Moral outrage in the digital age

Moral outrage is an ancient emotion that is now widespread on digital media and online social networks. How might these new technologies change the expression of moral outrage and its social consequences?

M. J. Crockett

Moral outrage is a powerful emotion that motivates people to shame and punish wrongdoers[1]. Moralistic punishment can be a force for good, increasing cooperation by holding bad actors accountable. But punishment also has a dark side — it can exacerbate social conflict by dehumanizing others[2] and escalating into destructive feuds.

Moral outrage is at least as old as civilization itself, but civilization is rapidly changing in the face of new technologies. Worldwide, more than a billion people now spend at least an hour a day on social media, and moral outrage is all the rage online. In recent years, viral online shaming has cost companies millions, candidates elections, and individuals their careers overnight.

As digital media infiltrates our social lives, it is crucial that we understand how this technology might transform the expression of moral outrage and its social consequences. Here, I describe a simple psychological framework for tackling this question (Fig. 1). Moral outrage is triggered by stimuli that call attention to moral norm violations. These stimuli evoke a range of emotional and behavioural responses that vary in their costs and constraints. Finally, expressing outrage leads to a variety of personal and social outcomes. This framework reveals that digital media may exacerbate the expression of moral outrage by inflating its triggering stimuli, reducing some of its costs and amplifying many of its personal benefits.

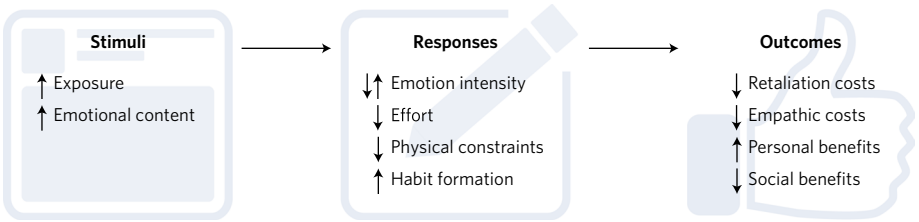

Fig. 1 | How digital media might transform moral outrage. Moral outrage is an emotion elicited by stimuli appraised as signifying moral norm violations. The subjective experience of outrage in reaction to such stimuli motivates the expression of behavioural responses such as gossip, shaming or punishment. Expressing outrage can lead to positive and negative outcomes for oneself and for society. Digital media may promote the expression of outrage by magnifying its triggers, reducing its personal costs and amplifying its personal benefits, while at the same time reducing its benefits for society.

Triggers of moral outrage

People become outraged when they think a moral norm has been violated[1]. A recent study conducted in the US and Canada suggests that encountering norm violations in person is relatively rare: less than 5% of reported daily experiences involved directly witnessing or experiencing immoral acts[3]. But the internet exposes us to a vast array of misdeeds, from corrupt practices of bankers on Wall Street, to child trafficking in Asia, to genocide in Africa — the list goes on. In fact, data from a study of everyday moral experience[3] show that people are more likely to learn about immoral acts online than in person or through traditional forms of media (Fig. 2a).

Before the internet existed, gossip served a purpose of spreading news about who could be trusted within local social networks[4]. By this logic, information should be shared as a function of its ability to reinforce trust and cooperation within the community. But online platforms have profoundly changed the incentives of information sharing. Because they compete for our attention to generate advertising revenue, their algorithms promote content that is most likely to be shared, regardless of whether it benefits those who share it — or is even true.

Research on virality shows that people are more likely to share content that elicits moral emotions such as outrage[5]. Because outrageous content generates more revenue through viral sharing, natural selection-like forces may favour 'supernormal' stimuli that trigger much stronger outrage responses than do transgressions we typically encounter in everyday life. Supporting this hypothesis, there is evidence that immoral acts encountered online incite stronger moral outrage than immoral acts encountered in person or via traditional forms of media (Fig. 2b). These observations suggest that digital media transforms moral outrage by changing both the nature and prevalence of the stimuli that trigger it. The architecture of the attention economy creates a steady flow of outrageous 'clickbait' that people can access anywhere and at any time.

The experience and expression of moral outrage

Moral norm violations cause people to experience moral outrage and to express it via gossip, shaming and punishment. Digital media might alter the subjective experience of outrage in several ways. By increasing the frequency and extremity of triggering stimuli, one possible long-term consequence of digital media is 'outrage fatigue': constant exposure to outrageous news could diminish the overall intensity of outrage experiences, or cause people to experience outrage more selectively to reduce emotional and attentional demands. On the other hand, studies have shown that venting anger begets more anger[6]. If digital media makes it easier to express outrage, this could intensify subsequent experiences of outrage. Future research is necessary to resolve these possibilities.

People express outrage in several ways that vary in terms of their effort and constraints. Offline, people can harm wrongdoers' reputations through gossip, or directly confront them with verbal sanctions or physical aggression. The latter two methods require more effort and also carry potential physical risks for the punisher. In contrast, people can express outrage online

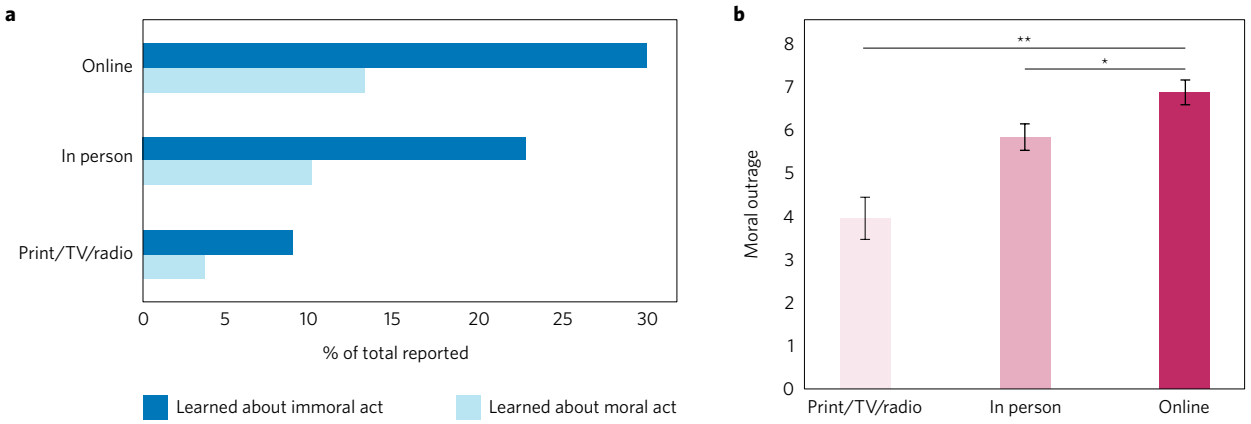

Fig. 2 | Exposure to immoral acts and resulting moral outrage, online versus offline. In a previous study[3], people's smartphones were signalled five times a day for three days to sample everyday moral experiences in a geographically and demographically diverse sample of North American adults ($N = 1252$). **a**, A reanalysis of this data revealed that participants were more likely to learn about immoral acts online than in person or via traditional forms of media (print, television and radio; $\chi^2 = 9.51$, $P = 0.009$). The figure displays the percentage of total reported moral/immoral acts that were learned about in each setting. **b**, For each immoral act, moral outrage was calculated by multiplying self-reported anger and disgust[1]. Moral outrage ratings were entered into a linear mixed-effects model using the MIXED command in SPSS with setting type (traditional media, in person or online) as a categorical BY variable. This revealed that immoral acts encountered online evoked more outrage than immoral acts encountered in person ($t_{(771)} = 2.723$, $P = 0.007$) or via traditional forms of media ($t_{(762)} = 4.555$, $P < 0.001$). Error bars represent s.e.m.; $^*P < 0.01$, $^{**}P < 0.001$.

with just a few keystrokes, from the comfort of their bedrooms, either directly to the wrongdoer or to a broader audience. With even less effort, people can repost or react to others' angry comments. Since the tools for easily and quickly expressing outrage online are literally at our fingertips, a person's threshold for expressing outrage is probably lower online than offline.

Furthermore, expressing outrage in person (that is, via verbal sanctions or aggression) is necessarily constrained by physical proximity to the wrongdoer. But expressing outrage online is not limited by location, time of day, or the probability of chance encounters with perpetrators. People can and do seek out targets of outrage, even if they are total strangers living across the globe in a different time zone. A paradigmatic example is the case of Justine Sacco, a woman who tweeted a comment about AIDS in Africa that many perceived to be racist. Within hours, she became the top trending topic on Twitter as millions of strangers around the world piled on the shaming bandwagon. The ease of piling on raises the intriguing possibility that in online settings, people may express moral outrage without actually experiencing the degree of outrage their behaviour implies.

A further, speculative hypothesis is that the design of digital media platforms may encourage habitual outrage expression. Offline, the stimuli that trigger outrage and the way people choose to respond are typically unique to the situation. But social media apps streamline triggering stimuli and available responses into a heavily designed 'stimulus–response–outcome' architecture that is consistent across situations. Clickbait headlines are presented alongside highly distinctive visual icons that allow people to express outrage with a tap of the finger. Positive feedback for these responses (likes, shares, and so on) is delivered at unpredictable times — a pattern of reinforcement well known to promote habit formation[7]. And just as a habitual snacker eats without feeling hungry, a habitual online shamer might express outrage without actually feeling outraged. Thus, when outrage expression moves online it becomes more readily available, requires less effort, and is reinforced on a schedule that maximizes the likelihood of future outrage expression in ways that might divorce the feeling of outrage from its behavioural expression.

Costs and benefits of moral outrage

Expressing moral outrage can be costly. Offline, moralistic punishment carries a risk of retaliation. But online social networks limit this risk. They enable people to sort themselves into echo chambers with sympathetic audiences[5]. The chance of backlash is low when you're only broadcasting moral disapproval to like-minded others. Moreover, they allow people to hide in a crowd. Shaming a stranger on a deserted street is far riskier than joining a Twitter mob of thousands.

Another cost of outrage expression is empathic distress: punishing and shaming involves inflicting harm on other human beings, which for most of us is naturally unpleasant. Online settings reduce empathic distress by representing other people as two-dimensional icons whose suffering is not readily visible. It's a lot easier to shame an avatar than someone whose face you can see.

Despite these and other costs, people are still obviously motivated to express moral outrage. One reason for this is that expressing moral outrage benefits individuals by signalling their moral quality to others[8]. That is, outrage expression provides reputational rewards. People are not necessarily conscious of these rewards when they express outrage. But the fact that people are more likely to punish when others are watching[9] indicates that a concern for reputation at least implicitly whets our appetite for moral outrage. Of course, online social networks massively amplify the reputational benefits of outrage expression. While offline punishment signals your virtue only to whoever might be watching, doing so online instantly advertises your character to your entire social network and beyond. A single tweet with an initial audience of just a few hundred can quickly reach millions through viral sharing — and outrage fuels virality[5].

Expressing moral outrage does not merely benefit individuals. It can also benefit society by holding bad actors accountable and sending a message to others that such behaviour is socially unacceptable. Online platforms put these tools in the hands of everyone, enabling traditionally disempowered groups to check the behaviour of more powerful interests. Expressing outrage online may heighten people's

adherence to a cause by publicly committing themselves to action. But digital media limits the potential social benefits of moral outrage in several ways. First, ideological segregation online prevents the targets of outrage from receiving messages that could induce them (and like-minded others) to change their behaviour. For politicized issues, moral disapproval ricochets within echo chambers but only occasionally escapes[5]. Second, by lowering the threshold for outrage expression, digital media may degrade the ability of outrage to distinguish the truly heinous from the merely disagreeable. Third, expressing outrage online may result in less meaningful involvement in social causes, for example through volunteering or donations. People are less likely to spend money on punishing unfairness when they are given the opportunity to express their outrage via written messages instead[10].

Finally, there is a serious risk that moral outrage in the digital age will deepen social divides. A recent study suggests a desire to punish others makes them seem less human[2]. Thus, if digital media exacerbates moral outrage, in doing so it may increase social polarization by further dehumanizing the targets of outrage. Polarization in the US is accelerating at an alarming pace, with widespread and growing declines in trust and social capital. If digital media accelerates this process further still, we ignore it at our peril.

Conclusions

If moral outrage is a fire, is the internet like gasoline? Technology companies have argued that their products provide neutral platforms for social behaviours but do not change those behaviours. This is an empirical question that behavioural scientists should address, because its answer has ethical and regulatory implications.

The framework proposed here offers a set of testable hypotheses about the impact of digital media on the expression of moral outrage and its social consequences. Digital media may promote the expression of moral outrage by magnifying its triggers, reducing its personal costs and amplifying its personal benefits. At the same time, online social networks may diminish the social benefits of outrage by reducing the likelihood that norm-enforcing messages reach their targets, and could even impose new social costs by increasing polarization.

Preliminary data support the framework's predictions, showing that outrage-inducing content appears to be more prevalent and potent online than offline. Future studies should investigate the extent to which digital media platforms intensify moral emotions, promote habit formation, suppress productive social discourse, and change the nature of moral outrage itself. There are vast troves of data that are directly pertinent to these questions, but not all of it is publicly available. These data can and should be used to understand how new technologies might transform ancient social emotions from a force for collective good into a tool for collective self-destruction. ❐

M. J. Crockett

The Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, USA.
e-mail: mj.crockett@yale.edu

DOI: 10.1038/s41562-017-0213-3

References

1. Salerno, J. M. & Peter-Hagene, L. C. *Psychol. Sci.* **24**, 2069–2078 (2013).
2. Fincher, K. M. & Tetlock, P. E. *J. Exp. Psychol. Gen.* **145**, 131–146 (2016).
3. Hofmann, W., Wisneski, D. C., Brandt, M. J. & Skitka, L. J. Replication data for: morality in everyday life. *Harvard Dataverse* https://doi.org/10.7910/DVN/26910 (2014).
4. Dunbar, R. I. *Rev. Gen. Psychol.* **8**, 100–110 (2004).
5. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. & Bavel, J. J. V. *Proc. Natl Acad. Sci. USA* **114**, 7313–7318 (2017).
6. Bushman, B. J. *Pers. Soc. Psychol. Bull.* **28**, 724–731 (2002).
7. Dickinson, A., Nicholas, D. J. & Adams, C. D. *Q. J. Exp. Psychol. Sect. B* **35**, 35–51 (1983).
8. Jordan, J. J., Hoffman, M., Bloom, P. & Rand, D. G. *Nature* **530**, 473–476 (2016).
9. Kurzban, R., DeScioli, P. & O'Brien, E. *Evol. Hum. Behav.* **28**, 75–84 (2007).
10. Xiao, E. & Houser, D. *Proc. Natl Acad. Sci. USA* **102**, 7398–7401 (2005).

Competing interests

The author declares no competing interests.