Feeding the Troll: Online Hate Speech as Communal Act

Matthew Overstreet, University of Pittsburgh

This paper presents online hate speech as a relational process. It explores some popular conceptions of the internet troll, and using the work of Diane Davis, argues that the troll should be understood not as an embodied actor, but as an intersubjective discursive effect. All communication, Davis argued, is underlain by a structure of exposure—a presymbolic openness to the other. Language and meaning, for Davis, are also inherently fluid and fickle. I argue that these ideas find their furthest expression in the contemporary social media environment, and that the troll is an effect of this environment. Seen in this way, the troll does not speak, but is instead spoken by community norms and technological affordances. From this premise, I argue that a nonreferential view of language can help temper the troll by robbing hate speech of its performative force.

What goes on inside the mind of a Twitter troll? A pinned tweet and accompanying video from @killanderson provide a hint. The video documents an encounter from August 2016. It starts with a screenshot of another user's tweet: "White people do not get to tell people of color what is and is not racist." We hear the click of keys as @killanderson types a response: "You need to go back. You don't deserve to live in a country run by white people." This is followed by a montage of angry responses, interspersed with screenshots of @killanderson's mentions page as the number of views—people who have seen this troll's troll—climbs past 100k, 200k, to nearly 300k. The montage is overlaid by steadily building female laughter. Above the video, the text of the tweet reads simply, "I win."

In the following pages, I'd like to try to understand this tweet—and the behavior of trolls in general—in relational terms. The troll, and his target, I'll argue, define, and are defined by, each other and the world from which they emerge. This means that we—you and I—are called to take responsibility for the troll. My hope is that in doing so, we can begin to demythologize his hate speech, and ultimately, rob it of its force.

To begin, let's consider the video we just watched. In what performance is the author engaged? What does he seek to accomplish? Well, on the surface, it's a demonstration of power, of control. It seeks to show that the rhetor understands social norms (racism is bad), how norms can be breached (blatant racism), and the way in which others will respond to a breach (outrage). It also shows that he seeks this outraged response. Why? Simply because. He needs no other reason. So this tweet is persuasive rhetoric in a very basic sense. It seeks to persuade the viewer that the troll can do with her as he wishes.

Now certainly, one can't reason with a troll like @killanderson. His original claim that you need to go back contains no statement of fact which could be disputed (the original poster identifies as Native American, so to where would she return?). In fact, any attempt to argue that she does not need to go back, or that such a claim is nonsensical, would only work to acknowledge the troll's larger claim: that he is master of the conversation, that he knows what buttons to push, and pushes them at his whim.

The troll's embrace of unreasonableness has led some to compare him with the Anti-Semite, as described by Jean-Paul Sartre. Of this figure, Sartre (1995) writes:

> Never believe that anti-Semites are completely unaware of the absurdity of their replies. They know that their remarks are frivolous, open to challenge. But they are amusing themselves, for it is their adversary who is obliged to use words responsibly, since he believes in words. The anti-Semites have the right to play. They even like to play with discourse for, by giving ridiculous reasons, they discredit the seriousness of their interlocutors. (p. 20)

On the surface, this description does seem to apply to the troll. His you need to go back, with its willing incoherence does, we could say, work to discredit the seriousness of the original, presumably sincere statement as to race relations. Rather than playing within the rules—presenting an argument that white people have much to add to definitions of racism, for example—the troll destabilizes the base on which the conversation rests. Unlike the cultural left, with its refrain that words matter, the troll seems to

believe that words do not matter. Or, more precisely, that they matter in a different way. To have rhetorical impact—which is all the troll seeks, remember—words need not be true or make sense. The goal is to amuse oneself; absurdity ups the enjoyment.

The comparison of the troll with Sartre's anti-Semite is enlightening, but it fails in one key respect. Sartre is clear that for the anti-Semite, hatred of the Jew is a matter of faith: he "has chosen to devaluate words and reason" and instead respect only a pre-cognitive repulsion (p. 19). This repulsion works on a bodily/affective level. Sartre writes, for example, of a lover who is stricken impotent when he learns that his partner is a Jewess. So for the anti-Semite, hatred is passionate, physical and localized. He hates the Jew. He is a true believer, and nothing can shake his belief. This faith, in turn, allows him to play with words, because his logic goes deeper than language.

For me, it's hard to conceive of the troll as a true believer. There's racism here, and sexism, but it often seems incidental. A quick review of @killanderson's tweets, for example, reveals the standard attacks on minorities and women, along with contempt for anime and Generation Y. Unlike Sartre's anti-Semite, in other words, the troll's hatred is not localized. What are we to make of this?

A recent Medium post by writer and web comic creator Dale Beran (2017) provided insight into the psychology of the troll. Beran focused on the trolls of the so-called alt-right. As he saw it, these trolls are young men (almost exclusively young men) who have "failed at the real world and have checked out of it and into the fantasy worlds of internet forums and video games." For these subjects, Beran wrote, "America, and perhaps existence itself is a cascade of empty promises and advertisements." Their response to this artifice is a retreat from all obligations, yes, and nihilism, yes, but also something more. The alt-right, Beran argued, know that the system is rigged, and *celebrate* that fact. Through symbols such as Pepe the frog, they embrace loserdom, reveling in deplorableness. They recognize that they are grotesque neckbeards, but do not care. Or, more precisely, want to be *perceived* as not caring.

This focus on perception is key. According to Beran, the troll is acutely aware of his lack of social status. He knows he is deplorable (in our eyes). And it hurts him deeply. He performs hate, in turn, because it allows him to display, as to ultimately transcend, this deplorableness. He turns weakness into strength by embracing it. Integrally though, this transcendence can only be achieved with the help of the community. He needs an audience for his performance, bodies capable of being affected by his hate. In short, the community has defined the troll. His trolling is an attempt at redefinition. But he is not the ultimate arbiter of his fate. The final say as to what the troll is, what his actions mean, still rests with the community.

The above analysis, I hope, hints at the thoroughly relational nature of the troll. Whereas he wants to be perceived as an autonomous agent, a master manipulator, doing with others as he pleases, he is instead deeply reliant on those he affects. This relationship is so intimate, in fact, that I'd say there is no troll apart from the systems in which he moves. Liberalism, capitalism, language, Twitter: the troll is more effect than cause. To understand him, therefore, we must understand that which defines him. The work of rhetorician Diane Davis can be of assistance.

In *Inessential Solidarity: Rhetoric and Foreigner Relations* (2010), Davis, following Emmanuel Levinas, suggested that all communication is underlain by a "fundamental structure of exposure" (p. 3). There is, she wrote, "an originary (or preoriginary) rhetoricity—an affect*ability* or persuad*ability*—that is the condition for symbolic action" (p. 2). Before we can communicate, in other words, we must be exposed, open to the alterity of the other. This means that We always comes before I. The individual, the singular subject, Davis wrote, "is exposed to an inappropriable outside that constitutes it, affects and alters it, prior to and in excess of symbolic intervention" (p. 7). Before we speak, to be capable of speech, we must be defined by that which is not us.

Twitter—@killanderson's weapon of choice—is notable for being the rare social media platform that allows users to contact each other without prior permission. Hypothetically, any user can contact any other user at any time, and can say anything. As with language itself, though, there's an admission cost: to speak, users must allow themselves to be spoken to. In this sense, the platform enacts, at a fundamental level, the preoriginary addressability that Davis argued underpins all meaning making. As language users, we are never alone. Twitter makes this apparent.

The platform also provides insight into the nature of meaning. In an earlier book, *Breaking Up (at) Totality: A Rhetoric of Laughter* (2000), Davis maintained that all symbolic action is suffused with a form of erotic energy she terms laughter. Reason and logic, along with conventional discursive forms, attempt to clean up this eroticism, to stabilize meaning. This project is doomed to failure though. As a result, subjects and social structures are fluid, fickle, and in a constant state of overflow. "To be spoken by a language contorted in laughter," she wrote, "is to be spoken by language on the loose: no/thing is excluded, censored, or negated" (p. 95).

Once again, social media proves Davis's analysis apt. In the hyperconnected digital realm meaning—any meaning—can be challenged, twisted, reshaped from without. Twitter, in particular, is a fecund semantic swamp. What is the meme economy, for example, but the unchecked proliferation of meaning? New forms emerge, and with them new logics, only to immediately be submerged by newer forms and logics. Reason, as embodied in traditional philosophical discourse, has no place here. Same with grammar. And morality. The old rules simply do not apply; instead, laughter in its most primal and yes, *erotic* form rules the day.

The semantic fecundity of the digital realm allows for great creativity. Foundations are innately unstable; perversion and mutation occur constantly. The possibility of (re)invention—I'd like to suggest that this is social media's ultimate public good. Ironically, though, the same conceptual fluidity which fuels invention allows for the emergence of the troll. Here is a subject uniquely able to exploit our digital nakedness—the fact that to speak we must be spoken. Here is a subject capable of feeding off the lack of fixed meanings, the excess inherent in a language on the loose. What gives him this power? The answer, I'd argue, is in the extent to which the troll is bound up with his tools. Social media platforms are designed to capture your attention, to keep you engaged. *The troll obeys the very same logic*. His redefinition is predicated on your attention, remember. And he does what is necessary to obtain it. In this sense, he uses social media exactly as it is intended to be used. He is, we can say, less a bug, than a feature.

So if the troll is one with the system, how do we temper him? How do we allow meanings to evolve, but also check their ability to cause harm? The first impulse for many is to try and limit potential meanings, re-erect some of the barriers that technology has torn down. On Twitter, this typically takes the form of appeals to authority (demands that certain forms of speech, or certain speakers, be banned). Now, this strategy can work. It can render digital environments safer, more family friendly. Censorship is dangerous though, I'd argue, because it risks robbing these environments of their *raison d'être*. The value of tools like Twitter, remember, is their ability to help us think in new ways, ways that we can't predict or control. Limitations on meaning—any meaning—risk compromising this.

Of course, many scholars have been thinking about ways to balance freedom and security in the digital realm. In a recent *Time* magazine article, Whitney Phillips and Ryan Milner, authors of *The Ambivalent Internet: Mischief, Oddity, and Antagonism Online*, suggest a renewed commitment to what could be called humanistic ethics. They argue that digital tools often hide what's really at stake in a given situation. Real people are "flattened to pixels on a screen," abstracted from "their full emotions, politics and history" (para. 6). This decontextualization allows users to act cruelly, thoughtlessly. As an antidote, Phillips and Milner argued for recognition of the serious real-world repercussions of our online actions.

I agree it's important to consider consequences. But can such an ethics temper the troll? The troll, as we've seen, wants to be perceived as a master manipulator. It seems that publicizing the real world repercussions of his actions would only bolster this image. It would paint him as prime mover, an active agent capable of asserting force beyond the systems which define him. Let's not give him that much credit. The troll, remember, rather than being above the game, is brought into being by the game. Its attributes are thus his own. And herein lies his weakness. He is unstable, excessive, exposed. There is no I without we, remember, no troll without tools and target. This puts him in a very precarious position.

So what does this sort of relational analysis mean in practical terms? Well, it seems to me that if we refuse to grant the troll agency—if we view him solely as an intersubjective effect—we must, in a sense, take responsibility for him. We exist in every hateful word he utters, Davis would claim. And we must acknowledge this trace. When we do, what happens? I don't know; this is something we need to work out.

I suspect, though, that when hate speech is viewed as relational, our response to it will change. Maybe we will stop trying to temper the troll by force. Maybe we will see that the more we protest, the more potential for lulz.

If the troll can't be shut down by force, what then? This is a question with no easy answer, but perhaps Davis's work offers something like a starting point. Quoting Victor Vitanza, she suggested an "antibody rhetoric" capable of "enhancing our abilities to tolerate the incommensurabilities" which make up the postmodern condition (p. 102). As I read it, such a rhetoric demands the utter rejection of foundations, a rejection of even the pretense of an objective (or intersubjective) referent. In short, it means we must come to view all language—even terrible, hurtful language—as a tool deployed to achieve certain ends. When dealing with the troll, in other words, we should focus not on truth, but consequences. And we should work to deprive him of the consequences he desires.

So, in conclusion, I've argued that we should see the troll as something we, the good people of social media, call into being. He is defined from without, and only through this process gains the power to affect. This means that by redefining him, and what his rhetoric means, we can rob it of its performative force. Of course, I recognize that this is easier said than done. It would require that we come to believe (or at least act as if) words do not matter. Or they matter differently. This is a hard position to accept. Indeed, word merchants of all stripes want us to believe the opposite. In a world without limits though—which for better or worse is our world—it may be our only option.

References

Beran, Dale. (2017, February 14). 4chan: The skeleton key to the rise of Trump. Retrieved February 26, 2017, from https://medium.com/@DaleBeran

Davis, Diane. (2000). *Breaking up [at] totality: A rhetoric of laughter.* Pittsburgh, PA: University of Pittsburgh Press.

Davis, Diane. (2010). *Inessential solidarity: Rhetoric and foreigner relations.* Pittsburgh, PA: University of Pittsburgh Press.

Phillips, Whitney, & Milner, Ryan. (2017, June 29). We are all Internet bullies. TIME. Retrieved July 3, 2017, from http://time.com/4839798/trump-troll-mika-tweet-online-ethics/

Sartre, Jean-Paul. (1995). *Anti-Semite and Jew: An exploration of the etiology of hate.* New York, NY: Schocken.