Only Human (In the Age of Social Media)[1]

Barrett Emerick and Shannon Dea[2]

ABSTRACT: This chapter argues that for human, technological, and human-technological reasons, disagreement, critique, and counterspeech on social media fall squarely into the province of non-ideal theory. It concludes by suggesting a modest but challenging disposition that can help us when we are torn between opposing oppression and contributing to a flame war.

1. Introduction

We write this chapter from the trenches of the culture wars, with particular attention to the battles that have been waged in professional philosophy in recent years. While those battles have often begun with journal articles, they have tended to play out and indeed to become considerably enlarged by public group letters of protest that go viral on social media. Each of us has signed one such online open letter, and both of us declined to sign a third that we were invited to join. Even though we largely agreed both with its sentiment and with the need under the circumstances for some kind of intervention, we separately decided not to sign that third

[1] Acknowledgements: Barrett is grateful for the helpful feedback he received when he gave talks at Southern Illinois University Edwardsville, Binghamton University, and the University of Houston that included parts of this chapter. He also thanks Audrey Yap who has deeply influenced his thinking on many of these issues. Shannon thanks Jenny Saul for conversations that helped shape her thinking on the matters discussed in this chapter. As well, she gratefully acknowledges funding support from the Social Sciences and Humanities Research Council of Canada. Finally, she co-authored this chapter as a settler scholar and uninvited, but grateful guest on the territory of the nêhiyawak, Anihšināpēk, Dakota, Lakota, and Nakoda peoples, and the homeland of the Métis/Michif Nation.

[2] Both authors contributed equally and each lists themself as first author on their own c.v. For the purposes of this volume, we let ChatGPT decide the author order.

letter because we worried that another social media flame war would just make things worse for the very people the open letter sought to support.

Knowing when and what to say within the public square is often messy and complicated. This is especially true in the age of social media. It is harder than ever to decide when to speak up, when to stay silent, and when to change the subject. When human beings — who are at once deeply social and biased — meet harmful algorithms, social media flame wars can result. However, when we witness false and oppressive speech online, we are often moved by a powerful impulse — perhaps even a felt duty — to respond. How should we show solidarity with the oppressed in the age of social media?

In a 2021 article, Jennifer Saul surveys the arguments for and against the view that there exists a moral obligation to counter oppressive speech and offers a careful analysis of the differences between counterspeech in traditional settings and on social media. On Saul's view, social media is distinct from other contexts due to three factors: a larger audience, an unknown audience, and unpredictable outcomes (in particular, if the counterspeech goes viral). The distinctive features of social media that Saul surveys make it possible that our counterspeech will backfire and work against those goals in a way and to an extent that we just can't predict in advance. Saul therefore concludes that any moral obligation to engage in individual counterspeech on social media is substantially lessened, although she stops short of saying that we ought never to engage in such counterspeech.

In a 2015 article that now seems quite prescient, Kathryn Norlock considers some of the *sui generis* aspects of online shaming and concludes that the moral intuitions we experience in such circumstances are not reliable. In the context of social media, the effects we bring about can wildly exceed our intentions, making it very difficult to assess our moral duty. Norlock counsels “concerted attention to the uncontrollable aspects of the tool we employ, and the effects that, though we may not intend them, we are complicit in inducing in others or entirely responsible for bringing about” (195).

In this chapter, we wish to bring to the fore a view that permeates both Saul’s and Norlock’s articles, even though it remains unstated there — namely, that disagreement, critique, and counterspeech on social media fall squarely into the province of non-ideal theory for reasons both human and technological. In her influential 2012 mapping of varieties of non-ideal theory, Laura Valentini usefully distinguishes between non-ideal theory as partial-compliance (vs. full compliance) theory, non-ideal theory as transitional (vs. end-state) theory, and non-ideal theory as (more or less) realistic (vs. utopian) theory (2012). We are here interested in that third category. Within the realistic vs. utopian dialectic, ideal theory starts with a “view from nowhere” and from there seeks to derive universal principles or conclusions that are then applied to somewheres. Ideal theory can lead us astray when thinking about actual speech in a real world that is far from ideal.

On social media, both the human and technological realities are far from ideal. In what follows, we weave back and forth between focusing on the human, the technological, and their interaction in order to demonstrate the need for a (more or less) realistic philosophical account of expression on social media. We conclude by suggesting not a remedy but a disposition that can help us when we are torn between opposing oppression and contributing to a flame war.

2. Human bias and harmful algorithms

Arguably, one of the biases to which we are most prone in the age of social media is false polarization (FP). FP leads people to overestimate the difference between their perspectives and other people's. Most of us regard our own perspectives as moderate and realistic, and regard views that diverge from ours as extreme and less connected to reality.[3] Moreover, since we regard our own perspectives as realistic, we tend to be quite certain about them, while underestimating the certainty of those with whom we disagree.[4] With such extreme views, we reason, they can't possibly be serious.

FP is older than social media, but the rise of social media has exacerbated it. Social media provides participants with a larger audience than most people ever get face-to-face. Further, the lack of body language in online interactions can disinhibit people such that they post things that they would never say — or wouldn't say nearly as bluntly — in person.[5]

Moreover, harmful algorithms lead to the most provocative posts getting the most views, creating a feedback loop that leads to more and more extreme views being expressed.[6] This in turn provides additional inductive support for the view that other people's views are more extreme than ours, thereby making FP even harder to shake.[7] More insidiously, these harmful algorithms have replaced the epistemic with the statistical. On social media, we are no longer in the realm of rational agents who weigh evidence while reckoning with biases, seeking the truth as best they can. What prevails isn't truths that have passed public scrutiny but whatever garners the most clicks. As Anjana Susarla observes, outrage is a powerful click magnet — "the perfect negative emotion to attract attention and engagement." She continues that while one person tweeting their outrage might not get much uptake, if that "one person is able to attract enough initial engagement, algorithms will extend that individual's reach by promoting it to like-minded

[3] See Robinson, Keltner, Ward and Ross, 1995, and Pronin, Gilovich and Ross, 2004.
[4] See Blatz and Mercier, 2017.
[5] See Lapidot-Leffler and Barak, 2012.
[6] See Munn, 2020.
[7] On the difficulty of debiasing for FP, see Kenyon, 2014.

individuals. A snowball effect occurs, creating a feedback loop that amplifies the outrage" (2020, n.p.).

The intersection of FP with harmful algorithms is at the root of the social media pile-ons that have come to be known as "cancel culture." In fact, "cancel culture" is a misnomer because what it describes is rarely a cancellation and it is not a culture. Multiple commentators have observed that so-called cancellation is too vaguely defined and, however you define it, extremely rare.[8] Frequently, figures who are characterized as having been "canceled" for their views continue to flourish professionally.[9] Less commonly noted, but just as crucial for our purposes here, is the fact that cancel culture isn't a culture at all. When internet pile-ons occur, they are a reflection not of a shared culture, but of discrete strangers who have been sucked in by the algorithm.[10]

It is an irony of our age that the type of pile-ons labeled as cancel culture are exacerbated by this use of the term. As cynical or naïve public figures and media outlets rail against cancel culture, "cancel culture" itself becomes a trending social media topic garnering outrage, clicks and more pile-ons, and making existing pile-ons worse.

It is far from clear whether viral tweetstorms caused by harmful algorithms constitute speech in any robust sense, but they are in any event remote from the idealized (often romanticized) metaphor of the free exchange of ideas in the town square. In this context, well-intentioned open letters and tweets in support of equity-denied groups can produce unintended

[8] Hobbes, 2022 provides an excellent overview of these critiques.

[9] Some even benefit from the so-called cancellation. For instance, country singer Morgan Wallen was temporarily suspended by his record label after a video of him using a racist slur was made public. The controversy led to astonishing commercial success for Wallen. See Caulfield, 2022.

[10] See Mishan, 2020.

effects — turning those who are “canceled” into martyrs and causes célèbres, and thereby directing more harm to the very groups the open letters and tweets were intended to support.

3. Social responsiveness, self-formation and imaginal relationships

When we refer to turning people into martyrs and causes célèbres, you might think that we are only talking about their reputations. But humans are deeply social beings who are responsive to others’ ways of regarding and treating them, such that when we are viewed in a particular way by other people — or when we think that we are viewed in that way — it can change not just our reputations but our selves.

Hilde Lindemann argues that persons are constructed in part via their social interactions (2014, 15). Her account is rich and complicated. For our purposes here, what is important is that there is a multi-stage process we are all constantly enacting in our interactions with others which helps to create our identities. First, we *act*. Next, others *recognize* our actions (either well or badly) and then *respond* to us in a way that we then must respond to (in one way or another). Often, the recognition and response of others is internalized — if girls are told they are naturally bad at math, at least some of them might believe it, not push themselves as hard, underperform, and then actually fulfill the prophecy by growing into the identity others have laid out for them.[11] Not all such prophecies are negative; if you feel the support and respect of your teachers or family when you are growing up you might come to push yourself to try things you otherwise wouldn’t have, and thereby become good at those things. And, of course there is room for resistance to those judgements; someone might become angry about being told that they are naturally bad at math and might study even harder, just to prove others wrong! The point is that

[11] See also Hacking, 1995 on what he calls “looping effects.”

those judgements very often play some role — whether negative or positive — in how we act and who we become.

The examples we have just given have to do with childhood development and the acquisition of various skills. Social responsiveness also plays out with regard to our moral character or political commitments. If you cheat on your partner and everyone in your moral community reaches the judgement that you are a cheater — that you can't be trusted and will always be a philanderer — then it won't be a surprise if you end up living in to that identity. Again, that's not deterministic; there is room for you to exercise agency and thwart the expectations of others. However, doing so might often be hard, and if others don't create space for you to go in a new direction, you might actually end up not having the opportunity to prove yourself trustworthy.

Consider another scenario. Let's say a work colleague tends to use a lot of sexist language. Having worked with them over the years you have come to believe that they genuinely care about being a good person but don't get what all the talk about sexism is about — what's wrong with thinking that men and women are just different and naturally good at some things and not others? How might you intervene in such a case?

A common distinction in activist circles is that between "calling out" and "calling in." If you decided to call your friend in, you might pull them aside one day and invite them to coffee to talk through why they should stop using sexist language. By contrast, if you decided to call them out, you would challenge their sexist views and remarks publicly. While there are often good reasons to make our disapproval public, a public call-out might be especially likely to encourage what in professional wrestling they call "the heel turn," in which someone has a choice about whether to change their wrongful ways or to embrace becoming the villain. Upon being called out publicly, the co-worker might hunker down and embrace being sexist. To be clear, if they do so that doesn’t absolve them of moral responsibility; they shouldn’t be sexist and are wrong for committing to it. But, if your goal as their colleague is to encourage them to change their ways, a public calling out likely isn’t going to get the job done.

While our examples so far have come from in-person contexts, social responsiveness of the kind we’re describing also plays out online. Consider so-called internet “trolls” — people who say offensive things in order to provoke a strong response (often outrage) from others. If the account we’ve been developing so far is true, it seems to support the common advice not to “feed the trolls”: when people counter or express outrage about what trolls say, they not only give them exactly the attention they are seeking, but thereby also help to shore up that identity — and some of the time, to become trolls when they otherwise wouldn’t have. It is hard to come back from thousands of people hating you, and many people at such a crossroads may choose to take the low road because it feels like the only live option. Again, none of this absolves people of moral responsibility; they shouldn’t do trollish things and the fact that doing so often leads to such backlash does not mean that they have no agency or are transformed into trolls by others against their will. Instead, the point is to recognize that, as profoundly social beings, such exchanges are sites where people exercise agency in forming their own and each other’s identities.

On the face of it, the troll example is similar to the examples of the philanderer and sexist co-worker. However, given the epistemic biases we discussed in Section 2, there is compelling evidence that some features of social media significantly exacerbate the effects of social responsiveness. The larger size and scope of the social media audience are among these features, as are the harmful algorithms we discussed in Section 2. Another distinctive feature of social media that can affect social recognition is imaginal relationships with online others.

Norlock argues that when we interact with others on social media, we form what psychologists term “imaginal relationships” with them (2017). In general, imaginal relationships may be with real or imaginary/fictitious persons; what makes them imaginal is not the reality of the relata, but our internal, imaginal development of aspects of the relationships. We have all had the experience of imagining ourselves in some sort of relationship with a character from film or literature. When we do, though, we are usually aware that the person isn’t real and the relationship is all in our heads. With imaginal relationships on social media, however, it’s trickier to know what’s real. On Norlock’s view, as human beings, “we are constituted to mentally conceptualize those with whom we communicate, and… cyberspace presents possibly insurmountable challenges to our capacities to control those conceptualizations. The speed and volume of online affirmation outmatches what the human mind evolved to manage” (2017, 194).

Norlock examines the phenomenon of online shaming and suggests that online pile-ons are motivated less by the desire to shame or punish putative wrongdoers, and more by the pleasure of imaginal relations with other shamers: “In many cases, the intent of shame justice seems to be to enjoy the company one has in cyberspace with so many approving others” (2017, 191). She notes that online “cancellations” often begin with jokes before moving on to a judging phase, with the highest volume of activity taking place during the joking phase. This makes sense. It is often easier to make a joke than to criticize openly. However, a thousand jokes can be much more cutting than a handful of criticisms.

Further, as Saul has noted, those thousand jokes can prompt sympathetic responses that bolster support for their target (2021, 148-49). She points to a psychological study that found that audience members’ view of an online commenter hinges on the number of commenters. Where a single comment about a putative wrongdoer might be viewed favorably, the very same comment is viewed negatively if there is a large group of commenters — a result that might be caused by natural sympathy for apparent victims of bullying.

From imaginal relationships and natural sympathy, heroes and villains are easily born.

4. Becoming heroes and villains

We started this chapter by referring to recent social media flame wars in philosophy and our struggles to determine the best course of action when sparks start to fly. In Sections 2–3, we bracketed the world of philosophy and described some of the more general ways in which human cognition and sociality, distinctive aspects of social media technology, and interactions between humans and social media make the context of online speech far from ideal. In this section, we return to the philosophical dust-ups with which we started in order to elaborate one of the ways in which the non-ideal character of social media interactions in general can feed back in deleterious ways within existing communities, turning community members into heroes and villains.

The reference here to existing communities might seem surprising given our discussion in Section 3 of discrete strangers who have been sucked in by the algorithm. Imaginal relationships are the key here. Consider: a handful of philosophers who don't know each other separately tweet about something. One of those tweets goes viral and people — many of them philosophers — start piling on to like, retweet, and reply. So far, we're still in the territory of discrete strangers and algorithms that we described above. However, as the virality starts to produce natural sympathy in some of the audience, audience members enter into imaginal relationships with the principals. Where a viral tweet attracts audience members and responders from all walks of life, those audience members may not make any further connection with each other. However, when virality occurs among members of a particular group — as in the case of philosophers — audience members and responders can start to get more personally involved. Although the details of such cases will vary significantly in light of contextual details, there are some elements that we think can often be found among them.

When colleagues are called out online for expressing oppressive views, they must make the choice about how to respond.[12] Either they can apologize and recant or double down and endorse further oppressive views. Of course, accepting responsibility can be painful and costly in various ways, but in addition to such costs wrongdoers often have a positive incentive to lean in to the views they express. One need only think of academics who would otherwise be unknown except for their expression of offensive views online or in their scholarship. The status and name-recognition that is born from having been at the center of an internet pile-on can be scary and harmful, but it can also lead to fame and influence — as well as the pleasure of simply being known, not just by those piling on, but by those who have greater status than the wrongdoers in question. Consider a minor academic who is defended by someone with greater name recognition. When that happens, the wrongdoer joins a new cohort from which they otherwise might have been excluded. The allure of flocking together with other, higher status birds can be hard to resist.

Throughout such cases, what we find are important dynamics at play in the back and forth between the initial action, the response from others on social media, and others in the larger community who then get involved. Such cases involve the opportunity for the initial actor to

[12] We are here taking it for granted that expressing oppressive views is wrong at least some of the time. For more on this see Emerick, 2021.

choose whether to recant, and each involves some incentive for the actor to double down and accept the identity that has been attributed to them.

Along the way, as supporter-groups form on either side, they can firm up through what we here term *hero/villain thinking*. Recall the heel turn from Section 3 — the way that someone, upon encountering epistemic friction or challenge from others, might choose to just embrace the label — to embrace being a villain. When flame wars play out within a community, they can similarly encourage embracing the identity of hero. We referred at the outset to the strong pull that we often feel in the face of oppressive speech to *do something*. Part of that pull is no doubt to help those who have been unjustly treated and part (as Norlock has shown) is to join in imaginal community with others. However, we have both experienced, and have witnessed in others, the desire to be one of the good guys — to be the hero in contrast with the villains who have engaged in wrongdoing. Of course, people ought to act rightly and in general it's better to aim to be a good guy than a bad guy. But we want to suggest that hero/villain thinking is to be avoided, regardless of whether the hat you aim to wear is white or black.

At bottom, the problem with hero/villain thinking is that it is ideal thinking in a far-from-ideal world. It gets human beings and the moral worlds in which they participate and shape each other badly wrong. Hero/villain thinking encourages a flattening of the other person, reducing them to either their wrongful or righteous actions, and obscuring the fullness of what it means to be a moral agent.[13] To be clear, we are not arguing that we should let people off the hook for the wrongful actions they commit or the beliefs that they hold. To the contrary, we think it's

[13] Notice the similarities between hero/villain thinking as false polarization. Hero/villain thinking is not always or only produced by false polarization. However, both phenomena obscure the broad swathe of moral and political ground between the extremes that is occupied by most people.

essential that we not use such flattening, reductionist language because doing so in fact undermines our ability to hold someone responsible and to pay attention to the larger social systems within which they act.

Feminist activists and theorists have done lots of work over the years to challenge the trope of the "ideal victim" of sexual assault and to say that many acts of sexual violence don't occur the way they do on crime shows or in movies, but instead are much more commonplace. In her paper, "Credibility Excess and the Social Imaginary in Cases of Sexual Assault," Audrey Yap builds off that work and argues that we should similarly be careful not to buy in to the idea of the ideal perpetrator when thinking about what sexual violence amounts to or who commits it (2017). If we want to be able to hold most perpetrators of sexual violence responsible for what they do, we have to be able to recognize that they committed the wrongful action in the first place. And, that means not thinking of them as monsters — or, we would argue — as only villains.

We want to follow Yap's move and further broaden the feminist critique of the ideal victim construct. We do not wish to "both-sides" the matter; again, it is better to try to be a hero than to try to be a villain. However, just as there is no ideal victim or ideal villain, there is no ideal ally. As social media flame wars have burned in philosophy, we have seen the emergence not only of putative villains, but also seeming heroes, who latch on to a particular fight and come to be identified with it. This can calcify them in black-or-white positions that efface complexity and nuance and can predispose them to certain kinds of responses (again, we think that some open letters might be among them) that risk making things worse. And, as we hope we have by now made clear, it's bad moral metaphysics.

Part of what it means to be an agent who is not programmed for one type of behavior or another is that people are complicated. Most people are capable of both cruelty and kindness, both violence and care. But, that’s a really uncomfortable thing to recognize, for at least two reasons. The first is that doing so means acknowledging that we are vulnerable. If we only have to be afraid of villains, if we only have to guard against monsters, then we can relax when we’re around people that don’t look like them or are in places where they don’t live. If the fact of agency means that lots of people are capable of cruelty and violence then perhaps the world comes to feel less secure, less safe, than it otherwise would. It also means acknowledging that we are fallible — it means acknowledging the ways that we ourselves are capable of serious wrongdoing — and recognizing that fact can be really existentially challenging. It can be hard to face that possibility in ourselves because of what doing so would force us to give up. Nevertheless, it is precisely this challenging work that we recommend in the non-ideal context of social media.

5. Only human

We began this chapter by saying that we are skeptical of well-intentioned open letters and tweets in support of equity-denied groups and are worried about the way that such efforts can have the opposite effect than what they intend. In sections 2–4, we elaborated our reasons for wanting a different kind of response to misleading and oppressive views. In short, we wish to avoid the following types of outcomes:

1. contributing to hero/villain formation;
2. drawing attention to the hero/villain and extending the issue’s shelf-life longer than it would otherwise have had if it had not secured any uptake from the larger community;

3. diverting attention away from those who need and deserve it;

4. directing additional harm to those with whom we wished to show solidarity;

5. elevating the issue in the larger public consciousness in a way that makes it a contender for a battle in the "culture wars", thereby further exacerbating 1–4.

That last point is especially worrisome because by elevating the issue in the public consciousness and making it a larger social issue whole groups of people might be harmed as a result. The question with which we conclude, then, is how to respond to oppressive views without adding fuel to the fire and leading to outcomes 1–5. Put simply: what is to be done in the far-from-ideal age of social media? Predictably, there is no ideal approach. But for that reason, and in light of all we have said above, we wish to counsel not a remedy but a disposition. We might pithily express it as such: be human.

When we say "be human," we are not counseling complacency or quietism. We are definitely not saying "it's all good, man." What we mean, rather, is that when we find ourselves feeling the call to respond to misleading and oppressive speech, we should consciously resist the pull of both social media algorithms and hero/villain thinking. These modes are, respectively, too big and too small for our moral agency. Social media algorithms are driven by statistical force in numbers that far exceed our capacity for moral reasoning or moral responsibility. They are too big for us. Hero/villain thinking is storybook thinking that reduces away the richness and complexity of human moral existence. It is too small for us.

When we see that someone is wrong on the internet, we need to remind ourselves that we belong in the "just-right" Goldilocks space in between those two extremes. That is a space for fallible beings navigating an unpredictable, confusing, non-ideal reality in a state of partial information, and in so doing forming ourselves, others, and our shared moral worlds. We hope that reminding ourselves of that will help us to be less reactive and more responsive.

This does not mean abandoning solidarity with equity-denied and oppressed groups. Rather, it means pausing to consider what form of solidarity in the particular context would actually help (or at least wouldn't harm) the folks with whom we wish to ally, and by what values our intended response should be animated. There is no shame in feeling attracted to the pull of the hero role; we are only human after all. But we need to balance that pull against more careful moral deliberation, and a recentering on the others we wish to support, the values we wish to enact, and the likely outcomes of our response. Where we cannot identify those likely outcomes, we should ensure that our deliberations are tempered by fallibilism.

The responses we might consider can take a number of forms. There is no rule that the modality needs to match the offense. That is, just because the offense occurred on social media does not mean that the response must also occur there. Some offline responses that we have seen or pursued include fundraising events for equity-denied groups, joyful community gatherings with music and celebration of equity-denied groups, or speaker or author invites for members of groups that have been targets of oppressive misinformation. We are, as humans, better judges of the possible effects of those sorts of activities than we are of social media phenomena.

This is not to argue that people should not be active on social media. As is often correctly pointed out, social media activity and relationships are especially important for people who are geographically isolated, disabled, poor, care-givers, or for various other reasons, unable to fully avail themselves of sufficient face-to-face social interactions. Increasingly, access to social media is a necessity for human flourishing. Further, some appropriate responses to misleading and oppressive social media speech are social media responses. The “be human” disposition is constitutively not a one-size-fits-all approach.

Finally, we note that social media is changing rapidly. Over the course of writing this chapter we saw Twitter become X and Facebook become Meta. As X has been (further) vitiated by Elon Musk’s changes to it, some people are exploring the possibilities of Mastodon’s non-profit, federated, cooperative model. Others are trying Threads or seeking invites to Blue Sky. It is too soon to tell what shape social media will take in the coming months and years. While climate collapse and rising fascism make it a tough time to be optimistic, we hope that the new social media platforms open new possibilities. On whatever platforms we participate, we will work hard to be human, to remain human, and to treat others as humans.

Reference list

Blatz, C.W. and Mercier, B. (2017). ‘False Polarization and False Moderation: Political Opponents Overestimate the Extremity of Each Other’s Ideologies but Underestimate Each Other’s Certainty’, *Social Psychological and Personality Science* 9(5), pp. 521–529.

Caulfield, K. (2022) ‘Morgan Wallen’s “Dangerous” Is Close to Breaking Longest Top 10 Record on Billboard 200’, *Billboard*, 26 August. https://www.billboard.com/pro/morgan-wallen-dangerous-top-10-record-billboard-200/.

Emerick, B. (2021) ‘The Limits of the Rights to Free Thought and Expression’, *Kennedy Institute of Ethics Journal* 31(2), pp. 133-152.

Hacking, I. (1995) ‘The Looping Effects of Human Kinds’, in Sperber, D., Premack, D. and Premack, A.J. (eds.) *Causal Cognition: A Multi-Disciplinary Approach.* Oxford:

Clarendon Press, pp. 351–82.

Hobbes, M. (2022) 'Is Cancel Culture Really a Threat to America?' YouTube, 27 February. https://www.youtube.com/watch?v=RkVYvp_CumI.

Kenyon, T. (2014) 'False polarization: debiasing as applied social epistemology', *Synthese* 191, pp. 2529–2547

Lapidot-Leffler, N. and Barak, A. (2012) 'Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition', *Computers in Human Behavior* 28(2), pp. 434–443.

Lindemann, H. (2014) *Holding and Letting Go: The Social Practice of Personal Identities*. New York: Oxford University Press.

Mishan, L. (2020) 'The Long and Tortured History of Cancel Culture', *The New York Times Style Magazine*, 3 December. https://www.nytimes.com/2020/12/03/t-magazine/cancel-culture-history.html.

Munn, L. (2020) 'Toxic by design: toxic communication and technical architectures', *Humanities and Social Sciences Communicatio*ns 7(53), n.p. https://doi.org/10.1057/s41599-020-00550-7.

Norlock, K. 'Online Shaming', *Social Philosophy Today* 33, pp. 187‒197.

Pronin, E., Gilovich, T. and Ross, L. (2004) 'Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others', *Psychological Review* 111, pp. 781–799.

Robinson, R. J., Keltner, D., Ward, A. and Ross, L. (1995) 'Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict', *Journal of Personality and Social Psychology* 68, pp. 404–417.

Saul, J. (2021) ‘Someone is Wrong on the Internet: Is There an Obligation to Correct False and Oppressive Speech on Social Media?’, in MacKenzie, A., Rose, J. and Bhatt, I. (eds.) *The Epistemology of Deceit in a Postdigital Era: Dupery By Design*. Cham, Springer, pp. 139–158.

Susarla, A. (2020) ‘Hate cancel culture? Blame algorithms’, *The Conversation*, 28 January. https://theconversation.com/hate-cancel-culture-blame-algorithms-129402.

Valentini, L. (2012) ‘Ideal vs. Non-ideal Theory: A Conceptual Map’, *Philosophy Compass* 7(9), pp. 654–664.

Yap, A. (2017) ‘Credibility Excess and the Social Imaginary in Cases of Sexual Assault’, *Feminist Philosophy Quarterly* 3(4), n.p. https://doi.org/10.5206/fpq/2017.4.1.