Misinformation and disagreement

Ritsaart Reimann
Mark Alfano

1 Introduction

This chapter addresses the relationship between misinformation and disagreement. We begin by arguing that one traditional bogeyman in this domain, ideological polarization, does not account for the many problems that have been documented. Instead, affective polarization seems to be the root cause of most of these problems. We then discuss the relationships between moral outrage, misinformation, and affective polarization. We next turn to the political implications of affective polarization and conclude by discussing some potential solutions to the problems that arise in this area.

A preliminary point worth noting is that we here focus on the case of the United States, partly because it is among the best-documented and partly because recent events such as the Trump Presidency and the 6 January 2021 insurrection (and the ongoing spread of the QAnon conspiracy theory) make it an object of pressing concern. It may be possible to generalize to other countries, especially countries such as Australia and the United Kingdom that have also been affected by Rupert Murdoch's media empire, but we are not in a position to say anything decisive about the rest of the world.

2 The relationship between misinformation and disagreement

In this section, we discuss two views of the relationship between misinformation and disagreement. We argue against the received view, which focuses on ideological polarization, and for what we call the revised view, which focuses instead on affective polarization.

2.1 The received view

Because a well-informed citizenry is critical for a healthy democracy, misinformation poses a threat to the democratic process (Kuklinski et al., 2000).[1] Without a shared

---

understanding of scientific, social, and political reality, the factual foundation on the basis of which democratic deliberation ought to proceed is compromised (Baker, 2001). Thus, the widespread circulation of falsehoods undermines our ability to reach political agreement (Benkler et al., 2018). Under these conditions, the policy positions of opposing partisans pull apart due to fundamentally different and incongruent interpretations of the state of the world (Tucker et al., 2018).

If this synopsis is right, and misinformation really is the primary driver of *ideological* polarization, then much of today's disagreement should be explicable in terms of our changing media diet. Indeed, the charge that misinformation proliferates online – and that social media in particular are to blame for our contemporary political climate – is commonplace. In support of this position, critics cite the lack of traditional media norms and gatekeeping mechanisms online (Tucker et al., 2018); the emergence of economic incentives for misinformation profiteers (Bakir & Mcstay, 2017); the rise of hyperpartisan online news outlets (Faris et al., 2017); the formation of virtual echo-chambers (Cinelli et al., 2021); and the ascendency of algorithms optimized for engagement as opposed to informational veracity (Pariser, 2011; Alfano et al., 2020).

Despite an initially flurry of enthusiasm for these types of explanations, recent studies give reason for pause. According to Benkler et al. (2018), traditional media are still the most prevalent source of political information for most of the US population. Relatedly, Shearer and Gottfried (2017) find that only 20% of US adults frequently receive news via social media. Moreover, a recent analysis by Alcott and Gentzkow (2017) established that the average social media user encountered just 1.14 fake-news articles during the run-up to the 2016 US Presidential election. Hence, it appears that both our dependency on social media as a source of political information and the prevalence of misinformation within this new media ecosystem have been overstated.

Likewise, with respect to echo-chambers and filter-bubbles, popular concerns seem exaggerated. Starting with filter-bubbles, Bakshy et al. (2015) argue that the Facebook algorithm filters out less than 10% of ideologically dissonant content and that one in five newsfeed articles expose users to opinions of the opposing party.[2] Relatedly, Flaxman et al. (2016) find that conservative users on average consume just 11% more conservative content than liberals, and

---

[2] We note, however, that Bakshy was a Facebook employee when this study was published.

that liberals similarly consume only 11% more liberal content than their conservative counterparts. In addition, online political discussions are rife with disagreement (Duggan & Smith, 2016), and there even appears to be more disagreement online than in face-to-face settings (Barnidge, 2017). Taken together, these studies support the more general observation that social media users frequently encounter cross-cutting content and are rarely embedded in ideologically segregated networks. In fact, according to a recent study by Dubois & Blank (2018), at most 8% of the online population is trapped in any kind of echo-chamber.

Yet other research challenges the very phenomenon that is purportedly being explained. According to Fiorina et al. (2004), today's average American citizen remains a centrist rather than an extremist on most issue points. Likewise, Lelkes (2016) and Finkel et al. (2020) find that, at the mass level, Americans have neither become more extreme nor more consistent in their policy positions. Various in-depth analyses of ANES and PEW data – in some cases dating back to the 1970's – confirm these results and show that contemporary Democrats and Republicans are no more polarized, ideologically, than they have been throughout the past few decades (Jamieson & Cappella, 2015; Mason, 2018).

Despite the temptation to thus conclude that democracy is alive and well, and that our contemporary media environment poses no imminent threat to the prospect of reaching political agreement, there is an alternative account to be considered.

2.2 The revised view

An assumption of the received view is that politics is a rational practice in which sophisticated, self-interested individuals vote for whichever party promises to enact policies that are most conducive to their preferences (Mason, 2018). In contrast to this policy-centric model, a growing number of political theorists and psychologists argue that much of our political behaviour is driven by a desire to form emotionally-bonded coalitions (Tajfel & Turner, 1979; Bavel & Pereira, 2018). The idea that our sense of self derives in large part from the groups to which we belong has become increasingly popular (Goleman, 2006). One implication of this view is that, as fundamentally social creatures, we are evolutionarily primed to notice and respond to cues of group membership (Kerr & Levine, 2008). Accordingly, our affective system elicits strong emotional responses to others based on whether those others belong to our group (Sherif et al., 1988). With respect to politics, this theory predicts that political behaviour is

driven primarily by affective party affiliation, and only in part, if at all, by policy considerations (Bavel & Pereira, 2018).

Thus, despite equivocal evidence for the *received* account of *ideological* polarization, it may be the case that Americans are becoming increasingly *affectively* polarized across party lines. We call this the *revised* view. In support of this position, Iyengar et al. (2019) cite that the rate at which partisan parents loathe the prospect of their child marrying someone from the opposing party has risen from just 4-5% in the 1960s, to between 30-50% in 2010. Utilizing trust and dictator games, Carlin and Love (2013) find that copartisans consistently award greater financial allocations to one another and are more willing to impose penalties on members of the opposing group. Similarly preferential tendencies have been found across a range of domains, including evaluations of job applicants (Gift & Gift, 2015) and online dating behaviour (Huber & Malhorta, 2017). Taken together, these studies reflect a general trend of increasing interpartisan dislike, which, when measured in terms of the "feeling thermometer" (ANES), has risen from 22.64 degrees in 1978, to over 40 degrees in 2016 (Iyengar et al., 2019).

This prompts three central questions. First, what is the relation between misinformation and affective polarization? Second, what are its political implications? Finally, what can be done to stem this rising tide of resentment? Before turning to each of these, we retrace several important regulatory, social, and technological transformations that transpired in the United States throughout the second half of the twentieth century

In the fall of 1950, the American Political Science Association published a report that lamented the lack of disagreement in America's largely centrist political landscape. Concerned with an disinterested and disengaged electorate, it recommended that the two major parties distance themselves from the centre and each other by developing more distinct and coherent ideologies (Mason, 2018; Finkel et al., 2020). Around this time both the Republican and Democratic party featured influential conservative and liberal factions. In fact, the dominant alignment of Republicans with conservatism and Democrats with liberalism didn't get underway until the 1960s.

A popular explanation of this realignment of party and ideology centres on the Civil Rights Act of 1964 and President Lyndon Johnson's decision to align the Democratic party with its liberal, Northern, pro-civil rights contingent (Sundquist, 1983). In so doing, Johnson alienated Southern and conservative Democrats who subsequently defected to the Republicans. This event

thus sorted Democrats and Republicans along three lines: racial, ideological, and geographic. In subsequent decades, greater religious sorting, increased urbanization among Blacks, and a growing differential in mean Republican and Democratic incomes further reinforced these partisan cleavages (Mason, 2018).

Mason (2018; Iyengar et al., 2019; Finkel et al., 2020) suggests that this process of partisan sorting is intimately related to affective polarization. Her argument proceeds in two steps. First, increased partisan sorting decreases interpartisan interactions through membership in 'cross-cutting-cleavages', resulting in fewer shared and positively charged superordinate identities. This is concerning insofar as it is these sorts of superordinate identities that have been shown to reduce out-group animus (Nordlinger, 1972). Second, when a series of social identities such as race and religion align with partisan affiliation, these identities reinforce one another and become enmeshed in a single, overarching identity.

From this, two things follow. First, the more our identities align and interlock, the more emotionally invested we become in maintaining them (Roccas & Brewer, 2002). Second, the fact that our identities are aligned in this way implies that we perceive as threatening anyone who deviates along any one of these party lines, transfiguring political animosity into social, religious, and racial resentment. In support of these hypotheses, Mason (2018) shows not only that overall levels of anger have increased in tandem with increased levels of sorting, but that these increases are especially pronounced among society's most well-sorted members. Specifically, whereas only 11% of people with cross-cutting identities voice any kind of political animosity, 49% of those with well-sorted identities report feeling angry at their political opponents.

Independently, the second half of the twentieth century witnessed the adoption of new communication technologies. Though the details of these developments escape the scope of this paper, we emphasize that they reshaped the economic incentives for media production in such a way that it became both *possible* and *profitable* to produce highly-targeted content for small and homogenous audiences. Prior to the introduction of these new technologies, media production was costly, and circulation was constrained by limited broadband capacity (Postman, 1985). Hence, there were only a handful of media outlets, each of which had an incentive to produce neutral and generally appealing content so as to capture the largest possible market share (Baker, 2001).

In addition, the American media ecosystem used to be heavily regulated. Recognizing the fourth estate as an indispensable organ of democracy, the Federal Communications Commission (FCC) explicitly pursued policy that it perceived to be in the public interest. For instance, the FFC fairness doctrine – introduced in 1949 – not only mandated that broadcasters discuss controversial issues of public importance, but that they present these issues in an honest, fair, and balanced fashion (Baker, 2001; Benkler, 2018). Under growing pressure from both private and political interests (Herman & Chomsky, 1988), the FCC and Congress loosened regulations. This process of deregulation culminated in the Telecommunications Act of 1996, which transformed the American media from a public good into a capitalist consortium.

In conjunction with the aforementioned process of social sorting, these technological and regulatory changes have had a profound impact both on the type of media content that is produced and the patterns in which it is consumed. Specifically, it has enabled, permitted, and incentivized media producers to target highly sorted and profitable niches with ideologically, politically, and culturally congenial content (Baker, 2001; Benkler, 2018). Moreover, and faced with increased competition in a profit-driven market, informational veracity has been sacrificed for entertainment value (Postman, 1985; Baker, 2001). Enter outrage and the proliferation of misinformation.

3 Moral outrage, misinformation, and affective polarization

Misinformation pertains to a much broader set of communications than blatantly false representations of facts (Postman, 1985; Tucker, 2018; Benkler et al., 2018). For our purposes, any type of communication that intentionally or inadvertently misleads or misrepresents the state of the world through patent falsehoods, exaggerations, or decontextualization counts as misinformation. While proponents of the received view are likely concerned with the *ideological* implications of *factual misinformation*, our analysis focuses on the *affective* implications of *partisan misrepresentations of the values and character of their adversaries*.

Specifically, we focus on what Berry & Sobierja (2018) have dubbed "The Outrage Industry" and argue that outrage purveyors contribute to affective polarization through moralization, vilification, and misrepresentation of their political opponents. In addition to enlisting falsehoods, these provocateurs have mastered the art of hyperbole and

decontextualization. Furthermore, any apparent policy positions that these pundits adopt are generally thinly-veiled segues into ad hominem attacks and outright character assassinations.

While it seems improper to characterize this style of political discourse as subtle, it does introduce a more subtle form of misinformation. Consider, for instance, the following statement by the late Rush Limbaugh: "[There are] gazillions of similarities between National Socialism in Germany and Obama's healthcare plan… Nobody is saying that Obama is Hitler… What we're saying is that his healthcare plan mirrors Nazi Germany's" (from Berry & Sobierja, 2018, p.49). With statements like these, assessing factual veracity is beside the point. Instead, what we want to highlight are the characteristic features and consequences of this style of communication. In addition to hyperbole (e.g., 'gazillions'), Limbaugh positions this attack on the Affordable Care Act in the context of one of history's most violent and morally charged atrocities. In so doing, he prepares his audience for a comparison between Obama and Hitler, which, even despite his explicit disclaimer, clearly amounts to an attack on Obama's character. More than this, it misrepresents and vilifies Obama not merely as a misguided politician, but as a *moral enemy.*

To fully appreciate the political implications of this kind of moral vilification, we note that moral beliefs are generally held with great conviction, elicit strong emotional responses, and tend to be evaluated in absolute terms (Hare, 1981; Skitka & Mullen, 2002). Accordingly, these beliefs are less subject to revision and compromise than non-moral beliefs, and increase animosity among attitudinally dissimilar others (Tetlock et al., 2000; Skitka et al. 2005). In a landmark study by Skitka et al. (2005), perceived moral difference was causally linked to intolerance and a desire for social distance. Hence, rather than encourage partisans to engage in constructive political dialogue, moral outrage galvanizes intergroup anger and conflict (Jamieson & Cappella, 2015; Mason, 2018). Moreover, in depicting members of the opposing party as immoral, it reinforces an 'us' versus 'them' dynamic that increases the salience of partisan identities.

On this score, research suggests that, as identity becomes more salient, identity-related needs increasingly outweigh the need for accuracy (Kruglanski, 2004). Under these conditions, beliefs about political figures, facts, and policies are likely to be distorted by identity-confirming biases (Bavel & Periara, 2018). Accordingly, people are prone to uncritically accept and dogmatically reject information that secures or threatens their identities, respectively (Kruglanski, 2004; Bavel & Periara, 2018). If this is right, mere exposure to moral outrage may

exacerbate our susceptibility to false beliefs regarding the moral values and character of our political opponents.

While readers outside the USA may be forgiven for thinking that Limbaugh must have been some sort of anomaly, he was the single most popular talk radio host in American history (Jamieson & Cappella, 2015). Besides drawing 14 million weekly listeners, his show introduced an 'outrage template' that has since been picked up by many equally outrageous personalities, at least five of whom now each draw more than 8 million listeners per week (Berry & Sobierja, 2018). In addition to providing a template for talk radio, Fox News – America's most watched cable news network – has amplified Limbaugh's moralized take on politics via cable TV.

More recently, the phenomenon of online outrage has come into focus (Crockett, 2017; Brady et al., 2017; Mooijman et al., 2018). Looking at Twitter, Brady et al. (2017) find that for each additional moral-emotional word contained in a tweet, the likelihood of that tweet being retweeted increases by 20%. These results have since been replicated across 27 studies by five independent labs, with an average effect size of 1.15 (Brady & Van Bavel, 2021). Worth noting is that these studies distinguish between purely emotional, purely moral, and mixed moral-emotional language and find that neither purely moral nor purely emotional language seems to drive message diffusion, suggesting that it is the unique combination of moral *and* emotional language – characteristics of moral outrage – that enthrals our passions.

To contextualize this outrage, it's worth reflecting on the fact that humans are hyper-social and tend to form coalitions that engage in intergroup conflict (Tajfel & Turner, 1979; Goleman, 2006). From this perspective, psychologists have advanced various theories of morality that posit moral norms as biological and cultural adaptations that help coordinate group behaviour (Haidt, 2013; Curry, 2016). A recurring theme is that these norms facilitate collective action by establishing clear rules about how individuals ought to behave in certain cooperative contexts.

Nevertheless, given the incentives for defection in such contexts (e.g., freeriding), cooperation remains a risky strategy. To mitigate this risk, individuals rely on signals of one another's normative commitments (Everett et al., 2016; Jordan & Rand 2020). Moreover, for these signals to be effective, they must be costly, which is to say, impossible to fake or more expensive to send when they are false than when they are true (Mercier, 2020). Expressing moral outrage is a costly signal that communicates our commitment to a shared set of moral and

cooperative norms. First, outrage is hard to fake because it is an emotional reaction, the sincerity of which humans are primed to detect (Levy, 2020). Second, expressing outrage at the moral transgressions of others incurs considerable cost: specifically, the risk of retaliation (Crockett, 2017; Jordan & Rand 2020).

Hence, moral outrage cuts both ways. On the one hand, it signals our commitment to ingroup members (Brady & van Bavel, 2021). At the same time, it signals our animosity toward and willingness to punish anyone who deviates from our conception of the good (Haidt, 2013). In the context of partisan politics, outrage thus emerges as a uniquely effective way of signalling and maintaining our social identity (Brady et al., 2020).

Reflecting on these mechanisms in the context of social media, Crockett (2017) argues that computer-mediated communication reduces the costs associated with outrage. Not only does expressing outrage online require less energy, it also circumvents the risk of physical retaliation. Relatedly, while our tendency to express outrage in face-to-face settings is attenuated by empathic concern, we tend to be emotionally indifferent to online avatars. Social media also increases the payoffs of expressing outrage by digitally encoding our moral righteousness for everyone in the network to see. According to Brady et al. (2020), social media platforms also feature inherent technological affordances that encourage this moralized style of discourse. Given that these sites are optimized for engagement, it stands to reason that morally provocative content is algorithmically prioritized in users' newsfeeds. Additionally, insofar as 'likes' and 'retweets' are rewarding, receiving feedback of other people's approval may reinforce the tendency to express these kinds of moral emotions.

While it is difficult to quantify the effects of moral outrage on outgroup animosity, Jamieson & Cappella (2015) find that increased exposure to Rush Limbaugh is positively correlated with interpartisan loathing. Operationalizing an intermediary construct known as perceived polarization (Lelkes, 2016; Enders & Armaly, 2019), various other studies have likewise established positive correlations between media exposure and affective polarization (Yudkin et al., 2019; Wilson et al., 2020). Even though these effects attend all types of media consumption, Yudkin et al. (2019) find that they are most pronounced among consumers of partisan media. Specifically, Republicans who rely on outrage outlets like the Drudge Report, conservative talk radio, and Fox News report feeling more antagonistic towards members of the opposing party.

4 The political implications of affective polarization

As outrage proliferates, America's political landscape begins to resemble a tribal battleground. While this may seem overblown, a recent study by Mooijman et al. (2018) found that expressions of outrage predict political violence. Scholars are increasingly concerned about the downstream consequences of interpartisan animosity (Westfall et al. 2015; Iyengar & Westwood, 2015; Wilson et al., 2020). A recurring theme is that, as identity supersedes policy, political compromise becomes nearly impossible.

First, increased salience of partisan identity may incentivise political elites to endorse ideologically popular as opposed to democratically negotiable policies (Westfall et al., 2015; Wilson et al., 2020). Second, this palpable hostility among rank-and-file partisans potentially signals that any representative found willing to cooperate across party lines will be seen as an appeaser (Iyengar & Westwood, 2015).[3] An in-depth analysis of the US Congress by Andris et al. (2015) buttresses these concerns in finding that the number of 'cross-party co-operators' in the House of Representatives is at an all-time low. This kind of political gridlock and unwillingness to compromise in turn impedes the government's ability to respond to emergent social, medical, environmental, and economic problems.

Before considering interventions, we draw attention to one further implication of outrage and affective polarization that has hitherto been treated as somewhat of an afterthought: trust. In line with our proposal that outrage is a costly signal of group membership, other research has found that moral outrage is a reliable signal of trustworthiness (Jordan & Rand, 2020). Furthemore, the authors determined that 'deliberativeness' attenuates our willingness to express outrage when there is no reputational benefit to be gained from doing so. This suggests that even though outrage may be a genuine emotional response under many circumstances (Haidt, 2013; Levy, 2020), people are also attentive to its instrumental value (Rom & Conway 2018). With respect to affective polarization, these results raise the possibility that moral outrage is strategically employed by both politicians and media provocateurs to induce interpartisan distrust. Doing so is in the interest of these actors because out-group distrust effectively inoculates in-group members from information that challenges the group's identity and leadership (Herman & Chomsky, 1988; Jamieson & Cappella, 2015; Benkler et al., 2018).

---

[3] Think, for example, of the epithet 'RINO', which stands for "Republican in Name Only."

Thus, moral outrage may exacerbate the effects of affective polarization by making partisans even less likely to reach compromise. In an environment characterized by distrust and a sense of moral as opposed to merely political or social antagonism, appeals to fact will not foster agreement. In important respects, trust precedes the very possibility of forming and updating beliefs on the basis of testimony (Adler, 1994). Thus, without the belief that others have some regard for our interests, we are disinclined to consider anything they say. This suggests that reversing the rising tide of affective polarization hinges on reshaping how partisans perceive one another, and rebuilding interpartisan trust.

5 Possible solutions

In view of equivocal evidence for increasing ideological extremity among the American electorate, we have argued that polarization is primarily an affective phenomenon. In so doing, we shelved the problem of factual misinformation and instead identified moral misinformation, in the form of the proliferation of moral outrage, as an important driver of political disagreement. Moreover, rather than focus exclusively on social media, we showed that this process of affective polarization can be traced to several regulatory, social, and technological transformations that occurred throughout the latter half of the twentieth century, at least in the well-studied US context.

In line with this revised account, the solutions reviewed here target a broader set of culprits, including political elites and traditional media outlets. Relatedly, the interventions we propose are less concerned with counteracting factual misinformation and aim instead at upending partisan sorting, dislike, and distrust.

Starting with the APSAs' 1950 report, we emphasize that even though increased ideological sorting helps voters navigate political complexity (Finkel et al., 2020) and increases political participation (Mason, 2018), it also impedes the prospect of political agreement. Thus, we suggest that political elites and parties moderate their policy positions (Iyengar & Westwood, 2015; Wilson et al., 2020). This is especially relevant for Republican elites, as it is the Republican party that has strayed furthest from the centre (Hare & Poole 2014; Benkler et al., 2018). Besides making compromise more likely at the institutional level, such moderation could also prompt reduced animosity among rank-and-file partisans, who have been shown to take cues from elite representatives (Enders & Armaly, 2019; Wilson et al., 2020).

With respect to elite cues, traditional media outlets play an equally central role (Benkler et al., 2018). A recent study by Levendusky & Malhotra (2016) found that news stories about polarization have increased by roughly 20% over the past 20 years. Making polarization more salient in the public eye has in turn been linked to increased partisanship and interpartisan animosity (Levendusky, 2013). On the bright side, other studies suggest that activating superordinate identities potentially increases positive sentiment across party lines (Van Bavel & Pereira, 2018; Brady et al., 2020). Relatedly, and even though debunking factual misinformation appears ineffective if not counterproductive (Walter & Murphy, 2018), correcting misperceptions about the ideological extremity of political opponents leads people to moderate their own policy positions and reduces outgroup animus (Ahler & Sood, 2018). Hence, we encourage media outlets to emphasize the average citizens' ideological neutrality and to reframe political issues in terms that highlight superordinate constructs such as 'American', 'human', etc.

Admittedly, the 'average ideologically neutral citizen' hardly makes for a compelling story. Yet this is precisely the point: addressing affective polarization requires that political journalism move away from sensationalism and towards norms that promote democratic deliberation. On this score, Benkler et al. (2018) emphasize that 'fair and balanced' reporting does not mean giving equal weight to all perspectives – including those who advocate conspiracy theories such as Pizzagate and QAnon. Rather, it implies giving equal weight to equally *valid* and *important* points of view. To this end, they suggest a shift from 'fair and balanced' to 'verifiable and accountable' reporting.

Regrettably, the economic incentives of a deregulated media ecosystem are unlikely to align with this vision of the fourth estate. While talk of media regulation invariably gives rise to legitimate free speech concerns, Baker (2001) outlines a battery of constitutionally admissible and politically desirable policies that at once safeguard the right to free speech and ensure that democracy does not succumb to the vitriol of profit-driven outrage media. Likewise, with respect to social media, free speech concerns should not get in the way of effective regulation. While America remains hesitant on this front, several European countries have begun holding social media giants such as Facebook accountable for the content posted on their platforms (Gorwa, 2021). In addition to encouraging other countries to follow suit, social media companies need to be forced to adjust their algorithms in accordance with democratic principles (for detailed

proposals, see Benkler et al., 2018; Brady et al., 2020), potentially via anti-trust regulations targeting their monopoly status (Alfano & Sullivan, 2021).

Finally, we echo Robert Putman's (2000) call for increased participation in non-political and ideologically cross-cutting social groups. Given that social sorting along party lines magnifies our political identities and fuels interpartisan animosity (Mason 2018; Finkel et al., 2020), we believe that escaping these partisan enclaves will help Americans set aside their differences (Nordlinger, 1972). More than this, increased interaction across the ideological divide can unfasten misconceptions of political opponents as moral enemies. In so doing, these cross-cutting cleaves can generate trust in our shared humanity, irrespective of ideology. Such interactions could be promoted via new housing and zoning regulations that encourage conservatives to move to denser, more diverse urban centers, as well as the construction of better public transport that would obviate the need for long commutes by car during which one listens to hours of far-right talk radio.

References

Adler, E. (1994). Testimony, Trust, Knowing. *Journal of Philosophy, 91*(5), 264-275.

Ahler, D. & Sood, G. (2018). The parties in our heads: misperceptions about party composition and their consequences. *The Journal of Politics, 80*(3), 964–981.

Alcott, H. & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 3*(2), 211-236.

Alfano, M., Carter, J. A., Ebrahimi Fard, A., Clutton, P., & Klein, C. (2020). Technologically scaffolded atypical cognition: The case of YouTube's recommender system. *Synthese*.

Alfano, M. & Sullivan, E. (2021). Online trust and distrust. In M. Hannon & J. de Ridder (eds.), *Handbook of Political Epistemology*. Routledge.

Andris, C., Lee, D., Hamilton, M., Martino, M., Gunning, C. & Selden, J. (2015). The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives. *PLoS ONE, 10*(4), 1-14.

Baker, E. (2001). *Markets, Media, and Democracy*. Cambridge University Press.

Bakir, V. & Mcstay, A. (2017). Fake News and the Economy of Emotions: Problems, Causes, Solutions. *Digital Journalism, 5*(10), 1-22.

Bakshy, E., Messing, S. & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132.

Barnidge, M. (2017). Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication 34*(2), 302-321.

Benkler, Y., Faris, R. & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics.* Oxford University Press.

Berry, J. & Sobieraj, S. (2018). *The Outrage Industry.* Oxford University Press.

Brady, W., Wills, J., Jost, J., Tucker, J. & Van Bavel, J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7318.

Brady, W., Crockett, M. & Van Bavel, J. (2020). The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. *Perspectives on Psychological Science, 15*(4), 978-1010.

Brady, W. & Van Bavel, J. (2021). The Effect Size of Moral Contagion. *OSF Preprint.*

Brady W. & Van Bavel, J. (2021). Social Identity shapes antecedents and functional outcomes of moral emotion expression in online networks. *OSF Preprint.*

Carlin, R. & Love, G. (2013). The politics of interpersonal trust and reciprocity: an experimental approach. *Political Behaviour, 35*(1), 43–63.

Cinelli, M., Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Science, 118*(9), 1-8.

Crockett, M. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771.

Curry, O. (2016). Morality as Cooperation: A Problem-Centred Approach. In T. Shackelford & R. Hansen (Eds.), *The Evolution of Morality* (pp. 27-51). Springer.

Dubois, E. & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729-745.

Duggan, M. & Smith, A. (2016). The Political Environment on Social Media. *Pew Research Center*. https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/

Enders, A. & Armaly, M. (2019). The Differential Effects of Actual and Perceived Polarization. *Political Behaviour, 41*(1)*,* 815-839.

Everett, J. A. C., Crockett, M. J, & Pizarro, D. A. (2016). Inference of Trustworthiness From Intuitive Moral Judgments. *Journal of Experimental Psychology, 145*(6), 772-787.

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E. & Benkler, Y. (2017). Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election. *Retrieved From:* https://ssrn.com/abstract=3019414

Finkel, E., Bail, C., Cikara, M., Ditto, P., Iyengar, S., Klar, S., Mason, L., McGrath, M., Nyhan, B., Rand, D., Skitka, L., Tucker, J., van Bavel, J., Wang, C. & Druckman, J. (2020). Political Sectarianism in America. *Science, 370*(6516), 533-536.

Fiorina, M., Abrams, S. & Popo, J. (2004). *Culture War? The Myth of a Polarized America.* Pearson Longman.

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

Gift K, Gift T. 2015. Does politics influence hiring? Evidence from a randomized experiment. *Political Behaviour, 37(*3), 653–75.

Goleman, D. (2006). *Social intelligence: The new science of human relationships.* Bantam Books.

Gorwa, R. (2021). Elections, institutions, and the regulatory politics of platform governance: The case of the German NetzDG. *Telecommunications Policy, 45*(6).

Haidt, J. (2013). *The Righteous Mind: Why Good People are Divided by Politics and Religion.* Penguin Press.

Hare, R. M. (1981). *Moral thinking.* Oxford University Press.

Hare, C. & Poole, K. (2014). The Polarization of Contemporary American Politics. *Polity, 46*(3), 411–429.

Herman, E. & Chomsky, N. (1988). *Manufacturing Consent: The Political Economy of the Mass Media.* Pantheon Books.

Huber, G. & Malhotra, N. (2017). Political homophily in social relationships: evidence from online dating behavior. *The Journal of Politics, 79*(1), 269–283

Iyengar, S. & Westwood, S. (2015). Fear and loathing across party lines: new evidence on group polarization. *American Journal of Political Science, 59*(3), 690–707.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhorta, N. & Westwood, S. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science, 22*(1)*,* 129-146.

Jamieson, K. & Cappella, J. (2015). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment.* Oxford University Press.

Jordan, J. & Rand, D. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology, 118*(1), 57–88

Kerr, N. & Levine, J. (2008). The Detection of Social Exclusion: Evolution and Beyond. *Group Dynamics: Theory, Research, and Practice, 12*(1), 39-52.

Kruglanski, A. (2004). *The Psychology of Closed Mindedness.* Psychology Press.

Kuklinski, J., Quirk, P., Jerit, J., Schwieder, D. & Rich, R. (2000). Misinformation and the Currency of Democratic Citizenship. *The Journal of Politics, 62*(3), 790-816.

Lelkes, Yphtach. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly, 8*(1), 392-410.

Levendusky, M. (2013). Why Do Partisan Media Polarize Viewers? *American Journal of Political Science, 57*(3), 611-623.

Levendusky, M. & Malhotra, N. (2016). Does Media Coverage of Partisan Polarization Affect Political Attitudes?" *Political Communication, 33*(2), 283-301.

Levy, N. (2020). Virtue signalling is virtuous. *Synthese.* Retrieved from: https://doi.org/10.1007/s11229-020-02653-9

Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity.* The University of Chicago Press.

Mercier, H. (2020). *Not Born Yesterday: The Science of Who we Trust and What we Believe*. Princeton University Press.

Mooijman, M., Hoover, J., Lin, Y., Ji, H. & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396.

Nordlinger, E. 1(972). *Conflict Regulation in Divided Societies*. Cambridge, MA: Center for International Affairs, Harvard University.

Pariser, E. (2011). *The Filter Bubble: What The Internet is Hiding From You.* Penguin Books.

Postman, N. (1985). *Amusing Ourselves to Death.* Penguin Books.

Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster.

Roccas, S. & Brewer, M. (2002). Social Identity Complexity. *Personality and Social Psychology Review, 6*(2), 88–106.

Rom, S. & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology, 74*(1), 24-37.

Shearer, E. & Gottfried, J. (2017). News Use Across Social Media Platforms 2017. *Pew Research Centre.* https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/

Sherif, M., Harvey, O., White, B., Hood, W. and Sherif, C. (1988). *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. Wesleyan University Press.

Skitka, L. & Mullen, E. (2002). The dark side of moral conviction. *Analyses of Social Issues and Public Policy, 2*(1)*,* 35–41.

Skitka, L., Bauman, W. & Sagris, E. (2005). Moral Conviction: Another Contributor to Attitude Strength or Something More? *Journal of Personality and Psychology, 88*(6), 895-917.

Sundquist, J. (1983). *Dynamics of the Party System: Alignment and Realignment of Political Parties in the United States*. Brookings Institution.

Tajfel, H. & Turner, J. (1979). An Integrative Theory of Intergroup Conflict. In W. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 83-93). Brooks/Cole.

Tetlock, P., Kirstel, O., Elson, B., Green, M. & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5)*,* 853–970.

Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Brendan, N. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *Retrieved From:* http://dx.doi.org/10.2139/ssrn.3144139

Van Bavel, J. & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences, 22*(3), 213-224.

Walter, N. & Murphy, S. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs, 85*(3), 423-441

Westfall, J., Van Boven, L, Chambers, J. & Judd, C. (2015). Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide. *Perspectives on Psychological Science, 10*(2), 145-158.

Wilson, A., Feinberg, M. & Parker, V. (2020). Polarization in the contemporary political and media landscape. *OSF Preprint*.

Yudkin, D., Hawkins, S. & Dixon, T. (2019). The Perception Gap: How False Impressions are Pulling Americans Apart. *OSF Preprint.*