Are We All 'Harmless Torturers' Now?

Paul Bloom and Matthew Jordan

August 9, 2018

There is a dial in front of you, and if your turn it, a stranger who is in mild pain from being shocked will experience a tiny increase in the amount of the shock, so slight that he doesn't even notice it. You turn it and leave. And then hundreds of people go up to the dial and each also turns it, so that eventually the victim is screaming in agony.

Did you do anything wrong? Derek Parfit, the influential British philosopher who died in January 2017, called this the case of the Harmless Torturer. Parfit first considered a simpler scenario in which a thousand torturers each turn the dial a thousand times on their own victim. This is plainly terrible. But then he explores a contrasting case where each of the torturers turns a dial a thousand times — each turn shocking a different one of the thousand victims. The end result is the same; a thousand people in agony. And yet morally it feels different, since nobody, individually, caused any real harm to any single individual.

This seems like the sort of clever technical example that philosophers love — among other things, it's a challenge to a utilitarian view in which the wrongness of an act is reduced to its consequences — but one with no actual real-world relevance. But the world has changed since Parfit published his scenario in 1986. Today, in 2018, the two authors of this article are Harmless Torturers, and you — regardless of which side of any particular issue you are on — probably are one, too.

Parfit's scenario unfolds all the time on social media. Someone writes something ugly about you on Facebook; depending on your relationship with that person, it may or may not be personally hurtful, but nobody notices it, and so it's not a big deal. But if a day later it has a thousand likes and several hundred mocking comments, you might well be crushed. Even though any particular comment on its own may have caused you little or no pain, the aggregate effect is far more severe.

In his 2015 book, "So You've Been Publicly Shamed," Jon Ronson explored the effects of internet mobbing, including the story of a woman whose ironic tweet about white privilege went terribly wrong, generating tens of thousands of angry tweets, leading her to lose her job and go into hiding. Since then, the mob has been busy: its focus has turned to a dentist who killed a lion, a series of white women who without apparent cause called the police on black people, a left-wing professor who asked her friends to expel a journalist from a protest, and many many others.

When we think of the savagery of social media, we often think of awful individual behavior — death threats and rape threats; the release of personal information, including home addresses and the locations of the victim's children; vicious lies; and the like. Harmless Torturers never go that far; we just like, retweet and add the occasional clever remark. But there are millions of us, and we're all turning the dial.

Parfit never tells us what motivates the torturers in his thought experiment, but there are a lot of considerations in everyday life. We are moral animals, after all. There is abundant evidence from laboratory studies and from real life that we wish to see immoral agents get their comeuppance. And this is grounded in sound evolutionary logic: If we weren't disposed to punish or exclude bad actors, there would be no cost to being a bad actor, and cooperative societies couldn't get off the ground.

There is also a sort of social credit that comes with being seen as a moralistic punisher; we want to show off our goodness to others, to signal our virtue. We are more likely to punish when others are watching, and there is evidence that third parties think more highly of — and are more likely to later trust — those who punish bad actors versus those who sit back and do nothing.

Moral and social motivations are hard to disentangle in the real world. When a philosopher — Bryan W. Van Norden, writing for The Stone — says, "Like most Americans, I spontaneously cheered when I saw the white nationalist Richard Spencer punched in the face during an interview," it is hard to tell how much of this is a report of genuine pleasure at a racist getting his just deserts and how much of it is the desire to be seen as anti-racist to an approving audience. If the conscious motivation for our condemnation is mostly signaling, the idea of making our victims suffer might never even occur to us. And the ease with which we can express moral outrage online — in most cases without any real world repercussions — makes this condemnation that much easier. As our Yale colleague Molly Crockett put it in an article recently, "If moral outrage is a fire, is the internet like gasoline?"

There is also a system of reward built into online shaming. In an essay at Quilette, "I Was the Mob Until the Mob Came for Me," a self-described former social justice warrior writing under the pseudonym Barrett Wilson described the thrill he felt in his mobbing days: "Every time I would call someone racist or sexist, I would get a rush. That rush would then be reaffirmed and sustained by the stars, hearts and thumbs-up that constitute the nickels and dimes of social media validation."

But isn't this death by a thousand cuts a good thing? If it were Hitler, wouldn't you be right to let him have it? Yes — but the problem is that when we are infused with moral outrage, acting as part of a crowd and operating in a virtual world with no fixed system of evaluation, law or justice, all our enemies are Hitler. There can easily be, as Ronson puts it, a "disconnect between the severity of the crime and the gleeful savagery of the punishment."

Certainly, public shaming can have positive effects; sometimes the angry mob gets it right — punching up and hitting the right target. But Harmless Torturers can just as easily swarm the weak; the mobbing can be based on lies and confusions or ignorantly encouraged by powerful celebrities and politicians, including, notably, the current president.

The Harmless Torturer effect isn't limited to social media; we can also see the effects of aggregation when it comes to more impactful individual actions. Likes and retweets bear a structural similarity to execution by stoning, particularly if the crowd is large: it's hard to see the victim, and nobody has good aim. Social shunning is another case, torture through the accumulation of omissions — individuals avoiding social contact with a certain person — as opposed to actions.

The writer Julian Sanchez, a senior fellow at the libertarian Cato Institute, has used the Parfit example in a discussion of behaviors like whistling at women on the street or jokingly using offensive language. He points out that a typical response to criticism of such behaviors is denial — many feel that there are no bad intentions in these actions and nobody is hurt by them. But even if this is true for certain individual actions, this situation changes when we consider them in the aggregate, happening over and over and over again, a thousand times from a thousand different people, their impact becomes more clear.

It's difficult to change the sorts of behaviors that Sanchez discusses, and probably even harder to get people to rethink internet mobbing, which, when we feel we have right on our side, can feel so good. Our minds have evolved to think about the effects of our individual actions; it's hard to consider aggregate effects. But the lesson of Parfit's Harmless Torturer is that if we want to be decent people, we should try.