.

Reintegrative shaming aids in the cooperative function of norms in at least four ways. First, the fact that a cooperative group responds to and evaluates wrongdoing as such can help to assure its members that others are committed to doing their part. When wrongdoing goes unanswered, there is always the danger that a cooperative enterprise will unravel. This is because it appears that much norm compliance is conditional: I will follow the rules only if I expect that enough others will do the same.[28] In this way, reintegrative shaming is like all sanctions imposed upon norm violators. Second, reintegrative shaming does not push the shamed toward even worse patterns of behavior. This is an important feature of Braithwaite's argument in favor of reintegrative shaming. If there is no coming back from shaming, there is no reason to reform after the humiliation and return to a pattern of cooperative behavior after a minor deviance.[29] To use an analogy: If the punishment is death for all crimes, you have no deterrence-based reason to stop at petty theft. Similarly, if the punishment is social death, you have no reason to try to integrate yourself back into a system of cooperation. Third, and relatedly, reintegrative shaming does not cut individuals off from a potential cooperator. An act of antisocial behavior does not imply that a person will not play by the rules in the future. In game theoretic terms, just because a person defects in one iteration of the game does not mean they will defect in future iterations. It all depends on the strategy the player is using, and how they respond to the strategies of others. Part of the effectiveness of tit-for-tat is that it can push particular individuals back into a cooperative equilibrium, assuming they are not permanent defectors (more on this in a bit). And, finally, reintegrative shaming allows individuals to hold themselves accountable through the actions of others.[30] Recall that social norms are a way of keeping at bay the temptations of self-interest. While a powerful feature of our normative psychology is that such norms can motivate on their own, sometimes this is not enough.[31] Reintegrative shaming offers a means by which we can remind ourselves of our own normative commitments beyond moments of weakness. In this way, reintegrative shaming forms a part of what Joseph Heath and Joel Anderson term "the extended will."[32]

[27] Stearns, *Shame: A Brief History*, 26.
[28] Bicchieri, *The Grammar of Society*, 11.
[29] Braithwaite, *Crime, Shame and Reintegration*, 102.
[30] Heath, "A Defense of Stigmatization," 20.
[31] On the motivational power of norms, see Daniel Kelly and Taylor Davis, "Social Norms and Human Normative Psychology," *Social Philosophy and Policy* 35, no. 1 (2018): 58–62.
[32] Joseph Heath and Joel Anderson, "Procrastination and the Extended Will," in Andreou and White, eds., *The Thief of Time: Philosophical Essays on Procrastination* (Oxford: Oxford University Press, 2010), 233–53.

In contrast to reintegration, public shaming can be a form of ostracism or dramatic stigmatization. Consider the case of criminal branding or maiming. One way to shame a thief is to cut off their hand. Similarly, branding might be utilized to mark criminals. Both branding and maiming are permanent in their imprint. Unlike a half-shaven head, there is no growing back a hand. Because of this permanent aspect, branding and maiming are cases of what Braithwaite calls "disintegrative shaming." They are ways of excising individuals from a community.

Disintegrative shaming does have a role in sustaining a community of cooperation, albeit a more limited one than reintegrative shaming. The first reason for believing that disintegrative shaming plays a role in promoting cooperation relates to the fact that ostracism is a sanction for wrongdoing. Ostracism, understood as a deterrent, then, appears to have a role to play in assuring cooperators that others are doing their part. However, cutting a person from a community of cooperators is a rather extreme deterrent relative to alternatives. Because of this, we ought to be careful about deploying disintegrative shaming for purely deterrent purposes. While reintegrative shaming focuses on promoting cooperative behavior in individuals for the sake of the community, disintegrative shaming focuses on protecting a community from particularly deviant individuals. And this leads to the second way that disintegrative shaming plays a role in cooperation. Ostracism is a way of excising individuals who will not cooperate and seek to defect whenever possible.[33] Think here of Hobbes's Foole:

> The fool has said in his heart: "there is no such thing as justice"; and sometimes with his tongue, seriously alleging that: "every man's conservation and contentment being committed to his own care, there could be no reason why every man might not do what he thought conduced thereunto, and therefore also to make or not make, keep or not keep, covenants was not against reason, when it conduced to one's benefit."[34]

The Foole is a pathological defector—he only cooperates when he has no other option (for example, he is worried about being caught). Hobbes argues that the proper response to the Foole is to kick him out of society. A person who will always take advantage of you is not someone you want to involve yourself with. Thus, ostracism is an instantiation of a grim trigger strategy: If you ever defect against me, I will never cooperate with you in the future.[35] There are times and places for this. This is what disintegrative shaming does.

[33] Carol Barner-Barry "Rob: Children's Tacit Use of Peer Ostracism to Control Aggressive Behavior," *Ethology and Sociobiology* 7, nos. 3–4 (1986): 281–93.

[34] Thomas Hobbes, *Leviathian*, ed. Edwin Curley (Indianapolis, IN: Hackett Publishing Company, Inc., 1994), 90.

[35] Michael Moehler, "Why Hobbes's State of Nature Is Best Modeled by an Assurance Game," *Utilitas* 21, no. 3 (2009): 319–26.

That being said, it is important to notice that ostracism is much more limited in its ability to produce cooperation than are reintegrative approaches. This is for the following reasons. First, as noted above, ostracism may push individuals to engage in antisocial behavior. If ostracized individuals find themselves with no way to redeem themselves, they may find it better to indulge in even worse behavior. This predictable result of ostracism also has the effect of making disintegrative shaming a kind of self-fulfilling prophecy: "It is a good thing we got rid of that guy—look at what he is doing now!" Second, and relatedly, ostracism deprives a community of a potential cooperator. This is not the same as an individual walking away from another particular individual. It is one thing for you to say no to a particular individual for the rest of your life. It is an order of magnitude more serious for a whole community to say no to that individual. To summarize: disintegrative shaming does not induce cooperative behavior from individuals like reintegrative shaming does. Instead, disintegrative shaming is a way of excising individuals from a community almost entirely. For these reasons, ostracism should be reserved for particularly awful cases in which redemption is not possible.

Fundamentally, then, reintegrative shaming helps promote cooperation by inducing cooperative behavior in individuals. Disintegrative shaming is not well suited to promoting cooperation in individuals, but for seeking and extricating from a community pathological defectors. A side effect of this type of shaming is that sometimes individuals who are not pathological defectors get turned into defectors. This is a loss for all parties, and so we should be very circumspect about disintegrative shaming relative to reintegrative shaming.

V. When Public Shaming Technologies Work Against Each Other

Speaking positively of public shaming practices like the pillory may strike some readers as bizarre. Is it not a good thing that we no longer parade our adulterers around with half-shaven heads? I agree. It is, of course, a good thing that we no longer parade adulterers in the street with half-shaven heads or put minor criminals in pillories for the mob to jeer and throw rotten food at. But I think it is important to understand where this negative judgment of public shaming comes from, historically speaking. Many readers will likely be familiar with Nathaniel Hawthorne's aforementioned novel *The Scarlet Letter*, which describes the public shaming of a woman, Hester Prynne, for adultery in a seventeenth-century Puritan town. Hawthorne's story depicts the practice in a critical light, as the reader naturally empathizes with the plight of Prynne. Hawthorne's book, first published in 1850, was a part of a growing consensus that public shaming through the pillory or the stocks was unacceptable. Earlier, in a 1787 polemic against public shaming, Benjamin Rush suggests that, "[i] gnominy is universally acknowledged to be a punishment worse than

death."[36] A striking fact is that by the twentieth century the vast majority of states in the United States outlaw common forms of public shaming punishments such as the pillory and the stocks.[37] Delaware stands as the last holdout, outlawing public whipping in 1952. Delaware's obstinacy in this matter did not go unnoticed. In 1876, the *New York Times* ran an editorial criticizing Delaware's failing to give up forms of public shaming.[38]

I do not think it is an accident that public shaming begins to face mounting public pressure in the eighteenth and nineteenth centuries. In this section, I will argue that changing circumstances decoupled the artifacts of shaming from the social technology of public shaming.[39] In particular, changing circumstances made reintegrative shaming more difficult to achieve. I will then suggest we are seeing something similar happening today: social media as an artifact of shaming is not well suited to producing reintegrative shaming and naturally drifts toward ostracism and the pathologies that come along with that.

Concurrent with the push to abolish public shaming is an enormous growth in population in many urban areas. As the legal scholar Adam Hirsch documents in Massachusetts, most cities (barring Boston and Salem) were no larger than 2,000 people as late as 1690.[40] By the mid-eighteenth century, nearly half of all towns in Massachussetts grew to more than 1,000 people, with thirty towns containing more than 2,000 people.[41] Perhaps more importantly, much of the population growth is related to immigration. Mobility grew in Massachusetts in eighteenth century, with Hirsch reporting that over ten percent of Boston residents in the late eighteenth century had been in the city for fewer than five years. Hirsch suggests that this growth in a mobile population rendered shaming less effective as a form of social control. In Hirsch's words: "[T]he penalty of a session on the pillory must have appeared far less daunting when performed before persons with whom the offender was unacquainted, and with whom he need have no further contact whatsoever."[42] Thus, it seems that part of the explanation for the pushback against public shaming was not limited to moral disapproval of the practices. Part of the story is that such practices were increasingly seen as ineffective in establishing social control. This helps explain why public shaming in the form of the pillory at

[36] Ronson, *So You've Been Publicly Shamed?* 54–55; Stearns, *Shame: A Brief History*, 62.
[37] Stearns, *Shame: A Brief History*, 62–65.
[38] Ibid., 62.
[39] This argument will be somewhat speculative. Whatever explains the decline of public shaming in Western societies in the eighteenth and nineteenth centuries is undoubtedly a complex, multicausal explanation. However, I want to emphasize some features of the shift that are illustrative for my purposes.
[40] Adam Hirsch, "From the Pillory to the Penitentiary: The Rise of Criminal Incarceration in Early Massachusetts," *Michigan Law Review* 80, no. 6 (1982): 1223.
[41] Ibid., 1228.
[42] Ibid., 1232.

least is not seen as a viable option—the pillory is ineffective in open societies such as our own where people can come and go within particular social groups. In such a context, the pillory as an artifact of shame is no longer well fit to the social technology of shaming, and so becomes an obsolete piece of technology, like the buggy whip or VHS tape. Worse still, the pillory applies the pain of humiliation without any of the benefit of reintegrative shaming. Thus, the pillory has no place in a modern, diverse society. It simply enacts pain and suffering without the particular benefits that attached to public shaming in the past.

The experience of eighteenth-century Massachusetts is important as it suggests that whether or not an artifact of shame is effective depends on how that artifact interacts with the prevailing social conditions. The pillory acts as a vehicle for reintegrative shame partially because of its temporary nature, but also because individuals subject to the pillory see themselves as connected to the onlookers. Reintegrative shaming implies that a person was already integrated within a community—you cannot reintegrate a person into a group who was never a member of that group in the first place. If a person sees himself as outside the community, shaming him will be relatively ineffective in promoting cooperative behavior. Worse still, shaming under these circumstances may further promote antisocial behavior to the extent that targets of shaming do not see themselves as bound to the people doing the shaming. If you try to shame me, and I see you as a complete stranger, I am more likely to be resentful toward you than to seek to reconcile with you. Therefore, in order to have the relevant effect, artifacts of shaming depend not just on their technical features, but on the social context in which they are deployed.

This allows me to return to the motivating issue of the essay: shaming through social media. Social media clearly can act as an artifact of shame. The question is: Does it live up to the function of shame in promoting cooperation? I suggest not for two reasons.

The first thing to notice about shaming through social media is its virtually permanent status.[43] Unlike the pillory, in which a person suffers for hours, a social media post can last as long as the servers it is stored on remain operating. Consider the opening case from the introduction: this woman's shaming still exists. You can find it right now if you would like. You can see the video of what she did. You can see the description of her wrongdoing. You can also find all the comments people have hurled against her. In this way, social media is a lot like more traditional publication of the written word than older, more ephemeral, artifacts of shame. As Jeremy Waldron points out in his concern with the publication of hate speech, "it is the

[43] See also Aitchison and Meckled-Garcia, "Against Online Public Shaming: Ethical Problems with Mass Social Media," 5.

enduring presence of the published word or the posted image that is particularly worrying."[44] So, too, for online shaming.

This permanence that attaches to shaming via social media posts undercuts its ability to act as a form of reintegrative shaming. Part of the point of reintegrative shaming is that it includes the possibility of returning to the fold. When the pillory unlocks, and you walk down from the platform, the shaming is complete and you have an opportunity to redeem yourself. With online shaming, the shaming is never marked as over. It is difficult to find an opportunity for redemption. Worse still, consider the phenomenon of how people rediscover old stories on the Internet.[45] People sometimes share old stories, and people see them as new and re-amplify the story as though it were new. When the story is a form of shaming, this leads to a renewal of the person's initial shaming. By its nature, then, social media shaming drifts toward disintegrative shaming. Thus, we have good reason to believe online shaming is not particularly effective at promoting pro-social behavior.

Some have tried to deal with this permanent aspect of online shaming through establishing a legal right to be forgotten. The European Union, for example, has a legal framework in place for individuals to request search engines to scrub certain information from their results. This kind of legal innovation is clearly a push in the right direction if we want to make online shaming functional. But there are three problems that need to be addressed if we wish to go down this path. The first is that the legal innovation has a mixed result in practice. This is because of the Streisand Effect—or the fact that trying to suppress information can backfire by drawing attention to the information you are trying to suppress.[46] In the EU, for example, Mario Costeja González was able to get the Court of Justice of the European Union to endorse his claim for Google to remove information involving his financial troubles.[47] However, in doing so, the court inadvertently made Costeja González and his woes famous. The second problem is that a legal right of this sort may conflict with the value of free speech. If the government has the power to compel search engines to remove certain information, this may create worrisome unintended effects later down the road. The third problem with adopting a legal right to be forgotten is that it assumes the primary problem with online shaming is its virtually permanent nature. If we are able to limit the temporal scope of online shaming, we will render online shaming more reintegrative and less disintegrative. Or so this approach

[44] Jeremy Waldron, *The Harm in Hate Speech* (Cambridge, MA: Harvard University Press, 2012), 37–38.

[45] I thank Nora Benedict for this observation.

[46] The Streisand effect is named after Barbara Streisand. Streisand sued a photographer for taking a picture of her house and making it publicly available. The lawsuit drew significant attention to the photo, which was largely unknown to the public. In an effort to suppress the photo, Streisand had, in fact, made the photo more famous.

[47] Antoon De Baets, "A Historian's View of the Right to Be Forgotten," *International Review of Law, Computers and Technology* 30, nos., 1-2 (2016): 63.

suggests. The more severe problem with online shaming rests in the second reason against online shaming as a functional response to wrongdoing.

The second reason to believe that online shaming drifts toward disintegrative shaming is that the Internet is a massively scaled environment where people who do not know each other interact.[48] The Internet is perhaps an open society par excellence; boundaries are fluid and almost anyone can acquire access to public spaces. As a result, you find individuals interacting with each other who do not know each other and have almost no interaction beyond, perhaps, a stray comment in a social media thread. In the context of shaming, then, social media shaming catalyzes the shaming of strangers by strangers. Because of this, online shaming almost necessarily drifts toward disintegrative shaming. Recall that a precondition for reintegrative shaming is that shamed people see themselves as connected to their shamers. Almost all of the people who piled on the woman in the introduction did not know her, and she did not know them. There is no opening for reconciliation. She only sees a mass of people who describe her as outside of the acceptable bounds of society.

The drift toward disintegrative shaming makes it unlikely for online shaming to be productive in the sense of promoting cooperative behavior. To be clear, sometimes disintegrative shaming is appropriate. For example, it strikes me as reasonable to conclude that Harvey Weinstein's pattern of sexual wrongdoing suggests that he is best excised from our community. But that is an extreme case. Most of the cases that go viral do no strike me as best subject to disintegrative shaming, but rather reintegrative shaming. These people are not monsters, but ordinary people who did something wrong. But, the very medium of shaming precludes respecting this fact. As a result, targets of shamers do not find themselves with any chance or hope for redemption. Instead, the targets of shame likely resent their treatment, and find themselves with limited opportunity to reintegrate themselves back into their ordinary lives.[49]

VI. Conclusion

I have argued that the problem of public shaming is fruitfully understood as arising from a tension within the technology of public shaming. On the one hand, public shaming occurs through the medium of technology or what I have called the artifacts of shame. Most recently, people shame through social media. On the other hand, public shaming is itself a social technology. Historically, public shaming has been most important in its role of promoting cooperation of individuals within communities. The problem arises because certain artifacts of shame do not mesh well with the social

[48] I explore more fully the relevance of scale to problems of public shaming in "The Problem of Public Shaming," *Journal of Political Philosophy* 22, no. 8 (2022): 188–208.

[49] In *So You've Been Publicly Shamed?* Ronson does an excellent job surveying the deep impact these episodes have on individual lives.

technology of shaming. In particular, I have argued that social media has a natural tendency to become a form of disintegrative shaming, which is ill suited to producing cooperative individuals. This drift toward disintegrative shaming latent in online shaming helps explain the tension in many people's attitudes (mine included) toward online shaming. It is clear that some targets of shaming did something wrong. But doing something wrong does not render one liable to excommunication from mainstream society. Unfortunately, public shaming on the Internet is prone to lead to this outcome, or so I have argued. Because of this, if we wish to preserve public shaming as a social technology, we need to find a way to realign the technology of public shaming. Perhaps, however, public shaming has become obsolete as a social technology. Much like the pillory and the stock is a relic from the past that is primarily of historic interest for us today, maybe it is time to put public shaming behind us a way of treating one another entirely.

Political Science, University of Georgia, USA