Journal of the American Philosophical Association (2018) 298–322 © American Philosophical Association
DOI:10.1017/apa.2018.27

Technological Seduction and Self-Radicalization

ABSTRACT: *Many scholars agree that the Internet plays a pivotal role in self-radicalization, which can lead to behaviors ranging from lone-wolf terrorism to participation in white nationalist rallies to mundane bigotry and voting for extremist candidates. However, the mechanisms by which the Internet facilitates self-radicalization are disputed; some fault the individuals who end up self-radicalized, while others lay the blame on the technology itself. In this paper, we explore the role played by technological design decisions in online self-radicalization in its myriad guises, encompassing extreme as well as more mundane forms. We begin by characterizing the phenomenon of technological seduction. Next, we distinguish between top-down seduction and bottom-up seduction. We then situate both forms of technological seduction within the theoretical model of dynamical systems theory. We conclude by articulating strategies for combating online self-radicalization.*

KEYWORDS: technological seduction, virtue epistemology, digital humanities, dynamical systems theory, philosophy of technology, nudge

1. Self-Radicalization

The 2015 Charleston church mass shooting, which left nine people dead, was planned and executed by white supremacist Dylann Roof. According to prosecutors, Roof did not adopt his convictions 'through his personal associations or experiences with white supremacist groups or individuals or others' (Berman 2016: 1). Instead, they developed through his own efforts and engagement online.

The same is true of Jose Pimentel, who in 2011 began reading online Al Qaeda websites, after which he built homemade pipe bombs in an effort to target veterans returning from Iraq and Afghanistan (Weimann 2012). Also self-radicalized online were Tamerlan and Dzhokhar Tsarnaev, perpetrators of the 2013 Boston Marathon massacre (Majeed 2016: ch. 11). In 2016, Omar Mateen killed forty-nine people in a Florida night club. Like Pimentel, Roof, and the Tsarnaev brothers, Mateen was radicalized through his activities online (Pilkington and Roberts 2016). More recently, Alex Minassian, sole perpetrator of the 2018 'incel'-related mass attack in Toronto, was found to be self-radicalized online, with his 'worldview . . . heavily influenced by the grievance culture that defines [online] message boards like r/Incel and certain sections of 4chan' (see Feldman 2018).

These are just a few examples of a phenomenon called lone-wolf terrorism (Weimann 2012). Many scholars agree that this form of terrorism is facilitated by

the Internet (Precht 2007; von Behr et al. 2013: §3). However, the specific mechanisms by which the Internet facilitates self-radicalization—in the extreme case of lone-wolf terrorism as well as in more mundane cases such as extremist attitudes and voting behavior—are disputed. According to Bjelopera (2010), it does so by normalizing the kind of behaviors and attitudes the discussion of which would typically be viewed as unacceptable and be met with disapprobation offline (see also von Behr et al. 2013). On another view (Silber et al. 2007), the Internet facilitates self-radicalization by generating an echo-chamber (aka 'filter bubble'): people interested in radical ideology tend to communicate directly or indirectly only with each other, reinforcing their predilections. A related opinion is that the phenomenon of group polarization, as facilitated by social media chat forums, is among the salient causes (Sunstein 2011, 2017).

Other researchers question the underlying assumption that the Internet is the sort of thing we should blame. Omotoyinbo (2014: 58), for example, argues that in these cases the 'Netizen [user of the Internet] is the criminal and that the Net is a guiltless accomplice.' Those inclined to turn an eye away from the medium itself have sought various alternative explanations for the phenomenon of self-radicalization, such as social alienation (Abrahams 2002; Torok 2011).

Surrounding the dark shadow of lone-wolf terrorism is the broader penumbra of self-radicalization that results in less dramatic but still worrisome actions and attitudes. The 'Unite the Right' rally in Charlottesville, Virginia, in 2017 brought together neo-Nazis, white supremacists, and white nationalistic sympathizers for one of the largest in-person hate-themed meetings in the United States in decades. Many of the participants in this rally were organized and recruited via the Internet.[1] Even more broadly, white supremacists and their sympathizers met and organized on r/The_Donald (a Reddit community) and elsewhere during the 2016 American presidential campaign that resulted in the election of Donald Trump.[2] The radicalization of Trump voters may be less extreme, but it demands an explanation just as much as the radicalization of Dylann Roof.

Omotoyinbo's analysis of online self-radicalization, according to which blame should not go to the technology, runs counter to the view of Langdon Winner (1986: 19–39) that artifacts and sociotechnical systems can 'have politics' (see section 2.2 below for examples involving web design). Winner identifies two main ways in which values, principles, and power relations can be embedded in such systems. First, the adoption of a technology could necessitate certain kinds of human relations while making others impossible. The sociotechnical system thus becomes a way of settling a political question with relatively permanent infrastructure. Second, a sociotechnical system could be strongly consilient with one set of political relationships and strongly in tension with another. In either case, there is a feedback loop between the political arrangements that lead to the adoption of the sociotechnical system in the first place and those facilitated by the

[1] As reported by the Anti-Defamation League, such online recruitment efforts had already begun increasingly transitioning to real-world meetups during the months leading up to Charlottesville; see https://www.adl.org/blog/alt-right-moving-from-online-to-real-world-activity.

[2] See https://fivethirtyeight.com/features/dissecting-trumps-most-rabid-online-following.

system once it is in place. We worry that the contemporary Internet reliably fosters political attitudes and associations that other sociotechnical systems for information-transfer and communication would not.

This is not to suggest that there is any one cause of Internet-facilitated self-radicalization or that there must be any particular primary cause for all cases. However, we do think that one overlooked influence in some prominent cases is the way gathering and delivery of information work on the contemporary Internet.

In what follows, we explore how various kinds of technological design decisions can and do play a role in online self-radicalization. If our argument is on the right track, then combating online radicalization must involve not merely investigating and implementing existing strategies, such as depolarisation (Vinokur and Burnstein 1978), community policing (Ndili 2016), and public education (Neumann 2013). It requires in addition that we allocate resources (philosophical, engineering, political) to technological design decision-making.

Here is the plan for this paper: section 2 introduces two varieties of what we call technological seduction: top-down seduction (section 2.1) and bottom-up seduction (section 2.2). Section 3 situates both varieties within the theoretical model of dynamical systems theory by showing how top-down and bottom-up seduction generate two kinds of feedback loops that pull technology users toward extremes. Section 4 concludes by articulating strategies for combating online radicalization in light of the conclusions drawn in sections 2 and 3.

2. Varieties of Technological Seduction

In this section, we characterize seduction as a multistage process. According to Forrester (1990: 42), 'the first step in a seductive maneuver could be summed up as, "I know what you're thinking"'. This gambit expresses several underlying attitudes. First, it evinces 'the assumption of authority that seduction requires'. The authority in question is epistemic rather than the authority of force. Seduction is distinguished from assault in that it aims at, requires, even fetishizes consent. However, as Conly (2004) argues, seduction can be said, at least in some cases, to 'interfere with the reasoning process,' sometimes 'subverting' it.[3] The seducer insists that he is better placed to know what the seducee thinks than the seducee himself is. Second, 'I know what you're thinking' presupposes or establishes an intimate bond. Nothing is more bound up with personal identity than someone's inner life—their thoughts, feelings, emotions, and values (Strohminger and Nichols 2014). The seducer's insistence that he knows what the seducee is thinking thus brings the two into close connection. Third, 'I know what you're thinking' blurs the line between assertion, imperative, and declaration. This is because human agency and cognition are often scaffolded on dialogical processes. We find out what we think by expressing it and hearing it echoed back in a way we can accept; we also find out what we think by having thoughts attributed to us and agreeing with those attributions (Alfano 2013, 2018; Doris 2015; Wong 2006).

[3] Like Forrester, Conly does not explicitly define seduction; she only characterizes it in relation to human reasoning.

The next step in a typical seduction, as we here conceive it, is an affirmative response by the seducee: 'Yes, you do know what I'm thinking.' The third step moves from the realm of language to the realm of action. The seducer now suggests, 'So let's do. . . '. And in the fourth and final step in a (successful) seduction the seducee agrees. Seduction is often (though not necessarily) modally robust (Pettit 2015). After someone has been seduced, in a wide range of counterfactual scenarios they will typically both continue to accept that the seducer knows what they are thinking and continue to be committed to acting in concert with the seducer.

In this section, we diagnose two ways in which information technologies play the functional role of the seducer by telling netizens, 'I know what you're thinking.' We label these top-down and bottom-up technological seduction for reasons that emerge below.

2.1 Top-down seduction

Top-down technological seduction is imposed by technological designers, who, in structuring technological architecture in particular ways (e.g., designing websites with certain kinds of menus and options), invite users to accept that their own thinking is similarly structured.[4] In so doing, designers encourage or 'nudge' the user toward certain prescribed choices and attitudes (Thaler and Sunstein 2008). In this section, we will outline this phenomenon, and in section 3 we will consider how nudged choices can (sometimes in tandem with bottom-up seduction) guide individuals in malignant directions.

First, though, some preliminaries. Nudging is a familiar phenomenon and one that need not be viewed as inherently problematic. As Thaler and colleagues (2014) point out, people 'do not make choices in a vacuum. They make them in an environment where many features, noticed and unnoticed, can *influence* their decisions' (2014: 1).

Consider the presentation of our choices in supermarkets and when selecting television stations (White 2013). Supermarkets must present goods in some way or another, with certain items closer to the front and to checkout stands; likewise, television stations must be assigned some distinct number, which will inevitably be less than or greater than other numbers and thus will involve more or less clicking effort relative to the default (low number) presentation. Sunstein (2015) calls this the inevitability of choice architecture. Choices (e.g., what to buy at the supermarket, which station to watch) will unavoidably be structured in some particular way; choices are not presented without a frame even by those who value the chooser's autonomy.

When the inevitability of choice architecture is paired with collateral information about human choice behavior (including systematic cognitive biases), decision makers, even when free to make their own choices, will nonetheless be steered toward particular choices in a way that reflects the manner in which the initial

[4] The contrast between the systems designer/engineer's and user's perspectives is a recurring issue in human-computer interaction, see Rogers et al. (2012).

space of choices is presented. Nudging can be intentional or nonintentional. For instance, if the store shelves at the supermarket are stocked randomly, individuals are more likely to purchase things closer to the front, whatever these things happen to be. Individuals can accordingly be nudged unintentionally (one might also say, negligently or recklessly) through thoughtless design of choice architecture.

But nudging can be and often is intentional, as when choice architecture is structured to steer in ways that align with the designer's goals. According to proponents of libertarian paternalism (Thaler and Sunstein 2003a, 2003b; Sunstein 2014: 430), policymakers not only can but should design choice architecture in a way that both makes it transparent to public scrutiny and nudges citizens toward positive decisions. In the supermarket example, this might involve laws that require supermarkets to place soda and cigarettes toward the back of the store and away from checkout stands (see Alemanno 2012) or that require employers to present a prudent retirement plan for employees as a default (see Thaler and Benartzi [2004] and Dayan and Bar-Hillel [2011]).

Not all intentional nudges are benevolent. A nudge could be calculated to benefit the nudger (e.g., certain kinds of marketing) or a third party (e.g., nudging people to donate to a charity). The *New York Times* presently gives subscribers a free choice to cancel their subscription, but exercising that option requires calling during business hours. While the choice to cancel in this case is a free one, the *Times* is capitalizing on information about human psychology to add friction that steers individuals toward continuing with the subscription, a goal that aligns with the choice architects' goals.

Creating friction is just one way to nudge people, and it need not rely on technological design in any serious way. But technological design offers additional ways to nudge. When gathering information online, for example, one encounters what is in effect a choice architecture chain that involves moving through a series of drop-down menus, app search tools, algorithm-generated key terms, friend suggestions, related stories, and other content. These processes feed back into one another, making someone's journey through the choice architecture increasingly path-dependent over time.

The design decisions that drive the various links in such a chain can be malignant if (among other things) they construct the relevant categories and subcategories so as to make it seem that some phenomena that deserve attention are not worthy of attention or that some phenomena that do not deserve attention do deserve attention (see Alfano and Skorburg [2018] for an example). Consider, for example, drop-down menu design (see Carter 2017). On the main page of the US political website *The Daily Caller*, you will find a category called 'The Issues'. There are myriad issues that might be of general concern or interest. Which specific ones do you (the user) want to know about? How do you decide?

The Daily Caller makes this very simple. Hover your mouse over 'The Issues' and three choices pop up: <DEFENSE>, <EDUCATION>, and < ENERGY>.[5] This design choice gives the illusion of carving news reality at its joints and engages what Kahneman (2011) calls the what-you-see-is-all-there-is (WYSIATI) mindset,

[5] Search conducted on August 5, 2017, in the United Kingdom.

an outlook that often leads people to accept uncritically one of the framed options as legitimate and to carry on searching (encountering further menus) having already made a choice within that framed space.

What is nonexplicitly communicated by *The Daily Caller*'s design choice is that defense, education, and energy are all the issues and the only ones the user will or should care about. Supposing the reader clicked on the first option, <DEFENSE>, and then clicked on the first story of the four juxtaposed by images, the reader will within seconds of thinking about something as general as issues find herself reading a story about how a US Army howitzer has 'covfefe' stamped on the gun of its M109 Paladin.[6]

The design of application interfaces can likewise engage the WYSIATI mindset by inducing users to perceive not only choice options but also experiences and values in misleading ways. As Google Design ethicist Tristan Harris (2014) notes:

> When we wake up in the morning and turn our phone over to see a list of notifications—it frames the experience of 'waking up in the morning' around a menu of 'all the things I've missed since yesterday.' . . . How empowering is this menu of choices when we wake up? Does it reflect what we care about?

Harris's case involves a smartphone feature that uses an algorithm to frame a narrative each morning around what the algorithm takes to be the key things one has missed during sleep. The kinds of notifications that frame such a narrative include information such as: who accepted an Instagram follower request, spam from a music app, a spam suggestion from an app called any.DO to 'take a moment to plan your day', information that someone liked an Instagram photo, and so on. Even in cases where there is a substantial disparity between the user's values and interests and the algorithm's imputation of these values and interests, a user may be inclined tacitly to accept the algorithm's narrative by selecting from its highlights.

Technological seduction in these two kinds of cases need not be a matter of saying or implying, 'I know that you're thinking that *p*.' The kind of seduction that features in these cases is structural. In the drop-down menu case, the seduction at issue involves suggestions like:

- 'I know that you think these are the most important categories when it comes to news' (e.g., main categories on menu).[7]
- 'I know that you think *X* is associated with *Y*' (e.g., above-chance association of tags or labels).[8]
- 'I know that you treat *X* as a subcategory of *Y*' (e.g., submenus).

[6] See http://dailycaller.com/2017/06/23/army-tank-has-covfefe-stamped-on-the-gun-photo/; accessed August 18, 2017.

[7] Instances of seduction taking this shape feature in consumption of 'personalized' news, as reported in literature on filter bubbles (Pariser 2011; Flaxman et al. 2016). For reports on the prevalence of such personalization, see Van Hoboken (2012) and Dillahunt et al. (2015).

[8] See Benkler et al. (2017) for evidence of this form of seduction during the 2016 US presidential campaign.

When netizens are confronted with these first steps in the seductive process, they may very well continue to the second, third, and fourth steps. In other words, they may—explicitly or implicitly—think, 'Yes, that is what I think'. This makes them susceptible to a suggestion to act in the third step. It seems highly likely that this process played a role in the 2016 US presidential elections, at least for some voters. The media with which they engaged impressed categorical associations on them, then used those associations to bolster exhortations to vote for one candidate over another (or not to vote at all).[9]

Compare what we are calling top-down seduction to Kant's deduction of the categories in the *Critique of Pure Reason* ([1781]1958: A84–85–B116–117). The analogy is, of course, only partial, but we think that a brief acknowledgment of the commonalities is instructive. Note that in much the same way that, for Kant, quantity, quality, relation, and modality are the given categories that structure cognition of objects in the empirical world, for readers of *The Daily Caller*, the given categories that structure cognition of the news are *Politics, US, World, Entertainment, Sports, Business, Outdoors,* and *The Issues* (subdivided into *defense, education*, and *energy*). In both cases, the categories are, in a straightforward sense, already provided and are so in a way that circumscribes what moves are available to the thinker. To use another news site as an example, netizens who get their news from *Breitbart* face a world structured by the categories of *Big Government, Big Journalism, Big Hollywood, National Security, Tech,* and *Sports*. Netizens who get their news from the *Guardian*, by contrast, encounter a world structured by the categories of *World, Politics, Sport, Football, Culture, Lifestyle, Environment, Economy, Media, Tech,* and *Travel*. Naturally, there are differences between the Kantian picture of the mind and the technological categorizations we are exploring. According to Kant, there is no way to get outside the categories. Netizens, however, can change their media diet. But if our arguments above about the stickiness of defaults and the WYSIATI mindset are on the right track, netizens are liable to get stuck in the categories imposed by technological design.

When netizens are invited to accept that their own thinking is structured in socially dangerous or hateful ways, choice architects can seduce users to embrace prejudiced attitudes. In other words, they end up following the second step in the formula for seduction articulated above. Just compare the *Daily Caller* drop-down menu entitled 'Issues' (one that seduces merely to simple-minded thinking) with the kind of semantic tags *Breitbart* associates with its content. One of the more objectionable tags at *Breitbart* is 'Black Crime'.[10] There is no corresponding tag for 'White Crime' or any other racialized crime.) As of October 2017, there were six *Breitbart* stories with this tag, all with sensationalistic headlines, such as, 'Black rape gangs violate two Detroit women in one night, hours apart,' and 'Black mob swarms Georgia Walmart to see 'how much damage' they could do.'

[9] See, for example, http://www.breitbart.com/2016-presidential-race/2016/11/05/hillary-clinton-jail-not-white-house/, which uses Breitbart's frequent association of Hilary Clinton with criminality to make a case against electing her.

[10] See http://www.breitbart.com/tag/black-crime/, accessed October 11, 2017.

If we compare the semantic tags on all stories published in 2016 by *Breitbart* with the tags used by other news organizations, it becomes clear that the world will look very different to a regular *Breitbart* reader than it does, for example, to a regular *NPR* or *Huffington Post* reader (Figures 1–3).[11]

Figures 1–3 show that the media world in which readers of *Breitbart* live looks very different from the media world in which readers of either *NPR* (a centrist media organization) or the *Huffington Post* (a liberal voice) live. *Breitbart* consumers see a world in which the most important news is that Mexican cartels commit atrocities in Texas, Muslim terrorist immigrants rampage through Europe, Barack Obama seeks to take citizens' guns, Russia and Turkey intervene in Syria to help fight the Islamic State, and Milo Yiannopoulos takes on feminist social media censorship via his self-styled 'Dangerous Faggot Tour'. Readers of other sites encounter some of the same phenomena (framed in less threatening ways) as well as a host of other phenomena that fit less easily into *Breitbart*'s Manichean worldview, such as right-wing attacks against the LGBT community, the sentencing of Dylann Roof, law enforcement agencies' showdown with technology companies over consumer privacy, and NASA's exploration of the solar system.

This phenomenon is well-characterized by 'agenda-setting' theory, which suggests that consumers of news 'learn how much importance to attach to a topic on the basis of the emphasis placed on it in the news' (McCombs 2000). In essence, the 'public agenda' for *Breitbart* readers can be said to be shaped by the emphasis and coverage given by *Breitbart* to certain topics and tags. We also note that modern news websites, including those we analyzed above, have facilities for commenting and sharing via social media outlets; arguably social media could serve to augment the 'agenda-setting' power of such news websites.

Moreover, as figures 1–3 indicate, the consumer of (for instance) *Breitbart* will have her experiences structured in a way that may naturally give rise to distinctive biases. Plausibly, a reader of *Breitbart Texas*, for example, will be led to associate Mexican concepts with crime concepts and negative valence. Such conceptual connections (e.g., Mexico = crime = bad), when ossified, are epistemically problematic because they lead people to make unjustified generalizations and act upon them (see Holroyd and Sweetman 2015).

We will discuss the effects of top-down seduction in connection with self-radicalization in more detail in section 3. First, though, we turn to a related but distinct form of technological seduction.

2.2 Bottom-up seduction

Whereas top-down technological seduction plays out through the agency of designers (whether they know it or not), bottom-up seduction can occur without the involvement of anyone's agency other than that of the seducee. Top-down technological seduction imposes a taxonomic structure of categories, subcategories, and associations on users, inviting them to accept that this is how their own concepts, subconcepts, and mental associations are structured. It is to

[11] This digital humanities methodology was pioneered in the context of philosophy by Alfano et al. (2018).

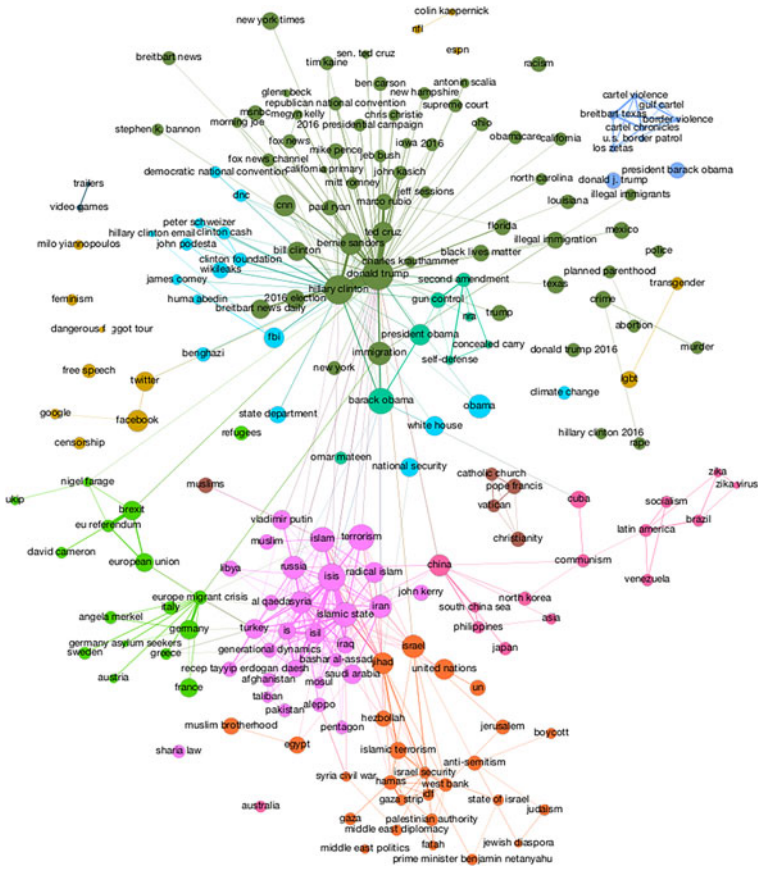

Figure 1. Network of the top 200 semantic tags on *Breitbart* in 2016. Layout = ForceAtlas. Node size = PageRank. Node color = semantic community membership. Edge width = frequency of co-occurrence.

some extent Procrustean: the same top-down structure is suggested to and imposed on all users.

Bottom-up technological seduction is different: it creates suggestions either by aggregating other users' data or by personalizing for each user based on their location, search history, or other data. It takes a user's own record of engagement as the basis for saying, 'I know what you're thinking'.[12] Engagement, in this context, refers to all recorded aspects of a user's individual online behavior. This includes the user's browsing history (which sites/links are visited, frequency of such visits, etc.), search history, record of sharing and 'liking' posts on social media, email record (if, for example, Google is used for both search and email),

[12] Although our descriptions of bottom-up seduction use agency-language in places, we remain neutral with respect to artificial agency. For a notable defence of artificial agency and the morality of artificial agents, see Floridi and Sanders (2014).

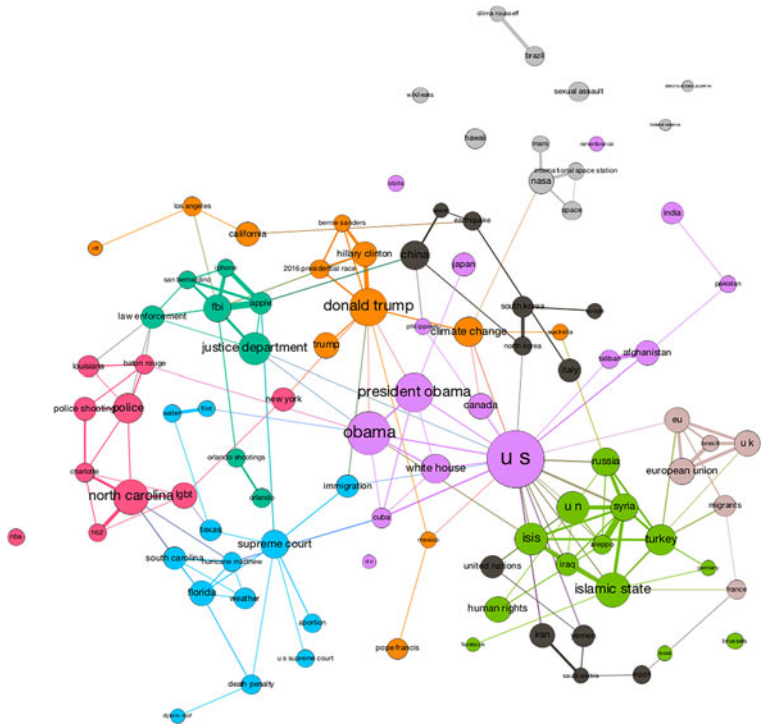

Figure 2. Network of the top 100 semantic tags on *NPR*'s 'The Two Way' news section in 2016. Layout = ForceAtlas. Node size = PageRank. Node color = semantic community membership. Edge width = frequency of co-occurrence. Only 100 semantic tags were chosen for *NPR* due to its overall sample size, which is one order of magnitude less than *Huffington Post* and *Breitbart*.

physical location and trajectory (if, for example, Google's location services are used for navigation or merely having the 'location history'[13] feature active), and so on. While it is possible to disguise these aspects of one's online signature in various ways, most users neglect to do so. In addition to the individual's own record of engagement, others' records of engagement can be used to profile that individual. For example, to the extent that your record of engagement—even in depersonalized aggregated form—is more similar to that of one set of users than that of another set of users, you are liable to be profiled among the former. Bottom-up seduction can say, 'I know what you're thinking' because it can justifiably say, 'I know what you and people like you thought, and what those other people went on to think'.

Bottom-up technological seduction occurs when profiling enables both predictive and prescriptive analytics to tell a netizen, 'I know what you're thinking', and prompt the sort of uptake that leads to action. This is especially worrisome when the process bypasses the user's capacity for reasoning. Following Koralus and Mascarenhas

[13] https://support.google.com/accounts/answer/3118687?hl=en; site accessed October 17, 2017, from Australia.

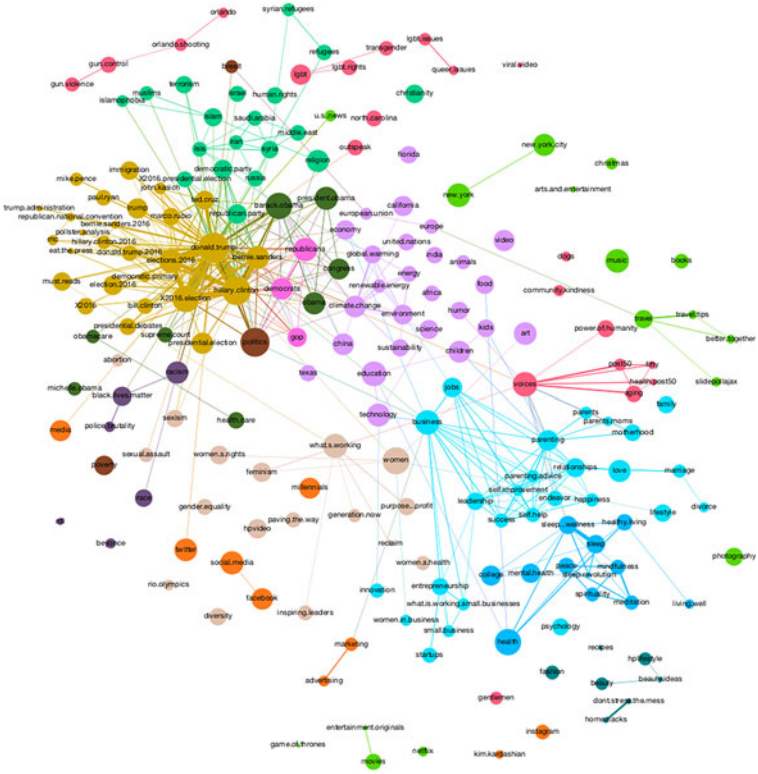

Figure 3. Network of the top 200 semantic tags on the *Huffington Post* in 2016. Layout = ForceAtlas. Node size = PageRank. Node color = semantic community membership. Edge width = frequency of co-occurrence.

(2013) on reasoning in general and Koralus and Alfano (2017) on moral reasoning in particular, we understand reasoning as the iterative, path-dependent process of asking and answering questions. Profiling enables online interfaces such as Google to tailor both search suggestions (using predictive analytics) and answers to search queries (using prescriptive analytics) to an individual user.

Consider a simple example: predictive analytics will suggest, based on a user's profile and the initial text string they enter, which query they might want to run. For instance, if you type 'why are women' into Google's search bar, you are likely to see suggested queries such as 'why are women colder than men', 'why are women protesting', and 'why are women so mean'.[14] And if you type 'why are men' into Google's search bar, you are likely to see suggested queries such as 'why are men jerks', 'why are men taller than women', and 'why are men attracted to breasts'.[15] These are cases of predictive analytics. The same predictive searches

[14] Search conducted June 10, 2017, in the Netherlands.

[15] Search conducted June 10, 2017, in the Netherlands.

conducted in another geographic location, at another time, by an account with a different history and social graph will yield different results.[16]

Prescriptive analytics in turn suggests answers based on both the query someone runs and their profile. In its most naïve form, a search for the query 'cafe' returns results for cafés nearest to the user; the top results will differ for someone in Amsterdam as opposed to Abuja. To continue with our prior examples: in response to 'why are women colder than men', one of Google's top suggestions is a post titled 'Why are Women Always Cold and Men Always Hot', which claims that differences between the sexes in the phenomenology of temperature are due to the fact that men have scrotums.[17] In response to 'why are men jerks', one of Google's top suggestions is a post titled 'The Truth Behind Why Men are Assholes', which contends that men need to act like assholes to establish their dominance and ensure a balance of power between the sexes.[18] And in response to 'why are women so mean', Google suggests posts answering questions about why beautiful women in particular are so mean, why women are so mean to each other, and why women are so mean to men. Most of these posts have a strongly misogynistic flavor, suggesting that beautiful women are mean because they are sick of being approached by weak men, that women are mean to each other to establish dominance hierarchies in pursuit of male attention, and that women are mean to men in general because feminism has corrupted their natural impulses. These misogynistic themes have recently been given an academic gloss by Jordan Peterson (2018), but he first articulated them in a post on *Quora*,[19] from which they quickly spread to *Reddit* and went viral.[20]

In a case of bottom-up technological seduction, Google suggests questions and then answers to those very questions, thereby closing the loop on the first stage of an iterative, path-dependent process of reasoning. If reasoning is the process of asking and answering questions, then the interaction between predictive and prescriptive analytics can largely bypass the individual's contribution to reasoning, supplying both a question *and* its answer. When such predictive and prescriptive analytics are based in part on the user's profile, Google is in effect saying, 'I know what you're thinking because I know what you and those like you thought'.

Consider next the path-dependency mentioned above. Which question you ask depends in part on both the questions you asked previously and the answers you accepted to those questions. If both the initial question and its answer are shaped by predictive and prescriptive analytics, then the first question-answer pair in the

[16] Depending on a user's profile, the content of search results can be subject to change, as in the case of Google's Personalized Search, which can 'customize search results for you based upon 180 days of search activity linked to an anonymous cookie in your browser'; see https://googleblog.blogspot.com.au/2009/12/personalized-search-for-everyone.html.

[17] https://www.qualityhealth.com/womens-health-articles/why-women-always-cold-men-always-hot; site accessed June 10, 2017, from the Netherlands.

[18] http://elitedaily.com/dating/sex/men-assholes/; site accessed June 10, 2017, from the Netherlands.

[19] https://www.quora.com/What-are-the-most-valuable-things-everyone-should-know.

[20] https://www.reddit.com/r/JordanPeterson/comments/75akno/dr_jordan_petersons_42_rules_for_life_the_origins/.

process of reasoning largely bypasses the human being's contribution. But that in turn means that subsequent questions and answers depend on this bypassing, potentially sending the user deeper into an epistemic and ethical morass.

To illustrate, suppose you were interested in anything beginning with the text string 'alt', such as 'alternative energy'. You type these first three letters into Google's search bar, and it suggests 'alt right'. Though you were not initially interested in this query, the suggestion piques your curiosity. You run the 'alt right' query, and several of the top results are videos on YouTube (a subsidiary of Google's parent company, Alphabet). The top result is a video by *The Atlantic* titled, 'Rebranding White Nationalism: Richard Spencer's Alt-Right'.[21] After watching this eleven-minute video, you allow the top suggested video (as determined by the video you clicked on and your own profile) to autoplay. It is a clip titled 'White nationalist Richard Spencer talks to Al Jazeera'.[22] When the video ends, you allow the next top suggested video to auto-play: 'BEST OF Richard Spencer vs. Hostile Audience at Texas A&M'.[23] This is a post by the white supremacist account Demography is Destiny. It celebrates Spencer's political positions and those like them. The first three letters of an innocent online search have been hijacked: in just a few steps, you went from the start of a query about alternative energy to Demography is Destiny.[24] The hypothetical netizen who follows this path may set off on the four-step process of seduction articulated above. First, YouTube tells them, 'I know what you're thinking'. Next, they accept that attribution. Third, they encounter propaganda that directs them to take action. Finally, they do so. This seems to have been precisely the steps followed by at least some of the participants in the 'Unite the Right' riot and white supremacist rally.[25]

YouTube is especially adept at this kind of bottom-up seduction because it uses unsupervised learning on Google's powerful artificial intelligence system to find patterns in individuals' and groups' preferences, then recommends clips that they are most likely to engage with (Newton 2017):

> 'We knew people were coming to YouTube when they knew what they were coming to look for,' says Jim McFadden, the technical lead for YouTube recommendations, who joined the company in 2011. 'We also wanted to serve the needs of people when they didn't necessarily know what they wanted to look for'.

McFadden's team succeeded. The vast majority of the time people spend watching videos on YouTube is now driven by algorithmic recommendations rather than search or linking. Whistleblower Guillaume Chaslot, who was fired by YouTube in 2013 for raising this criticism, has shown that YouTube recommendations are systematically biased in favor of bizarre, violent, and extremist content (Lewis 2018).

[21] Search conducted June 10, 2017, in the Netherlands; https://www.youtube.com/watch?v=kVeZo_Lhazw.

[22] https://www.youtube.com/watch?v=kVeZo_Lhazw.

[23] https://www.youtube.com/watch?v=FxfDOOY2H28.

[24] See also Tufekci (2018) for a recent discussion of this phenomenon.

[25] https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd.

This is what we mean by the path-dependency of bottom-up technological seduction: one's initial question may be shaped by predictive analytics, the answer to it determined by prescriptive analytics, and the character of one's subsequent questions and the answers to them shaped by the first round of question-and-answer. While the process may not bypass human cognition entirely, it can bypass critical reasoning in troubling ways. Feedback loops between predictive and prescriptive analytics are liable to make the human user a largely passive consumer and observer rather than an inquisitive and critical questioner.

3. Technological Seduction: A Theoretical Model

In this section, we first outline the concept of attractors as they feature in dynamical systems theory, then use this notion to illuminate how (in the course of the kinds of top-down and bottom-up technological seduction articulated in section 2.1 and 2.2) it can be difficult to avoid being pulled into the direction of hate or extremism online. This will be especially so for those who have established certain kinds of engagement histories. For such individuals, steering away from hate or extremism while navigating the Internet is like trying to thread a needle.

Dynamical systems theory (DST) is a mathematical framework for studying the behavior of systems whose endogenous states influence systemic parameters over time (see Palermos [2016] for helpful discussion; cf. Abraham et al. [1990] and Beer [1995]). Dynamic systems themselves are characterized in DST by a set of state variables, x, and a dynamical law, L—that is usually a set of differential equations that take states and parameters as arguments—where L regulates the change of those state variables over time.

Sometimes no matter what the initial state of the system is, the system will end up gravitating to some set of points—called a limit set—over time. Attractors are defined as limit sets that gravitate trajectories (i.e., sequences of states that arise from some initial state x_0 and the law, L) that are passing through all the nearby states. Each attractor is surrounded by what is called its '*basin of attraction*'. (Consider, for example, the gravitational field surrounding a star. Nearby objects, once in contact with the gravitational field, are then pulled in toward the star.) By contrast, repellers are defined as limit sets that push away trajectories that are passing through nearby states. Understanding the limit sets (attractors and repellers) of a system is crucial to understanding the regularities of the system's behavior. In the case of both attractors and repellers, feedback loops play a crucial role. An object on a trajectory through an attractor's basin of attraction will receive a little push toward the attractor, which will place it more firmly within the basin of attraction for another, bigger push at the next stage, which will place it even more firmly within the basin of attraction at the stage after that. Likewise for repellers: an object on a trajectory through a repeller's basin of repulsion will receive a little push away from the repeller, which will place it further from the repeller at the next stage, and so on.

Attractors as such are evaluatively neutral. Some tend toward beneficial consequences. In the case of transactive memory systems featuring individuals in close relationships (e.g., as when two individuals together recall something neither

would have recalled individually; see Wegner [1987] and Wegner et al. [1991]), the process of reciprocal interaction through mutual memory sharing draws such individuals toward memories that neither would easily (or at all) produce individually (see also Hollingshead 1998).

However, while the kinds of feedback loops that make transactive memory systems work are helpful, the feedback loops we have already seen in the case of technological seduction are pernicious. With reference to DST, we may appreciate how, when someone seeks news online, the two forms of technological seduction we have articulated function as distinct kinds of attractors that can make it difficult to avoid hateful attitudes and content. And as with the analogy of gravitational fields—once the searcher is pulled toward the attractor, it is all the more difficult not to get sucked in.

Take, for instance, the Kantian analogy discussed in the example of *Breitbart*'s design decisions in section 2.1. While searching from within the initial choice space *Breitbart* offers, certain concepts will be available to the user and others not; certain associations will be salient and others not; certain concepts will appear as species of superordinate concepts. As one continues to navigate within this frame, one's trust and distrust will be directed and shaped so that one tends further to confirm (for oneself) the *Breitbart* worldview. If one falls into its orbit, one is liable to get stuck.

This latter point—that one is liable to get 'stuck' in the circumstances described—comports with the way philosophers of education theorize about a notorious challenge to the efficacy of exemplars and other role models (see Zagzebski 2017). As Tanesini (2016) notes, the effectiveness of a message does not exclusively depend on the strength of the arguments contained, but also on audience receptiveness:

> [Those] who are the furthest away from intellectual virtue are precisely those who are less likely to pay attention. Exposure to exemplars might work only if it stimulates emulation. It is counterproductive if it leads to demoralisation or if it fans an already inflated conception of the self. (2016: 524)

Though Tanesini's point concerns entrenched attitudes in education, it generalizes to our case of top-down attractors. Once an individual's concepts and thinking have been initially furnished, shaped, and (through repeated confirmation and reinforcement) ossified through interaction with choice architecture like *Breitbart*'s, he is liable to become less receptive to the very kinds of influences that would otherwise push him away from the attractor. To the extent that one's cognitive and conceptual capacity for certain kinds of experiences is already formed through one's interaction history and subsequent cognitive and conceptual alignment with the relevant choice architecture, one is vulnerable to capture by the attractor. In this sense, one becomes genuinely stuck—intellectually rigid and immovable through typical discursive and rhetorical mechanisms (Roberts and Wood 2007: ch. 7).

One might try to avoid falling into this basin of attraction in one of three ways: first, one could ignore online news entirely. While this strategy would prevent one from being sucked in by a hateful attractor, it has the obvious downside of leaving

one uninformed. Second, one could make an explicit attempt to curate the sources of one's online news by avoiding problematic sources entirely (Levy 2017) and seek information from multiple reliable sources. However, if these multiple sources are not independent, then they will only serve to amplify the attractive power of the worldview they endorse. And establishing the independence of one's sources is no easy matter. Third, one could eschew curated news media and instead attempt to get directly to stories that are reported by accurate, reliable, and responsible sources. This could be done using, rather than a news website, a search engine such as Google.[26] While this third strategy might sidestep *Breitbart*-style choice architecture, one is not yet in the clear. By searching for news on Google, in an attempt to receive the news in a more autonomous fashion, one risks getting stuck in an uncurated, bottom-up predictive-and-prescriptive analytics feedback loop of the sort outlined in section 2.2, especially if one already has a particular history of searches and other types of engagement online. One may end up (despite attempting to avoid one kind of attractor) sucked in by a different attractor that enhances and confirms one's preexisting biases.

If our arguments in this section are on the right track, inhabitants of the contemporary online news ecosystem face a daunting challenge. Unless they are content to be uninformed, they need to find a way to avoid getting sucked into the worldview promoted by curated news media while also avoiding getting sucked into a self-confirmatory spiral of confirmation bias. For the netizen whose dispositions and patterns of behavior online are not naturally virtuous and innocent to begin with (in other words, for almost everyone), managing to avoid the kinds of attractors noted in section 2.1 and 2.2 would take a great deal of luck.

Here it is worth noting two interesting disanalogies between the way we encounter and interact with, on the one hand, mundane, nontechnological choice architecture (of the sort Sunstein takes to be inevitable in the ordinary course of things) and, on the other, with the online news ecosystem, replete with the top-down and bottom-up feedback loops that we have shown to be characteristic of it. First, in the former case, choice architecture (noticed and unnoticed) can influence our decisions, so as to nudge us toward particular choices. Imagine, for instance, a small town with two schools, one Protestant, the other Catholic. This is a generic form of what we are calling top-down seduction, as explained in section 2.1. There is, however, no obvious analogue in the generic (nontechnological) case to bottom-up seduction; predictive and prescriptive analytics are distinctive of the technological case. This is important because, as we have suggested in this section, it is precisely the interplay between top-down and bottom-up seduction that generates a special kind of cyclical problem for online news users, one that does not have an obvious nontechnological analogue. A second disanalogy between the two cases concerns a comparative difference in friction. The online case is frictionless or at least has less friction. It also features push notifications, pop-up ads, and autoplay of recommended content (e.g., on YouTube) and in many cases

[26] However, the 'neutrality' of algorithms as perceived by general Internet users, especially those used by search engines, are debatable from an anthropological perspective; see Byrne and Cheong [2017].

personalization. These features remove friction almost entirely, by design, in a way that ordinary choice architecture does not do to the same extent. This second disanalogy, accordingly, is a disanalogy of degree.[27]

4. Countering the Seduction: Technological and Political Solutions

This paper began with the problem of online self-radicalization, a problem with at least three aspects: philosophical, political, and technological. In this concluding section, we survey potential solutions.

4.1 Top-down solutions

An initial obstacle to countering or mitigating the process whereby top-down seduction lures people to self-radicalisation is a discriminatory one. Recall that attractors as such are evaluatively neutral. Some tend toward beneficial consequences, some toward neutral or mixed consequences, and some toward harmful consequences. Any design principle that targeted attractors as such would be treating beneficial attractors that feature in the sort of choice architecture encouraged by nudge theorists as not relevantly different from the kinds of nefarious attractors surveyed in section 2.1. Technological design that simply aims to eliminate attractors or reduce the scope of their basins is not a viable way to combat top-down seduction.

An initial way forward is diagnostic and draws on the resources of cognitive ergonomics and cognitive systems engineering (CSE; Hollnagel and Woods 1983, 2005). One straightforward description of the difference between positive nudging choice architecture and the kind of top-down design we find in choice architecture such as *Breitbart*'s is this: *Breitbart*'s design aims to nudge readers in the directions it does regardless of any other relevant interests that should be taken into account. In this respect, nudging choice architecture is starkly different. Such architecture nudges the user to take into account what is in the user's own interest in a way that is predicated upon the fact that individuals will often act against their own interests (Thaler and Sunstein 2003a: 1159).

CSE is useful here: in CSE, the user and the system are seen as a 'single interacting system' situated within a work context defined by the user's aims, and within this work context CSE models how people perceive, process, attend to, and use information to achieve their goals (Ritter et al. 2014: 39). In the specific case where users are interacting with website architecture on news sites, the relevant work context is pretty clearly defined. The aim is conveying relevant news to the user in ways that are as friction-free as possible and meet journalistic thresholds of informativeness and accuracy. By modeling how actual use lines up with goals, researchers employing CSE have improved information technology design in multiple sectors, such as healthcare (Bisantz et al. 2014). For example, researchers take into account both the needs and the cognitive limits and pressures of

[27] Thanks to an anonymous referee for encouraging expansion on this point.

diagnosticians, surgeons, emergency technicians, and so on in the design and implementation of visual displays, interactive communication systems, and other technologies. We expect that such models would be similarly fruitful in distinguishing which kinds of choice architecture lend themselves to use that runs contrary to users' goals.

In sum, CSE modeling offers a way of identifying one respect in which top-down design is problematic, and this is useful insofar as we want to identify clearly and eliminate the problem cases in a way that avoids, to the extent feasible, imposing our own values as designers on those who use a particular system. However, a pair of prescriptive problems remains: (1) what sort of ameliorative guidance is appropriate? (2) How can compliance be ensured?

First, regarding guidance: there is some hope from the human-computer interaction (HCI) field (Card et al. 1980). HCI proponents recommend adhering to various concrete principles in the design process. For example, in the case of designing a news website, such principles include: 'Have we made the most important information in this interface stand out [by means of font size, flow, etc.]?' 'Will the user get the information they need in a timely fashion if there is an emergency?'(Ritter et al. 2014: 39). Designers adhering to such principles will not structure the user's experience of news acquisition so that it is constrained within the navigation space of *Big Government, Big Journalism, Big Hollywood, National Security, Tech,* and *Sports*.

Unfortunately, such guidance in the form of HCI principles will be useless against reckless and malicious designs that either disregard the prospect of technological seduction or actively aim to seduce users. The latter are what Joanna Bryson calls 'evil programmers'.[28] They employ technical design in an ideologically motivated way with the aim of exploiting technology design to their own ends. After all, the imperatives that guide technical design are merely hypothetical in character: *if* designers have certain benevolent ends, *then* they ought to implement certain kinds of specific design strategies. It is of course doubtful that designers of the kinds of sites that exploit aggressive technological seduction have such ends.

This bring us to the issue of compliance. In the face of 'evil programmers', viable solutions to top-down seduction move from the technological to the political. One form is imposed by International Organization for Standardization (ISO), a global body that regulates a wide range of products and services. While compliance with ISO is voluntary, the network-effect benefits of compliance make it advisable for all but the smallest firms. Consider in particular ISO standard 9241-210, which covers the ergonomics of human-computer interaction. According to the revised 9241-210, the following activities are requirements for human-centered design:

1. Understanding and specifying the context of use,
2. Specifying the user requirements in sufficient detail to drive the design,

[28] https://joanna-bryson.blogspot.co.uk/2017/07/three-very-different-sources-of-bias-in.html. Bryson's thinking is especially important in the case of bottom-up design (see section 4.2 below)

3. Producing design solutions that meet these requirements, and
4. Conducting user-centered evaluations of these design solutions and modifying the design to take into account the results[29]

ISO 9241-210 can also be adopted in conjunction with ISO 9241-151, which covers 'Guidance on World Wide Web user interfaces'. ISO 9241-151 specifically targets the design of websites, including best practices for 'high-level design decisions and design strategy.'[30] In our view, the 'Black Crime' tag on *Breitbart* (see above, section 2.1) violates 9241-151.[31] However, it may be necessary to formulate and implement new ISO standards to cope with the problems caused by top-down technological seduction. For example, perhaps in addition to conducting user-centered evaluations only after a design has been implemented, it should be necessary to enlist a neutral third-party to survey (potential) users before a design is finalized and implemented.

However, enforcing compliance raises further political issues. In 2010, ISO 9241-210 changed the above four activities to 'requirements' from mere 'suggestions'. Such language constitutes progress to the extent that local political bodies make compliance with such international standards mandatory—viz., where costs for noncompliance are significant.

One note of optimism here concerns the EU's General Data Protection Regulation (GDPR), which safeguards EU citizens against privacy and data breaches. The updated regulations, which went into effect in May 2018 and protect only Europeans (unlike ISO, which is global), come with heavy penalties for breaches of GDPR regulations. For instance, serious infringements carry a maximum penalty of 'up to 4% of annual global turnover or €20 million (whichever is greater)'.[32] In the case of both Alphabet (Google's parent company) and Facebook, this fine would add up to more than 1 billion euros. Although violations of ISO standard 9241-210 (and more generally, technological design decisions that have harmful consequences) are a different matter from data and privacy breaches of the sort targeted by GDPR, the precedent of imposing such penalties indicates that an updated GDPR may also ban and punish negligent and malicious technological designs that foster self-radicalization. Just this year (2018), Germany imposed a stringent new law against hate speech on social media, with associated fines of up to €50 million.[33]

A final point: while our focus here has been technological design (and corresponding political interventions to such design), recent work on the social epistemology of the Internet (see, for example, Heersmink 2016; Miller and

[29] See https://www.iso.org/standard/52075.html; for discussion, see Rau 2013 .

[30] See https://www.iso.org/standard/37031.html.

[31] In a nutshell, the 'Black Crime' example used in *Breitbart*'s navigation can be said to go against the specific aspect of 'Appropriateness of content for the target group and tasks', among others. A brief outline of the standard can be found in https://www.iso.org/obp/ui/#iso:std:37031:en.

[32] See http://www.eugdpr.org/the-regulation.html.

[33] See http://www.bbc.com/news/technology-42510868.

Record 2013, 2016) suggests that a further mitigating solution may lie with user education in the form of online intellectual virtue inculcation. Such an approach has been pioneered in recent work by Richard Heersmink (2017), who defends a view according to which traditional epistemic virtues such as 'curiosity, intellectual autonomy, intellectual humility, attentiveness, intellectual carefulness, intellectual thoroughness, open-mindedness, intellectual courage and intellectual tenacity' (2017: 1) may be fruitfully deployed when interacting with search engines. For example, regarding intellectual carefulness, Heersmink writes that the careful user, rather than accepting the first result, will

> keep pursuing her enquiry until she has reached a proper level of understanding. This entails consulting and cross-checking at least a number of different online sources. It may also entail using different search terms such as, for instance, 'Why did the dinosaurs go extinct?'. Or include Boolean operators, for example: 'dinosaurs + extinct−creationism' or 'dinosaurs + extinct + scientific knowledge'.

Such a careful user will be better off epistemically than one who manifests online carelessness. As Heersmink points out, the 'featured snippet' that Google shows if one simply types 'What happened to the dinosaurs' is the following paragraph of misinformation:

> The Bible gives us a framework for explaining dinosaurs in terms of thousands of years of history, including the mystery of when they lived and what happened to them. Dinosaurs are used more than almost anything else to indoctrinate children and adults in the idea of millions of years of earth history. (2016: 66; see also Heersmink 2017)

Heersmink's call for virtue cultivation online is echoed by Shannon Vallor (2016: 10), who argues for the cultivation of 'technomoral virtues explicitly designed to foster human capacities for flourishing with new technologies'. One way to accomplish this is by studying individuals who 'speak up on important and controversial matters of global justice, security, and community civic engagement' (2016: 185) online, despite being attacked by trolls, as their 'technomoral virtues of honesty, courage, civility, and perspective' (2016: 185) could lead to better design of social platforms.

4.2 Bottom-up solutions

Our recommendations for responding to bottom-up technological seduction mirror those for preventing and curbing top-down technological seduction. In the first instance, it is helpful to distinguish between cases of negligence, where the relevant agents can be expected to amend their ways when they receive suitable guidance, and cases of recklessness and malicious intent, where regulation and enforcement are the only way to prevent and stop bad behavior.

To curb inadvertent bottom-up seduction, several solutions are promising. Since this phenomenon often involves the usurpation of people's capacities to engage critically as both questioners and answerers of questions, technology could be designed in such a way that—instead of feeding people both questions and answers—it prompts people to reflect on the questions they are asking or the answers to them. This will be especially important in cases where someone is already one or more steps into a path-dependent inquiry. On the question side, search and recommender systems could be programmed to stop suggesting queries after a small, finite number of previous suggestions — at least until the user has gone through a 'time out' period. On the answer side, systems could be programmed to return a more diverse array of results after a small, finite number of previous answers. Both of these interventions should, in principle, help to reduce the gravitational power of attractors. It is, of course, not guaranteed that this sort of intervention would have stopped the radicalization of someone like Minassian. However, interventions of this sort promise to choke the flow of netizens down the four-step road from, 'I know what you're thinking' to extremism.

Next, in addition to nudging the individual user to avoid problematic attractors, search and recommendation systems could be automatically monitored for the presence and advent of attractors as measured by, for example, a very high ratio between in-degree and out-degree. Since many attractors are evaluatively neutral or even beneficial, a specially trained team of computer scientists, social scientists, and humanists would then be tasked with investigating the basins of attraction identified in this way to determine whether any of them constitute hate-speech or are liable to lead to self-radicalization. Such content could then be de-indexed, pushed lower in people's newsfeeds, and so on.

Solutions along these lines are already being implemented. For instance, users who search for content related to the Islamic State on YouTube are being targeted with content that expresses skepticism about the aims and methods of the Islamic State (Holley 2017; Manjoo 2017). In the same vein, Facebook recently implemented a 'related articles' widget that displays content from independent sources just below trending topics in users' newsfeeds (Su 2017).

While it may be possible to rely on well-intentioned and well-resourced firms to implement such design principles when given the chance, not all firms have sufficient good will and resources. For reckless and malicious actors, regulation with sufficient enforcement power will be needed. For well-intentioned firms that lack the technological skills or the resources to put those skills to use, a taxpayer-funded repository of code-reviewed open-source algorithms that embody best practices may be the optimal solution.

Naturally, our recommendations in this section presuppose normative standards. These standards involve epistemic norms (e.g., conditions for harnessing the wisdom of crowds, which requires among other things decentralization and independence of sources), moral norms (e.g., related to the epistemic conditions for practical rationality), and political norms (e.g., related to the epistemic grounds for democracy). How best to manage the threat of technological seduction will always

be a problem that requires both scientific expertise and humanistic reflection on the values we want our societies to embody.

MARK ALFANO
TU-DELFT & AUSTRALIAN CATHOLIC UNIVERSITY

J. ADAM CARTER
UNIVERSITY OF GLASGOW
adam.carter@glasgow.ac.uk

MARC CHEONG
MONASH UNIVERSITY

References

Abraham, F. D., R. H. Abraham, and C. D. Shaw. (1990) *A Visual Introduction to Dynamical Systems Theory for Psychology*. Santa Cruz, CA: Aerial Press.

Abrahams, Max. (2002) 'What Terrorists Really Want: Terrorist Motives and Counterterrorism Strategy'. *International Security*, 32, 86–89.

Alemanno, Alberto. (2012) 'Nudging Smokers: The Behavioural Turn of Tobacco Risk Regulation'. *European Journal of Risk Regulation*, 3, 32–42.

Alfano, M. (2013) *Character as Moral Fiction*. Cambridge: Cambridge University Press.

Alfano, M. (2018) 'The Nietzschean Precedent for Anti-reflective, Dialogical Agency'. *Behavioral & Brain Sciences*, e37, 12–13.

Alfano, M., and G. Skorburg. (2018) 'Extended Knowledge, the Recognition Heuristic, and Epistemic Injustice'. In J. A. Carter, A. Clark, J. Kallestrup, O. Palermos, and D. Pritchard (eds.), *Extended Knowledge* (New York: Oxford University Press), 239–65.

Alfano, M., A. Higgins, and J. Levernier. (2018) 'Identifying Virtues and Values through Obituary Data-mining'. *Journal of Value Inquiry*, 52, 59–79.

Beer, R. D. (1995) 'A Dynamical Systems Perspective on Agent-environment Interaction'. *Artificial Intelligence*, 72, 173–215.

Benkler, Yochai, Robert Faris, Hal Roberts, and Ethan Zuckerman. (2017) 'Study: Breitbart-led Right-wing Media Ecosystem Altered Broader Media Agenda'. *Columbia Journalism Review*, March 3. https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php.

Berman, Mark. (2016) 'Prosecutors Say Dylann Roof "Self-Radicalized" Online, Wrote Another Manifesto in Jail'. *Washington Post*, August 22. https://www.washingtonpost.com/news/post-nation/wp/2016/08/22/prosecutors-say-accused-charleston-church-gunman-self-radicalized-online/?utm_term=.8a561f1dcb24.

Bisantz, A. M., C. M. Burns, and R. J. Fairbanks, eds. (2014) *Cognitive Systems Engineering in Health Care*. Boca Raton, FL: CRC Press.

Bjelopera, Jerome P. (2010) 'American Jihadist Terrorism: Combating a Complex Threat'. CRS Report for Congress. https://fas.org/sgp/crs/terror/R41416.pdf.

Byrne, Joanne, and Marc Cheong. (2017) 'The Algorithm as Human: A Cross-disciplinary Discussion of Anthropology in an Increasingly Data-driven World'. Talk given at the AAS/ASA/ASAANZ Shifting States Anthropology Conference. University of Adelaide.

Card, Stuart K., Thomas P. Moran, and Allen Newell. (1980) 'The Keystroke-level Model for User Performance Time with Interactive Systems'. *Communications of the ACM*, 23, 396–410.

Carter, J. A. (2017) 'Intellectual Autonomy, Epistemic Dependence and Cognitive Enhancement'. *Synthese*. doi: 10.1007/s11229-017-1549-y.

Conly, Sarah. (2004) 'Seduction, Rape, and Coercion'. *Ethics*, 115, 96–121.

Dayan, Eran, and Maya Bar-Hillel. (2011) 'Nudge to Nobesity II: Menu Positions Influence Food Orders'. *Judgment and Decision Making*, 6, 333–42.

Dillahunt, T. R., C. A. Brooks, and S. Gulati. (2015) 'Detecting and Visualizing Filter Bubbles in Google and Bing'. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (New York: Association for Computing Machinery), 1851–56.

Doris, John. (2015) *Talking to Ourselves: Reflection, Ignorance, and Agency*. New York: Oxford University Press.

Feldman, Brian. (2018) 'Posts Indicate Suspect in Toronto Attack Was Radicalized on 4chan'. *New York Magazine*, April 24. http://nymag.com/selectall/2018/04/suspect-in-toronto-posts-about-incel-rebellion.html.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. (2016) 'Filter Bubbles, Echo Chambers, and Online News Consumption'. *Public Opinion Quarterly*, 80, 298–320.

Floridi, Luciano, and Jeff W. Sanders. (2014) 'On the Morality of Artificial Agents'. *Minds and Machines*, 14, 349–79.

Forrester, John. (1990) *The Seductions of Psychoanalysis: Freud, Lacan, and Derrida*. New York: Cambridge University Press.

Harris, Tristan. (2014) 'How Technology Hijacks People's Minds'. http://www.tristanharris.com/essays/.

Heersmink, Richard. (2016) 'The Internet, Cognitive Enhancement, and the Values of Cognition'. *Minds and Machines*, 26, 389–407.

Heersmink, Richard. (2017) 'A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues, and Education'. *Social Epistemology*. doi: 10.1080/02691728.2017.1383530.

Holley, P. (2017) 'YouTube is Tricking People who Search for ISIS Videos'. *Washington Post*, July 24. https://www.washingtonpost.com/news/innovations/wp/2017/07/24/youtube-is-tricking-people-who-search-for-isis-videos/.

Hollingshead, A. B. (1998) 'Retrieval Processes in Transactive memory Mystems'. *Journal of Personality and Social Psychology*, 74, 659.

Hollnagel, E., and D. D. Woods. (1983) 'Cognitive Systems Engineering: New Wine in New Bottles'. *International Journal of Man-Machine Studies*, 18, 583–600.

Hollnagel, E., and D. D. Woods. (2005) *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: CRC Press.

Holroyd, Jules, and Joseph Sweetman. (2015) 'The Heterogeneity of Implicit Bias'. In Michael Brownstein and Jennifer Saul (eds.), *Implicit Bias and Philosophy* (Oxford: Oxford University Press), 80–103.

Kahneman, Daniel. (2011) *Thinking, Fast and Slow*. London: Macmillan.

Kant, Immanuel. ([1781] 1958) *Critique of Pure Reason*. Translated by Norman Kemp Smith. London: Macmillan.

Koralus, P., and M. Alfano. (2017) 'Reasons-based Moral Judgment and the Erotetic Theory'. In J.-F. Bonnefon and B. Trémolière (eds.), *Moral Inferences* (New York: Routledge), 77–106.

Koralus, P., and S. Mascarenhas. (2013) 'The Erotetic Theory of Reasoning: Bridges between Formal Semantics and the Psychology of Propositional Deductive Inference.' *Philosophical Perspectives*, 27, 312–65.

Levy, N. (2017) 'The Bad News about Fake News'. *Social Epistemology Review and Reply Collective*, 6, 20–36.

Lewis, P. (2018) '"Fiction is Outperforming Reality": How YouTube's Algorithm Distorts Truth'. *The Guardian*, February 2. https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

Majeed, Khader, Loo Seng Neo, Gabriel Ong, Eunice Tan Mingyi, and Jeffrey Chin, eds. (2016) *Combating Violent Extremism and Radicalization in the Digital Era*. Hershey, PA: IGI Global.

Manjoo, F. (2017) 'A Hunt for Ways to Combat Online Radicalization'. *New York Times*, August 23. https://www.nytimes.com/2017/08/23/technology/a-hunt-for-ways-to-disrupt-the-work-of-online-radicalization.html.

McCombs, Maxwell. (2000) 'The Agenda-Setting Role of the Mass Media in the Shaping of Public Opinion'. Unpublished manuscript. https://www.infoamerica.org/documentos_pdf/mccombs01.pdf.

Miller, Boaz, and Isaac Record. (2013) 'Justified Belief in a Digital Age: On the Epistemic Implications of Secret Internet Technologies'. *Episteme*, 10, 117–134.
Miller, Boaz, and Isaac Record. (2016) 'Responsible Epistemic Technologies: A Social-Epistemological Analysis of Autocompleted Web Search'. *New Media & Society*. doi:10.1177/1461444816644805.
Ndili, Alex. (2016) 'The Role Of Community Policing In The Prevention Of Youth Radicalization in the Eastleigh Area of Nairobi County'. University of Nairobi. http://erepository.uonbi.ac.ke:8080/xmlui/handle/11295/100166.
Neumann, Peter R. (2013) 'Options and Strategies for Countering Online Radicalization in the United States'. *Studies in Conflict & Terrorism*, 36, 431–59.
Newton, C. (2017) 'How YouTube Perfected the Feed'. *The Verge*, August 30. https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-recommendation-personalized-feed.
Omotoyinbo, Femi Richard. (2014) 'Online Radicalization: The Net or the Netizen?' *Socialnès Technologijos*, 1, 51–61.
Palermos, S. O. (2016) 'The Dynamics of Group Cognition'. *Minds and Machines*, 26, 409–40.
Pariser, Eli. (2011) *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin.
Peterson, J. (2018) *12 Rules for Life: An Antidote to Chaos*. New York: Random House.
Pettit, P. (2015) *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect*. Oxford: Oxford University Press.
Pilkington, Ed, and Dan Roberts. (2016) 'FBI and Obama Confirm Omar Mateen Was Radicalized on the Internet'. *The Guardian*, June 14. https://www.theguardian.com/us-news/2016/jun/13/pulse-nightclub-attack-shooter-radicalized-internet-orlando.
Precht, Tomas. (2007) 'Home Grown Terrorism and Islamist Radicalisation in Europe: From Conversion to Terrorism'. http://www.justitsministeriet.dk/sites/default/files/media/Arbejdsomraader/Forskning/Forskningspuljen/2011/2007/Home_grown_terrorism_and_Islamist_radicalisation_in_Europe_-_an_assessment_of_influencing_factors__2_.pdf.
Ritter, F. E., G. D. Baxter, and E. F. Churchill. (2014) 'User-centered Systems Design: A Brief History. In *Foundations for Designing User-centered Systems* (London: Springer), 33–54.
Roberts, Robert, and Jay Wood. (2007) *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.
Rogers, Y., H. Sharp, and J. Preece. (2012) *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley.
Silber, Mitchell D., and Arvin Bhatt. (2007) 'Radicalization in the West: The Homegrown Threat'. https://sethgodin.typepad.com/seths_blog/files/NYPD_Report-Radicalization_in_the_West.pdf.
Strohminger, Nina, and Shaun Nichols. (2014) 'The Essential Moral Self'. *Cognition*, 131, 159–71.
Su., S. (2017). 'New Test with Related Articles'. *Facebook Newsroom*, April 25. https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/.
Sunstein, Cass R. (2011) *Going to Extremes: How Like Minds Unite and Divide*. Oxford: Oxford University Press.
Sunstein, Cass R. (2015) 'The Ethics of Nudging'. *Yale Journal on Regulation*, 32(2), 413–15.
Sunstein, Cass R. (2017) *Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.
Tanesini, A. (2016) 'Teaching Virtue: Changing Attitudes'. *Logos & Episteme*, 7, 503–27.
Thaler, Richard H., and Cass R. Sunstein. (2003a) 'Libertarian Paternalism is not an Oxymoron'. *The University of Chicago Law Review*, 70, 1159–1202.
Thaler, Richard, and Cass Sunstein. (2003b) 'Libertarian Paternalism'. *The American Economic Review*, 93, 175–79.
Thaler, Richard H., and Shlomo Benartzi. (2004) 'Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving'. *Journal of Political Economy*, 112, S164–S187.
Thaler, Richard H., and Cass R. Sunstein. (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Thaler, Richard H., Cass R. Sunstein, and John P. Balz. (2014) 'Choice Architecture'. In Eldar Shafir (ed.), *The Behavioral Foundations of Public Policy* (Princeton, NJ: Princeton University Press), 428–39.

Troumbley, Rex. (2015) 'Coercive Cyberspaces and Governing Internet Futures'. In Jenifer Winter and Ryota Ono (eds.), *The Future Internet: Alternative Visions* (Cham: Springer International Publishing), 17–40.

Tufekci, Zeynep. (2018) 'YouTube, the Great Radicalizer'. *New York Times*, March 10. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Vallor, Shannon. (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Van Hoboken, J. V. J. (2012) *Search Engine Freedom: On the Implications of the Right to Freedom of Expression for the Legal Governance of Search Engines*. Alphen aan den Rijn, Netherlands: Kluwer Law International.

Vinokur, Amiram, and Eugene Burnstein. (1978) 'Depolarization of Attitudes in Groups.' *Journal of Personality and Social Psychology*, 36, 872.

Von Behr, Ines, Anaïs Reding, Charlie Edwards, and Luke Gribbon. (2013) 'Radicalisation in the Digital Era: The Use of the Internet in 15 Cases of Terrorism and Extremism'. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=266610.

Wegner, D. M. (1987) 'Transactive Memory: A Contemporary Analysis of the Group Mind'. In Brian Mullen and George R Goethals (eds.), *Theories of Group Behavior* (New York: Springer), 185–208.

Wegner, D. M., R. Erber, and P. Raymond. (1991) 'Transactive Memory in Close Relationships'. *Journal of Personality and Social Psychology*, 61, 923.

Weimann, Gabriel. (2012) 'Lone Wolves in Cyberspace'. *Journal of Terrorism Research*, 3. doi:10.15664/jtr.405.

White, Mark. (2013) *The Manipulation of Choice: Ethics and Libertarian Paternalism*. Dordrecht: Springer.

Winner, Langdon. (1986) *The Whale and the Reactor*. Chicago: University of Chicago Press.

Wong, David. (2006) *Natural Moralities*. Oxford: Oxford University Press.

Zagzebski, L. (2017) *Exemplarist Moral Theory*. Oxford: Oxford University Press.