The Duty to Promote Digital Minimalism in Group Agents

Chapter 7 in *Kantian Ethics and the Attention Economy: Duty and Distraction* (Palgrave)

Timothy Aylsworth
Florida International University
taylswor@fiu.edu
ORCID iD: 0000-0001-8164-3451

Clinton Castro
University of Wisconsin
cgcastro@wisc.edu
ORCID iD: 0000-0003-4740-0055

[Human beings'] propensity to enter into society … is combined with a thoroughgoing resistance that constantly threatens to break up this society. The predisposition for this obviously lies in human nature … Now it is this resistance that awakens all the powers of the human being … to obtain for himself a rank among his fellows, whom he cannot *stand*, but also cannot *leave alone*.
Kant, Idea 8:20-21

"Our algorithms exploit the human brain's attraction to divisiveness."
Facebook, internal memo to senior executives[1]

Abstract. In this chapter, we turn our attention to the effects of the attention economy on our ability to act autonomously *as a group*. We begin by clarifying which sorts of groups we are concerned with, which are *structured groups* (groups sufficiently organized that it makes sense to attribute agency *to the group itself*). Drawing on recent work by Purves and Davis (2022), we describe the essential roles of *trust* (i.e., depending on groups to fulfill their commitments) and *trustworthiness* (i.e., the property of a group that makes trusting them fitting) in autonomous group action, with particular emphasis on democratic institutions (which we view as group agents) and democratic legitimacy (which depends on trust and trustworthiness). We then explain how engagement maximization promotes polarization, which is detrimental to trust and trustworthiness and, in turn, democratic legitimacy and democratic institutions. We close by considering what groups might do to protect themselves from the threat posed to them by the attention economy.

7.0 Introduction

So far, we have canvassed the moral reasons we have to restructure our relationship with technology in virtue of the effects it has on us as individuals. But if we restrict our focus to the ways that technology can harm us as individuals, we overlook some morally significant effects for groups. In this chapter, we argue that addictive technology weakens our capacity to act autonomously *as a group.* We defend this claim by arguing that the certain features of the attention economy (e.g., that it contributes to polarization[2]) threaten to undermine the legitimacy of political institutions.

We begin by explaining what is distinctive about group-level harms. In short, these are harms that cannot be fully explained without reference to a group agent. There are many cases where a harm to a group is both constituted by and reducible to harms to individuals. For instance, a group of pensioners might be harmed by embezzlement because the pension fund balance is reduced. This is not a genuinely group-level harm, since there is no harm over and above harms to the individuals (e.g. each pensioner receives a smaller disbursement). Group-level harms, on the other hand, involve harms to structured groups. Insofar as a group

[1] Quoted in Orlowski (2020).
[2] See Rathje et al. (2021).

collectively pursues goals and acts on the basis of shared intentions,[3] they are vulnerable to being harmed in ways that undermine the group's capacity to achieve those ends.

We draw on this account of group-level harms to demonstrate how the legitimacy of democratic institutions is threatened by the attention economy. We argue that legitimacy is partly a function of citizens' trust in institutions, as well as those institutions' competencies in the relevant domain, sensitivity to citizens' needs, and ability to signal competence and sensitivity to citizens' needs.[4] But the corrosive effects of fake news, polarization, echo chambers, and a motley of other features of the attention economy place a drag on all of these factors.[5] Thus, the attention economy not only harms us individually. It harms us collectively as well.

7.1 Group Autonomy and Group Harms

To better understand the kind of group-level harms at issue here, it is helpful to contrast them with related (but importantly distinct) harms. For instance, there are situations where a harm to an individual is thought to be a harm to the group of which that individual is a member. This is commonly pointed out in cases of racial injustice and genocide. When an individual is targeted for violence because of their racial identity, it is often said that this act harms the racial group *as a group*. The Nazi's anti-Semitism is an obvious example. Acts of violence against individual Jews on Kristallnacht harmed not only those in Germany and Austria, who were most immediately affected; it harmed European Jews as a group insofar as it put them all in danger. This notion of group harms could be seen as the underpinning of Martin Luther King's famous claim that "Injustice anywhere is a threat to justice everywhere" or the workers' slogan: "An injury to one is an injury to all."[6] A similar concept is invoked in justifications of international criminal law when actions such genocide are characterized as "crimes against humanity."[7] This issue also arises in bioethics, as practices like genetics research are frequently discussed in terms of risks to groups or communities.[8]

---

[3] We discuss various accounts of group agents in section 2. For the most part, our account of group agents is similar to List and Pettit (2011).

[4] Our argument in this section draws on Purves and Davis (2022).

[5] See Rubel et al. (2021) and Pham et al. (2022).

[6] The famous quote about injustice comes from Martin Luther King Jr., "Letter from Birmingham Jail." The "injury to all" slogan is used by the Industrial Workers of the World and has been attributed to David C. Coates. See Haywood (1929, 186).

[7] The idea here is that the harm of an action like genocide extends beyond national borders because it harms all of humanity. This idea is often deployed in the justification of humanitarian interventions and international criminal punishments. Larry May argues that an act qualifies as a crime against humanity when it targets people for their group affiliation (e.g. ethnicity, religion, etc.) rather than some property that is unique to the individual (87). May claims that such acts demonstrate a "callous disregard for the individuality of the person" and that this constitutes "an assault on what is common to all humans and hence to all of humanity" (84). See May (2000). In addition to May's philosophical justification, the concept of "crimes against humanity" has been codified into international law. The 1998 Rome Statute, which established the International Criminal Court (ICC) requires that all parties to the Statute recognize "that such grave crimes threaten the peace, security and well-being of the world." It outlines a list of such crimes in Article 7.

Others have criticized this justification. Andrew Altman and Christopher Wellman argue that "harm to humanity is a convenient but ultimately unpersuasive fiction" Altman and Wellman (2004, 42).

[8] See Davis (2000) and Weijer et al. (1999).

Although group affiliations play an important role in examining the moral implications of things like racial injustice and genocide, they do not harm a group *qua* group. They are what Dan Hausman (2007) calls "group-mediated harms." In situations of this kind, the harms are suffered by individuals, even though the individuals are targeted because of their membership in a group. Consider the case of a person whose application for a federally insured mortgage is rejected because he is Black. In a scenario like this, the harm is mediated by his group membership (it happens to him in virtue of his race), but it is the *individual* who was wronged, not the group. His interest in securing this particular loan is not one that Black Americans hold collectively. It is not a purely collective interest that was disrespected, but an individual one.

By contrast, there are many cases where a group has shared interests that it pursues collectively. Groups that pursue interests through shared agency are susceptible to being harmed in a distinctive way. Corporations, nation states, and sports teams have interests of this kind. For instance, if their star player is hit by a bus, Manchester United will be less likely to win the European Cup.[9] Not only is the player harmed as an individual, the club is harmed as a group. Winning the cup is a shared interest that no player on the team would be capable of pursuing on his own. In cases of group-mediated harms, when we say that the group is harmed, this simply amounts to the claim that individual members of the group have become more susceptible to a certain kind of harm (e.g., individual Jews were subject to violence during the Holocaust, individual Black Americans discriminated against, etc.). Group-level harms, by contrast, must involve groups that have collective interests, aims, or intentions.

Appiah (2011) draws a similar distinction when discussing group rights.[10] In his view, "collective rights" are ones that are exercised by groups, whereas "membership rights" are rights that individuals have in virtue of their membership in a group. He says, for example, that a democratic state has a collective right to self-determination. A right of this kind can be held only by a group, since it is not a right that any individual citizen is capable of exercising on her own (2011, 268).[11] By contrast, a person's right to vote in the American presidential election is a membership right that she has in virtue of her U.S. citizenship. The right is mediated by her membership in the group, but it is not a right that is held by a group.

---

[9] Not every harm to a member of the team is necessarily a harm to the team itself. It depends on the player's contribution to the team's collective pursuits. Hausman writes, "An injury to a star player, a crucial executive, or an important leader is also an injury to the team, corporation, or tribe. This is not because any injury to any member of a structured group is automatically an injury to the group. If some of the least able players on a baseball team could be replaced with equally good players from the team's farm club, then an injury to them would not be an injury to the team" Hausman (2007, 357).

[10] In order to avoid the controversy surrounding group rights, we have chosen to discuss the issue here in terms of group harms. We can have obligations to groups even if groups do not have rights. Appiah defends the notion of group rights in the article, but he is responding to those who are critical of group rights. The standard objection to group rights is that group agents are not the kinds of entities that can bear rights and that all putative cases of group rights ultimately dissolve into individual rights. See, for example, Narveson (1991). Appiah is responding most directly to James Sterba's skepticism: "Moral entitlements are not held by groups … Rights are possessed by persons. As when persons are entitled to be made whole for some injury earlier done to them, the duty owed is … to them as individuals." (Sterba 2009, 57–58).

[11] See also Appiah (2005a) and (2005b).

Given the controversial status of group rights in the philosophical literature, we tend to frame our argument here in terms of group harms. The argument of this chapter does not depend on the existence of group rights. In our view, we have obligations to respect the *interests* of certain groups even if groups do not have *rights*. To push the claim further, we believe such obligations would not be undermined even by the claim that groups have no moral status whatsoever—a rejection that is stronger than denying group rights. It is not uncommon for philosophers to defend thoroughgoing individualism about moral status; they argue that our moral concerns should be limited exclusively to individuals.[12] Even on this view, it is still possible to make sense of the claim that we have prima facie obligations to respect the interests of certain groups. The interests of groups would have moral significance whenever the group's ability to pursue its collective interests is something that matters to individuals.[13] Even if Manchester United has no intrinsic moral status as a group, undermining the club's capacity to win the European cup would run afoul of the interests of many individuals (the players, the coaches, the fans, etc.). Of course, not all instances of causing harm should count as moral wrongs. If Liverpool FC were to sign a contract with Manchester's star player, this would harm Manchester United (or, for the individualist, we could say that it harms those to whom the club matters), but it would not necessarily be a moral wrong. In order for a harm of this kind to constitute a moral wrong, we must have a sufficient reason to respect the interests that are at stake.

When the Acme Corporation creates a product that is superior to and cheaper than Biffco's equivalent product, there is no question that this harms Biffco. But, other things being equal, this does not seem to be a moral wrong. Even though Acme's action goes against the collective interests of Biffco (and against the interests of individual shareholders, employees, etc.), there is no reason to think that Acme has an obligation to respect those interests. The moral evaluation seems quite different, however, when we consider a case of one state interfering with another state's capacity for self-government. If Arendelle were to undermine a democratic election in Blefuscu, this would rightly be seen as a moral wrong. Arendelle and its citizens have a moral obligation to respect Blefuscu's capacity for self-determination.[14] We do not necessarily

[12] There are many examples of this kind of individualism. See the discussion of Sterba and Narveson above. Similarly, Thomas Pogge claims that "the ultimate units of concern are human beings, or persons—rather than, say, family lines, tribes, ethnic, cultural, or religious communities, nations, or states." (Pogge 1992, p. 48).

[13] As noted in earlier chapters, according to Kant, beneficence is an imperfect duty to promote the ends of others as if they were our own. But there is an important qualification; we are obligated to promote only their morally permissible ends. When we extend this idea of beneficence to groups, we can mitigate the worry that we are under an obligation to promote the ends or respect the interests of groups that are pursuing morally impermissible goals.

[14] Although this claim might be intuitively obvious to most readers, providing a full justification of it would be a difficult philosophical task and one that lies outside of the scope of this paper. Perhaps the simplest route would be to point to certain historical realities. It is widely accepted that states have this right. Formal recognitions of a nation's right to sovereignty are least as old as the Treaty of Westphalia (1648). The right to self-determination was enshrined in Article 1 of the United Nations Charter, which states that one of the purposes of the UN is "To develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples, and to take other appropriate measures to strengthen universal peace." Self-determination was also acknowledged as a "right" in the Atlantic Charter signed by the US and the UK.

have obligations to respect the interests of sports teams and corporations (though we might in certain cases),[15] but we almost always have obligations to respect a democratic state's interest in self-determination.[16]

We will develop our account of democratic legitimacy more fully in section 3, but before moving on, we should provide one final clarification about the nature of group-level harms. Thus far, we have specified only a single condition of what makes a group susceptible to group harms: a collectively held interest. More needs to be said, however, about what is distinctive about collective interests. Someone might object that the putative cases of group-mediated harms, which we discussed above, actually *do* involve shared interests. It could be argued that Black Americans shared an interest in putting an end to housing discrimination; European Jews shared an interest in the downfall of the Nazi party, etc. What distinguishes such groups from "structured"[17] groups, however, is that racial and ethnic groups as such lack the structure to pursue their shared interests collectively. They do not act *as a group*. In addition to having shared interests, corporations, states, and sports teams are as such organized in such a way that they can pursue their interests collectively. In short, they are group agents.

There are several competing accounts of group agency, and our argument in this paper is not necessarily committed to a particular view.[18] At a minimum, agents must have the capacity to represent states of affairs in their environment, motivational states, and the ability to intervene in the environment in accordance with their motivations (List and Pettit 2011, 20). In order for a group to qualify as an agent, we must be able to attribute some form of representational states, motivations, and capacities to the group. Some critics worry that this requires us to posit something ontologically extravagant like a "group mind," but most accounts in the literature are, in some sense, reductionist.[19] We can attribute representational states and motivations to groups

---

Skeptics might argue, however, that an existing social contract is not a sufficient foundation for a moral right. Another approach would be to ground a state's right to self-determination in the rights of individuals. For instance, one could argue for a Lockean view according to which the democratic legitimacy of the state is necessary to protect the natural rights of its citizens. For a more thorough defense of the right to self-determination as a concept of international law, see Margalit and Raz (1990).

[15] For instance, the CEO of a corporation may have an obligation to promote the company's interests in virtue of commitments she has made. The manager of the football club has a similar obligation.

[16] The justifiability of humanitarian interventions is controversial. See Walzer (1977), ch 6. It can be argued, in extreme cases, that a state loses its claim to sovereignty when it fails to provide for the basic security of its citizens. In such a case, the foreign power who intervenes might argue that they are not interfering with the affairs of a sovereign state. Such interventions ought to be rare, however. The history of colonialism is one important reason why international law promotes reluctance to intervene. Tasioulas and Verdirame write: "By embracing this notion [of self-determination] international law gave recognition to the moral and political value of self-government, accepting that people prefer to be ruled by their own bad rulers rather than foreigners, including those foreigners with some claim to greater competence" (2022).

[17] Hausman (2007) uses the term "structured" groups to refer to groups that are susceptible to harm. For reasons that we explain below, we prefer to frame the discussion in terms of "group agents."

[18] See List and Pettit (2011), Ludwig (2016), and Tuomela (2013).

[19] John Searle says that talk of group minds is "at best mysterious and at worst incoherent. Most empirically minded philosophers think that such phenomena must reduce to individual intentionality" (1990, 404). All three of the accounts mentioned above are ontologically reductionist. They do not posit an independently existing "group mind." Group agents are nothing more than sets of individuals acting in a particular way. List and Pettit argue that

insofar as these are defined functionally; groups are capable of jointly engaging in rational deliberation, for instance. There is nothing mysterious about this kind of group deliberation; it happens all the time in department meetings and boardrooms. The department's conclusion that it would be good to hire a new faculty member is a representational state that can be attributed to the group. The same can be done with intentions. It seems perfectly sensible to say that the orchestra intentionally played Beethoven's 9th symphony. But the intention to play the symphony cannot be attributed to any particular individual.[20] A flutist in the orchestra is not capable of playing Beethoven's 9th Symphony by herself, though she is certainly capable of doing her part.[21] Talking about groups in this way (i.e., in terms of shared beliefs, desires, and intentions) makes it possible to attribute states to them that are characteristic of rational agency. And to the extent that groups are indeed capable of rational agency, they can be said to have the capacity for autonomy. Understood in this way, one way to harm a group *qua* group would be to violate or undermine its autonomy. The racial and ethnic groups mentioned above, which are susceptible only to group-mediated harms, are not group agents (though advocacy organizations such as the NAACP and the Shoah Foundation are).

Insofar as groups are capable of acting as group agents, we can think of them as either possessing or lacking autonomy. Groups can have or lack the capacity to set and pursue their own ends. The rational agency of groups can be exercised well or it can function poorly. Structured groups, such as sports teams and corporations, typically have an interest in functioning well. Undermining the capacities of a group agent can therefore be understood as a harm to the group itself. You could harm a group agent by undermining its autonomy (its capacity to set and pursue its own ends) in ways that are perfectly analogous to ways that you could harm an individual agent. You could undermine the group agent's capacities or its

---

they can reject eliminativism about group agents without rejecting methodological individualism—the claim that all explanations of the social world ultimately boil down to facts about individuals (2011, 3-6).

[20] A claim of this kind is defended by Velleman (1997); he argues against those like Bratman and Searle who hold views of group agency that do not involve genuinely shared intentions. On their view, group agency is nothing more than individuals who intend to coordinate their actions and intentions with other individuals. Velleman's view is closer to that of Margaret Gilbert (1990), who talks about a "plural subject" brought about by a "pool of wills." Velleman's point, which we concur with, is that it would not make sense to talk about an individual intending to perform the symphony since it is not clear that an agent can intend to do something that she is not capable of doing herself. For Bratman, an individual has a "we-intention" whose content may be something like this: I intend that *we* play the symphony.

[21] On fully reductionist views, the collective intention is to be understood simply in terms of a set of individual intentions (see, for example, Bratman 1992, 1993). Each individual merely intends to do her part and to "mesh" her sub-plans with other members of the group.

authenticity.[22] In some cases, harming a group in this way constitutes a moral wrong. As we suggested earlier, one of the clearest examples of this is democratic legitimacy.

There are at least two ways of understanding the nature of this moral wrong. Some might believe that the group has what Appiah calls a "collective right" to self-determination, and such an account would make the role of the group agent ineliminable. Others might reject the notion of group rights and argue that only individuals have been wronged. When a state's democratic legitimacy is undermined, this wrongs the individual citizens who have an interest in their state's legitimacy or in its ability to function properly. We prefer the view that some group agents have moral status and that some moral wrongs should be understood, at least partially, in terms of harms to groups. But we believe that the main conclusion of this section (i.e.., that there is a moral obligation to respect the autonomy of certain groups) does not depend on this view. The argument is open to the reductionist view as well.

Even according to the individualistic account, the moral wrong cannot be fully understood without reference to the group agent. The only way to make sense of the claim that individuals were wronged is to point out that they had an interest in the group agent's autonomy. So it does not ultimately matter whether the group agent's autonomy has intrinsic or instrumental moral value; either way, there are cases where it is morally wrong to undermine a group agent's autonomy. The important conclusion here is that some groups have shared interests, which they pursue collectively, and there are situations where it would be morally wrong to undermine the group's ability to pursue its ends. In the next section, we demonstrate what this would look like by considering the group agency of the state.

Kant was particularly concerned with the proper functioning of the state. As we saw in the previous chapter, Kant believed that it would be impossible for groups of human beings to enjoy freedom without creating a state to safeguard their rights. He also worries that the threat of violence would always loom in the absence of a state. In a footnote from *Toward Perpetual Peace,* he writes,

> Within each state it [human malevolence] is veiled by the coercion of civil laws, for the citizens' inclination to violence against each one another is powerfully counteracted by a greater force, namely that of the government, and so not only does this give the whole a moral veneer (*causae non causae*) but also, by its

[22] For example, Blefuscu's army is a group agent. And as we saw in chapter 2, autonomous capacities require the ability to form coherent plans, revise those plans in light of new information, etc. If a spy from Arendelle were to disrupt communication between the units of Blefuscu's army, that would undermine the group's baseline capacities. It might be helpful to use Bratman's (1993) model of group agency to clarify this example. On Bratman's view, the members of a group agent can jointly intend an action only if the agents intend to mesh their sub-plans. Rosa and Ray can be said to share the intention to paint the house only if they intend to mesh their sub-plans (such as the sub-plans involving color choice). If Rosa is painting the house green, and Ray is painting it red, then we must reject the claim that they are painting the house together. Shared agency requires means-end coherence. If the spy successfully disrupts communication between units of Blefuscu's army, then she would undermine that group agent's autonomy by making it impossible for them to mesh sub-plans. Similarly, the spy could interfere with the Blefuscu army's authenticity. She might disseminate false information that leads the group to attack the wrong targets. This kind of manipulation undermines the group's autonomy in much the same way that Iago undermined Othello's autonomy by means of deceit.

> checking the outbreak of unlawful inclinations, the development of the moral predisposition to immediate respect for right is actually greatly facilitated" (8:375-76).

As we will see in what follows, the attention economy poses a variety of threats to this stability. Kant's moral philosophy gives us compelling moral reasons to be concerned about the proper functioning of the state and the capacity of groups to set and pursue their own ends.

7.2 Trust, Trustworthiness, and Democratic Legitimacy

We'll now show how trust and trustworthiness play an important role in preserving a democratic state's legitimacy. Many of the state's aims can be accomplished only if those who are subject to its rule see it as trustworthy. This means that undermining public trust (and/or the state's trustworthiness) hinders the state's ability to fulfill its responsibilities. For instance, criminal justice systems depend crucially on citizens' willingness to serve on juries, testify as witnesses, cooperate with the police, etc. And citizens are far less willing to cooperate when they do not trust the institutions or the individuals charged with carrying out these tasks.[23] We will argue that states are vulnerable to group-level harms insofar as their group agency is susceptible to being undermined by threats to public trust.

As we suggested above, respecting a democratic state's right to self-determination is one of the most compelling examples of an obligation to respect group autonomy.[24] In the absence of a compelling reason to intervene, it is typically seen as morally wrong to interfere with a state's ability to govern itself. This right to self-determination can be seen as having both an external dimension, which involves being free from external control (e.g., colonialism, puppet governments, etc.), and an internal dimension which concerns "the rights of people to pursue democratic governance domestically" (Tasioulas and Verdirame 2022).[25] When dealing with

[23] See Purves and Davis (2022) for an interesting discussion of how algorithmic opacity threatens the legitimacy of criminal justice institutions by eroding the public's trust. For a compelling example of how declining public trust undermines institutions of criminal justice, see Fenton (2021). He discusses how years of corruption and violence had made it nearly impossible for the Baltimore police to get citizens to cooperate: "While the police department leadership begged citizens to cooperate, some of its elite officers were running roughshod on Black men in poor neighborhoods, creating a free-fire zone for anyone seeking to exploit them" (2021, 269).

[24] Many of the most complicated debates in the literature on self-determination involve questions of secession and the formation of new states. It is notoriously difficult to pin down exactly when a nation, people, or ethnic group has a legitimate claim to secede and form an independent state. See Crawford (2006) for an extensive discussion. Crawford begins by acknowledging the widely held view that the formation of new states is "a matter of fact, not of law" (2006, 4). For this reason, we tend to speak in terms of a *state's* right rather than a nation's right or a people's right.

[25] Tasioulas and Verdirame cite Franck (1992) for a defense of the importance of the internal dimension. Franck writes, "Since self-determination is the oldest aspect of the democratic entitlement, its pedigree is the best established. Self-determination postulates the right of a people organized in an established territory to determine its collective political destiny in a democratic fashion and is therefore at the core of the democratic entitlement" (1992, 52)

democratic states, whose rightful authority derives from the consent of the governed, this issue is often framed in terms of "legitimacy."

Philosophers tend to discuss legitimacy in normative terms; they ask questions about the justificatory grounds of a state's claim to exercise authority and use coercive power. But it is important not to lose sight of descriptive legitimacy, which concerns citizens' attitudes and beliefs about the institutions that govern them.[26] Simply put, a state is legitimate, in the descriptive sense, just in case the citizens *believe* that the state is legitimate.[27] This kind of legitimacy is important because a state would be unable to function properly if citizens see the state as illegitimate and thus do not comply with its directives. This means that maintaining descriptive legitimacy requires some amount of public trust. Citizens will see their government as legitimate only if they trust state institutions.

7.2.1 Trust and Trustworthiness

Given the abundant philosophical literature on the nature of trust, it may seem difficult to determine what exactly is distinctive about trust and trustworthiness. But many of the accounts share certain basic features. Following Annette Baier (1986), it has become common to begin the discussion by recognizing that trust involves accepting one's vulnerability to another person.[28] If you trust your friend to watch your laptop while you go to the bathroom, you must accept the fact that this makes you vulnerable to certain risks. She could steal your laptop or deliberately pour coffee on it. If you trust her, you are *relying* on her to refrain from betraying you in those ways. What is more, if she were to betray you, it would be appropriate for you to feel disappointment or resentment; you could rightfully demand an apology. To some, the aptness of these reactive attitudes is one of the things that makes trust distinct from "mere reliance."[29] If you are relying on your car to get me to work and it breaks down, it would be inappropriate to feel betrayed by the car.[30]

---

[26] Our argument here (as well as this way of framing legitimacy) is heavily indebted to the discussion in Purves and Davis (2022).

[27] As Tom Tyler puts it, legitimacy is "the belief that authorities, institutions, and social arrangements are appropriate, proper, and just" (Tyler 2006, 376). Discussions of descriptive legitimacy are usually traced back to Weber. See Peter (2017) for an overview.

[28] See Baier (1986, 235).

[29] Pinning down exactly what distinguishes trust from reliance is the source of most of the controversy in the literature. Baier (1986) argues that trust involves relying on the trustee's goodwill toward the trustor. Jones (1999) also provides an account in which the trustee must be motivated by goodwill. Hardin (2002), on the other hand, offers a self-interest account. When you trust your tax preparer to do your taxes, you might not believe that he acts out of goodwill for you; you may think only that your interests converge insofar as it is profitable for him to maintain a relationship with you. Hawley (2014) criticizes motives-based accounts of trust, like Hardin's and Jones's, on the grounds that they cannot easily make sense of distrust. As we discuss below, Hawley defines trust as relying on someone to fulfill a commitment.

[30] See Hawley (2014). Most accounts of trust use the fittingness of "betrayal" to distinguish reliance from trust from mere reliance. Nguyen (forthcoming) is one of the few exceptions to this consensus. He argues that we might feel betrayed by an object (like a smartphone that fails to give us a calendar reminder that we were relying on), but this kind of "betrayal" is not exactly identical to the kind of betrayal we experience when we are let down by an agent. Even if we feel betrayed by the artifact, we do not ascribe moral notions of blame toward it.

Karen Jones (2012) expands the discussion of trust by analyzing the related concept of "trustworthiness," the property someone could have that would make it fitting for us to trust her. Trustworthiness, like trust, is usually restricted to a particular domain.[31] For instance, you trust your tax preparer to do your taxes, but you would not trust her to grade upper-division philosophy papers. As Katherine Hawley puts it: "Trust is a three-place relation, involving two people and a task" (2014, 1). The domain sensitivity of trustworthiness is what leads Jones to conclude that competence plays a key role in determining whether or not someone is trustworthy. Our students trust us to grade their philosophy papers only if they believe that we have competence in this domain (Purves and Davis 2022). According to Jones, if the trustee is to be regarded as "richly trustworthy" then she must also signal her competence to those who trust her.

Trust is typically understood as an interpersonal phenomenon, however, and there are some who are skeptical about applying concepts of trust and trustworthiness to groups or institutions.[32] Some of the reluctance to attribute trustworthiness to groups stems from the fact that accounts of trustworthiness often involve attributing certain mental states to the trustee (goodwill, self-interest, adopting certain reasons, etc.) and this may seem implausible when applied to groups or institutions.[33] In *Cooperation without Trust?* Karen Cook, Russell Hardin, and Margaret Levi argue that motivations and intentions cannot be ascribed to institutions or groups, and they claim that we cannot trust the individual members of these groups because we are not sufficiently familiar with their motives.[34] For these reasons, it might be preferable to adopt Hawley's view, since she defines trust as relying on someone to fulfill a commitment (Hawley 2014).[35] Even if we cannot attribute mental states to groups, it certainly seems

---

[31] See Jones (2012).

[32] See Budnik (2018). Budnik argues specifically that democratic governance does not require trust. Budnik takes it as given that trust is a strictly interpersonal phenomenon, so he is not talking about trust in institutions at all. Budnik focuses instead on placing trust in individual governmental officials. This means that the target of his opposition is not exactly the same as our claim in this paper. We are not suggesting that trust in institutions necessarily requires trusting particular governmental officials. Rather, we believe it requires trust in the group agents that are constituted by the institutions themselves.

Hawley (2017) is more open to the idea of describing groups as trustworthy, but she ultimately concludes that the standard distinction between trustworthiness and reliability does not apply to groups and that this makes it more appropriate to talk about groups as "reliable." She argues that the reactive attitudes that are appropriate in breakdowns of trust (anger, resentment, betrayal, etc.) apply to individual members of groups rather than to groups themselves.

Outside of philosophy, it is fairly common for those in the social sciences to talk about groups in terms of trust and trustworthiness. Political science titles like *The Oxford Handbook of Social and Political Trust* are commonplace (Uslaner 2018).

[33] For instance, Jones's account requires that the trustee treat the fact that the trustor is counting on her as a reason to act accordingly. Some may be skeptical about the capacity of groups to treat such facts as reasons. On our view of group agency, which draws on List and Pettit's view, this is not so problematic. Groups are certainly capable of something that is functionally akin to rational deliberation. So it does not seem terribly problematic to talk about groups taking facts as reasons.

[34] See Cook et al. (2005). Cf. Purves and Davis (2022).

[35] Kirby et al. 2018 discuss trustworthiness as it applies to corporations. They adopt Hawley's account on the grounds that it can plausibly be applied to corporations and that it makes sense of the appropriateness of trust in terms of commitments. They then cite Gilbert's many defenses of the claim that groups make commitments. See Gilbert 1996, 2006, and 2013). See Hawley (2017) to understand how she applies trustworthiness to groups and organizations.

reasonable to claim that groups and institutions make commitments.[36] For instance, the Preamble to the U.S. Constitution expresses the government's commitment to "provide for the common defence" and "promote the general Welfare." To say that the citizens of the U.S. "trust their government" simply amounts to the claim that they rely on the government to fulfill its commitments. To claim that the U.S. government is trustworthy, we must believe that it would be *appropriate* for us to rely on it to fulfill its commitments.

But this is a fairly thin conception of trustworthiness. What are the conditions under which it would be appropriate for citizens to rely on their government to fulfill its commitments? Here, it would be useful to return to the criteria suggested by Jones (2012), even if they must be amended in order to apply to groups. Jones explained why it is important that trustees signal their ability and their willingness to act in the ways that we are counting on them to act. If an institution fails to signal its reliability, its trustworthiness would be undermined. For instance, imagine a country that kept its military a secret. The citizens know nothing about the armed forces or its arsenal; they do not even know that their country has a military. How could they possibly trust their government to protect them? It would not be appropriate for the citizens to rely on their government to provide for their defense. If they were threatened by foreign invasion, the citizens might stockpile their own weapons, train their own militias, or flee the country entirely. They would have no reason to trust their government because the government failed to signal that it was able and willing to defend them from invasion.

Jones (2012) also includes a criterion about the trustee being responsive to the right sort of reasons. On her view, the trustee must be responsive to the fact that others are counting on her and she must take this fact as a compelling reason to act as counted on (Jones 2012, 71). Someone who is typically reliable in fulfilling their commitments might be seen as untrustworthy if they are not responsive to the fact that someone is counting on them. Perhaps you want to know if you should trust your colleague to finish a project that you have committed to work on together. You know that she is competent to finish the work, and she has adequately signaled her competence to you. But you also know that her work is motivated entirely by her desire to get a promotion. Say that she goes up for promotion on Monday and you are counting on her to finish the project on Wednesday. The fact that you are counting on her to finish the work means nothing to her, however. She cares only about the promotion. In this case, you would conclude that she is not trustworthy; you should not rely on her to finish the project. If she gets the promotion on Monday, she will no longer be motivated to fulfill the commitment. This is precisely why Jones (2012) argues that trustworthiness requires the trustee to take the fact that you are counting on her as a compelling reason to act. The robustness of the trustee's motivation matters.

Once again, this complicates the account of trustworthiness when applied to groups. While individuals can respond to facts as reasons, it is less clear that groups have this capacity (although we are less troubled by this worry, since we believe groups can be seen as agents). Purves and Davis (2022) express this concern as follows: "It is at best unclear whether

[36] See Kirby et al. (2018). Cf. Gilbert (1996), (2006), and (2013).

institutions possess such a capacity and hence whether they can take the fact of someone else's dependency as a direct and compelling reason to act" (2022, 7). Nevertheless, Purves and Davis want to apply the concept of trustworthiness to institutions, so they reframe Jones's responsiveness criterion in such a way that it can be applied to groups. Instead of "taking a fact as a reason," they suggest that trustworthy institutions must be "non-accidentally responsive" to the fact someone is counting on them. Institutions may have all sorts of mechanisms that make them responsive in this way (democratic elections could play this role, for instance). This mitigates the worry about ascribing mental states to groups while preserving Jones's core idea that trustworthiness requires a certain kind of responsiveness to the dependence of trustors.

Like Kirby et al. (2018), we think that Hawley's "commitment" account may be the most appropriate view when talking about placing our trust in groups. This means that institutional trust should be thought of as relying on a group to fulfill its commitment. We would also like to supplement Hawley's view of trust with Jones's account of trustworthiness.[37] As Purves and Davis show, Jones's criteria can be fruitfully extended to institutions, *mutatis mutandis*. This yields the following definitions of trust and trustworthiness when applied to groups:[38]

> **Trust in groups:** An individual *I* trusts a group *G* in a domain *D* if and only if *G* has made a commitment in *D* and *I* is relying on *G* to fulfill its commitment.
>
> **Trustworthiness of groups:** A group *G* is trustworthy to an individual *I* in a domain *D* if and only if (1) *G* is competent to fulfill its commitments in *D*, (2) *G* is non-accidentally responsive to the fact that *I* is counting on G to fulfill its commitments, and (3) *G* provides adequate reason for *I* to believe that *G* is competent to fulfill its commitments and that *G* is non-accidentally responsive to the fact that *I* is counting on *G*.

A trustworthy group must be competent in the relevant domain. And it must signal its competence and responsiveness to us in such a way that we are justified in believing that it will fulfill its commitments.

7.2.2 Democratic Legitimacy

[37] Hawley (2014) provides a rather thin account of trustworthiness. Hawley suggests that individuals can be seen as trustworthy in virtue of the simple fact that they reliably fulfill their commitments: "On the commitment account, trustworthiness requires us to ensure that our commitments do not outstrip our actions. This requires judiciousness in acquiring commitments as well as doggedness in fulfilling commitments already acquired, independent of others' expectations. Trustworthy people must sometimes disappoint up-front by refusing new commitments, rather than violate trust later on: this is the moral 'power of no'" (2014, 15).

[38] Our definition of trustworthiness is nearly identical to that of Purves and Davis (2022) except that we have added Hawley's language about commitments. It is helpful to think about institutional trustworthiness in terms of commitments because this makes the relevant domains explicit.

Now that we have these definitions of trust and trustworthiness in hand, we are in a position to explain how they are connected to legitimacy. The state's legitimacy depends crucially on its ability to fulfill its commitments. This could be demonstrated easily enough through the lens of something as narrow as the Hobbesian view according to which the sovereign has a legitimate claim to authority as long as it provides for the basic security of its citizens. Imagine a king who refuses to enforce conscription and relies entirely on a volunteer army. He remains committed to securing the defense of his subjects, however. He offers a variety of incentives in order to attract recruits to the army: high pay, generous benefits, and so on. This works at first. But then rumors begin to circulate that the treasury has fallen on hard times and the soldiers won't receive full pay. This narrative becomes so widespread that the soldiers no longer trust the king; they do not rely on him to fulfill his commitment to pay them. They stop showing up for service, and the erosion of public trust means that there are no new recruits to take their place. The lack of trust also affects the king's trustworthiness. Recall that the king is trustworthy only if he is competent to fulfill his commitment. Now that he has no army, he is unable to guarantee the basic security of his subjects. The erosion of trust undermined his trustworthiness because it made him unable to fulfill his commitments. This undermines his legitimacy as sovereign. The social contract that granted him authority was contingent on his ability to provide for the security of his subjects.

Purves and Davis (2022) highlight the importance of this feedback loop between trust and trustworthiness in the context of criminal justice. The proper functioning of the criminal justice system depends on "compliance and help-seeking behavior" (19).[39] When citizens no longer trust the criminal justice system, it becomes less able to do its job (i.e., to meet its commitments). When it becomes less able to fulfill its commitments, it becomes less trustworthy and citizens have even less reason to rely on it:

> Notice that this can (and surely does) generate the following feedback loop: decreasing legitimacy within a given population causes lower levels of compliance and help-seeking behavior by those in that population. The inability to secure voluntary compliance means it is less capable of achieving desired outcomes and delivering on its distinctive mandate. This failure gives rise to a further decrease in legitimacy, as individuals find themselves less trusting of a less capable institution. And the cycle starts anew. Declining descriptive legitimacy therefore negatively affects the institution's ability to achieve the aims that define the institution, which in turn erodes trust and confidence. Purves and Davis (2022, 19-20).

In order for the government to be able to meet its commitments, it must be trustworthy. Preserving the government's competence and capacities also requires that citizens have some

[39] They cite the work of Tyler and Jackson (2014) to provide an account of descriptive legitimacy. Cf. Tyler 2006 and Tyler and Huo 2002.

trust in the government.[40] This means that a trustworthy government must effectively signal both its competence and its responsiveness to the citizens.

7.2.3 How the Attention Economy Undermines Trust and Trustworthiness

The example given above evaluated legitimacy in Hobbesian terms, but the conditions of democratic legitimacy are more expensive. A legitimate democratic state must do more than simply provide basic security for its citizens. This makes democratic states more susceptible to being undermined by the erosion of public trust. For instance, citizens might not see their elected officials as legitimate if they believe that their elections are not free and fair. In this section, we show how certain noxious elements of the attention economy contribute to the erosion of citizens' trust and of the government's trustworthiness. One way this might happen is that the spread of fake news could disrupt the government's ability to signal its competence in some domain (e.g., misinformation about vaccines undermines public trust in health agencies and regulatory bodies). But this is certainly not the only threat. A growing body of evidence shows how social media contributes to polarization and radicalization, and this can lead to legislative gridlock by making it harder to achieve compromises.[41] We show how some of these negative effects are unintentional byproducts of how the algorithms are designed, but we also explain how and why social media has been weaponized by bad actors who are making a deliberate effort to undermine the efficacy of democratic states.

Because of their dependence on ad revenue, tech companies are incentivized to maximize engagement. This has led, naturally enough, to the development of algorithms that aim to direct users to content they might enjoy. This business model, when combined with certain features of human psychology, is what makes the attention economy such a powerful social force. Unfortunately, it is also implicated in some of the more pernicious aspects of our relationship with technology.

---

[40] There is a danger here of making it sound like less trust is always bad for democracy and more trust is always good. But this is obviously not the case. While a lack of trust undermines the government's capacities, an excess of trust leads to other problems. Hetherington (2008) points out, for instance, how support for military action positively correlates with political trust. Having too much trust in the government could result in citizens supporting unjust wars or using questionable tactics to win those wars. He writes: "A compelling case for the pernicious effects of high trust could be made today. Specifically, many believe that the Bush administration will not be remembered kindly by history because of its willingness to pursue extra-Constitutional means to battle terrorism, such as the use of wiretaps without first obtaining warrants; its reluctance to ban torture; and its desire to jail suspected nonmilitary enemy combatants without habeas corpus rights. Republicans' high levels of trust may be at the heart of their support of these initiatives and, consequently, the Bush administration's willingness to pursue them" (2008, 23).

Hetherington and Husser (2012) and other political scientists have shown that declining political trust has a variety of effects on the abilities of government to accomplish goals; it has a particularly negative effect on the government's ability to pursue liberal domestic policies. Klein and Robison put it even more broadly in terms of democratic health: "Trust in government is a crucial indicator of democratic health as trust enables governments to tackle difficult policy problems" (2020, 47).

[41] Hetherington and Rudolph (2017) argue that the breakdown of trust is an especially important part of the problem of polarization. They "explain how the polarization of political trust has contributed to ongoing political dysfunction in Washington" (2017, 579).

For instance, whether online or offline, people generally like to associate with others who are similar in some way. A great deal of research has shown how this tendency toward "homophily" in social media networks contributes to the existence of echo chambers.[42] When like-minded individuals cluster in online communities, especially ones with positive and negative reinforcement mechanisms (likes, upvotes, shares etc.), certain views are promoted, while other perspectives are either ignored or actively discredited. While much social media advertises itself as a platform for connecting, unfriending in the face of political posts plays a large role in pruning our online networks.[43]

Here it is important to address some skepticism around what we have called echo chambers. Some have argued that claims about echo chambers are false or overblown, citing, for example, the evidence that political polarization is often highest among older populations, a group that uses social media the least (Boxell et al. 2017). In response to this specific argument, we follow Van Bavel et al. (2021) in thinking that the balance of evidence suggests otherwise. Boxell et. al (2017) was based on observational data; however, other studies—such as Allcott et al. (2020) and Asimovic et al. (2021)—have shown that there does, indeed, seem to be a causal effect on polarization borne of social media use. For instance, Allcott et al. (2020) had some users but not others deactivate their account leading up to the 2018 U.S. election. Those who deactivated were, at the end of four weeks, both less belief and affectively polarized than participants in the study who did not deactivate their accounts. Similarly, Asimovic et al. (2021) had some users but not others from Bosnia and Herzgovnia deactivate their Facebook accounts during Genocide Remembrance week, and those who did deactivate had lower feelings of ethnic outgroup animosity than those who did not (cf. Van Bavel (2021)).

Further, we think that there is a complication in research about echo chambers. When the focus is on which news stories we see, there seems to be little effect (cf. Haidt and Bail (ongoing)). However, when we focus on networks, we get a different answer. It is undeniable that there is partisan social sorting on social media sites.[44] This is facilitated by, among other things, the ability to "unfriend" as discussed above. Zynep Tufekci notes that the fact that our networks are sorted can do the work that we might think seeing different news stories might have, as they tell us *how* to read the news stories:

---

[42] See, for example, Cinelli et al. (2021). They write: "Social media may limit the exposure to diverse perspectives and favor the formation of groups of like-minded users framing and reinforcing a shared narrative, that is, echo chambers… Our results show that the aggregation of users in homophilic clusters dominate online interactions on Facebook and Twitter" (1). Similarly, Finkel et al. (2020): "Social-media technology employs popularity-based algorithms that tailor content to maximize user engagement, increasing sectarianism within homogeneous networks (SM), in part because of the contagious power of content that elicits sectarian fear or indignation." (Finkel et al., 2020, p. 534).

Talisse (2019) argues that much of the recent polarization is the result of widespread ideological segregation, and this extends well beyond social media. People increasingly tend to associate only with those who share their political views. And he argues that politics has permeated spheres of social life where it used to be absent.

[43] See Sasahara et al. (2021) and Goyanes et al. (2021).

[44] See, e.g., Cinelli et al. (2021); Barberá (2015); Hong and Kim (2016); Mosleh et al. (2021); and Halberstam and Knight (2016).

> [W]hen we encounter opposing views in the age and context of social media, it's not like reading them in a newspaper while sitting alone. It's like hearing them from the opposing team while sitting with our fellow fans in a football stadium. Online, we're connected with our communities, and we seek approval from our like-minded peers. We bond with our team by yelling at the fans of the other one…. Belonging is stronger than facts. (Tufekci 2018)

There is empirical support for what is being said here. As Bail et al. (2018) have shown, being exposed via social media to information from the other side can *increase* belief polarization and Tufekci gives a plausible explanation as to why.

This is not to say that the selective exposure to news stories that social media and other digital tools enable is not a concern when it comes to polarization. Indeed, social media users are disproportionately exposed to like-minded political information because they tend to have relationships with like-minded users.[45] Confirmation bias also contributes to the rise of these informational environments as well. People usually prefer to see content that confirms their views and news feed algorithms can pick up on this fingerprint to reinforce their views by feeding them congenial information.[46] To make matters worse, once people come to inhabit an echo chamber, it is also fairly natural for them to adopt more extreme positions over time.[47] This makes them more likely to believe content that is false, inflammatory, or conspiratorial.[48] All of these phenomena (echo chambers, polarization, and fake news) have a deleterious effect on our capacity for democratic governance.

We explain each of these issues (and their interconnectedness) in greater detail below. We will also consider the ways that those who have an interest in undermining democratic governments have seized upon these vulnerabilities, effectively weaponizing the tools of the attention economy. Western democracies have become acutely aware of the ways that certain foreign powers (Russia, most notably) are using social media to foment polarization and spread misinformation.[49] This provides additional evidence for the claim that social media poses a threat to democracy.

To begin, it would be helpful to provide a fuller account of what polarization is and how the attention economy contributes to it.[50] This is a natural place to start, given the many ways

---

[45] See Halberstam and Knight (2016)

[46] See Cho et al. (2020).

[47] See Sunstein (1999).

[48] See, e.g., Brady et al. 2021; Rathje et al. 2021, and Vosoughi et al. (2018). They respectively show that posts using emotional language increases their diffusion, that posts about political out-groups are twice as likely to be shared than those about in-groups, and that falsehoods diffuse "significantly farther, faster, deeper, and more broadly than the truth in all categories of information"(Vosoughi, Roy, & Aral (2018, p. 1146)).

[49] In the UK, the Intelligence and Security Committee of Parliament released its report on Russian interference in 2020. The US Senate Intelligence Committee reached similar conclusions in the reports they released in 2019 and 2020. Both reports detailed the ways that Russia has been using social media to undermine democracy, as we explain below.

[50] We will generally avoid saying that social media "causes" polarization, because social media is certainly not the only cause of polarization. The causal story is multi-faceted, and the trend toward polarization began long before the rise of social media.

that polarization is connected to the mechanisms mentioned above (echo chambers, fake news, inflammatory content, etc.). The first thing to note is that polarization is not a singular phenomenon; there are several different senses in which American society is becoming increasingly polarized.[51] Talisse (2019) distinguishes between "political polarization," which concerns things like the distance of political parties, from "belief polarization," which refers to the way that an individual's belief becomes more extreme over time (106).

Within these two broad categories, there are even more fine-grained distinctions. One way for a belief to become more "extreme" is to increase the agent's confidence or credence in the belief. For instance, someone might go from *thinking* that Biden's spending is the cause of inflation to *feeling certain* that Biden's spending is the cause of inflation. Or someone might shift from being *somewhat opposed* to the bill to being *strongly opposed*. Other instances of belief polarization might shift the content of the belief entirely. Someone might go from thinking that *vaccines cause autism* to thinking *vaccines contain mind-control chips*. Both of these anti-vaccination beliefs are false, but their content is quite different.

Talisse also distinguishes between different kinds of political polarization. For most people, this brings to mind the idea that the distance between political parties has shifted. Either Democrats have moved further left; Republicans have moved to the right, or both. Partisan polarization, by contrast, involves the "ideological uniformity" of the party (Talisse 2019, 98-99). This reinforces the importance of being part of the ingroup, as those who stray from the party's ideological commitments are treated with derision (e.g., calling someone a "RINO," a Republican In Name Only). Finally, there is "affective polarization," which is characterized by high levels of trust within the partisan group, as well as "distrust and antipathy toward the members of opposing groups" (99).

When ideologically uniform groups stick together, deliberate on issues, and discuss their attitudes toward the outgroup, the more they become susceptible to all forms of polarization. Talisse cites a considerable body of evidence that shows how easily this happens. For example, Sunstein et al. (2000) conducted a study of mock jurors in which groups who were disposed toward large punitive damages deliberated together. After deliberating as a group, many individuals agreed to numbers that were substantially higher than their pre-deliberation judgment. This kind of belief polarization happens all the time. Whenever like-minded individuals confer with one another, they are prone to becoming more confident in their beliefs and adopting more extreme forms of those beliefs.

As we explained above, some of the underlying mechanisms here come from features of our psychology. People want to be seen positively by members of the ingroup; they want to have their beliefs confirmed, etc. Social media combines these tendencies with algorithms that show people content they will probably like, which often means showing them things that align with their ideology.[52] It also gives users tools that make it very easy to measure whether or not one's

[51] Although our comments here might apply more broadly, we will generally discuss polarization in the US. Similar things are happening in the UK and Europe, but most of the examples we give will pertain to the US.

[52] See, for example, Halberstam and Knight (2016).

ingroup approves of something (likes, upvotes, retweets, etc.).[53] In some cases, belief corroboration among peers has positive epistemic value. For instance, you might be uncertain whether or not Brad Pitt was in a movie you saw, so you ask Gabriel and he tells you that Pitt was in the movie. You then ask Valeria who also confirms that Pitt was in the movie. In each case, your confidence that Pitt was in the movie goes up, and this seems reasonable. As Nguyen (2020) points out, the same cannot be said of confirmation that happens in an echo chamber or filter bubble. He cites Wittgenstein's example of confirming the belief that *p* by looking at identical newspapers that all say *p* (2020, 144). The belief is not legitimately corroborated here because the papers are mere copies. In an echo chamber or filter bubble, people are prone to the same kind of bootstrapped corroboration. This might happen in cases where everyone in the bubble is getting their information from the same source. And even when they do not have the same source, if the informational ecosystem is the product of selective forces that omit certain perspectives, then confirmations from within the bubble should be discounted.

But when combined with affective polarization, this is the opposite of what happens inside an echo chamber. Because affective polarization involves increased trust of the ingroup and increased distrust of the outgroup, corroboration within the group counts for far more than external confirmation. This is why Nguyen (2020) argues that echo chambers are more dangerous than filter bubbles. An informational bubble can be popped simply by exposing someone to the omitted perspectives. But those who are trapped in an echo chamber are predisposed to reject any contrary evidence, and they are likely to distrust any sources that come from outside of the chamber.[54]

In addition to increasing confidence, these informational environments also contribute to radicalizing members of the group, as they are frequently exposed to content that is exaggerated, hyperpartisan, or simply false.[55] But the spread of fake news and conspiracy theories is not limited to echo chambers. As noted previously, Alfano et al. (2020) were able to show how

---

[53] See, for example, Goyanes et al. (2021).

[54] As Nguyen (2020) explains, filter bubbles are defined as information contexts in which certain views or ideas are omitted. Those who inhabit bubbles of this kind are simply not exposed to alternative perspectives. Social media can produce such informational environments by means of algorithmic filtering or by users' self-sorting. Nguyen cites those like Cass Sunstein (2009) and Eli Pariser (2011) who attribute much of the recent polarization and extremism to filter bubbles. Nguyen points out, however, that bubbles can easily be popped. One simply has to expose people to the omitted perspectives.

Echo chambers, on the other hand, do not simply ignore alternative views, opposing viewpoints (those held by the outgroup) are actively discredited. This is what makes echo chambers more dangerous and harder to eradicate. When someone is in an echo chamber, exposing them to alternative views does *not* weaken their conviction. On the contrary, it tends to bolster it.

The existence of echo chambers helps explain the empirical findings of Bail et al. (2018). They conducted an experiment in which people were exposed to opposing political perspectives for a month. Far from weakening subjects' confidence, this increased the strength of their convictions. They found that the effect was more pronounced on the right than it was on the left. Republicans became "substantially more conservative" after being exposed to opposing viewpoints on Twitter for a month (2018, 9216).

[55] Ross et al. (2021) show how there is an excessive focus on "fake news" which involves flagrantly false content. They argue that fake new is less widespread than "hyperpartisan" news, which depicts real events but with a strong partisan bias. See also Brady et al. (2021); Rathje et al. (2021); Alfano et al. (2020); and Ribeiro et al. (2020). They present compelling evidence that inflammatory content is more successful at being spread across social media.

watching certain categories of content on YouTube prompted the algorithm to start suggesting videos about conspiracies and other problematic content, "[sending] viewers down a conspiratorial rabbit hole" (Alfano et al. 2021, 840). Another study showed how incredibly common it is to be exposed to fake news. In just one month leading up to the 2016 US presidential election, the average American encountered between one to three fake news stories.[56] We have also learned that false information spreads faster than true information, "especially when the topic is politics" (Vosoughi et al. 2018).

What is more, the spread of fake news is harmful even when the content is not believed. Merten Reglitz (2022) explains why this is case:

> [O]nline fake news threatens democratic values and processes by playing a crucial role in reducing the perceived legitimacy of democratic institutions. This decrease in perceived legitimacy is the outcome of the primary effect that fake news has on citizens: even if its content is not believed, fake news can be a major cause of a loss of citizens' epistemic trust in each other's political views and judgment. Such a loss of trust in each other is problematic for democratic institutions since these rely for their acceptance and functioning on citizens seeing them as morally justified. (Reglitz 2022, 164).

The spread of fake news undermines our ability to trust one another and to engage in collective deliberation. People are less willing to engage with others when they are under the impression that the other person's beliefs are the product of fake news. Our capacity to engage in public reason is dependent on our ability to recognize that other people's reasons have normative force for us as well.[57] The existence of fake news, whether it is believed widely or not, threatens our ability to reason with each other.[58]

Misinformation is not the only kind of polarizing content that spreads quickly on social media, however. Content that sparks outrage is far more likely to spread than content that does not.[59] Once again, the effect is especially pronounced when it comes to politics, and one of the most popular forms of this content is that which depicts political opponents in such a way that it

---

[56] Lazer et al. (2018) point out how this is probably a conservative estimate, since the study that produced this number tracked a limited set of fake news stories.

[57] As Korsgaard puts it: "[I]f personal interaction is to be possible, we must reason together, and this means that I must treat your reasons…as reasons, that is, as considerations that have normative force for me as well as you, and therefore as public reasons" (2009, 192).

[58] Although we do not have the space here to explore the question of whether or not fake news is widely believed, there is at least some reason that belief in fake news has had a considerable impact. For instance, there is no credible evidence of widespread voter fraud in the 2020 presidential election, but polls show that as many as 75% of Republicans believe that Trump has a legitimate claim to the presidency because "real cases of fraud changed the results." See Montanaro (2021). It is clear that some people believe fake news. And it is certainly the case that people spread fake news and share it widely. Vosoughi et al. (2018) show that although bots play a role, it is mostly humans who spread fake news: "false news spreads more than the truth because humans, not robots, are more likely to spread it" (2018, 1146).

[59] See Brady et al. (2017), Rathje et al. (2021); Pew Research Center (2017); León and Trilling (2021); Corbu et al. (2020); Frimer et al. (2022); and Wang and Inbar (2022).

sparks moral outrage.[60] Nguyen and Williams (2020) argue that seeking gratification through this kind of moral outrage is problematic for several reasons. One problem with it is that we become incentivized to seek satisfying representations of political opponents rather than truthful ones. This is yet another way that people might corroborate their beliefs with evidence that should be discounted.[61] They write: "This invites a problematic form of circularity—where one picks one's sources based on agreement with one's antecedent beliefs, and then goes on to use those sources to buttress one's antecedent beliefs" (Nguyen and Williams 2020, 162). In addition to making individuals worse off epistemically, when we are consumed with outrage and incivility, we weaken our ability to govern ourselves democratically.[62]

Some of the most extreme instances of this can be seen by considering recent events in Myanmar, a country that recently went through a rapid adoption of social media that coincided with explosive violence against an ethnic minority, the Rohingya. The ongoing crisis in Myanmar has been recognized by the U.N. as a genocide in which social media—Facebook, in particular—played a "determining role" (Miles 2018). As we will soon recount, similar events have happened world over.

The seeds of anti-Rohingya sentiment in Myanmar are at least 100 years old. In the early 1900s, the British colonial rulers of Myanmar—then Burma—imported large numbers of Muslim subjects as part of a "divide and rule" scheme in the majority Buddhist country (Fisher 2022). After the British left in 1948 and a newly independent Burma sought to establish itself, suspicions of the Muslim minority lingered.[63] This sentiment coalesced into a long standing, government supported, persecution of Rohingya Muslims, who were denied citizenship and as recently as 2014 were not included in the census.[64]

In 2016 these simmering tensions exploded into what the U.N. has recognized as a, "textbook example of ethnic cleansing." In his reporting on the atrocity, Max Fisher describes the stomach-turning violence in grim detail:

> The soldiers, sent to exterminate the impoverished minority that many of Myanmar's leaders and citizens had come to see as an intolerable enemy within, would arrive at a village, then begin by setting rooftops afire. They lobbed grenades through hut doorways

[60] See Brady et al. (2017).

[61] Nguyen and Williams write: "But when one is engaged with moral outrage porn, one is seeking out representations of moral outrage for the sake of the resulting gratification, and so one is incentivized to preselect those representations with which one agrees. This invites a problematic form of circularity—where one picks one's sources based on agreement with one's antecedent beliefs, and then goes on to use those sources to buttress one's antecedent beliefs" (2020, 162).

[62] Frimer et al. 2022 conducted a study that looked at civility of politicians on Twitter. It found a considerable increase in incivility in the last decade, and they highlight the ways that this might pose a danger to democracy: "[P]olitical incivility can undermine respect for alternative viewpoints, erode public trust in the political process, and incite other forms of uncivil, undemocratic behavior" (2022, 1).

Jonathan Haidt, a well-known critic of the effect of social media on democracy summarizes this point nicely in an interview: "So long as we are all immersed in a constant stream of unbelievable outrages perpetrated by the other side, I don't see how we can ever trust each other and work together again" (Illing 2018).

[63] See Fisher (2022).

[64] See BBC (2020).

> and sent rockets slamming into the walls of longhouses. They fired into the backs of peasants fleeing across the surrounding fields. As the houses burned, the men of the village would be arrayed in a line and shot to death. Families streamed by the hundred thousand toward the border. The soldiers attacked these too. They hid land mines in the refugees' paths. Survivors who made it to relative safety in Bangladesh detailed horror after horror to journalists and aid workers who picked their way through the overcrowded camps (Fisher 2022, 158).

With the fraught history between the Rohingya and the government, it is sadly not surprising that the Rohingya were persecuted in 2016 and still are to this day. But the timing and explosion of violence still call out for an explanation, with many—including the U.N., as we have seen—pointing their finger at social media.[65]

To tell this side of the story, we need to go back in time once more. As late as 2011, Myanmar was heralded as "the last 'unphoned' country in the world", with less than 1% of the population having internet access (Blah 2018). Myanmar's information landscape changed dramatically in 2012 when, as part of an economic liberalization scheme, its telecommunications market was released from a long-standing state-owned monopoly.[66] As early as 2013, the recently "unphoned" country had its state-run newspaper saying that "a person without a Facebook identity is like a person without a home address" (McLaughlin 2018). By 2015, nearly 40% of the population was online, with nearly all internet being accessed via smartphone in a media environment where Facebook was "synonymous with the internet" (Ibid.). And even then, alarm bells were already ringing.

In 2014, the ultra-nationalist Buddhist monk Wirathu was able to use his Facebook megaphone to help turn a rumor—about a Muslim shop owner raping his Buddhist employee—into a deadly riot (Ibid.). Government officials desperate to stop the riots scrambled to get in touch with Facebook about controlling the hate speech on their platform that was fueling the violence. Eventually, the government—not knowing what else to do—decided to simply temporarily block access to Facebook, which almost immediately put an end to the mayhem (Ibid). Those trying to reach Facebook before the site was blocked reported getting a response from the company only after the violence subsided; the companies representatives were concerned over the site being unreachable and looking to get it back online (Ibid).

Immediately after the riots, there was talk of improving content moderation. At the time of the riots, the language community standards had not even been translated into Burmese (and still weren't as late as 14 months after the incident).[67] But problems with hate speech clearly weren't adequately addressed. Muslims, and Rohingya in particular, were still being referred to in derogatory terms and being attached to larger than life rumors to rationalize islamophobia. Examples from 2015 include viral memes telling Rohingya to "keep out" in idiomatic language typically reserved for dogs, and fake news "revealing" Burmese Muslims to be heavily armed.

[65] See Reuters (2018).
[66] See McLaughlin (2018).
[67] Ibid.

One popular meme—"liked" 8,400 times—features a false picture of a stockpile of guns and ammunition (later revealed to be taken in Cairo in 2012). Another meme calling Rohingyas Bengalis (perpetuating the myth that they are illegal immigrants from Bangladesh) and cannibals (accompanied by cartoonishly fake pictures of human butcher shops) gained over 9,000 "likes" and 40,000 "shares."[68]

By 2016, Facebook had many signs that it had an unresolved hate speech problem in Myanmar. But it pressed on with its expansion into this growing market by launching its "Free Basics" program, which allowed free access to data so long as it was accessed via the Facebook app (Fisher 2022). Soon after, 38% of people living in Myanmar were getting all or most of the news via the app (Ibid).

By 2017, thousands of Rohingya, hundreds below the age of five, had been killed. The U.N. stated that social media, on Facebook in particular, drove the violence:

> It was used to convey public messages but we know that the ultra-nationalist Buddhists have their own Facebooks and are really inciting a lot of violence and a lot of hatred against the Rohingya or other ethnic minorities … I'm afraid that Facebook has now turned into a beast, and not what it originally intended (Miles 2018).

In 2018, the site reaffirmed its efforts to control hate speech (Stecklow 2018). Four months later, however, a Reuters report found that Myanmar's corner of Facebook was still awash in anti-muslim propaganda. The report found over one thousand posts attacking Rohingya and other Muslims. One post mentioned in the report—still up in 2018, despite being posted in 2013—says that the Rohingya must be fought "the way Hitler did the Jews" (Ibid.). Another calls Rohingya "Non-human … dogs" and "Bengalis" (Ibid). The report mentions that "In early 2015, there were only two people at Facebook who could speak Burmese reviewing problematic posts" and that in 2018, the company didn't have a single employee in Myanmar. It goes on to state that Facebook "continues to rely heavily on users reporting hate speech … because its systems struggle to interpret Burmese text" (Ibid.).

We should add, here, that it isn't clear how many content moderators would be enough, though more would certainly be better. At the time that Reuters wrote their report, Myanmar had 18 million users. And, by Facebook's own admission, their algorithms—like those of many others geared towards promoting engagement—promote and reward divisive content at a very large scale.[69]

In response to all this, one might think that Myanmar is a special case. To show that it is not, let us consider a few other episodes.[70]

Sri Lanka—like Myanmar—is a country that has seen rapid internet adoption via *zero-rating* programs that, like Free Basics, involve generating a user base by exempting certain

[68] See C4ADS (2016).
[69] See Horowitz and Seetharaman (2020).
[70] For coverage of these episodes and others that we could not include for reasons of space, see Fisher (2022).

data from billing (such as data used in-app).[71] Like Myanmar, Sri Lanka has a Muslim ethnic minority that has a fraught relationship with the (Sinhalese, Buddhist) ethnic majority.

A few years into rapid adoption of the internet via free access to social media, familiar patterns began to emerge in Sri Lanka. In 2018, a video of a misunderstanding between a non-Sinhalese speaking Muslim shop owner and an angry Sinhalese customer was recorded. The Sinhalese customer accused the shop owner of putting sterilization pills in his soup, this being one of the many Islamophobic rumors circulating on social media. The shop owner, trying to placate the customer and not fluent in Sinhalese, inadvertently admitted to the accusation. A video of the altercation went viral on Facebook. Eventually, riots erupted. The government could not contain them until they shut down access to social media. As in Myanmar, this ended the violence and also caught the attention of Facebook representatives who had ignored concerns over hate speech. The representatives wanted to know why they had lost traffic to their site.[72]

Sri Lanka and Myanmar are not the only places where these sorts of dynamics have played out. Similar events have unfolded in Ethiopia, for example, with Frances Haugen, a whistleblower at Facebook, saying "What we saw in Myanmar and are now seeing in Ethiopia are only the opening chapters of a story so terrifying, no one wants to read the end of it." Haugen states—and we agree with her—that "engagement-based ranking" is "fanning ethnic violence" (Akinwotu 2021).

Her suspicion seems to have been corroborated by Karsten Müller and Carlo Schwarz, two researchers studying anti-refugee attacks and Facebook use in Germany. The team found that towns where Facebook use was higher, so was anti-refugee violence. In fact, the relationship was disturbingly strong. The New York Times sharply summarizes one of their key findings as follows: "Wherever per-person Facebook use rose to one standard deviation above the national average, attacks on refugees increased by about 50 percent" (Taub and Fisher 2018).

Further, Müller and Schwarz were able to provide evidence of a causal relationship. German internet infrastructure is localized, enabling a study of connections between (localized) internet outages and anti-refugee violence. And there is a tight connection between the two. In areas of high Facebook usage, internet outages were associated with precipitous drops in anti-refugee violence (about 35%).[73] This effect, however, was *not* present in outages among communities where internet usage is high but Facebook usage is not, singling out the social network as a key variable in the drop in violence.

When we zoom out a little, the story of social media and democracy might seem so multifaceted that it may seem impossible to connect all of these disparate phenomena into a cohesive narrative. But this is why our argument in this paper has been couched in terms of trust and trustworthiness. Framing these issues with an eye toward their effect on trust and trustworthiness makes it easier to see how the attention economy threatens democratic

[71] See Fisher (2022, 169).
[72] Ibid., 175.
[73] Ibid.

legitimacy.[74] Almost all of the issues described in this section connect to trust or trustworthiness in some way; in many cases, distrust is partially constitutive of the phenomenon in question. For instance, echo chambers involve attitudes of distrust toward outside sources, and affective polarization is characterized by a distrust of members of the outgroup and trust of the ingroup. And the spread of fake news and conspiracy theories undermines the government's ability to effectively signal its competence or its responsiveness to citizens (or to even be competent and responsive in the first case).

What is more, there are several feedback loops in the process. Not only is there the loop between trust and trustworthiness, which we discussed above, the different forms of polarization also feed into one another. It is not hard to see how belief polarization might contribute to affective polarization and vice versa. The more that someone distrusts the outgroup and trusts the ingroup (i.e., the more they are affectively polarized), the more likely they are to become belief polarized as well (their beliefs become more extreme and they become more confident in those beliefs). Both forms of polarization make it harder for us to engage in public reason. This is why Talisse comes to the conclusion that polarization strikes at the heart of some of democracy's core commitments: "This is the fundamental problem posed by polarization. Belief polarization directly attacks our capacities to properly enact democratic citizenship, dissolving our abilities to treat our fellow citizens as our political equals. Moreover, belief polarization is part of a larger dynamic by which partisan divisions expand and extremity intensifies, all within a structure of self-perpetuating social dysfunction" (Talisse, 2019, 123).

Figure 7.1 is meant to demonstrate how all of these issues are connected to democratic legitimacy.

[74] Sabatini and Sarracino (2019) evaluated the effect of social media on three kinds of trust: institutional, trust in strangers and trust in neighbors. They found that "all the forms of trust significantly decrease with participation in online networks" (229).

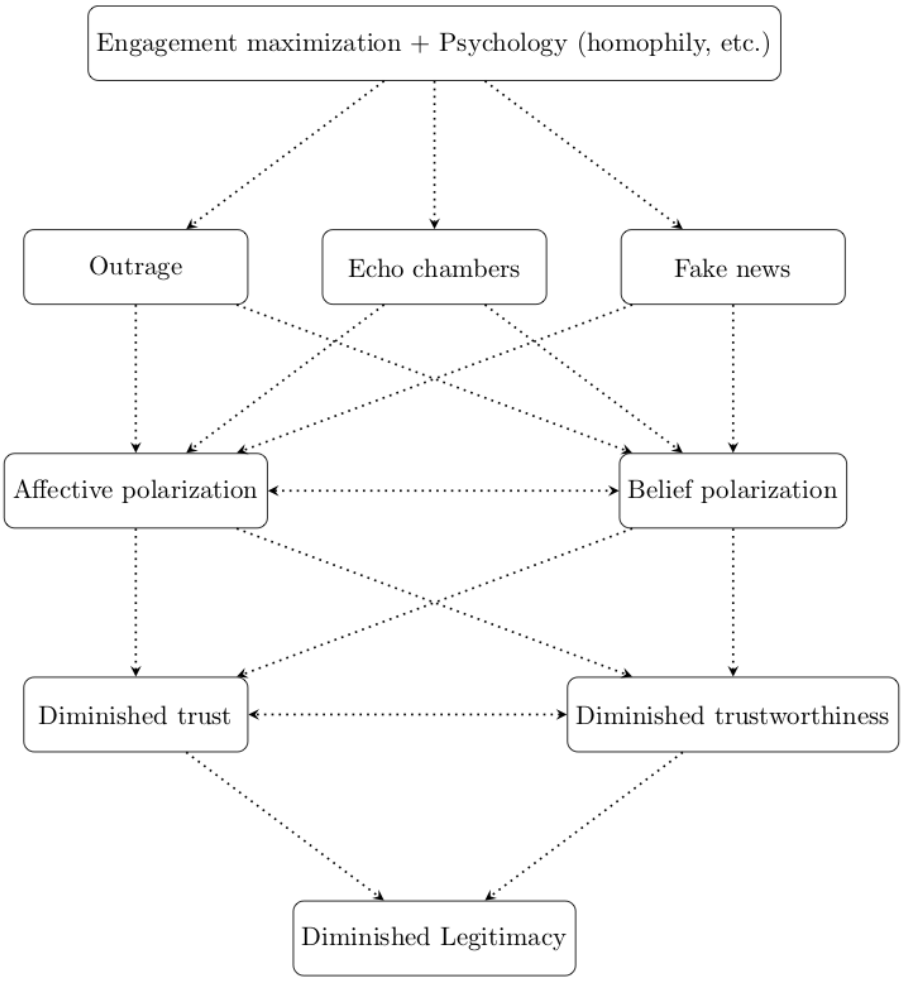

Fig 7.1
How the attention economy affects democratic legitimacy

Although this diagram does not capture every component of the causal story linking social media to democracy, it does illustrate our understanding of the connection between some of the key issues.

First, social media combines with certain features of human psychology in such a way that it exacerbates polarization. All of this coalesces to undermine both trust in the government and the government's trustworthiness.

As further evidence of this claim, we will conclude this section with a brief discussion of the ways that these vulnerabilities have been seized upon by those who want to undermine the efficacy of democratic governments. The fact that Russia's Internet Research Agency (IRA) has engaged in this kind of behavior would seem to bolster the claim that polarization can be exacerbated through social and that this weakens our democracy. Recent reports released by the governments of the UK and US have left little doubt about the IRA's playbook. They also paint a rather bleak picture about the efficacy of these efforts. Between 2015 and 2017, 30 million users of facebook, Twitter, and Instagram shared, liked, or reacted to content that originated from the

IRA.[75] But disseminating fake news is not their only tactic. The majority of their efforts were dedicated toward stoking polarization.[76] The US Senate report offered the following conclusion:

> The preponderance of the operational focus, as reflected repeatedly in content, account names, and audiences targeted, was on socially divisive issues-such as race, immigration, and Second Amendment rights-in an attempt to pit Americans against one another and against their government. The Committee found that IRA influence operatives consistently used hot-button, societal divisions in the United States as fodder for the content they published through social media in order to stoke anger, provoke outrage and protest, push Americans further away from one another, and foment distrust in government institutions. (6)

This was consistent with the findings of the UK report, as they concluded that one of Russia's aims was the "general poisoning of the political narrative in the West by fomenting political extremism and 'wedge issues', and by the 'astroturfing' of Western public opinion; and general discrediting of the West" (10).

Once again, the evidence of the efficacy of these efforts is disheartening. The IRA promoted and organized hundreds of rallies, with a particular emphasis on divisive issues (promoting protests for both Black Lives Matter and Blue Lives Matter, for example). One of the most striking successes took place in Houston, Texas on May 21, 2016, when the IRA organized two events to take place across the street from each other. On one side of the street, there was a "Stop the Islamization of Texas" rally which was organized by a group called "Heart of Texas." On the other side of the street, there was a rally called "Save Islamic Knowledge," organized by "United Muslims of America." Both groups turned out to be IRA accounts.[77] The entire event was orchestrated from St. Petersburg. To make matters worse, the Heart of Texas post encouraged rally goers to bring guns. Luckily, no one was hurt.

It is very clear *what* Russia's IRA is trying to do. They are using fake news, divisive issues, and inflammatory content in an attempt to polarize western democracies like the US and UK. It is somewhat less clear *why* they are trying to do this. The UK report concludes that Russia's approach is "fundamentally nihilistic" as they believe "any actions it can take which damage the West are fundamentally good for Russia." Perhaps the idea is that weakening the US and the UK will make it easier for them to accomplish certain foreign policy aims. We can only speculate about the motivation for their activities. But we do not need to speculate about the activities themselves. And it is also clear what kind of effect these actions have. Senator Richard Burr chaired the Senate Intelligence Committee that authored the report on Russia's social media

[75] See Howard et al. (2019).

[76] See, for example Freelon and Lokot (2020): "State-sponsored disinformation agents have demonstrated success in infiltrating distinct online communities. Political content attracts far more engagement than non-political content *and appears crafted to exploit intergroup distrust and enmity*…Our results make it clear that group identity lies at the core of the IRA's attack strategy. Political audiences were addressed as liberals, conservatives, and Black people to *provoke anger against oppositional outgroups* (2020, 2, emphasis added).

[77] See Timberg and Dwoksin (2018).

campaign. Rather fittingly, he identified the erosion of trust as one of their fundamental aims: "Russia is waging an information warfare campaign against the U.S. that didn't start and didn't end with the 2016 election. Their goal is broader: to sow societal discord and erode public confidence in the machinery of government. By flooding social media with false reports, conspiracy theories, and trolls, and by exploiting existing divisions, Russia is trying to breed distrust of our democratic institutions and our fellow Americans."[78]

Of course, problems like polarization and echo chambers cannot be attributed entirely to the actions of malicious groups like the IRA. They also cannot be attributed entirely to social media, as the evidence shows that these trends began much earlier. But, as we hope to have shown in this paper, there is very good reason to believe that our relationship with the attention economy has poured gasoline on these fires. As Harinda Dissanayake, a presidential adviser in Sri Lanka, put it, "We don't completely blame Facebook. The germs are ours, but Facebook is the wind" (Dissanayake 2018). Putting out these fires will not be an easy task, but it is high time for us to start taking the problem seriously. Our capacity for democratic governance hangs in the balance.

7.3 Implications of the Group-Level Duties

Let us now take stock of what we have learned about group-level duties.

First, consider the obligations to be digital minimalists. Our application of this duty to groups cannot apply as straightforwardly to groups as it does to people, since it is controversial whether groups have "humanity." But, as we think that we have established, groups can be or fail to be autonomous. That is, they can succeed or fail to set and pursue their own ends. And—in the case of the groups we are concerned with in this chapter (i.e., democratically governed government bodies)—there is something moral at stake in their autonomy. There are two ways to ground the moral weight of the group's autonomy. It could have intrinsic or final value (which would be a controversial standpoint for us to take, but not one that we necessarily want to reject). Or it could matter instrumentally, which would be much less controversial. Either way, these groups do have an obligation to safeguard their own autonomy and, as we have shown, their autonomy is threatened by the attention economy.

What, then, should groups do to protect themselves? We will have little new to say here, because many of the recommendations have been mentioned previously in this book, albeit under different headings. Groups should think about how their members are educated, for instance, as this, as we have shown, can curb some of the excesses of the attention economy (including excess polarization).[79] Groups can also think about how their members relate to the attention economy, in part by thinking about how that economy is structured and also, for example, how workers and children might be asked to interact with it. Some of these aims might be achieved

[78] US Senate Press Release (2019).
[79] See Lees and Cikara (2021).

through regulation (as we argued in chapter 6), but others might be accomplished through lighter touch means, such as public service announcements.

We can also quickly address groups' obligations to one another to be attention ecologists. This can have both positive and negative sides to it. We have seen how the IRA has worked towards undermining our trust in each other through its active measures. This is clearly problematic, as it undermines the morally important autonomy of groups that is at least partly grounded in its subjects' legitimate interests in self-governance. Perhaps less obviously, it also follows from this that corporate entities such as Facebook have moral reasons to think about how its activities affect other agents. Such companies might not act on these obligations absent being so compelled, but we take it that we have laid the groundwork for thinking that certain groups (such as states and nations) have justification for compelling corporate entities to refrain from undermining them. Finally, entities such as Facebook also have wide duties to promote the well functioning of nations and states, as these group agents—or their members—have an obligation to *promote* the autonomy of others.

Lastly, we would like to consider the interesting case of the duties that individuals might have to groups. We might have the intuition that individuals just as any other agent—group or otherwise—at least have the duty not to undermine group agents. However, we also likely have the thought that there isn't much we can actually do to promote or undermine such agents, at least not through ordinary activities. They are so large and we are so small.

There is some truth in this—no nation will ever crumble because of how either of us use our phones—but it might not follow from the fact that we are relatively ineffective and that we have no duties here. There are many philosophical puzzles that bear resemblance to the problem we are discussing here. Climate change is a problem, but each of us, considered individually, won't make the problem any better or any worse. Likewise, factory farming is one of the major moral atrocities of our times. But none of us, individually, can stop it, whether or not we eat meat. Does this mean that we have no individual-level obligations to, say, drive less or become vegetarians?

This issue, which is sometimes referred to as the "problem of collective harm," has become a hot topic in contemporary ethics. It is particularly difficult for consequentialists to explain why individuals have moral obligations to perform certain actions (avoid eating meat, refrain from joyrides in gas-guzzling SUVs, vote in large-scale elections), when it appears that each contribution is causally and morally insignificant.[80] If moral rightness depends *entirely* on

[80] Each of these have been discussed in depth in the collective harm literature. On vegetarianism, see Singer (1980), Norcross (2004), Kagan (2011), Harris and Galvin (2012), Budolfson (2019), and Aylsworth and Pham (2020). On climate change and gas-guzzling joyrides, see Sinnott-Armstrong (2005), Kingston and Sinnott-Armstrong (2018) and Broome (2012). On voting, see Brennan (2011) and Barnett (2020)

the consequences of your actions as an individual, why should you do any of those things?[81] The outcomes appear to be fixed no matter what you do. If we hold the rest of the world fixed and reduce your carbon emissions to zero, there is very little reason to think that this will have any effect on climate change. The same goes for factory farming or large-scale elections.

Given that Kant's ethics does not rely on consequences in the same way, we may be tempted to believe that the categorical imperative yields better results when applied to these situations. But the situation is not so simple. An undergraduate student of Kant's ethics might think that we can solve the problem simply by reflecting on what happens when everyone acts this way. They think of the "universalizability" test like this: We ask what would happen if everyone eats meat and drives gas-guzzling SUVs, and then we realize that such actions are immoral because the consequences are disastrous.

Unfortunately, that is not how the formula of universal law works. We cannot simply evaluate the disastrous consequences of everyone performing some action and rule it out. First, that is closer to rule utilitarianism than it is to Kant's ethics. The categorical imperative requires us to find a contradiction that results from the universal maxim; it is not enough to simply find bad results. Second, if we use the test this way, we could use it to derive a wide variety of faulty prescriptions. As Wood (1999) explains, the test would yield both false positives and false negatives. It would show permissible actions to be wrong and it would allow us to act on impermissible maxims.

The problem turns on the specificity of the circumstances designated in the maxim. This is a routine exercise in any introduction to ethics course. You could ask yourself whether or not you could universalize the maxim of getting coffee at the cafe on Bird Road at 67th Ave, Miami FL at 8 am on Monday morning, and you would immediately realize that this yields a contradiction. If everyone in the world showed up at this cafe at this time, you would not be able to achieve your end. On the flip side, you could ask whether or not it is permissible to make a false promise on "Tuesday, August 21, to a person Hildreth Milton Flitcraft" Wood (1999, 102). This maxim is so specific that it is possible for you to universalize it. There are so few people who would find themselves in this situation that it does not yield a contradiction when willed to be a universal law.

When we return to the collective action problem, we see that the same issue arises.[82] As Andrew Chignell (2015) points out, there is a crucial difference between asking what would happen if *everyone* performed some action versus asking whether *anyone in relevantly similar circumstances* performed some action. He argues that this difference allows opportunistic carnivores to act on their maxim. By hypothesis, the opportunistic carnivore is someone who eats

[81] This problem is particularly pressing for consequentialists. As Shelly Kagan explains: "The problem, in effect, is this: consequentialism condemns my act only when my act makes a difference. But in the kind of cases we are imagining, my act makes no difference, and so cannot be condemned by consequentialism—even though it remains true that when enough such acts are performed the results are bad. Thus consequentialism fails to condemn my act. In cases of this sort, therefore, consequentialism seems to fail even by its own lights" (2011, 108). Kagan goes on to defend a version of the expected utility argument. But his view is challenged by Nefsky (2011) and Budolfson (2019).

[82] Shafer-Landau (1994) raises this issue for the universalizability test as it applies to vegetarianism.

meat when they are reasonably sure that their action will not make a difference. Surely there is no contradiction in willing that anyone in that circumstance act on such a maxim.

As we saw earlier, the formula of universal law is successful when it shows us that we are making an exception of ourselves. We behave one way while expecting everyone else to behave differently when in similar circumstances. But this is harder to accomplish in cases of collective harm. As Christopher Kutz explains, "The CI [categorical imperative] test works when an individual's maxim can be realized only when it is exceptional, not when it, on the contrary, owes its success to the fact that others act in precisely the same way" (2000, 8). Typically the test works when you are performing an immoral action and you realize that your end can be achieved only if others refrain from performing this action. But in cases of collective harm, the opposite is true. Your action appears to be permissible precisely because you are doing what everyone else is doing. Your action is not exceptional at all.

But this does not mean that Kant's ethics has nothing to say about what we should do in these situations. In a recent paper, Maike Albertzart argues that Kant's ethics can respond to these issues by drawing on the duty of beneficence. As you may recall from previous chapters, this duty requires us to adopt the happiness of others as one of our ends, and Kant's theory of agency tells us that willing an end involves willing the necessary means to that end. Albertzart writes, "To adopt the happiness of others as an end implies willing the necessary means for achieving this end. Given the negative impact climate change is expected to have on human happiness, combating climate change qualifies as one of these necessary means" (2019, 844).

Albertzart asks us to imagine an agent who wants to fight climate change as a way of promoting the happiness of those who will be negatively affected by it. In order to will that end, she wills one of the necessary means (*viz.* that everyone avoid unnecessary car trips). When she takes an unnecessary car trip, she acts in a way that is inconsistent. She cannot universalize the maxim of her action because it "contradicts her chosen means for combating climate change as part of the obligatory end of the happiness of others" (Ibid.).

We believe that Albertzart's solution is on the right track, but it seems to move a little too quickly when determining the necessary means to the end of fighting climate change. Strictly speaking, to fight climate change, we do not need *everyone* to refrain from unnecessary carbon-emitting activities. We only need a critical mass of people to do so. It is not true that we must require everyone to avoid unnecessary car trips. In order to will the end of mitigating the climate crisis, we need to will only that a sufficiently large number of people reduce their carbon emissions.

But this is where the formula of universal law demonstrates the problem with making an exception of oneself. Consider the case of overfishing—another collective harm problem. If one person ignores the fishing limit, this is unlikely to have any effect on the stability of the fish population. However, if a sufficiently large group of people take too many fish, the population will cross a critical threshold and collapse. The fish will not be able to repopulate the lake. In order to sustain the ecosystem, we do not have to will that *everyone* refrain from taking more than the limit. Instead, we must will that a sufficiently large set of anglers refrain from doing

this. Imagine that you are fishing on the lake and you have willed the end of a sustainable fish population. When you decide to exceed your limit, you make an exception of yourself. You want a critical number of people to refrain from doing precisely what you are doing. You realize that the success of the end you have willed depends on other people not performing this action.

That is the Kantian response to the problem of collective harm. First, it is routed through the duty of beneficence. We can think of an end that we are required to adopt, and then we develop an understanding of the necessary means to that end. Second, we come to see how defecting when we expect others to cooperate involves making an exception of ourselves and is thus inconsistent with the formula of universal law.

When applied to the problems in this chapter, we can see how useful this would be. Given the massive scope of polarization, moral outrage, and fake news, we may be tempted to think that each person plays an insignificantly small role. The action of a single individual is too small to make a difference. But now we see the error of thinking exclusively about the consequences of an individual action. Instead, we should think about how we are committed to certain social ends. We want healthy democratic governance; we know that this involves trust, trustworthiness, and social cooperation. In order to evacuate efficiently, we need people to follow directions when the National Hurricane Center tells them to evacuate or stay put. If we want to control pandemics, then we need them to trust regulatory bodies who are telling us that a vaccine is safe and effective.

If we will those ends, we must also will the necessary means to those ends. And that requires us to avoid spreading misinformation on the internet, to refrain from sharing inflammatory content that provokes moral outrage and distrust, and possibly to disengage from the attention economy in general. Given the empirical evidence on the radicalizing effect of the attention economy, willing the ends of a rightful civil condition requires us to will that a critical mass of people unplug from social media. There is a common thread uniting the genocide in Myanmar, the ethnic violence in Sri Lanka and Germany, and the distrust of health information about the Zika virus in Brazil.[83] Each of these were fueled, in part, by the fact that huge populations were radicalized by social media and the algorithms' aim of capturing as much attention as possible.

As individuals, we may not be able to put out these forest fires of collective harm. But that does not mean that we have no obligation to do our part. If we will the end, we must will the necessary means. And we must not make exceptions for ourselves. Our duty to be digital minimalists is not just something that we owe to ourselves. It is something that we owe to the groups to which we belong. In some cases, the autonomy of those groups matters. We want to live in functional, democratic states, and we want those institutions to succeed at setting and pursuing their own ends. If we succumb to the temptations of the attention economy, we will go along with everyone else in pouring gasoline on the fires of polarization, outrage, and distrust. We will undermine the necessary means to our chosen ends.

[83] See Fisher (2022).

7.4 Conclusion

In this chapter, we have argued that the attention economy threatens to harm us not only as individuals but also as a group. Structured groups, like group agents, are susceptible to a distinctive form of harm insofar as they have shared interests that they pursue collectively. Democratic states are a paradigmatic case of a group agent that is liable to this kind of harm. And given the nature of democratic legitimacy, trust and trustworthiness are key components of the state's agency. Without trust, the government is unable to achieve its aims or fulfill its commitments. Various features of the attention economy threaten to undermine trust in the government and the government's trustworthiness. Western democracies (like the US and the UK) caught on to this problem only after serious damage had been done. We would do well to remain vigilant about this threat in the future. That is why it is so important to understand what is at stake and why it is of vital moral significance.

7.5 References

Akinwotu, Emmanuel. 2021. Facebook's role in Myanmar and Ethiopia under new scrutiny. *The Guardian*. https://www.theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny Accessed 17 July 2023.

Alfano, Mark, Amir Ebrahimi Fard, J. Adam Carter, Peter Clutton and Colin Klein. 2021. Technologically scaffolded atypical cognition: the case of YouTube's recommender system. *Synthese* 199: 835–58.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110: 629–76.

Altman, Andrew and Wellman, Christopher Heath. A defense of international criminal law *Ethics* 115: 35–67.

Appiah, Anthony. 2005a. *The ethics of identity*. Princeton: Princeton University Press

———. 2005b. *Thinking it through*. Oxford: Oxford University Press.

———. 2011. 'Group rights' and racial affirmative action. *The Journal of Ethics* 15:265–80.

Asimovic, Nejla, Jonathan Nagler, Richard Bonneau and Joshua A. Tucker. 2021. Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences* 118: e2022819118.

Aylsworth, Timothy and Castro, Clinton. 2021. Is there a duty to be a digital minimalist? *Journal of Applied Philosophy* 38: 662-73.

———. 2022. On the duty to be an attention ecologist. *Philosophy and Technology* 35: 1-22.

Aylsworth, Tim, and Adam Pham. 2020. Consequentialism, collective action, and causal impotence. *Ethics, Policy & Environment* 23: 336–49.

Baier, Annette, 1986. Trust and antitrust. *Ethics* 96: 231–60.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout und Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115: 9216–21.

Barberá, Pablo. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23: 76–91.

Barnett, Zach. Why you should vote to change the outcome. *Philosophy & Public Affairs* 48: 315-447.

BBC. 2020. Myanmar Rohingya: What you need to know about the crisis. BBC. https://www.bbc.com/news/world-asia-41566561

Blah, Dan. Access and openness: Myanmar 2012. *Open Technology Fund*. https://www.opentech.fund/news/access-and-openness-myanmar-2012/

Brady, William J., Killian McLoughlin, Tuan N. Doan, and Molly J. Crockett. 2021. How social learning amplifies moral outrage expression in online social networks. *Science Advances* 7: eabe5641.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2017. Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences* 114: 10612–17.

Bratman, Michael. 1992. Shared cooperative activity. *The Philosophical Review* 101: 327–41.

———. 1993. Shared intention. *Ethics* 104: 97–113.

Brennan, Jason. 2011. *The ethics of voting*. Princeton: Princeton University Press.

Broome, John. 2012. Climate matters: Ethics in a warming world. New York: W. W. Norton & Company.

Budnik, Christian. 2018. Trust, reliance, and democracy. *International Journal of Philosophical Studies* 26: 221–39.

Budolfson, Mark Bryant. 2019. The inefficacy objection to consequentialism and the problem with the expected consequences response. *Philosophical Studies* 176: 1711–24.

C4ADS. 2016. Sticks and Stones: Hate speech narratives and facilitators in Myanmar. https://c4ads.org/reports/sticks-and-stones/ Accessed 17 July 2023.

Carr, Nicholas G. 2010. *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company; Newport.

Castro, Clinton and Pham, Adam. 2020. Is the attention economy noxious? *Philosophers' Imprint* 20 (17): 1-13.

Cho, Jaeho, Saifuddin Ahmed, Martin Hilbert, Billy Liu and Jonathan Luu. 2020. Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization. *Journal of Broadcasting & Electronic Media* 64: 150–72.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118: e2023301118.

Cook, Karen S., Russell Hardin, and Margaret Levi. 2005. *Cooperation without trust?* New York: Russell Sage Foundation.

Corbu, Nicoleta, Alina Bârgăoanu, Raluca Buturoiu, and Oana Ştefăniţă. 2020. Does fake news lead to more engaging effects on social media? Evidence from Romania. *Communications* 45: 694–717.

Crawford, James. 2006. *The creation of states in international law*. Oxford: Clarendon Press.

Davis, Dena S. 2000. Groups, communities, and contested identities in genetic research. *Hastings Center Report* 30: 38–45.

Dissanayake, Harindra. Quotation of the day: Where Facebook rumors fuel thirst for revenge. *The New York Times*. https://www.nytimes.com/2018/04/21/todayspaper/quotation-of-the-day-where-facebook-rumors-fuel-thirst-for-revenge.html Accessed 17 July 2023.

Fazelpour, Sina and Danks, David. 2021. Algorithmic bias: Senses, sources, solutions. *Philosophy Compass* 16:e12760.

Fenton, Justin. 2021. *We own this city: A true story of crime, cops, and corruption*. New York: Random House.

Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary McGrath, Brendan Nyhan, David Rand, Linda Skitka, Joshua Tucker, Jay Van Bavel, Cynthia Wang, and James Druckman.. 2020. Political sectarianism in America. *Science* 370: 533–36.

Fisher, Max. 2022. *The chaos machine: The inside story of how social media rewired our minds and our world*. New York: Little, Brown and Company.

Franck, Thomas. 1992. The emerging right to democratic governance. *American Journal of International Law* 86: 46–91.

Freelon, Deen und Tetyana Lokot. 2020. Russian disinformation campaigns on Twitter target political communities across the spectrum. Collaboration between opposed political groups might be the most effective way to counter it. *Harvard Kennedy School Misinformation Review*

https://misinforeview.hks.harvard.edu/article/russian-disinformation-campaigns-on-twitter/ Accessed 24 July 2023.

Frimer, Jeremy A., Harinder Aujla, Matthew Feinberg, Linda J. Skitka, Karl Aquino, Johannes C. Eichstaedt, and Robb Willer. 2022. Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science* 194855062210838.

Gilbert, Margaret. 1992. *On social facts*. Princeton: Princeton University Press.

———. 2006. *A Theory of political obligation: Membership, commitment, and the bonds of society*. Oxford: Oxford University Press.

———. 2013. *Joint commitment: How we make the social world.* Oxford: Oxford University Press.

Goyanes, Manuel, Porismita Borah, and Homero Gil de Zúñiga. 2021. Social media filtering and democracy: Effects of social media news use and uncivil political discussions on social media unfriending. *Computers in Human Behavior* 120: 106759.

Great Britain and Intelligence and Security Committee. 2020. *Russia: Presented to Parliament pursuant to section 3 of the Justice and Security Act 2013*. https://isc.independent.gov.uk/wp-content/uploads/2021/03/CCS207_CCS0221966010-001_Russia-Report-v02-Web_Accessible.pdf Accessed 24 July 2023.

Haidt, Jonathan and Chri Bail. n.d. Social media and political dysfunction: A collaborative review. https://jonathanhaidt.com/social-media/ Accessed 17 July 2023

Halberstam, Yosh, and Brian Knight. 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics* 143: 73–88.

Hardin, Russell. 2002. *Trust and trustworthiness*, New York, NY: Russell Sage Foundation.

Harris, John Richard, and Richard Galvin. 2012. 'Pass the cocoamone, please': Causal impotence, opportunistic vegetarianism and act-utilitarianism. *Ethics, Policy & Environment* 15: 368–83.

Hausman, Daniel M. 2007. Group risks, risks to groups, and group engagement in genetics research. *Kennedy Institute of Ethics Journal*. 17.4: 351–69.

Hawley, Katherine. 2014. Trust, distrust and commitment. *Noûs*, 48: 1–20.

———. 2017. Trustworthy groups and organizations. In *The Philosophy of Trust,* edited by Faulkner, Paul and Thomas Simpson, 52–70. Oxford: Oxford University Press.

Haywood, William Dudley. 1929. *The autobiography of Big Bill Haywood*. New York: International Publishers.

Hetherington, Marc J. 2008. Turned off or turned on? How polarization affects political engagement. *Red and Blue Nation*, 54.

Hetherington, Marc J., and Jason A. Husser. 2012. How trust matters: The changing political relevance of political trust. American Journal of Political Science 56: 312–25.

Hetherington, Marc J., and Thomas J. Rudolph. 2017. Political trust and polarization. In *The Oxford handbook of social and political trust*, edited by Eric M. Uslaner. Oxford: Oxford University Press.

Hong, Sounman, and Sun Hyoung Kim. 2016. Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly* 33: 777–82.

Horwitz, Jeff, and Deepa Seetharaman. 2020. Facebook executives shut down efforts to make the site less divisive. *The Wall Street Journal*

https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499. Accessed 17 July 2017.

Howard, Philip, Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. 2019. The IRA, social media and political polarization in the United States, 2012-2018. U.S. Senate Documents, October. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs

Illing, Sean. 2016. Why social media is terrible for multiethnic democracies. *Vox*. November 15, 2016. https://www.vox.com/policy-and-politics/2016/11/15/13593670/donald-trump-social-media-culture-politics.

Jones, Karen. 2012. Trustworthiness. *Ethics*, 123: 61–85.

Kagan, Shelly. 2011. Do I make a difference? *Philosophy & Public Affairs* 39: 105–41.

Kant, Immanuel. 2007a. *Anthropology, history, and education*, eds. Günter Zöller and Robert B. Louden. Cambridge University Press.

Korsgaard, Christine. 2009. *Self-Constitution: Agency, identity, and integrity*. Oxford: Oxford University Press.

Kingston, Ewan, and Walter Sinnott-Armstrong. 2018. What's wrong with joyguzzling? *Ethical Theory and Moral Practice* 21: 169–86.

Kirby, Nikolas, Andrew Kirton, and Aisling Crean. 2018. Do corporations have a duty to be trustworthy? *Journal of the British Academy* 6: 75–129.

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyham, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein, Emily Thorson, Duncan Watts, and Jonathan Zittrain. 2018. The science of fake news. *Science* 359: 1094–96.

Lees, Jeffrey, and Mina Cikara. 2021. Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376: 20200143.

León, Ernesto de, and Damian Trilling. 2021. A sadness bias in political news sharing? The role of discrete emotions in the engagement and dissemination of political news on Facebook. *Social Media + Society* 7:20563051211059710.

List, Christian and Pettit, Philip. 2011. *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.

Ludwig, Kirk. 2016. *From individual to plural agency: Collective action: Volume 1*. Oxford: Oxford University Press.

Lo Re, Stefano. 2022. The glowing screen before me and the moral law within me: A Kantian duty against screen overexposure. *Res Publica*:1-21.

Margalit, Avishai, and Joseph Raz. 1990. National self-determination. *The Journal of Philosophy* 87: 439–61.

May, Larry. 2000. *Crimes against humanity: A normative account*, Cambridge: Cambridge University Press.

McLaughlin, Timothy. How Facebook's rise fueled chaos and confusion in Myanmar. *Wired*. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/ Accessed 17 July 2023.

Miles, Tom. 2018. U.N. investigators cite Facebook role in Myanmar crisis. *Reuters*. https://www.reuters.com/article/us-myanmar-rohingya-facebook-idUKKCN1GO2PN Accessed 17 July 2023.

Montanaro, Domenico. 2021. Most Americans trust elections are fair, but sharp divides exist, a new poll finds. *NPR*. https://www.npr.org/2021/11/01/1050291610/most-americans-trust-elections-are-fair-but-sharp-divides-exist-a-new-poll-finds. Accessed August 17, 2022

Mosleh, Mohsen, Cameron Martel, Dean Eckles, and David G. Rand. 2021. Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences* 118: e2022761118.

Narveson, Jan, 1991. Collective rights? *Canadian Journal of Law and Jurisprudence* 4: 329–345.

Nguyen, C. Thi. 2020. Echo chambers and epistemic bubbles. *Episteme* 17: 141–61.

———. (forthcoming). Trust as an unquestioning attitude. *Oxford Studies in Epistemology*.

Nguyen, C. Thi, and Bekka Williams. 2020. Moral outrage porn. *Journal of Ethics and Social Philosophy* 18:147-72.

Newport, Cal. 2019. *Digital minimalism: Choosing a focused life in a noisy world.* New York: Portfolio.

Norcross, Alastair. 2004. Puppies, pigs, and people: Eating meat and marginal cases. *Philosophical Perspectives* 18: 229–245.

Odell, Jenny. 2019. *How to do nothing: Resisting the attention economy*. United Kingdom: Melville House.

Orlowski, Jeff. 2020. We need to rethink social media before it's too late. We've accepted a Faustian bargain. *The Guardian*. https://www.theguardian.com/commentisfree/2020/sep/27/social-dilemma-media-facebook-twitter-society Accessed 17 July 2023.

Pariser, Eli. 2011. *The filter bubble: What the Internet is hiding from you*. London: Penguin UK.

Peter, Fabienne. 2017. Political legitimacy. In *The Stanford Encyclopedia of Philosophy* edited by Edward N. Zalta. https://plato.stanford.edu/archives/sum2017/entries/legitimacy/ . Accessed August 17, 2022.

Pew Research Center. 2017. Critical posts get more likes, comments, and shares than other posts. https://www.pewresearch.org/politics/2017/02/23/partisan-conflict-and-congressional-outreach/pdl-02-23-17_antipathy-new-00-02/ Accessed 17 July 2023.

Pogge, Thomas W. 1992. Cosmopolitanism and sovereignty. *Ethics* 103: 48–75.

Purves, Duncan and Davis, Jeremy. 2022. Public trust, institutional legitimacy, and the use of algorithms in criminal justice. *Public Affairs Quarterly* 36:136-62

Reglitz, Merten. 2022. Fake news and democracy. *Journal of Ethics and Social Philosophy* 22: 162-187.

Rathje, Steve, Jay J. Van Bavel und Sander van der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118: e2024292118.

Reuters. 2018. Myanmar: UN blames Facebook for spreading hatred of Rohingya. *The Guardian*. https://www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya . Accessed 17 July 2023.

Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida und Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–141.

Ross, Robert M, David Gertler Rand und Gordon Pennycook. 2019. Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. Preprint. PsyArXiv.

Rubel, Alan, Clinton Castro, and Adam Pham. 2021. *Algorithms and autonomy: The ethics of automated decision systems*. Cambridge: Cambridge University Press.

Sabatini, Fabio, and Francesco Sarracino. 2019. Online social networks and trust. *Social Indicators Research* 142: 229–60.

Sasahara, Kazutoshi, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2021. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science* 4: 381–402.

Searle, John, 1990. Collective intentions and actions. In *Intentions in communication*, edited by P. Cohen, J. Morgan, and M. Pollack, 401–415. Cambridge: MIT Press.

Singer, Peter. 1980. Utilitarianism and vegetarianism. *Philosophy & Public Affairs* 9: 325–337.

Sinnott-Armstrong W. 2005. It’s not my fault: global warming and individual moral obligations. In *Perspectives on Climate Change: Science, Economics, Politics, Ethics*, eds. Sinnott-Armstrong W., Howarth R., 285–307. Amsterdam: Elsevier.

Stecklow, Steve. 2018. Hatebook: Inside Facebook’s Myanmar operation. *Reuters*. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ Accessed 17 July 2023.

Sterba, James P. 2009. *Affirmative action for the future*. Ithaca: Cornell University Press.

Sunstein, Cass R. 1999. The law of group polarization. SSRN Scholarly Paper. Rochester, New York.

———. 2009. *Going to extremes: How like minds unite and divide*. Oxford: Oxford University Press.

Sunstein, Cass, Daniel Kahneman, and David Schkade. 2000. Deliberating about dollars: The severity shift empirical study. *Columbia Law Review* 100: 1139-75.

Talisse, Robert B. 2019. *Overdoing democracy: Why we must put politics in its place*. Oxford: Oxford University Press.

Tasioulas, John and Guglielmo Verdirame. 2022. Philosophy of international law. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/sum2022/entries/international-law/ . Accessed 24 July 2023.

Taub, Amanda and Max Fisher. Facebook fueled anti-refugee attacks in Germany, new research suggests. *The New York Times*. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html Accessed 17 July 2023.

Timberg, Craig and Dwoskin, Elizabeth. 2018. Russians got tens of thousands of Americans to RSVP for their phony political events on Facebook. *The Washington Post*. Accessed August 15, 2022. https://www.washingtonpost.com/news/the-switch/wp/2018/01/25/russians-got-tens-of-thousands-of-americans-to-rsvp-for-their-phony-political-events-on-facebook/. Accessed 24 July 2023.

Tufekci, Zeynep. How social media took us from Tahrir Square to Donald Trump. *MIT Technology Review*. https://www.technologyreview.com/2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/ Accessed 17 July 2023.

Tuomela, Raimo. 2013. *Social ontology: Collective intentionality and group agents*. Oxford: Oxford University Press.

Tyler, Tom. 2006. *Why people obey the law*. Princeton: Princeton University Press.

Tyler, Tom and Huo Y. 2002. *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.

Tyler, Tom and Jackson, J. 2014. Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law* 20: 78–95.

Uslaner, Eric M. 2018. *The Oxford handbook of social and political trust*. Oxford University Press.

U.S. Senate. Committee on Intelligence. 2020. Russian active measures campaigns and interference in the 2016 U.S. election. Volume 2: Russia's use of social media with additional views. URL= https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures

U.S. Senate. Press Release. Senate Intel Committee Releases Bipartisan Report on Russia's Use of Social Media. URL= https://www.intelligence.senate.gov/press/senate-intel-committee-releases-bipartisan-report-russia%E2%80%99s-use-social-media Accessed August 17, 2022.

Van Bavel, Jay J., Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. 2021. How social media shapes polarization. *Trends in Cognitive Sciences* 25: 913–16.

Velleman, J. David, 1997. How to share an intention. *Philosophy and Phenomenological Research* 57: 29–50.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359: 1146–51.

Walzer, Michael. 1977. *Just and unjust wars*. New York: Basic Books.

Wang, S.-Y. N., & Inbar, Y. 2022. Re-examining the spread of moralized rhetoric from political elites: Effects of valence and ideology. *Journal of Experimental Psychology: General* 151: 3292–3303.

Weijer, Charles; Goldsand, Gary; and Emanuel, Ezekiel. 1999. Protecting communities in research: Current guidelines and limits of extrapolation. *Nature Genetics* 23: 275–80.