
INCREASING THE TRANSACTION COSTS OF HARASSMENT

WOODROW HARTZOG* & EVAN SELINGER**

Wouldn't it be nice if the rules, agreements, and guidelines designed to prevent online harassment were sufficient to curb improper behavior? As if. Wrongdoers are not always so easily deterred. Sometimes these approaches are about as effective as attacking tanks with toothpicks.

As Danielle Citron contends in her critically important work, *Hate Crimes in Cyberspace*, the design of the Internet facilitates vitriol and abuse, even when it is legally, contractually, and normatively prohibited. Communicating almost effortlessly at distance—sometimes anonymously and typically with minimized body language—can heighten emotional detachment and blunt moral sensitivity. Tragically, when a mediated environment makes it easy to harass others, harassment occurs, all things being equal.

Fortunately, there's hope. Since mediated environments can fuel harassment, designing online spaces to make harassment difficult—or, in economic terms, costly—should diminish it. But as with anything important, the devil is in the details.

Citron attests to the wisdom of the “Designing for Better Selves” approach by arguing smart design choices can “nudge users to treat others as deserving respect rather than as objects that can be mistreated.” She also maintains that online intermediaries—such as website and app designers—should adopt conscience design principles and strategies along with “clear policies prohibiting cyber harassment” and “robust enforcement” of them (239-40).¹

* Associate Professor, Samford University's Cumberland School of Law; Affiliate Scholar, Center for Internet and Society at Stanford Law School.

** Professor of Philosophy, Rochester Institute of Technology.

¹ By supporting civil design strategies, Citron joins scholars from across the disciplines in either advocating embedding ethics and etiquette into code (or code-like structures) or discussing the ramifications of doing so. See LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (1999); PETER-PAUL VERBEEK, *MORALIZING TECHNOLOGY* (2011); JONATHAN ZITTRAIN, *THE FUTURE OF THE INTERNET AND HOW TO STOP IT* (2008); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules through Technology*, 76 *TEX. L. REV.* 553 (1997-1998); BATYA FREEDMAN, *VALUE SENSITIVE DESIGN*, <http://www.vsdesign.org/>; Langdon Winner, *Do Artifacts Have Politics?*, 109 *DAEDALUS* 121 (1980); Bruno Latour, *Where are the Missing Masses? The Sociology of a Few Mundane Artifacts*, in *SHAPING TECHNOLOGY/BUILDING SOCIETY: STUDIES IN SOCIOTECHNICAL CHANGE* 225-48 (W. E. Bijker and J. Law eds., 1994); Helen Nissenbaum, *Values in Technical Design*,

In order to advance the conversation Citron started, it's worth exploring what this strategy entails. Since Citron's vision for leveraging design to fight harassment needs a locus, we are proposing transaction costs.

TAKING TRANSACTION COSTS SERIOUSLY TO PREVENT ABUSE

In economic theory, transaction costs refer to a range of expenses that are required for participating in market exchanges.² But the concept can be expanded to cover the expense required to do anything. For example, time and effort are valuable resources, and we often evaluate how desirable possibilities are by calculating how much of these goods are required to actualize them.

Consider how companies limit the amount of complaints they need to attend to by hiding contact information or only providing consumers with limited contact information such as street mailing addresses and PO box numbers.³ While an e-mail demanding a refund is quick to construct and easy to share, the effort required to compose and send a comparable physical letter—write it on (or print it to) paper, put it in an envelope, address the envelope, get a stamp, affix the stamp, and put the envelope in a mailbox—can be an effective deterrent. This is especially the case if you are suspicious that the note will be ignored the first time around and you will eventually have to go through the laborious process again. By comparison, it is much easier to forward an archived e-mail missive.

Citron validates a comparable strategy for using time and effort as deterrents by citing Professor Nancy Kim's nudge proposal: "Companies could nudge users to think about others' humanity by slowing down the posting process in the hopes that posters would use the time to think more carefully about what they say"(240-41).⁴

ENCYCLOPEDIA OF SCIENCE, TECHNOLOGY AND SOCIETY, (C. Mitcham ed., 2005); Helen Nissenbaum, *How Computer Systems Embody Values*, *COMPUTER* 34.3, 119-20 (2001); Helen Nissenbaum, *Values in the Design of Computer Systems*, *COMPUTERS AND SOCIETY* 28.1, 38-39 (1998).

² See, e.g., Eduardo González Fidalgo, *Transaction Cost Economics*, *INTRODUCTION TO BUSINESS*, <http://intobusiness.weebly.com/transaction-cost-economics.html>; DOUGLAS W. ALLEN, *TRANSACTION COSTS* (1999), <http://www.sfu.ca/~allen/allentransactioncost.pdf>; ALEXANDRA BENHAM AND LEE BENHAM, CHAPTER 11: THE COSTS OF EXCHANGE, IN *THE ELGAR COMPANION TO TRANSACTION COST ECONOMICS* (Peter G. Klein and Michael E. Sykuta eds., 2010), <http://www.elgaronline.com/view/9781845427665.00018.xml>.

³ Zoya Sheftalovich, *Sick of Companies Playing Hide and Seek?*, *CHOICE* (Sept. 5, 2014), <https://www.choice.com.au/shopping/consumer-rights-and-advice/taking-action-and-making-a-complaint/articles/companies-that-are-difficult-to-contact> ("With some companies making it so difficult to find their contact details, or leaving you on hold or waiting for a reply email for so long, it's hard not to conclude they're intentionally trying to prevent their customers from being able to contact them.").

⁴ Nancy S. Kim, *Web Site Proprietorship and Online Harassment*, 2009 *UTAH L. REV.* 993, <http://ssrn.com/abstract=1354466>.

We believe online abuse can be mitigated by a range of related strategies that manipulate transaction costs. In the remaining sections, we emphasize three key areas companies have targeted: speech, access, and defense.

THE COST OF SPEECH

Online harassment usually requires communication, and so the most direct way to limit it is to make harmful speech costly to conduct. For example, social media messaging systems can be restricted to designated users. Twitter users can limit private messages to their “followers” and Facebook users can do the same with their “friends.” Additionally, nearly every social platform allows users to block others. Blocking restrictions impact behavior by forcing those who are blocked to comply with or else expend the effort required to work around (e.g., deception) or override (e.g., hacking) the constraints.

More modest interventions can effectively nudge civility, too. Content filters on the anonymous social media app Yik Yak aim to prevent users from posting someone’s full name. This intervention makes it harder for abusers to locate and learn about potential targets.

Yik Yak also targets potentially problematic content by prompting users (under triggering circumstances) with the following message: “Pump the brakes, this yak may contain threatening language. Now it’s probably nothing and you’re probably an awesome person but just know that Yik Yak and law enforcement take threats seriously. So you tell us, is this yak cool to post?” This notification imposes additional time and effort for offenders to process and respond to what has been conveyed, and thus increases the cost of speaking. Is it a foolproof plan? Of course not. But it just might let cooler heads prevail over temporarily heated emotional reactions.

THE COST OF ACCESS

Blocking features can make it harder for abusers to access their intended victims’ information. Users can be blocked at various levels, ranging from being unable to share, tag, and upvote other’s posts, to being fully unable to access any aspect of content associated with another user’s profile.

The efficacy of and desire for strong access to blocking restrictions was made explicit in 2013. During this period, Twitter briefly altered its policy. For a limited time, blocked users could follow, retweet, and favorite a public user who had blocked them, and blockees stopped being notified when someone decided they merited blocked status.⁵ Harassment victims responded swiftly and loudly. As a result, Twitter reversed course.

Harassment can even be deterred by raising transaction costs and making information unsearchable. In previous work, we argued that when information

⁵ Kashmir Hill, *Blocking People On Twitter Now Just Mutes Them (Update: Psych!)*, FORBES (Dec. 12, 2013), <http://www.forbes.com/sites/kashmirhill/2013/12/12/blocking-people-on-twitter-now-just-mutes-them/>.

is hard to find, it is relatively safe.⁶ Indeed, this logic underlies Europe's so-called "Right to Be Forgotten"—which is, in reality, a right to hide from search engines.⁷ It is also why harassment and abuse victims worry when Facebook makes all profiles searchable regardless of privacy settings.⁸

THE COST OF DEFENDING

Transaction costs also are important for defending against harassment. Unlike the previous two examples, however, the goal here is to facilitate action. For example, it should be easy to report abuse to social media administrators. Recognizing this, most popular social media have a report button in close proximity to users' posts. Yik Yak even implemented a voting system that immediately removes any post that receives five "downvotes." Crucially, all users easily can downvote a post without jumping through bureaucratic hoops (like satisfying registration requirements) or engaging in additional clicks.

By contrast, systems designed to make abuse difficult to report show little respect for users. Professor Mary Anne Franks thus criticized Twitter's previous abuse reporting system, alleging that it "set up a game that targets of abuse can never win."⁹ Under Twitter's old policy, parties conveyed harassment complaints on forms that took more time to complete than reporting spam, which required only the click of a readily available button.¹⁰

In *Hate Crimes in Cyberspace* Citron makes it clear that harassment cannot be eradicated solely by policies that penalize bad behavior. Systems and

⁶ Woodrow Hartzog and Evan Selinger, *Obscurity: A Better Way to Think About Your Data Than 'Privacy,'* ATLANTIC (Jan. 17, 2013), <http://www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/>.

⁷ Evan Selinger and Woodrow Hartzog, *Google Can't Forget You, But It Should Make You Hard To Find,* WIRED (May 20, 2014), <http://www.wired.com/2014/05/google-cant-forget-you-but-it-should-make-you-hard-to-find/>.

⁸ Samantha Allen, *How Facebook Exposes Domestic Violence Survivors,* DAILY BEAST (May 20, 2015), <http://www.thedailybeast.com/articles/2015/05/20/how-facebook-exposes-domestic-violence-survivors.html>; James Vincent, *Facebook Tells Users They Can't Hide From Searches,* INDEPENDENT (Oct. 11, 2013), <http://www.independent.co.uk/life-style/gadgets-and-tech/facebook-tells-users-they-cant-hide-from-searches-8874562.html>.

⁹ Mary Anne Franks, *The Many Ways Twitter Is Bad at Responding to Abuse,* ATLANTIC (Aug. 14, 2014), <http://www.theatlantic.com/technology/archive/2014/08/the-many-ways-twitter-is-bad-at-responding-to-abuse/376100/>.

¹⁰ *Id.* ("Contrast this to the procedure for reporting spam. To report spam, a user must click a button that says 'This account is spam.' That's it. Twitter is oddly unconcerned about false or unauthorized reports of spam: There are no questions about the user's involvement with the alleged spam, no requirement to provide links or explain how the content qualifies as spam, no requirement of a signature, no need to fear retaliation from the reported spammer.").

technologies must be designed to protect against it as well. To this end, we have argued that transaction costs associated with speech, access, and defense are the right starting points for determining how design can play a critical role in improving online interactions.

Of course, these categories are fluid and imprecise. For example, barriers to access might simultaneously limit speech. And, transaction costs might even be unrecognizable in the extreme. Consider authentication requirements for websites that limit access to users with particular backgrounds, like medical training. You could classify the years it takes to acquire such training as a transaction cost investment. But such semantics would stretch the idea beyond useful limits.

And let's not forget, design strategies aren't magical techno-fixes. For example, increasing the transaction costs for communicating online can lead to unpopular opinions being censored and strongly worded convictions being watered down.¹¹ Furthermore, designing potent large-scale civility nudges might constitute a form of techno-social engineering that adversely impacts people's judgment and character.¹²

It is also important to determine whether the companies that are committed to identifying promising new design strategies will need to run new experiments on their users to determine optimum transaction cost levels. If so, questions arise as to whether corporate approaches to changing user experiences will be ethically sound.¹³

Finally, more attention needs to be given to the role of the law in encouraging or prohibiting certain kinds of design. Not all companies will embrace good design as a matter of individual discretion. And this means mandates might be preferable in some instances.

It is hard work to construct and assess a transaction-cost framework for fighting abuse. But the finished product will be a useful tool for helping companies and policy makers focus their design efforts on effective and wise options. A mature version can give us a sense of how committed companies really are to preventing harassment and fostering civility.

¹¹ Evan Selinger, *When Nudge Comes to Shove*, SLATE (July 7, 2013), http://www.slate.com/articles/health_and_science/new_scientist/2013/07/nudge_critiques_is_nudging_behavior_unethical_infantilizing_coercive_or.html.

¹² See Brett Frischmann, *Human-Focused Turing Tests: A Framework for Judging Nudging and Techno-Social Engineering of Human Beings*, Cardozo Legal Studies Research Paper No. 441 (Sept. 22, 2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2499760.

¹³ See Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97 (Sept. 3, 2013), <http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-review-boards>.