

The Interpersonal is Political: Unfriending to Promote Civic Discourse on Social Media

Alexis Elder

(This is a pre-proof draft. Please cite the published version, in *Ethics & Information Technology*: <https://doi.org/10.1007/s10676-019-09511-4>)

Abstract

When it comes to social media's impact on civic discourse, opinions are divided. Some think social media is divisive because it creates echo chambers in which the politically like-minded reinforce each other's beliefs while encountering only caricatures of alternative viewpoints. Others claim that social media, at least when appropriately constructed, has the potential to enhance civic discourse by enabling conversations among very different individuals. What both sides overlook is that civic conversations on social media occur in the context of interpersonal relationships. When we look at the role of interpersonal difference in relationships, and the impact of disagreement on individuals' willingness to engage, the importance of both selective affiliation and exposure to alternative viewpoints can be seen as complementary rather than conflicting concerns. I develop an account of healthy disagreement in interpersonal relationships, then supplement it with an account of how social media technologies of disconnection can mediate relationships so as to support this healthy disagreement.

Introduction

Social media technologies connect us, and this fact has been a major focus of both design decisions and ethical appraisals. Some accounts of social media presume that this would be good for public discourse, by permitting widespread, lively, bottom-up discussion of matters of civic interest. But this optimism has not always been borne out.

In her book *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Shannon Vallor notes that the relationship between social media and civic engagement has turned out to be much more complicated than was initially thought. She makes the case that

we cannot count on even well-meaning technocrats to build technologies that will reliably shape human responses so as to bring about virtuous behavior. We will also need virtuous characters in order to engage well with these technologies.

I am sympathetic to this analysis, but think that it misses an important element by focusing primarily on individuals and societies. Doing so misses an influential intermediate kind of social entity: the interpersonal relationship. Friendships, families, and loose associations of friendly acquaintances stand somewhere between ourselves as individuals, and enormous, diffuse associations such as nations, where the addition or loss of an individual goes unnoticed. People share (or do not share) political opinions on social media within a context of interpersonal relationships, and understanding this can help us to see what's at issue for users, as well as discover a surprising source of resistance, something that in fact has been criticized for contributing to social discord: technological disconnection options.

My goal, then, in this paper, is to start with an account of healthy interpersonal engagement, one that encourages both sharing of a variety of opinions and that supports open-minded, respectful consideration of differing views. I reach the surprising conclusion that individual users' ability to easily disconnect, although often considered part of the problem, may in fact contain the seeds of a solution.

Examples of Disconnection

To start with, it is important to recognize the wide variety of disconnection options available, in order to better evaluate their impact. I begin with a survey of varieties of technological disconnection.

Facebook allows users to “unfriend”, which severs a mutual connection; once one party selects the “unfriend” option, neither appears in the other’s news feed, any private content is no longer visible, and neither appears in the other’s friend list.

Twitter and Instagram, by contrast, have no such reciprocal severing, but allow users to “unfollow”: Alice can choose to “unfollow” Bob and will no longer see what he posts, but if Bob is following Alice’s content, his access to her content remains. Twitter and Instagram users can also “block” users, preventing them from initiating a connection (if they are not already followers) and preventing them from continuing to follow the user’s content.

Neither option actively notifies the other user(s), but the disconnection can be inferred in various ways: by looking to see if someone is still on your list of Facebook friends or Instagram followers, or in Twitter’s case by trying to re-follow someone, which generates a notice that you have been blocked.

Less visibly, there are options on Twitter and Facebook that allow users to stop seeing updates in their news feed, by “unfollowing” on Facebook or “muting” on Twitter, that give no indication to the other party that their messages have lost an audience member.

In the related field of communication technologies, users’ email addresses, phone numbers, and SMS or other text messages can be blocked so they do not reach their intended recipients, and features like email filters can be used to send incoming emails directly to a dedicated folder and bypass the inbox.

Users can be banned from using a platform entirely, whether by a moderator (on many message boards or in comment sections on websites) or social media network, like Twitter. In such cases, whether permanently or temporarily, the user can no longer access the network at all, to view, comment, send messages, or interact with other users. (For platforms that do not require

a user to sign in, content may still be visible to the banned user.) This is commonly but not always justified by appealing to user behavior that violates the platform's terms of service, as when someone posts a commercial link in a noncommercial discussion, or shares pornography or threatening content on many social media platforms.

Individual user content can also be removed without the author's involvement: for example, Twitter might remove an offending tweet, or Facebook might delete a user's post or comment. In some instances, users can also delete others' contributions to space they control: for example, I might be able to delete a comment on my post, or content that is shared to my wall. Some platforms allow for pre-screening of content, especially from new users, sending contributions to a moderation queue to be approved before posting.

On the other side of the coin, users sometimes leave platforms and delete their profiles, or set them to an inactive status, or simply cease to use them, which can be a trigger for their account to go dormant, if the platform is designed to account for this. And in an option requiring even less platform intervention, users may ignore messages or requests for connection in many platforms, even sometimes engaging in a practice known as "ghosting", when someone simply stops responding to someone's messages and attempts at interaction, suddenly or gradually.

Ethical Pros and Cons

These different forms of disconnection come with a variety of pros and cons. On the beneficial side, they can be important tools to combat abuse and harassment online, as well as the distribution of problematic content, whether copyright violations, privacy-violating images or information, or threatening or harmful material.

They can be used to maintain a particular focus or conversational tone, as when off-topic posts are removed from a discussion forum on a specific issue, or a support community for survivors of abuse blocks users whose behavior runs contrary to the mission of supporting victims.

They can give individuals greater control over their social connections, including how much, how often, or under what circumstances they connect with others. And this greater control, in turn, can help people to protect boundaries of various kinds. People can use these tools to protect the boundary between work and home life by preventing co-workers from seeing photos of family and friends, or evidence of participation in unusual or controversial hobbies, or sexual orientation, gender identity, and so forth. They can also establish boundaries in particular interpersonal relationships, as for example when grown children prohibit parents from viewing parts of their lives they do not wish to share.

At the same time, a variety of ethical concerns have been raised. In corporate contexts, information silos, where subgroups are incapable of interacting with other units of the organization, are important obstacles to organizational functioning. In informal contexts, as noted above, concern has been directed at echo chambers, epistemic bubbles, and filter bubbles. These various forms of social self-isolation, enabled or enhanced by technology, have been linked to a variety of social and political ills.

On the personal level, concerns have been raised about the relative ease with which we can disconnect from others online, both literally, as with the click of a button, and emotionally, as when we can ignore even a direct communication without either social pressure or the emotional cost of doing so under the gaze of another: we often have a plausible deniability shield we can

appeal to, that we've been busy or haven't had time to get to a message (Vallor 2010, 2012; Turkle 2015, 2011).

This last is an example of more general concerns about the deceptiveness that is enabled by the relative lack of transparency in digitally mediated interactions. If I can't see whether you've seen the text message, I can't hold you accountable for failing to respond. As communication technologies have evolved, the prevalence of read receipts, send times, and other means of making message uptake and exchange more visible have proliferated, perhaps reflecting a widespread desire for greater transparency. If this is correct, then the relative invisibility (to others) of many forms of disconnection warrants closer examination.

How should we assess the value of disconnection?

The range of disconnection options is quite diverse. It seems silly to think one should accept a simplistic, uniform judgment on all instances of the phenomenon. One might consistently adopt different conclusions about different cases: perhaps ghosting on an acquaintance is cowardly, but blocking abusive Twitter users is acceptable. Nevertheless, it would be helpful to have a background framework against which to assess particular cases.

One starting place would be to begin with the thought that relationships and personal connections are valuable, and that technological affordances of disconnection are, all else being equal, ethically problematic (Vallor 2012; Turkle 2011). After all, personal relationships require patience, steadfastness, and loyalty, even when an interaction is not immediately rewarding. Technologies that make disconnection easier and less visible – thereby sparing the person the social penalty that might be garnered by an evasion – can pose a moral hazard, a temptation to

behave badly. And if acted on often enough, this sort of evasion can become habitual, gradually degrading individual character and interpersonal relationships.

On the other hand, one might think that (technological) disconnection is, in general, a good thing. Such a premise could be motivated by the thought that today's hyperconnected world is overwhelming for the psychologies of creatures like us. Technologies that enable disconnection would be valuable insofar as they help us carve out room to cultivate ourselves in an environment with reduced distractions. They might also encourage us to replace technologically-mediated interactions with face-to-face ones. This viewpoint is common to the growing array of "quit lit" – testimonials by people who have left various social media platforms behind (Fierberg 2018)- and the rising popularity of "social media cleanses" (Chang 2017).

Then again, one might think that, on its own, tools of disconnection are neither good nor bad. What is valuable is the autonomy enabled by them. Insofar as users are empowered to choose for themselves which connections to nurture and sustain, and when to cut ties, the technologies are therefore valuable. Something like this view is supported by danah boyd's work on use of technology to support autonomy among teenagers (boyd 2014).

Each of these is plausible. And yet each seems to involve important limitations.

For example, relationships involving harassment and abuse seem to be those from which it is not merely permissible, but actually good, to disrupt through disconnection. To the extent that these technologies can help people to extract themselves from unhealthy relationships, by lowering barriers to exit in precisely the ways noted earlier as worrisome, they seem valuable. This is especially important for those who are socialized to assume the bulk of emotional labor in relationships, or whose relatively low social status attaches heavy penalties to avoiding even intrusive people of higher status. (Elder 2018)

At the same time, the pro-disconnection position seems to overlook many real benefits of technologically mediated connection, and to idealize a pre-technological social era in ways that are both unrealistic, given the prevalence of today's communication technologies, and overlook the challenges of social interaction absent today's technologies. For the remainder of this paper, therefore, I will set aside such presumptions, and focus on responsible and appropriate use of disconnection without assuming that the ultimate goal is or should be a digital-connection-free life.

And while it might be good for people to have the freedom to choose whether to connect or disconnect, noting the value of autonomy seems inadequate for a full accounting of the ethics of disconnection. Surely some of our choices to connect or disconnect can be better or worse. Someone's choice to disconnect from a beloved friend over a minor disagreement seems petty, while the choice to maintain a connection that facilitates abuse (of oneself or another) seems ethically questionable at best. Furthermore, such choices, whether or not they work out for the best, seem as though they can be based on better or worse reasons. And as the ongoing discussion over free will demonstrates, even uncoerced choices can be shaped and influenced by a variety of features of context, which are themselves often subject to ethical evaluation.

Psychological and sociological factors affecting social media use

Vallor notes the failure of early techno-optimists to predict just how silencing it would turn out to be to civic discussion when we connect millions or, now, billions of people together in social media networks (2016). Even though our physical capacity to share information has been enhanced, in many ways this has been psychologically discouraging to healthy civic discourse, in ways that turn out to be reinforced by flat-footed technocratic attempts to overcome it. In

particular, she points to a Pew Institute study on people's willingness to discuss politics on social media. Here, researchers found a precipitous drop-off (from over 80% to just over 40%) in people's willingness to discuss political issues offline versus on-. The problem is not just that people are shouting loudly at each other, or circulating inaccurate information online, but that many people, seemingly in response to these extremes, are simply opting out of the conversation altogether, creating a chilling effect on civic discourse. She suggests we need to cultivate virtuous characters in order to use social media wisely and well, and that sometimes this will involve not just working with value-neutral technology in positive ways, but actively resisting temptations presented by these technologies. (Vallor 2016)

Diagnosing the appropriate temptations to resist, however, turns out to be more difficult than one might think. Attempts to explain failures of political discourse on social media have often focused on epistemic bubbles, especially so-called filter bubbles. These arise from the fact that social media platforms not only share user content, but permit selection or actively do so selectively (for example, via algorithms that use browsing and "likes" history to prioritize content likely to engage a particular user's interest), and result in different users being exposed to very different content. So a liberal user would encounter more liberal-leaning political content than a conservative one, and visa-versa. This would give different social media users different information to reason with, and would result in radically different belief systems (hence "epistemic bubbles"). These epistemic bubbles can be attributed either to the decisions of individual users to block content from those they disagree with, and/or to automatic content-filtering mechanisms (hence "filter bubbles").

But data about users does not support this somewhat simplistic explanation, which focused largely on technical features of networks. C. Thi Nguyen points out, "It is possible to

pop an epistemic bubble by exposing a member [of the bubble] to relevant information or arguments that they have missed” (Nguyen 2018, 5). On the other hand, what Nguyen characterizes as “echo chambers”, in which people of similar views reinforce each other's interpretations of opposing evidence in ways that reinforce their own initial beliefs (for example, by taking a news report that conflicts with their view to be evidence of a “big media” conspiracy) are much more stable and difficult to change.

More troublingly, both balkanization and reluctance to engage at all seem to persist even when a variety of political views are on display – and often in ways that suggest that the display of other views is itself part of the problem. For example, researchers found that when Twitter users were deliberately and voluntarily exposed to content expressing opposing views, this tended to reinforce polarization (interestingly, more so for Republican than Democratic users, but measurable for both). (Bail et al. 2018)

Meanwhile, Vallor points out that a Pew study on user reluctance to share political content online postulated “heightened self-censorship might be tied to social media users' greater awareness of the opinions of others in their network”, an explanation consistent with a phenomenon known as a spiral of silence effect, to be discussed later in this paper. If one important goal of civic discourse is to facilitate sharing of information and ideas that might help people to share ideas, attention needs to be paid to factors that seem to increase people's willingness to hear evidence with an open mind, rather than using it to double down on pre-existing beliefs, and this is a substantially harder problem.

One might wonder whether changing minds via discussion on social media is a plausible goal at all – perhaps this is simply not something social media is good for. A Pew Institute study on this topic indicated that 14% of social media users report having changed an opinion about a

social or political issue because of content shared on social media. This sounds relatively bleak, although it does at least offer the possibility of change. However, the rate for different users varies by a number of factors, being for example double that rate among young, male social media users and a third that rate among social media users over 65. Race and ethnicity also turned out to have significant impact: only 11% of white respondents said they had changed their minds in this context, while 19% of black users did so, and 22% of Hispanic users. This suggests that social factors, including perhaps the character of one's social network, may significantly influence how one uses these media to discuss issues of civic importance. (Bialik, n.d.)

In the same vein, despite worries about the abilities of bots to spread disinformation on Twitter and Facebook, research indicates that false news spreads fastest among human users – and amongst humans *only*, it spreads faster than true news stories. “Contrary to conventional wisdom,” report the researchers, “robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.” But not all humans at the same rates: users with fewer connections and lower activity levels, and who had been on social media platforms for significantly less time than “power users”, turned out to be significantly *more* likely to circulate misinformation, suggesting again that problems with civic engagement are fundamentally connected to interpersonal relationships, including sometimes ill-founded trust in the fact-checking abilities of one's immediate social circle, not mechanical or technical features of information-sharing or connection considered in isolation. (Vosoughi, Roy, and Aral 2018)

Interpersonal relationships can also themselves be affected by the type of content shared. Rising use of social media to discuss social and ethically loaded issues has led many people to consider disconnecting from someone. A recent study indicates that “39% of social media users

have taken steps to block another user or minimize the content they see from them because of something related to politics”. (Duggan and Smith 2016) Although this might seem to be a symptom of the vice of cowardice, and possibly even a contributor to the ongoing balkanization of online discourse, I think these explanations miss valuable reasons to disconnect, reasons that ultimately support healthier and more varied public discourse.

From what should one disconnect?

There seems to be good reason both for the claim that we ought to be free to choose whether and when to disconnect, and for the claim that doing so will tend to shield us from difference, even when these differences are valuable or worth respecting. The ethical advantages of disconnection emphasize the good reasons we have to disconnect from people and communities. But the ethical disadvantages involve the perils of acting on bad reasons for disconnecting. And, crucially, the most hazardous are not those based on a desire for isolation, but rather, the temptations of social homogeneity and, more generally, cowardice.

This suggests the need for an account of the value of social difference, guidance about what makes a social connection worth maintaining, and an analysis of how individual character can influence actions, and be shaped in turn by those actions. For these reasons, I opt for an Aristotelian analysis of the ethics of disconnection. It requires updating his original theory to accommodate modern technologies and concerns. But the core resources of a rich, detailed analysis of friendship and the value of social connections that includes everything from civic friendship to intimate companions, and a complex account of individual character and its relation to the goods of social connection, are well-suited to the task.

Aristotle's account is often subjected to a line of criticism that might, if true, prove a significant impediment to the use of Aristotelian ethics to better understand ethical issues associated with disconnection. This is the thought that, on an Aristotelian analysis of ideals of friendship, the best people are highly similar, especially in values, which would end up supporting the use of disconnection to create things like echo chambers, and more generally, promote homogeneity. But as it turns out, closer examination of his view reveals resources useful precisely for identifying value in difference without inviting abuse.

The seemingly problematic interpretation comes from two of his main claims about the ideal form of friendship: that it occurs only when both friends are *virtuous* (for this reason, he refers to these friendships as friendships of virtue), and that the best friends are “other selves” to each other. This leads some to believe that Aristotle thinks friendships are better when friends are highly similar. The similarity interpretation is a relatively common one amongst those critical (to some degree) of Aristotle's theory of friendship. It can be found, for example, in Dean Cocking and Jeanette Kennett's description of his theory as a “mirrors” account of friendship, where friends are ideally mirrors of each other (Cocking and Kennett 1998), and in Bernard Williams' comment in “Persons, Character and Morality” that

“Once one agrees that a three-dimensional mirror would not represent the ideal of friendship, one can begin to see both how some degree of difference can play an essential role, and, also, how a commitment or involvement with a particular other person might be one of the kinds of project which figured basically in a man's life... something which would be mysterious or even sinister in an Aristotelian account.” (Williams 1981, 15–16) But I think this is not the best interpretation of Aristotle's account of virtue friendship.

Instead, a closer look at both what he says about individual character and virtue, and what he says about difference in virtue friendship, allow and in fact support the idea that our social lives can be enriched by difference.

In discussing heterosexual romantic relationships, he has this to say:

...human beings live together not only for the sake of reproduction but also for the various purposes of life; for from the start the functions are divided, and those of man and woman are different; so they help each other by throwing their peculiar gifts into the common stock. It is for these reasons that both utility and pleasure seem to be found in this kind of friendship. But this friendship may be based also on virtue, if the parties are good; for each has its own virtue and they will delight in the fact. (Aristotle 1999, 1162a20)

While much of this is both biologically essentialist and heteronormative in ways that may strike modern readers as problematic, the passage contains a number of structural features that can be reconstructed without importing the difficulties of his example. Aristotle's account of virtues make several things interesting about this passage. He starts with the idea of "living together". Throughout his discussion of friendship, he says reportedly that friendship consists of "living together" but not "grazing in the same field, like cattle". Instead, they do so by sharing "conversation and thought" – the distinctively human activities in which we perform our peculiar species function (Aristotle 1999, 1170b10-25). Lest one still worry that this makes the account overly biologically-essentialist, he returns to the idea in terms of sharing what one thinks makes one's life worthwhile later in the *Nicomachean Ethics*: "...whatever each group of people loves most in life, in that activity they spend their days together. For since they wish to live together with their friends, they follow and share in those pursuits which, they think, constitute their life together." (Aristotle 1999, 1172a1-7).

For Aristotle, both a good life and a virtuous character are closely connected to the idea of function, as in his famous *Ergon* Argument. What it is to be *eudaimon*, happy or flourishing, is for a person to function well as the kind of thing one is, and functioning well requires the features that help one to perform that function: just as a knife's sharpness helps it to function well as a knife, so human excellences like courage help us to function well as the sort of creatures we are.

This idea of function is central to his account of human goodness: a thing's function is both necessary for its identity (a "knife" that failed to cut at all would simply not be a real knife) and what it means for something to be good is just for it to function as the kind of thing it is. This is sometimes read as requiring that all human beings perform exactly the same function, in virtue of their all being the same kind of thing. But this passage invokes a more complex account of human function: one where different people can have different functions under the broader umbrella of humanity, and that these different functions can complement each other in making up complex human social lives. Each of us, then, can function differently and still count as human, and be excellent in our own ways without resembling each other in every respect: "each has its own virtue" in which others can delight, *and* these differences can enrich the lives we lead together, "throwing their peculiar gifts into the common stock". One can have friends with different perspectives on social and civic issues and see this as contributing to life's rich tapestry, just so long as one thinks these people are good in their own way.

This idea that human beings can excel in different ways is apparent also in his discussion of the nature of human virtue, as well as modern virtue-theoretic accounts of so-called "moral style" (Axinn 1990) – individual expressions of morality that can vary from person to person, just as one's personal style can be unique to the individual while recognizing a general spectrum of good to bad style. This comes up in his discussion of what he means when he characterizes (most) virtues as means between extremes, as the sweet spot between too much and too little of some human trait. Courage, he says, is the right amount of fear, neither too much fear (which would be cowardice) nor too little (which would be rashness) (Aristotle 1999). But we should not expect the same quantity of fear to constitute virtue for every person: rather, what counts as the sweet spot varies by individual. We can gain some insight from recognizing widely shared

extremes it is good for all of us to avoid, but we should not thereby conclude that the same target is appropriate for each of us: "...if ten pounds [of food] are too much for a particular person to eat and two too little, it does not follow that the [athletic] trainer will order six pounds; for this also is perhaps too much for the person who is to take it, or too little – too little for Milo [the professional wrestler], too much for the beginner in athletic exercises." (Aristotle 1999, 1106b)

Thus, it is consistent with Aristotle's account to hold both that good character is important for good friendship, and to resist the idea that good friends ought to resemble each other.

Instead, he describes a different relationship between character and friendship.

"Friends", he says,

...do and share in those things which give them the sense of living together. Thus the friendship of bad men turns out an evil thing (for because of their instability they unite in bad pursuits, and besides they become evil by becoming like each other), while the friendship of good men is good, being augmented by their companionship... (Aristotle 1999, 1172a10)

In order to live well, we need to function well, and failure of character impacts both the quality of one's own life and that of the shared life amongst friends. At the same time, friends remain susceptible to each other's influence, and people who associate with those of poor character are likely to both pick up on each other's bad habits and vices, and support each other in pursuit of goals that undermine their ability to live well.

We end up, then, with an Aristotelian account of friendship as one that advises us to avoid bad influence and strive to become better people in order to enjoy good lives and good friendships. We can do so, in part, by keeping in mind at least three considerations: (1) that some people are more likely than others to tempt us to act badly, and we have reason to avoid those who are especially likely to do so; (2) that we ought to avoid those people and interactions that

bring out the worst in us, in order to help us avoid vice and become better people, and (3) we have reason to seek out interactions with those who bring out the best in us, and are themselves good, even if not in all the same ways as us.

It looks like Aristotle would advise us, then, to unfriend, not on the basis of disagreement over politics, but whether these disagreements contribute to good character and healthy interactions, or the reverse. And this is an important difference.

How can disconnection technologies help achieve healthy diversity in social networks?

In order to put this Aristotelian framework to use, it is helpful to think about the technologies themselves, and the role they play in mediating human social experiences online. I do so using a framework developed by Asle Kiran (Kiran 2015), according to which technologies can mediate human experience along at least four dimensions: magnification/reduction, revealing/concealing, enabling/constraining, and involving/alienating. Note that these contrasting pairs are not intended to be mutually exclusive: rather, the idea is that each dimension of mediation includes both elements. Putting one's eye to a telescope, for example, reveals to the viewer details on the surface of the Moon that are not visible to the naked eye, but conceals the landscape immediately adjacent to the viewer.

For my purposes, it makes sense to begin with the last dimension first: involving/alienating. While disconnection technologies and uses of technology to disconnect are clearly alienating, the dual nature of this dimension becomes clearer when we consider phenomena such as the spiral of silence (Noelle-Neumann 1974; Hampton et al. 2014). This is a social and psychological process whereby a vocal minority can come to dominate a public

conversation. Roughly, a minority group in a larger conversation can, in virtue of being more vocal than other participants, come to give others the impression that their view is substantially more popular than it is. Many people are reluctant to voice their own views when they perceive these views to be in the minority, and so other contributors to the conversation become less willing to share opposing views, which amplifies the perceived dominance of the minority's viewpoint still further, which in turn makes others less likely to voice their own dissenting views, which leads the vocal minority view holders to increase the volume of their contributions on the assumption that they are voicing the majority view, feeding back into the process indefinitely.

This process where the amplification of some views leads to the reduction of others does not require technologies to exist, but can be supported by them. These technologies, then, mediate human interactions in ways that are both involving (to some) and alienating (to others). But it would seem that this can cut both ways.

Suppose, for example, that I am a member of the increasingly-silent majority on an issue of social importance, and am feeling alienated by the dominant discourse. I wish to become more socially engaged, but find it increasingly difficult to do so. One thing I might do is to exercise “block”, “mute”, or “unfollow” features of my social media platform, especially to remove contributions from especially antagonistic, inconsiderate, or hurtful social connections that increase my alienation from the conversation. Doing so could help me to become more active about speaking up, as well as doing so in ways that are less antagonistic and hurtful, and more respectful, to others, thus helping to transform the character of the conversation.

Given the polarizing nature of much online discourse, and the disaffection so many report as a result, disconnecting technologies could have an important role to play in counter-acting

these tendencies, and producing more involvement in social discourse about controversial political, social, and cultural issues.

Talking through this example can also bring out how the other dimensions Kiran identifies can also be involved in disconnection technologies. Blocking someone on Facebook obviously reduces the presence of content like that shared by them in your news feed, which amplifies the proportion of other contributions by contrast, and it is this feature that makes it a potential tool for combatting the spiral of silence.

But this is not just about magnifying, for example, the proportion of politically conservative versus politically liberal content in one's news. Disconnection technologies do not just reduce, they conceal. When someone is blocked on a social media platform, their contributions are no longer visible at all. And doing so, in turn, can reveal other voices that would not otherwise exist. This can occur not just in the context of interpersonal blocks and bans, but when platforms prohibit particular users or behaviors: doing so can empower others to contribute when they otherwise would not, revealing new parties to a conversation.

Furthermore, the disappearance of (especially) threatening contributions or contributors can change not just others' willingness to participate, but to listen. In a phenomenon known as the backfire effect (Nyhan and Reifler 2010), it has been documented that feeling threatened can make one less open to new possibilities, while conversely, feeling safe can make one more willing to consider alternative possibilities. Although the reliability of this effect has been challenged (Wood and Porter 2018), the mere possibility that someone might undergo such an effect seems an ethically significant factor, especially when supported by one's own reflections on one's tendencies.

A piece of testimony from philosopher Mary Midgley provides a pertinent example of the different kinds of conversations that seem to occur when and because some contributors are removed. After centuries of male-dominated academic philosophy, the mid 20th century saw the rise of several prominent women philosophers, including Mary Midgley, Philippa Foot, Elizabeth Anscombe, Mary Warnock, and Iris Murdoch. These women all attended Oxford University together, which had been accepting women since 1920. But when these women were attending, World War II had broken out and the vast majority of their male peers were drafted into the military. According to Midgley, their academic success was due at least in part to the absence of their male peers. In a letter to the *Guardian*, Midgley wrote:

As a survivor from the wartime group, I can only say: sorry, but the reason was indeed that there were fewer men about then. The trouble is not, of course, men as such – men have done good enough philosophy in the past. What is wrong is a particular style of philosophising that results from encouraging a lot of clever young men to compete in winning arguments. These people then quickly build up a set of games out of simple oppositions and elaborate them until, in the end, nobody else can see what they are talking about. All this can go on until somebody from outside the circle finally explodes it by moving the conversation on to a quite different topic, after which the games are forgotten. Hobbes did this in the 1640s. Moore and Russell did it in the 1890s. And actually I think the time is about ripe for somebody to do it today. By contrast, in those wartime classes – which were small – men (conscientious objectors etc) were present as well as women, but they weren't keen on arguing.

It was clear that we were all more interested in understanding this deeply puzzling world than in putting each other down. That was how Elizabeth Anscombe, Philippa Foot, Iris Murdoch, Mary Warnock and I, in our various ways, all came to think out alternatives to the brash, unreal style of philosophising – based essentially on logical positivism – that was current at the time. And these were the ideas that we later expressed in our own writings. (Midgley 2013)

In other words, supposing Midgley's surmise is accurate, it was not merely the case that each of these women independently was philosophically talented (although surely this is also true). Nor was it sufficient that they all be classmates at Oxford – other talented women have been there before and since, and have thus had the opportunity to connect with each other and

converse... but in a context in which the academic conversation involved a large number of male cohorts continuing a conversational style that constrained the sorts of ideas that could emerge from these conversations. It was only by (in this historical case, accidentally) removing these other voices from their conversational sphere that new topics and methods could emerge from these women's discussions. And this case is interesting because various members of this cohort disagreed deeply about a variety of foundational issues: for example, Foot was an atheist, Anscombe a devoted Catholic. Murdoch's philosophy was developed primarily in fiction, while others wrote more traditional academic papers. What enabled their fruitful collaboration was not intellectual homogeneity, nor a shared commitment to something like a common religion. Instead, their different starting points seem to have contributed to a rewarding discussion that enabled each of them to refine their views in ways that would not have been possible without their ongoing conversations.

The last of Kiran's dimensions of mediation involves enabling and constraining. Most obviously, user-controlled disconnections enable those who disconnect but constrain the disconnected from continued engagement. Of course, freedom of speech does not entail that anyone else listen, and blocking someone from participating with particular others does not amount to constraining them from speaking altogether, nor should it be taken to be equivalent to censorship. In fact, facilitating user-controlled disconnections can provide an alternative to top-down enforcement of uniform norms by an authority, whether in the form of a social media platform administrator or a government agent.

Setting aside this issue, the ethical significance of the enabling/constraining dimension of disconnection technologies, in this context, is that constraining harmful actions wisely and well can enable healthier interactions to take place.

The question is how to use them well.

Best practices for disconnection

These different dimensions can each be implemented well or badly.

Enabling disconnection, for example, can be an act of cowardice, if one does so to avoid encountering different views, or to “keep the peace” rather than have a difficult but necessary conversation with an acquaintance behaving badly. It can also, however, be an act of courage to enforce boundaries and refuse to engage. Larger features of context than the mere fact that something disconnects will be necessary to determine whether something expresses or reinforces a virtuous or vicious trait of character. One contextual factor that can influence this, as noted above, is whether maintaining a connection ends up contributing to a spiral of silence or helps encourage discussion, and the degree to which a disconnection enables one to celebrate difference versus creating an intellectual echo chamber.

These are not merely factors for individual decision-makers about individual actions. Design decisions about how disconnection works are also impacted. Consider the decision of whether to make disconnection visible or invisible to the people from whom one disconnects, whether the connection is severed in both directions, and how the disconnection affects sharing user content with others. Each of these factors can shape the context of decision-making for individuals, morally speaking, as well as influence what sorts of decisions people are likely to make. For example, a recent discussion on Twitter involved the ethics of muting versus blocking a follower: muting is invisible to the follower and everyone else observing their Tweets, and they can continue to interact with the account in ways that may make it easier for them to invite others to contribute – a “pile on” – or can make it easier for them to redirect aggressive content toward

one's other followers. (Kendall 2018) Thus, even if one is willing to tolerate some verbal abuse, or at least mute it, one might have moral reason to shield others from this same person. In other cases, muting rather than blocking might be the more appropriate response – for example, if you are not personally and psychologically in a good space to read particular content, but think that it could still contribute valuably to a larger conversation.

Or consider the fact that most contemporary platforms render individual disconnection relatively obscure to the other user(s). This can facilitate low-conflict disconnection, but may itself contribute to concerns about offering an “easy out” from difficult conversations, ones that both call for and help us cultivate virtues such as patience and courage. (Vallor 2012) While surely there are times when discretion is the better part of valor, perhaps it would be good to have more visible disconnection options, both to facilitate courageous expression of one's reasons for disconnection (“I can't remain friends with you because...”) and to make the social cost of belligerent or vicious behavior more visible to people. While this, too, could be abused, a wider range of options might, when paired with helpful social practices and baseline good character, help foster better conversations. One important rule of thumb seems to be asking oneself character-based questions about the decision itself, “would it be courageous or cowardly to disconnect from this person, and if it might be courageous, would it be kind or cruel to do so visibly, or explain why?” The details that determine the answers to these questions will doubtless vary by context, but framing these questions in terms of virtue can help direct us to salient features of that particular situation. It may matter whether one is disconnecting from an elderly relative or bombastic co-worker, whether one intends to carry on in person as if nothing is wrong or one is in a position to cut ties altogether, and whether the person ought to be given the chance to hear why the relationship is curtailed, or whether doing so would only cause more harm.

This raises another question: when does a concern about another user's conduct or content become a matter, not just for individual concern, but a more general reason to disconnect, such as by banning a social media user from that media platform? Many of the factors I have pointed to as ethically salient involve moderating individual psychological responses, like the backfire effect, or interpersonal relationships at the level of just a few individuals. But some public online behavior seems likely to affect a wide range of people. While a systematic justification of user bans is beyond the scope of this paper, perhaps some platforms might have reason to ban some users in order to encourage others to participate more.

Conclusion

My goal in this paper is not to deliver a uniform policy recommendation as to when someone ought to be blocked from a platform, or when you ought to unfriend versus unfollow someone, or how best to design disconnection technologies. Instead, it is to make mechanisms for disconnection, and their possible effects on people, relationships, and conversations explicit, and to articulate social goals at which we aim in our interactions, goals that can be both supported and undermined by disconnections. By making both the tools and the target more accessible, my hope is that we will be better equipped to foster civic discourse.

Paying attention to the psychological effects of political conversations on social media reveals a number of ethically salient phenomena, including echo chambers, the spiral of silence and the backfire effect. At the same time, using an Aristotelian framework to consider the goals, ideals, and practices of good interpersonal relationships reveals how such relationships can affect individuals. This, in concert with an account of how technological mediation affects user experience, can be thoughtfully deployed to promote healthy discourse by fostering good

friendship. And this in turn requires both courage to accept difference, and wisdom to know when to leave a relationship behind.

References

- Aristotle. 1999. *Nicomachean Ethics*. Translated by Terence Irwin. 2nd ed. Indianapolis, Indiana: Hackett Publishing Company. <https://books.google.com/books?id=-Mf5XV8q6CgC&dq>.
- Axinn, Sidney. 1990. "Moral Style." *The Journal of Value Inquiry* 24 (2): 123–33.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences* 115 (37): 9216–21. <https://doi.org/10.1073/pnas.1804840115>.
- Bialik, Kristen. n.d. "14% of Americans Have Changed Their Mind about an Issue Because of Something They Saw on Social Media." <http://www.pewresearch.org/fact-tank/2018/08/15/14-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/>.
- boyd, danah. 2014. *It's Complicated: The Social Lives of Networked Teens*. New Haven, CT, USA: Yale University Press. <https://books.google.com/books?id=p9u8AgAAQBAJ>.
- Chang, Mahalia. 2017. "Is 2017 the Year We All Quit Social Media?" Harper's BAZAAR. January 5, 2017. <https://www.harpersbazaar.com.au/culture/how-to-quit-social-media-6235>.
- Cocking, Dean, and Jeanette Kennett. 1998. "Friendship and the Self." *Ethics* 108 (3): 502–27. <https://doi.org/10.1086/233824>.
- Duggan, Maeve, and Aaron Smith. 2016. "2. The Tone of Social Media Discussions around Politics." *Pew Research Center: Internet, Science & Tech* (blog). October 25, 2016. <http://www.pewinternet.org/2016/10/25/the-tone-of-social-media-discussions-around-politics/>.
- Elder, Alexis M. 2018. *Friendship, Robots, and Social Media: False Friends and Second Selves*. 1 [edition]. Routledge Research in Applied Ethics 9. New York: Routledge, Taylor & Francis Group.
- Fierberg, Emma. 2018. "I Quit Social Media for a Month — and It Was the Best Choice I've Ever Made." *Business Insider*. January 10, 2018. <https://www.businessinsider.com/i-quit-social-media-for-one-month-it-changed-my-life-facebook-instagram-snapchat-twitter-millennial-2018-1>.
- Hampton, Keith, Lee Rainie, Weixu Lu, Maria Dwyer, Inyoung Shin, and Kristen Purcell. 2014. "Social Media and the 'Spiral of Silence.'" *Pew Research Center: Internet, Science & Tech* (blog). August 26, 2014. <http://www.pewinternet.org/2014/08/26/social-media-and-the-spiral-of-silence/>.
- Kendall, Mikki. 2018. "So Hey Can We Talk Again about Being Visible Online as a Woman in Media and Ways You Can Insulate Yourself from Things like This? Because a Lot of Folks Are Awful & Unfortunately Many Platforms Fail to Be Worth a Damn so You Have to Find Solutions for

- Yourself. <https://twitter.com/Nytimesarts/status/1031884222661844993>” Tweet. @Karnythia (blog). August 21, 2018. <https://twitter.com/Karnythia/status/1031917070517248000>.
- Kiran, Asle H. 2015. “Four Dimensions of Technological Mediation.” *Postphenomenological Investigations*, 123.
- Midgley, Mary. 2013. “The Golden Age of Female Philosophy | @guardianletters.” *The Guardian*, November 28, 2013, sec. World news. <https://www.theguardian.com/world/2013/nov/28/golden-age-female-philosophy-mary-midgley>.
- Nguyen, C. Thi. 2018. “Echo Chambers and Epistemic Bubbles.” *Episteme*, September, 1–21. <https://doi.org/10.1017/epi.2018.32>.
- Noelle-Neumann, Elisabeth. 1974. “The Spiral of Silence a Theory of Public Opinion.” *Journal of Communication* 24 (2): 43–51. <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>.
- Nyhan, Brendan, and Jason Reifler. 2010. “When Corrections Fail: The Persistence of Political Misperceptions.” *Political Behavior* 32 (2): 303–30. <https://doi.org/10.1007/s11109-010-9112-2>.
- Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York, NY: Basic Books. <https://books.google.com/books?id=J2ine5sIIkgC>.
- . 2015. *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York, NY: Penguin Press. <https://books.google.com/books?id=OUHZCwAAQBAJ>.
- Vallor, Shannon. 2010. “Social Networking Technology and the Virtues.” *Ethics and Information Technology* 12 (2): 157–70. <https://doi.org/10.1007/s10676-009-9202-1>.
- . 2012. “New Social Media and the Virtues.” In *The Good Life in a Technological Age*, edited by Philip Brey, Adam Briggie, and Edward Spence, 376. Routledge Studies in Science, Technology and Society. New York, NY: Routledge. <https://books.google.com/books?id=KNLngc2XBigC>.
- . 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York, NY: Oxford University Press. <https://books.google.com/books?id=RaCkDAAAQBAJ>.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359 (6380): 1146–51. <https://doi.org/10.1126/science.aap9559>.
- Williams, Bernard. 1981. “Persons, Character, and Morality.” In *Moral Luck: Philosophical Papers 1973-1980*, edited by James Rachels, 1–19. Cambridge Paperback Library. Cambridge, UK: Cambridge University Press. <https://books.google.com/books?id=wMGW2Ehldp8C>.
- Wood, Thomas, and Ethan Porter. 2018. “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence.” *Political Behavior*, January. <https://doi.org/10.1007/s11109-018-9443-y>.