doi:10.1017/S0003055421000885

The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis

ALEXANDER BOR *Aarhus University, Denmark*
MICHAEL BANG PETERSEN *Aarhus University, Denmark*

Why are online discussions about politics more hostile than offline discussions? A popular answer argues that human psychology is tailored for face-to-face interaction and people's behavior therefore changes for the worse in impersonal online discussions. We provide a theoretical formalization and empirical test of this explanation: the mismatch hypothesis. We argue that mismatches between human psychology and novel features of online environments could (a) change people's behavior, (b) create adverse selection effects, and (c) bias people's perceptions. Across eight studies, leveraging cross-national surveys and behavioral experiments (total N *= 8,434), we test the mismatch hypothesis but only find evidence for limited selection effects. Instead, hostile political discussions are the result of status-driven individuals who are drawn to politics and are equally hostile both online and offline. Finally, we offer initial evidence that online discussions feel more hostile, in part, because the behavior of such individuals is more visible online than offline.*

In 2010, Mark Zuckerberg was named "Person of the Year" by *Time* magazine. In the magazine's coverage, the mission of Zuckerberg was described in this way: "Facebook wants to populate the wilderness, tame the howling mob and turn the lonely, antisocial world of random chance into a friendly world" (Grossman 2010). Just a decade ago, the hope was that citizens would use social media to engage in civil discussions about important matters. These hopes were particularly high regarding discussions about politics (see e.g., Dahlberg 2001) as the anonymity of the Internet was seen as finally providing people with a context for the power-free communication emphasized in theories of deliberative democracy.

However, online discussions about politics turned out to be nasty, brutish, and not nearly short enough. According to a recent survey, two thirds of Americans have witnessed online harassment, and politics is one of the most toxic subjects on the Internet (Duggan 2017). Such findings are mirrored across Western democracies. For example, 80% of Danes, find that online discussions are dominated by "very extreme viewpoints," and 57% agree that some debaters are "so aggressive that you don't know what they are capable of" (TrygFonden 2017). The very features of online environments that we hoped would facilitate peaceful deliberations are now widely believed to be a core trigger of political hostility in online discussions. There appears to be a *hostility gap* where online discussions are felt as significantly more hostile than offline discussions.

In this article, we provide a comprehensive investigation of the potential individual-level causes of the hostility gap. We define political hostility as the use of intimidation in political discussions and ask why people experience more political hostility in online versus offline discussions. As online forms of political engagement become more and more widespread, this question is of increasing importance. Mapping the causes of online political hostility has thus been identified as a "key question for achieving impact on online harassment" (Matias 2016, 1).

This article focuses on one of the most common hypotheses informing both academic and popular discussions of online political hostility, what we refer to as *the mismatch hypothesis* (e.g., Baek, Wojcieszak, and Delli Carpini 2012; Cheng et al. 2017; Stein 2016). The mismatch hypothesis emphasizes the psychological consequences of differences between online and offline contexts. In other words, it focuses on how individuals behave differently in fundamentally similar political discussions online versus offline. In one of its simplest forms, the hypothesis entails that when people cannot see their discussion partner, even otherwise agreeable individuals struggle to contain their emotions, especially on contentious topics such as politics. Wolchover (2012), for example, argues that online hostility emerges from "a perfect storm" of "virtual anonymity and thus a lack of accountability, physical distance and the medium of writing." Or, in the words of Cheng et al. (2017), "anyone can become a troll."

While such narratives are popular, there is a dearth of empirical evidence. Thus, most previous treatments of online hostility have exclusively focused on online behavior, making it impossible to identify the unique role of the online context (e.g., Cheng et al. 2017; Gorrell et al. 2018). Meanwhile, none of the previous works comparing online and offline political communication has investigated hostility (Baek, Wojcieszak, and Delli Carpini 2012; Bisbee and Larson 2017; Gibson and Cantijoch 2013). Therefore, the objective of this manuscript is to take the first steps in the production of a viable, empirical research program on the psychological origins of online political hostility.

Alexander Bor, Postdoctoral Researcher, Department of Political Science, Aarhus University, Denmark, alexander.bor@ps.au.dk.
Michael Bang Petersen, Professor, Department of Political Science, Aarhus University, Denmark, michael@ps.au.dk.

Received: July 06, 2020; revised: March 30, 2021; accepted: July 12, 2021.

To facilitate this, we offer a sweeping approach, directed at drawing the overall contours of the research landscape and providing initial empirical evidence to guide further research. First, we place the mismatch hypothesis within a novel, overarching theoretical framework for the study of the hostility of online and offline political discussions. Second, focusing on the mismatch hypothesis, we outline the universe of possible theoretical mechanisms whereby the online environments can shape the psychology of political hostility by influencing either senders or receivers of political content. Third, we leverage eight empirical studies, documenting the cross-national existence of the hostility gap and provide a comprehensive test of the proposed mechanisms. Overall, however, these studies provide little support for the mismatch hypothesis. On the sender side, we demonstrate that individuals who are hostile online are just as hostile offline and just as eager to talk about politics online as offline. On the receiver side, we demonstrate that there are no systematic biases in people's perceptions of online messages. In tandem, this suggests that online hostility is not because online environments induce accidental failures of emotion regulation. Finally, we provide initial empirical tests of the predominant alternative to the mismatch hypothesis, which we refer to as the connectivity hypothesis. Consistent with this hypothesis, we find evidence that political hostility reflects a deliberate strategy among individuals with particular personality traits. These individuals participate intensively in political debates both online and offline, and the hostility gap simply reflects that their online activities are substantially more visible due to the public nature of online platforms.

A FRAMEWORK FOR RESEARCH ON THE HOSTILITY OF ONLINE AND OFFLINE POLITICAL DISCUSSIONS

Social media has created a unique ecosystem characterized by "a personalized, quantified blend of politically informative *expression*, *news*, and *discussion* that is seamlessly interwoven with non-political content" (Settle 2018, 15). Within this multitude of politically relevant online behaviors, we focus on discussions, understood as the direct exchange of arguments between a sender and one or more receivers. Discussions, defined in this way, occur in both online and offline environments. As argued by Settle (2018, 66), even if "there are very marked differences between face-to-face conversations and the interactions that appear [online], both capture the same fundamental construct: people exchanging ideas and opinion about politics with hopes of validating their own political views or informing or persuading others about their opinions." The notion of a hostility gap in both media and research takes this comparability of online and offline political discussions as the starting point (e.g., Baek, Wojcieszak, and Delli Carpini 2012).

We contend that any framework for explaining the hostility gap, and other potential differences between online and offline behavior and perceptions, needs to distinguish between two broad types of effects of online environments. The first class—constituting the core focus of this article—we call *mismatch effects.* As we elaborate below, this class of effects imply that the "perfect storm" of novel online features (e.g., anonymity and rapid text-based communication) induces fleeting psychological changes that increase the likelihood of certain psychological *states* that undermine civil discussions (Baek, Wojcieszak, and Delli Carpini 2012; Cheng et al. 2017). Simply put, when people log online their level of empathy is reduced or they become more aggressive than usual.

Narratives that emphasize mismatch effects often reference the entire package of features that differentiates online versus offline discussions rather than pinpointing a single feature at the expense of others. To provide an initial comprehensive examination, we follow suit and assess the overall differences between political discussions that occur in online rather than offline environments and leave it to future research to contrast single features.

Instead, we contrast the mismatch hypothesis with existing research on hostility more broadly. Thus, it is notable that while mismatch-oriented explanations of online (political) hostility emphasize the role of fleeting *states*, research on offline hostility often emphasizes the role of stable psychological *traits* such as status seeking (Bartusevičius, van Leeuwen, and Petersen 2020; Petersen, Osmundsen, and Bor 2020). In this alternative view, the personality of discussants matters more for the hostility of a discussion than the platform where it takes place.

But what can explain the hostility gap, if "people are people" no matter where they discuss? If antisocial personality is the main source of online (and offline) hostility, the hostility gap is likely to be an artefact of more mechanical effects of online environments' *connectivity.* Online environments are unique in creating large public forums, where hostile messages may reach thousands including many strangers, could stay accessible perennially, and may be promoted by algorithms tuned to generate interactions (Brady, Crockett, and Van Bavel 2020; Ribeiro et al. 2019; Settle 2018). From this perspective, online environments do not shape how people are motivated but shape what they can accomplish given a specific set of motivations. The hostility gap may thus emerge as a direct consequence of the larger reach of those already motivated to be hostile.

Our focus in the present article is to thoroughly test the role of mismatch effects by (1) comparing discussions in online and offline environments and (2) assessing the role of individual differences. Some of these findings point to the importance of connectivity and, accordingly, we return to and directly assess the role of connectivity toward the end of the article.

A MISMATCH: HUMAN PSYCHOLOGY AND ONLINE ENVIRONMENTS

To understand the potential importance of mismatch effects for the hostility gap, we need to appreciate the differences between offline and online interactions. In our offline lives, we know and directly interact with around 150 people (Hill and Dunbar 2003) and discuss politics with only nine of them (Eveland and Hively 2009). Multiple perspectives within psychological science converge on the argument that human psychology—including the mechanisms regulating aggression and hostility—is tailored to the intimate face-to-face interaction that characterizes offline environments—for example, emphasizing how human psychology is adapted to life in ancestral small social groups (Hill and Dunbar 2003) or how social strategies are calibrated by direct interactions with parents and peers in early life (Simpson and Belsky 2008). Empirically, a number of studies have also shown that social decisions are heavily shaped by intimate social cues such as facial expressions (Scharlemann et al. 2001) and eye contact (Kurzban 2001). Therefore, it is likely that the mechanisms responsible for both activating and restraining hostile responses rely on the wealth of cues available in face-to-face interactions.

Online environments are different. Written interactions on social media platforms or comments sections clash with inbuilt assumptions of human social cognition in, at least, four interrelated ways. First, there is a lack of the vivid social cues that are available in face-to-face interaction. Despite the abundance of emoticons, gifs, and memes, these remain only crude tools to communicate and understand emotions compared with a smiling face, a raised voice, or a defeated posture. Second, there is an exceptional possibility of privacy vis-à-vis discussion partners. People on the internet may choose to remain completely anonymous or to display a heavily curated presentation of themselves. Meanwhile, our psychology is adapted to an environment where people carry the burden of their reputation wherever they go. Third, relational mobility is significantly higher in online than in offline environments. Given the large number of potential discussion partners online, people can easily choose to leave one community and join another. This is not easily done in most offline circles. Finally, online interactions are often significantly more public, with other users being able to access the discussion even years after their occurrence, whereas discussions that occur on the savannah or over the dinner table have significantly fewer witnesses.

Following psychological research, we use the term *mismatch* to refer to differences between a given environment and the environment to which our psychology has adapted (Li, van Vugt, and Colarelli 2018). Prior research in judgment and decision making suggests that mismatches are consequential. For example, research in management has shown that teams that primarily interact using computer-mediated communication are significantly less able to coordinate emotionally (Baltes et al. 2002). Also, research in behavioral economics shows that feelings of anonymity activate more selfish impulses (Bohnet and Frey 1999) and that decision making in situations that do not resemble face-to-face interactions deactivates neural circuits related to emotional processing (Petersen, Roepstorff, and Serritzlew 2009). Investigating how specific features contribute to such mismatches when it comes to online political discussions is a fruitful avenue for future research. As the initial step, however, we focus broadly on the potential individual-level psychological consequences of the totality of these mismatches.

FROM MISMATCH TO ONLINE POLITICAL HOSTILITY

To provide a comprehensive basis for both present and future research endeavors, we seek to outline the contours and provide initial tests of all principal processes through which mismatches could create the hostility gap. Understanding political hostility as residing in exchanges between a sender and a receiver that can occur either online or offline, we contend that mismatches of the online environment (1) could induce behavioral *changes* in the sender for the worse, which increase the frequency of hostile messages online (the change hypothesis); (2) could attract particular senders, increasing the risk that online discussions contain individuals predisposed for hostility (the selection hypothesis); and (3) could induce perceptual changes in the receiver, by undermining the ability to correctly *perceive* the intentions of others, raising the likelihood of attributing hostility in online compared with offline contexts (the perception hypothesis). We now expand on each of these processes and the associated hypotheses in turn.

The Change Hypothesis

Mismatch-induced change builds off the large corpus of research that shows (1) that empathy is one of the key antidotes to aggression (e.g., Lim, Condon, and De Steno 2015) and (2) that social and physical proximity is a key trigger of empathy (Bohnet and Frey 1999): "Face-to-face settings might generate empathy and increase perspective taking ability to greater extent than online settings, because interlocutors are physically present and interact on an interpersonal level" (Baek, Wojcieszak, and Delli Carpini 2012, 367). In a nutshell, the hostility gap may emerge from emotion regulation problems in online contexts: it is significantly more difficult to contain hostile emotions in online than offline settings, especially upon discussing contentious topics such as politics. In a paradigmatic study, Cheng et al. (2017) argue that negative mood shifts and toxic discussions can turn most ordinary people into Internet trolls. Similarly, Coe, Kenski, and Reins (2014) find that most hostile comments in news forums come from infrequent users. In this view, online political hostility is an accidental failure of emotion regulation.

The Selection Hypothesis

Whereas the change hypothesis suggests that online political hostility is largely unintended, the selection hypothesis suggests that instrumental motivations play a significant role in it, even more so than offline. People make an active choice to opt in or out of online discussions (Settle and Carlson 2019). Hostile individuals may be more attracted to online environments, because aggression-based strategies can be pursued with less cost online due to anonymity, the lack of physical proximity, and, consequently, a low possibility of retaliation. From this perspective, one can think of online discussion environments as junk food or pornography: a culturally novel package of features that highjack the reward centers of particular individuals.

This version of the mismatch hypothesis is illustrated perfectly in a *Time* story on trolling: "The Internet is the realm of the coward. These are people who are all sound and no fury" (Stein 2016). Adding some support to this perspective, Rowe (2015) found that anonymous discussion forums are more hostile than nonanonymous ones. Also consistent with the importance of selection, Matias (2019) found that announcing community rules in online discussion forums not only changed the behavior of established users but also encouraged more civil individuals to enter these forums, leading to less hostility. Finally, there is some evidence that Internet trolls exhibit higher scores on dark personality traits such as sadism, suggesting that heated online interactions are particularly attractive to only a subset of individuals (Buckels, Trapnell, and Paulhus 2014).

The Perception Hypothesis

Both the change and the selection hypotheses focus primarily on quantitative differences in sending messages in a discussion. Meanwhile, mismatch-induced perception focuses on qualitative differences in the eye of receivers. This version of the mismatch hypothesis entails that political hostility puts a heavier psychological burden on receivers online than offline even if there is, in fact, little difference in the senders' original intentions.

Decades of research within management demonstrates that online interactions can limit trust and generally recommends that conflicts in teams are settled face-to-face (Hertel, Geister, and Konradt 2005). This recommendation reflects, in part, the danger of misunderstandings in online communication. As concluded by Olaniran (2002, 213) "Misrepresentations appeared to be brought about in part by the medium's lack of nonverbal cues and the fact that sometimes receivers have the tendency to be more serious when interpreting a message" (see also Holmes 2013). Furthermore, written online messaging is slower than verbal communication and does not require people to be present for the same discussion at the same time and place (Hesse, Werner, and Altman 1988). Consequently, online discussions could drag on longer, and people may find themselves getting involved in or unable to drop online interactions they would prefer to stop. This may be exacerbated by the public nature of the online discussions, which drags people into discussions based on their group identities (Brady, Crockett, and Van Bavel 2020). In general, each of these features may make people perceive that the resolution of conflicts is more difficult online than offline and, by implication, contribute to an increased perception of online relative to offline political hostility.

Alternatives to the Mismatch Hypothesis

The change, selection, and perception hypotheses all fall within the broad class referred to as mismatch effects of online environments. As noted in the proposed framework for the study of political hostility, the hostility gap may also emerge from more mechanical effects wherein the connectivity and the publicity of online discussions in tandem make hostility more visible to social media users. On this alternative account, the primary role of psychological processes is to generate individual differences in who, irrespective of environments, engages in hostility.

Research on the psychological roots of dominance reveals that "induc[ing] fear, through intimidation and coercion" is a primary strategy for attaining status (Cheng et al. 2013, 105), and one of the most consistent psychological findings is that individuals preoccupied with attaining higher status are much more likely to commit aggressive and hostile acts (including homicide) in everyday life (Wilson and Daly 1985). Recent research has extended this to the political domain and found that status-seeking (both at the individual and the group-level) is a strong empirical predictor of support for and engagement in aggression, even violence, for a political cause around the world (Bartusevičius, van Leeuwen, and Petersen 2020; see also Kalmoe 2014). As summarized by Bartusevičius, van Leeuwen, and Petersen (2020) status-seeking is "a—if not *the*—key predictor of disruptive political behavior."[1]

In short, against the mismatch hypothesis, which emphasizes the psychological effects of online environments for the emergence of online political hostility, a competing psychological explanation entails that specific individuals are hostile across all environments, both online and offline. If status-oriented traits are highly and equally predictive of hostility across contexts, this may imply that connectivity rather than mismatch effects lie at the heart of the hostility gap.

As is clear from this discussion, the mismatch and the connectivity explanations for online political hostility may, to some extent, be pitted against each other using the classical psychological distinction between states (i.e., ephemeral, context-induced motivations) versus traits (i.e., stable motivations grounded in personality)

[1] These insights may seemingly contradict prominent findings by Cheng et al. (2017) that situational factors trump individual differences in predicting online trolling. However, it is important to remember that this study was "primarily interested in studying the effects of mood and discussion context on the general population" (defined as people who are trolling not too often), thus it "filter [ed] banned users (of which many tend to be clearly identifiable trolls), as well as any users who had all of their posts deleted" (11). Cheng et al's (2017) work therefore cannot inform us how individual differences shape a *discussant*'s average likelihood to become hostile, but it offers an important reminder that situational factors affect which *discussions* turn hostile.

(Chaplin, John, and Goldberg 1988). At the same time, it is fruitful to think of these as extreme end points on a continuum. The change and perception hypotheses highlight how online contexts trigger detrimental psychological states and thus are close to the states end of the continuum. Conversely, the connectivity hypothesis implies that hostile predispositions will have similar effects in both online and offline contexts and thus lies at the traits end of the continuum. Yet, the selection hypothesis is an example of an intermediate case, where online contexts are particularly attractive to those with hostile personality traits and, therefore, describes states of political hostility as an interaction between context and traits.

OVERVIEW OF STUDIES

The causal claim of the mismatch hypothesis is that the bundle of features that distinguishes online and offline environments increases experiences of political hostility in online environments. While researchers often test causal psychological arguments using laboratory experiments, this particular causal claim is difficult to test in such an artificial setting. Laboratory experiments could be designed to manipulate single features of online versus offline interactions (e.g., anonymity or text-based discussion). Yet, it is hard to reproduce the entire bundle of features that constitute real-world online interactions in ecologically valid ways (e.g., their repeated and public nature).

Instead, we rely on other study designs: Studies 1–4 present a series of approximately representative online surveys in which we observe the same individuals' experiences with political hostility in both online and offline contexts. In these within-subject designs, we also examine how a primary psychological trait (specifically, individual differences in status-seeking) relates to political hostility. Our objective here is first to provide empirical evidence for the hostility gap and second to test whether people in general report changed behavior, select into discussions, or perceive attacks to be more harmful online than offline.

While increasingly sophisticated techniques for analyzing online behavior are being developed, we rely on self-reports to provide comparable within-subject measures of online and offline behavior. Because of the within-subject design, bias in these measures threatens the validity of causal estimates only if the bias is asymmetrical between online and offline self-reports. For example, while arguably the "holier than thou" effect (Epley and Dunning 2000) could bias self-reports of hostility downward, as long as assessments are equally self-serving both online and offline, it does not affect the validity of our tests. In Study 5, we estimate the size of a plausible source of asymmetric bias—namely, differences in social norms. Study 5 also offers an alternative test of the perception hypothesis relying on a mental simulation exercise.

Studies 1–5 find little evidence for the mismatch hypothesis, but present several findings that are broadly consistent with the connectivity hypothesis. Therefore, Studies 6–8 seek to validate these findings. Most importantly, Study 6 relies on behavioral measures to validate our self-reported measures of political hostility and replicates the finding that highly status-driven respondents are drawn to political discussions. Study 7, in turn, asks whether people—especially those who are highly status driven—have the ability to carefully calibrate the tone of their comments in online political discussions. Finally, Study 8 offers a direct test of the connectivity hypothesis by measuring whether people witness a disproportionately large share of political hostility against third parties in online contexts.

Again, it is relevant to note that the chosen design does not allow us to isolate the effects of single features of online environments. For example, it is not unlikely that the fast-paced nature of online communication increases hostility, whereas its public nature decreases it. We designed our studies to assess the average effect of the entire bundle of ecologically valid features that characterizes online versus offline environments. The psychological effects of this "perfect storm" of features are the very focus of the mismatch hypothesis.

All anonymized data, scripts, and materials that are necessary to replicate and reproduce our findings are available at the American Political Science Review Dataverse (Bor and Petersen 2021).

STUDIES 1–3: WITHIN-SUBJECT TESTS OF THE MISMATCH HYPOTHESIS

Studies 1–3 rely on original data from the United States and Denmark, respectively. These two countries constitute polar opposites on a number of variables relevant for our investigation. The United States is a high-polarization, high-conflict, low-trust, low-participation country, whereas Denmark is a low-polarization, low-conflict, high-trust, high-participation country (Nelson and Shavitt 2002). Exploiting these variations, our analysis follows a most different systems design logic and argues that—conditional on finding similar trends in both countries—our results should generalize to other advanced Western democracies.

Original online survey data bring a number of important benefits to our analysis. First, we could interview approximately representative samples of Internet users in the United States and Denmark. Study 2 (and Study 4, below) further improve the quality of our samples by screening out respondents who have no direct experience with following or participating in political discussions (either online or offline).

While studying various subpopulations (e.g., Twitter or Reddit users) is obviously important, we must rely on diverse samples to answer questions about the social prevalence of hostile behavior and the perceptions of these environments. Thus, our data are not bounded to any single platform and offer an overview of citizens' experiences, whether they are active on Facebook, 4chan, or somewhere else. Second, this feature also circumvents a primary challenge of studying hostility on social media platforms such as Twitter, where hostile messages violate the rules of conduct and, thus, are deleted. Third, with surveys, we can rely on psychometrically validated personality measures, which provide nuanced measures of complex psychological constructs (e.g., status-driven risk taking).

Prediction

The notion of the hostility gap contends that online political discussions are perceived as more hostile than offline political discussions. This prediction can be directly tested by measuring people's perceptions of the average level of hostility in online and offline discussions.

To test the relevance of mismatch-induced change, we focus on the most basic implication: there should be an asymmetry between hostility in online and offline political debates. People who are seldom hostile offline could still send hostile messages online, but not the other way around. We add nuance to this basic prediction by considering the role of status-seeking motivations. The connectivity hypothesis suggests that such individual differences consistently predict hostility across both online and offline contexts. In contrast, if online hostility is a result of mismatch-induced failures in emotion regulation, we would expect the relevance of instrumental, status-oriented concerns to decrease in online compared with offline acts of hostility.

Mismatch-induced selection implies that people predisposed for hostility intentionally select into political discussions of politics in online rather than offline environments in order to indulge in hostile debates. To test the selection hypothesis, we focus on the most basic observable implication: individuals who are motivated to engage in hostility—as captured by individual differences in status seeking—are more likely to engage in online than in offline political discussions.

Finally, to test the mismatch-induced perception hypothesis, we test whether people are more likely to become entangled in futile, draining, and frustrating debates in online contexts than in offline contexts. In Study 5, we turn to an even more basic test of the perception hypothesis: overattribution of hostility.

Methods and Materials

We collected data through YouGov and Lucid and fielded online surveys to approximately nationally representative samples of Americans and Danes. These agencies sample from a large subject pool and employ quota sampling to match population characteristics on age, gender, education, and region in both countries and also on race in the United States. For more information on our samples, see online appendix B.

Study 1 is an American sample of $N = 1{,}515$ testing the change and selection hypotheses using a within-subject design. Study 2 is a Danish sample testing all three hypotheses using a within-subject design. A total of 1,434 people participated in Study 2, but our focal questions comparing online and offline behavior and experiences were presented only to the 1,041 respondents who reported at least minimal experience with following or participating in political discussions in a prescreening question. Finally, Study 3 is a preregistered follow-up study in the United States ($N = 998$) testing the change and perception hypotheses, in a split-ballot question-order experiment allowing both within- and between-subject comparisons of online and offline hostility levels, mitigating potential concerns about common methods bias. The preregistration is available at https://aspredicted.org/3ny87.pdf.

To test the hostility gap hypothesis, we created a 0–1 additive index of perceived hostility of discussions online and offline by rescaling and averaging over ratings for whether these discussions are perceived to be *aggressive*, *uncivil*, and *hostile.* (Online appendix A reports full question wordings for all survey measures in our manuscript).

Our main dependent variable for testing the change hypothesis is self-reported hostility in online and offline discussions. In designing a novel battery of questions to tap into these constructs, we employed best practices from survey research on cyber bullying (Griezel et al. 2012) and online political behavior (Guess et al. 2019). We listed activities constituting important archetypes of hostile behavior (e.g., making "comments that could be taken as offensive or aggressive"). For the online battery, we also added two items, more specific to online interactions (e.g., "getting blocked or banned from a website for violating its guidelines").[2] We asked participants to indicate how often these things happen to them on a scale from "Never" to "Several times a day." To establish within-subject comparisons in the two contexts, we repeat the same questions twice, clearly distinguishing whether the questions concern "face-to-face political discussions" or "occur on the Internet, including social media and comments sections."[3] In Studies 2 and 3, we added several minor improvements to the scale to increase measurement validity.[4]

[2] Note that our measures provide a lenient test of the change hypothesis by including items that measure "regretted" and online-specific behavior. Dropping these items does not change the conclusions of our analysis.

[3] Our assessment of the satisfactory validity of our novel political hostility scale rests on three legs. First, in Studies 1 and 2, we included a psychometrically validated measurement of trait aggression (Diamond and Magaletta 2006). In both samples, participants scoring higher on trait aggression are also more likely to report political hostility (S1 USA online $r = 0.41$, offline $r = 0.42$; S2 Denmark online $r = 0.46$, offline $r = 0.52$), providing evidence of convergent validity. Second, in a recent study, Rasmussen et al. (2021) analyze behavioral data from Twitter and shows that participants with higher scores on our online political hostility scale post more tweets labeled as politically hateful, toxic, or highly negative by various machine learning algorithms (Perspective API toxicity scores $r = 0.30$, AFINN negative sentiment scores $r = 0.25$, Political hate word embedding scores $r = 0.33$). Finally, our own behavioral experiment indicates that people who report higher online political hostility are prone to write slightly more hostile comments, even in an artificial one-shot online experiment (Cohen's $d = 0.3$; see more details in Study 6).

[4] Specifically, we followed the recommendations by Guess (2015), we added a sentence asking participants to think specifically about the past 30 days, thereby reducing measurement error due to asymmetries in memory. Also, we added a final item to the online hostility battery stating, "I had a difficult time tempering my emotions" to

In Studies 1 and 2, we asked the online questions before the offline questions. In Study 3, we combined the within-subject design with a between-subject design, randomizing the order of the question batteries. In all studies, we average over all items to form two indices called *offline hostility* and *online hostility.* After rescaling, the two indices have a theoretical range from 0 to 1. We report descriptive and scale-reliability statistics for all indices in online appendix B.

To investigate the selection hypothesis, we adapt a measure from Valenzuela, Kim, and Gil de Zuniga (2012), tapping into the frequency of political conversations with members of various groups: family and friends, coworkers and acquaintances, strangers, people with whom they agree, and people with whom they disagree. Again, we repeat these questions twice, first for offline discussions and then for online discussions. We form two reliable indices, which we refer to as *talking about politics online* and *offline.*

Finally, to test the perception hypothesis, we designed eight new items tapping into people's perceptions about the severity of political conflicts online and offline. Five items concern whether people feel they become "involved in discussions they do not feel like having," whether they "continue a discussion even though they do not enjoy it," and so on. Meanwhile, three items focus on the resolution of conflicts, addressing the length of discussions, the ease with which they are discontinued, and finally whether offenses are followed by apologies. We report the results for a scale combining all items.

Our main independent variable is status-driven risk taking (SDRT), a concept developed to tap into competitive risk taking. It measures individual differences in the "tendency to seek and accept great risks … in pursuit of great rewards involving material wealth or social standing" (Ashton et al. 2010, 735). Respondents rate their agreement with the items on a seven-point scale. All items are averaged to form a single *status-driven risk taking* index. As alternative measures of hostile personality, we also included the Difficulties in Emotion Regulation Scale (Gratz and Roemer 2004) and a short version of the Buss-Perry Aggression Questionnaire in our surveys (Diamond and Magaletta 2006). Online appendix D reports results with these measures, replicating and extending the findings discussed below.

To test our predictions, we employ simple OLS regressions. Unless otherwise noted, our models adjust (or "control") for basic demographic covariates: age, gender, education, and income in both countries; a standard seven-point partisan identity scale in the US; a three-level (red-block, blue-block, or neither) partisanship variable in Denmark; and also an indicator for identifying as white in the United States. All variables are scaled to 0–1. Although these OLS regression models are effective tools for estimating partial correlations between our variables, they do not directly allow us to contrast online and offline political discussions. Formal tests for the statistical significance of the differences in coefficients rely on structural equation models, reported in detail in online appendix D.

Results

Is there a hostility gap? Yes. Figure 1 depicts the negativity of online (*dark gray*) and offline (*light gray*) discussions. Respondents in both countries rate online discussions—USA: M (SD) = 0.57 (0.34), Denmark: M (SD) = 0.65 (0.22)—much more negatively than offline discussions—USA: M (SD) = 0.38 (0.29), Denmark: M (SD) = 0.34 (0.22), contributing to a substantively large and statistically significant difference between the two platforms—USA: ΔM = 0.19, $t(1514) = 23.9$, $p < 0.001$, Cohen's $d = 0.61$; Denmark: ΔM = 0.32, $t(978) = 34.8$, $p < 0.001$, Cohen's $d = 1.1$.[5]

Do respondents report more online than offline hostility? No. Figure 2 below reports predicted levels of online hostility as a function of offline hostility (*thick line*) along with 95% confidence intervals (*gray shade*) and the observed values for each respondent (*jittered points*). These models omit demographic controls in order for the intercept to correspond to no offline hostility. The findings reveal a number of important observations.

First, we see relatively low levels of hostility in both online—S1 US: M (SD) = 0.07 (0.17), S2 DK: M (SD) = 0.06 (0.16), S3 US: M (SD) = 0.13 (0.21)—and offline settings—S1 US: M (SD) = 0.07 (0.17), S2 DK: M (SD) = 0.07 (0.16), S3 US: M (SD) = 0.16 (0.21). Put differently, between one third (S3 US offline) and three quarters (S2 DK online) of our samples report no instances of political hostility. Second, online and offline hostility are strongly related in both the United States and Denmark. Indeed, offline hostility alone explains 64% to 79% of the variance in online hostility. Third, to the extent that we see any asymmetry between the two forms of hostility, it is toward more offline hostility among highly hostile individuals. The intercepts in all models are very close to zero (S1 US: $\alpha = 0.006$, $p < 0.01$; S2 DK: $\alpha = 0.003$, n.s.; S3 US: $\alpha = 0.008$, n.s.). As additional robustness checks, we perform paired *t*-tests and equivalence tests for all samples (online appendix D2). We also replicate our results with between-subject comparisons in the split-ballot design we implemented in Study 3 by randomizing the order of the question batteries (online appendix D3). All our models indicate that it is unlikely that the behavior of a meaningful part of our sample changes for the worse in online political discussions. An objection in this respect could be that these analyses primarily address the frequency of online and offline hostility but not whether the hostility is qualitatively worse in

further increase the scale's sensitivity to minor asymmetries between online and offline hostile behavior.

[5] We also asked participants to rate the discussions on three positive items: peaceful, respectful, and constructive. As the six items do not form a single reliable scale, we report positive items separately in online appendix, section D. The results closely mirror the patterns reported here.

FIGURE 1. Distribution of Perceived Negativity of Online and Offline Discussions in the United States and Denmark

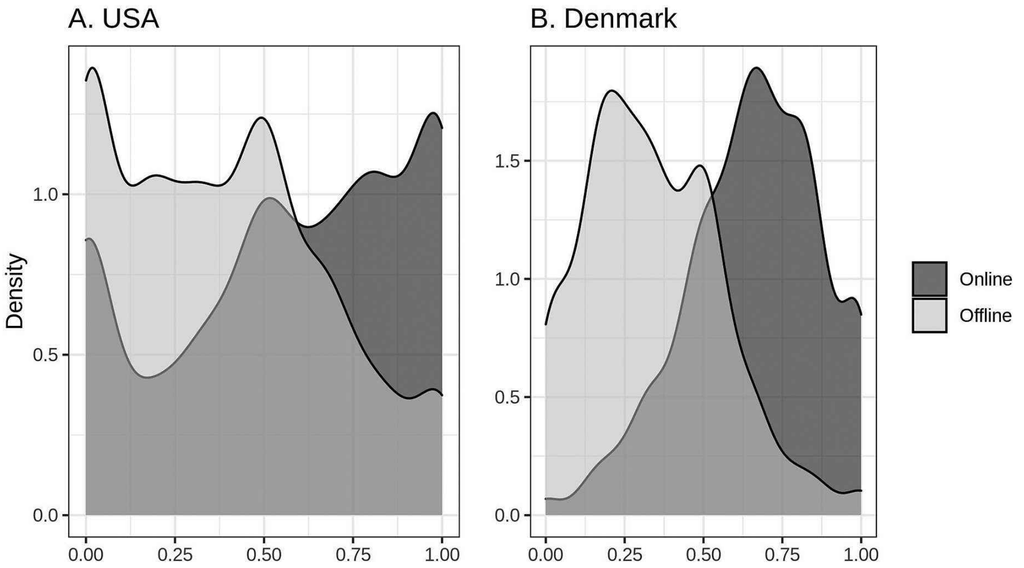
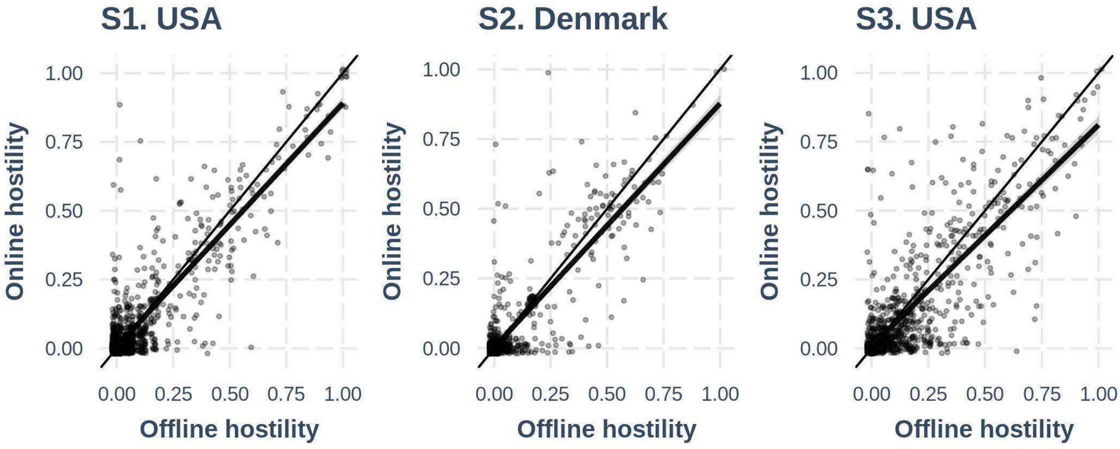


FIGURE 2. Relationship between Offline (*x*) and Online (*y*) Political Hostility



online contexts. We address this objection when testing the perception hypothesis below.

Is status-driven risk taking a less important correlate of online than offline hostility? No. The association between political hostility and status drive remains constant no matter which environment we look at (US Offline: β = 0.29, US Online: β = 0.29; Denmark Offline: β = 0.44, Denmark Online: β = 0.40, all *p*s < 0.001). This pattern is obvious from Figure 3: The predicted hostility sharply increases with higher values of status-driven risk taking, but the predictions are completely overlapping for online and offline environments (US: Contrast = 0.004, *p* = 0.77, DK: Contrast = –0.03, *p* = 0.12).

Are respondents higher on status-driven risk taking selecting into online as opposed to offline political discussions? To some extent. Specifically, we regress the two indices for talking about politics in online and offline environments on status-driven risk taking, adjusting for demographic controls. Under the selection hypothesis, we would expect that people higher on status-driven risk taking prefer online discussions more than offline discussions. Figure 4 presents predicted values of talking about politics online (*dark gray*) and offline (*light gray*) across the range of observed values of status-driven risk taking, keeping demographic covariates at their mean. First, it appears that status-driven risk taking is correlated with general interest in political discussions—all lines have positive slopes (for similar findings, see Sydnor 2019). However, we see no evidence that highly status-driven individuals prefer online interactions. Indeed, at the maximum of the

FIGURE 3. Political Hostility as a Function of Status-Driven Risk Taking

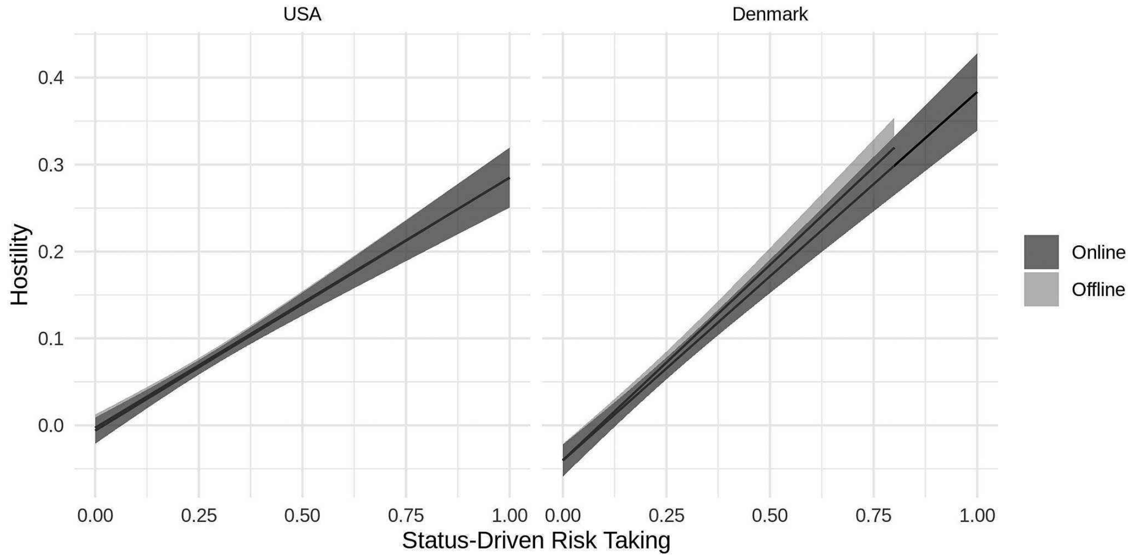
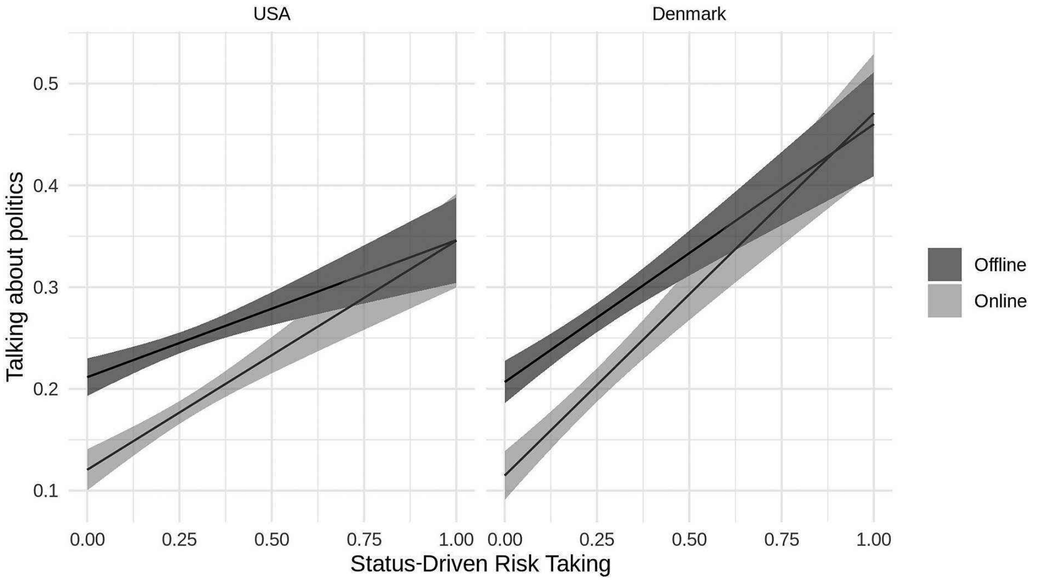


FIGURE 4. Talking about Politics as a Function of Status-Driven Risk Taking



scale, respondents report similar levels of discussion frequency.

However, unlike with the hostility measures, the two environments are not exactly similar. At lower levels of status-driven risk taking, peaceful respondents report talking more about politics offline than online. Formally, this is evident in the substantial difference between the intercepts of our two sets of models, reflecting the levels of online and offline political discussions at the lower end of status-driven risk taking (US Online: α = 0.16, US Offline: α = 0.21, Contrast = –0.05, $p < 0.1$; DK Online: α = 0.08, DK Offline: α = 0.17, Contrast = –0.09, $p < 0.001$). This difference translates to a 6–7-percentage-point increase in the chance that one discusses politics with someone above the median of the status-driven risk-taking scale (ps < 0.01; see details in online appendix D4).[6]

[6] One interpretation of the selection effect is that peaceful people are driven away by the hostility of online discussions. If this were the case, they would be most likely to select out of online discussions most likely to turn hostile: discussions with strangers and those with whom they disagree (Settle and Carlson 2019). However, Figure D5 in the online appendix shows that they never discussed politics with these individuals to begin with. Instead, peaceful individuals mostly discuss

Are political conflicts perceived to be more severe online than offline? No. To test the perception hypothesis, we investigate perceived differences in the severity and resolution of conflicts in offline and online discussions, relying on the novel scale introduced above and included in the Danish survey (Study 2) and the second US survey (Study 3). Against the perception hypothesis, we find that conflicts are perceived as more or equally severe in offline compared with online environments (US Offline: M = 0.42, US Online: M = 0.42, ΔM = 0.002, n.s.; DK Offline: M = 0.35, DK Online: M = 0.32, ΔM = –0.04, p < 0.001). Equivalence tests probing for the smallest effect size of interest (Cohen's d = 0.1) show that these effects are unlikely, assuming the perception hypothesis was true (ts > 2.2, ps < 0.05). Importantly, the finding that online discussions are not experienced as more severe also speaks against a quality-oriented version of the change hypothesis, which claims that the primary difference between online and offline hostility is not how often it happens but how severe it is when it happens.

These findings might seem paradoxical given the documented existence of the hostility gap. However, it is crucial to note that the evidence for the hostility gap reflects assessments of the overall discussion climates (with respect to their aggression, incivility, and hostility), whereas the perception questions narrowly relate to assessments of the discussions in which the participants themselves participate. Furthermore, the present findings are consistent with those from some previous research. For example, in a series of qualitative interviews, a group of feminist political activists stated that they find social media "a relatively *safer* and *easier* space to engage in feminist discussions than in … offline contexts" (Mendes, Ringrose, and Keller 2018, 243). Face-to-face disagreements might not only entail greater repercussions but also bombard people with a plethora of nonverbal cues signaling the discussion partner's anger.

STUDY 4: REPLICATION AND EXTENSION OF WITHIN-SUBJECT TESTS

The primary ambition of Study 4 is to test the change hypothesis on a broader set of hostile political behaviors. The scale developed for Studies 1–3 sought to strike a balance between the most severe forms of political hostility (motivating our inquiry) and milder and more pervasive forms of hostility. Yet, in the two highest quality samples (Studies 1 and 2), about two thirds of respondents claim to be innocent of political hostility, raising concerns that this drives the high correlations. Here, we measure political hostility with a more comprehensive list of items, spanning the entire range of political hostility from offensive jokes to true harassment. Study 4 also seeks to replicate the selection findings from Studies 1 and 2 that were ignored in Study 3 due to space limitations. Unless otherwise noted, all predictions and tests of Study 4 are preregistered at https://aspredicted.org/3ah9y.pdf.

Predictions

Our predictions mirror the analyses described in Studies 1–3. Importantly, at this point, the theoretical expectations of the mismatch hypothesis and our expectations based on previous evidence diverge. Against the mismatch-induced change hypothesis but consistent with the assumptions underlying the connectivity hypothesis, we predicted to find (1) no more self-reported hostility in online (vs. offline) political discussions, (2) a very high (r > 0.75) correlation between self-reported online and offline political hostility, and (3) no evidence that status-driven risk taking is a better predictor of online than offline political hostility. Similarly, against the mismatch induced selection hypothesis, we predicted (4) that respondents high in status-driven risk taking are not more likely to participate in online than in offline political discussions. Given the findings of Studies 1 and 2, we finally predicted (5) that respondents low in status-driven risk-taking select out of online (but not offline) political discussions.

Methods and Materials

A total of 1,317 American adults were interviewed by YouGov and selected via quota sampling to ensure resemblance of the population. Consistent with Study 2, we screened out participants who never participate in or follow political discussions online or offline, leaving us with a final N of 770. To self-reported political hostility, we include the same question as before but add a broader set of eight items ranging from "I made fun of my political opponents" to "I threatened or harassed my political opponents" (see OA Section A) for complete battery). We measure political participation with the same question and items as in Study 1 and 2. As in Study 3, we implemented a question order experiment, where participants randomly encountered either the block pertaining online or offline behavior first.[7]

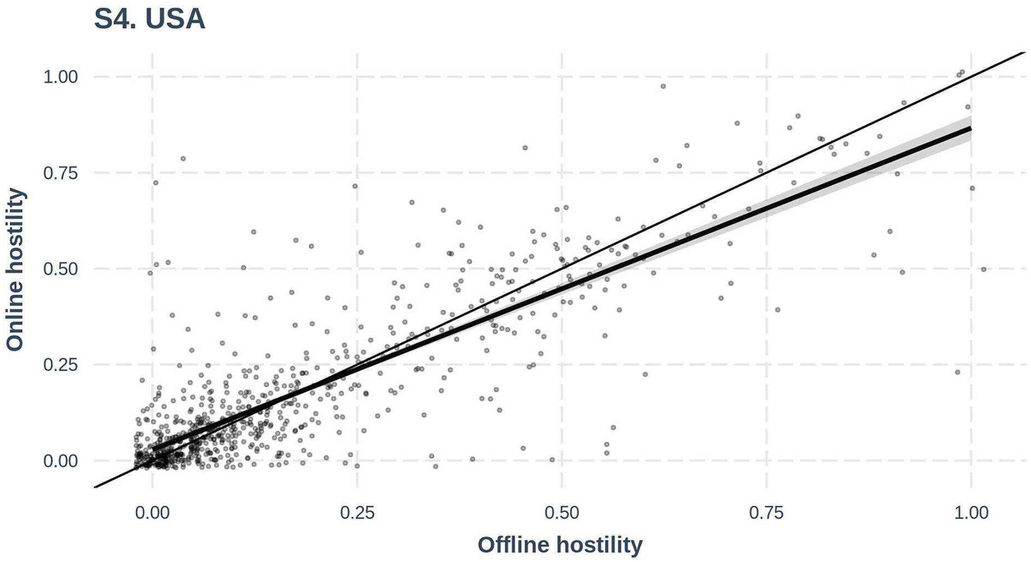
Results

Do people self-report more hostility in online political discussions than in offline political discussions? No. We

politics offline with friends and people with whom they agree, and the drop we see in online discussions is also driven by friends and similarly minded people. These findings are more in line with an alternative interpretation that the selection effect is driven by a preference among nonhostile people to avoid all online political discussions rather than hostile political discussions specifically.

[7] Each block was introduced with a short clarification on how we define online (offline) discussions. Because of the realities of the COVID-19 pandemic, we defined online discussions as "text-based" communications on the Internet and offline discussions as "discussions that occur in situations where you could hear and see the other person(s)" thereby including also video calls.

FIGURE 5. Preregistered Replication of High Correlation between Offline (x) and Online (y) Hostility, Relying on an Extended Hostility Battery from Study 4



find that respondents self-report equal levels of hostility online and offline—Online M (SD) = 0.17 (0.21), Offline M (SD) = 0.17 (0.20), ΔM = 0.00, $t(769) = 0.24$, n.s. With the revised scale, only 22% and 18% of respondents report absolutely no political hostility across online and offline discussions, respectively. In other words, it appears that most people who participate or follow political discussions occasionally engage in (milder forms of) incivility.

Is there a very high correlation between self-reported online and offline political hostility within individuals? Yes (see Figure 5). We find that an individual's offline political hostility is highly predictive of their online political hostility (Pearson’s $r = 0.84$).

Is status-driven risk taking a better predictor of online versus offline political hostility? No. The association between status drive and hostility is identical across the two contexts (Offline: $\beta = 0.52$, Online: $\beta = 0.51$, $ps < 0.001$).

Are respondents with high status-driven risk taking more likely to participate in online than offline political discussions? Are respondents with low status-driven risk taking selecting out of online political discussions? We find no evidence for either. Against the mismatch hypothesis and our previous findings (Studies 1 and 2), we find identical levels of participation across offline and online contexts along the entire range of SDRT. Participants low on SDRT participate just as little both online ($\alpha = 0.16$, $p < 0.001$) and offline ($\alpha = 0.18$, $p < 0.001$). Meanwhile, the effect of SDRT is large and positive across both contexts (Offline: $\beta = 0.35$, Online: $\beta = 0.38$, $ps < 0.001$).

To summarize, Study 4 replicated previous results regarding the change hypothesis with a novel battery spanning a broader and more comprehensive list of hostile behaviors. Meanwhile, we failed to replicate the selection effects found in the previous two studies.[8]

STUDY 5: VIGNETTE EXPERIMENT

The evidence against the mismatch hypothesis generated by Studies 1–4 relied on within-subject tests. This individual differences approach may raise concerns, especially regarding the perception hypothesis, as it ignores the most straightforward implication of the perception hypothesis: the same message feels more hostile when it is uttered in an online (vs. an offline) setting. Study 5 was designed to provide a direct, experimental test of this.

In addition, and with implications for especially the change hypothesis, the within-subject tests assume that self-reported hostility online and offline are directly comparable. Yet, a possible concern is whether people have different norms about online and offline conversations. If this were the case, people could make objectively more hostile statements online than offline but self-report equal levels of hostility because their online messages feel just as appropriate for the context, despite their content. The effects of such potential norm differences, however, are not straightforward because norm asymmetries also could make people *more* willing to admit to a hostile behavior online, where it is not a norm violation. To assess this potential complicating

[8] Given that Study 4 was collected during the COVID-19 pandemic, we cannot rule out that the failure to replicate this effect reflects changes in respondents’ discussion patterns as a response to the pandemic.

factor, Study 5 also tests whether offensive statements are considered more appropriate online than offline.

Predictions

To test the perception hypothesis, we predicted that hostile messages are considered more offensive in online (vs. offline) contexts. Importantly, against the perception hypothesis and consistent with the connectivity hypothesis and the previous findings, we expect that we will not find evidence for this hypothesis.

To assess the potential existence of a norm asymmetry, we use a distinction between injunctive and descriptive social norms (Gerber and Rogers 2009). Injunctive norms refer to behavior people perceive to be normatively appropriate or desirable. If there is a norm asymmetry in terms of injunctive norms, we expect that hostile messages would be considered less inappropriate in online (vs. offline) contexts, given the strong norms of civility in regular face-to-face discussions. Descriptive social norms refer to what other people actually do. While hostility is relatively rare in offline discussions, the hostility gap implies that it is much more frequent online. If there is norm asymmetry in terms of descriptive norms, we would expect that hostile messages are considered less rare in online (vs. offline) contexts.

Methods and Materials

To maximize internal validity, Study 5 relies on the popular experimental vignette methodology, inviting participants to imagine themselves in a situation described in a carefully manipulated yet highly realistic scenario. Social and organizational psychologists have found that such mental simulation exercises offer a good approximation of real experience (Aguinis and Bradley 2014). We presented participants with four hostile messages on controversial social topics (immigration, abortion, COVID-19, Capitol siege) balanced on ideological stance (the former two being conservative, the latter two being liberal). The order of the messages was randomized. We manipulated the context in which the message was uttered in a 2 (online vs. offline) × 2 (private vs. public) within-participant design. As our predictions highlight our main focus is the online versus offline effect, but we included the public versus private manipulation for exploratory purposes (see OA Section E) and to reduce experimenter demand effects. As an illustration, the immigration message read (highlighting the manipulated text with bold) as follows:

> Imagine that you participate in a political discussion at a **dinner party with 5 other people/public town hall meeting/private chat group with 5 other people/public Internet forum.** The discussion turns to the topic of **the siege of the US Capitol.** Someone makes the following comment: "Folks who think that the current levels of immigration are sustainable are stupid and irresponsible. Crime and unemployment due to immigration hurt hypocrite Democrats as much as everyone else. Wicked and sadistic immigrants harass innocent people. They should respect the culture of people living in this country more."

Participants were asked three questions after each message: (1) how offensive they found the message, (2) how appropriate or inappropriate the comment would be in the given context, and (3) how common or rare it is that such a comment is made in the given context. Each question was answered on a 1–7 scale with "not offensive at all" versus "extremely offensive," "perfectly appropriate" versus "extremely inappropriate," and "very common" versus "very rare" employed as anchors.

Study 5 was fielded on the same sample as Study 4; however, importantly, it included all respondents, irrespective of whether they ever participate or follow political discussions (N = 1,317 US Americans). Because each respondent answered these questions four times, our analyses rely on 5,268 observations. We regress offensiveness, inappropriateness, and rarity on an indicator for online context while including fixed effects for respondents and stories to account for the within-respondent design. We preregistered all our decisions at https://aspredicted.org/p2cy6.pdf.

Results

Are hostile messages considered more offensive in online (vs. offline) contexts? No. Consistent with previous findings, saying something online is not perceived to be inherently more offensive ($\beta = -0.01$, n.s.). Judging from the 95% confidence interval, even a vanishingly small effect (>0.006) is inconsistent with our data.

Are hostile messages considered less inappropriate in online (vs. offline) contexts? Yes, although the difference is very small ($\beta = -0.03$, $p < 0.001$).

Are hostile messages considered less rare in online (vs. offline) contexts? Yes, although the difference is even smaller than for inappropriateness ($\beta = -0.02$, $p < 0.001$).

Overall, consistent with our previous findings, we find no evidence for the perception hypothesis. While we find some evidence for norm asymmetries, these appear to be substantially small. To assess whether they could exert a bias on the self-reports of political hostility, we ran post hoc analyses using the difference between the perceived inappropriateness of online versus offline comments in Study 5 as an individual difference measure of perceived norm asymmetry (higher values indicate that the respondent believes it is more permissible to be hostile online). We find no evidence that this perceived inappropriateness gap is associated with being hostile online, nor does it moderate the relationship between offline and online hostility. To summarize, we find it unlikely that the self-reported hostility scales are differentially biased by norm asymmetries.

STUDIES 6–7: BEHAVIORAL EXPERIMENTS

Studies 1–5 find little evidence for the mismatch hypothesis but are consistent with the primary assumption underlying the connectivity hypothesis: That individual-level aggressive traits drive political hostility equally both online and offline. In the next two studies, we test the robustness of and add nuance to these findings, relying on original behavioral experiments.

To increase the measurement validity of our conclusions, Study 6 asks whether our self-reported hostility scale correlates with observed hostility in a political discussion. Meanwhile, Study 7 seeks to discern between different explanations for the behavior of those high in status-driven risk taking. The literature on status-driven hostility suggests that this behavior is fueled by strategic motivations to increase personal or group-based status (Petersen, Osmundsen, and Bor 2020). At the same time, however, the present evidence cannot rule out that the behavior of those predisposed for hostility reflects a form of mismatch-induced perception, where their hostility is grounded in an inability to correctly perceive the tone of messages in political discussions. Study 7 provides a direct test of this.

Both experiments emulate discussions on Facebook using between-subject experiments and behavioral measures of hostility. Participants read a Facebook post on immigration, which was manipulated in a 2 (position: pro-immigration or anti-immigration) × 8 (hostility) between-subjects factorial design. We chose immigration to the United States because it is hotly debated policy issue.

Predictions

The main ambition for Study 6 is to validate our original self-reported hostility measure. Therefore, we expect that participants who self-report being hostile more often in online political discussions will also write more hostile messages in our experiment.

Study 7 asks whether respondents, and particularly respondents high in status-driven risk taking, are able to match the tone of political messages following a simple written instruction to do so. Study 7 thus provides a direct test of whether some respondents are incapable of correctly perceiving the tone of political discussions. While Studies 1–4 leave little doubt that online environments do not contain systematically more cues inducing “accidental” hostility, we offered no evidence for the argument that hostility is strategic.[9]

Methods and Materials

A total of 2,137 and 2,089 participants living in the United States were recruited on Amazon’s Mechanical Turk for Studies 6 and 7, respectively; 206 (Study 6) and 449 (Study 7) participants were excluded from the study for failing simple comprehension checks. Thus, our final samples constitute 1,923 and 1,640 participants for the two studies (mean age 39 and 37 years, 58% and 60% females), respectively. Participants gave informed consent (see OA Section F) and were reimbursed with $1.50. The median response time was 12 minutes with both surveys.

Participants were asked to read a Facebook post on immigration. Participants in Study 6 also saw a nonpolitical post and selected the one that they would be most likely to reply to in real life. Next, all participants rated the hostility of the (selected) post on a scale from 0 to 100. The consecutive page prompted participants in both studies to write a comment to the post. To reduce the likelihood that participants simply parrot the target, we hid the target post from this screen. Finally, participants answered a question about their comprehension of the task. Those who failed to select either of the two correct answers from the six available options were excluded from the study.

To judge the hostility of political posts and comments, we relied on crowd sourcing to rate all materials and data involved in our experiments. Specifically, we invited MTurk workers to rate the perceived hostility of messages (median response time was under five minutes; raters were reimbursed with $0.60). Relying on untrained raters to assess hostility means we could let regular people define what they find hostile, instead of coming up with a coding scheme that may or may not capture peoples’ own experiences.[10]

Results

Offering participants an opportunity to choose between a political and a nonpolitical post in Study 6 allows us to increase ecological validity by investigating the behavior of respondents most likely to engage with the given text in real life. In this study, we find that people higher in status-driven risk taking are more likely to pick the political post ($\beta = 0.11$, $p = 0.08$). Consistent with the results of Studies 1, 2, and 4, hostile people have an enhanced preference for participating in political discussions.

Does our self-reported political hostility scale correlate with hostile behavior in Study 6? Yes. We validate our self-reported measure of online hostility from Studies 1–3 by observing a significant and substantial correlation ($p < 0.001$) between this measure and the hostility of the participants’ written responses, corresponding to about 30% of a standard deviation in the

[9] These two experiments also allow us to test a number of additional implications relevant for our study of online political hostility—these are reported in detail in OA Section G.

[10] First, we relied on 1,430 raters to design the 16 versions of the target Facebook post, ensuring that they are reasonably evenly paced along the hostility scale. Each participant rated only one comment to mimic the design of the experiment and to avoid demand effects due to seeing multiple versions of the stimulus. We ran bootstrapped simulations on the target post ratings to estimate the ideal number of raters per message for the participant comments. We found that the interrater reliability across two independent sets of ratings reaches acceptable levels after five raters (Krippendorff’s alpha > 0.8), and adding extra raters after 10 ratings has quickly diminishing returns. Accordingly, to rate respondents’ responses from Studies 5 and 6, we invited 787 and 1,459 participants, respectively, each rating 12 random comments (see more details in OA Section F). Workers were only eligible to participate in a single phase of the study as either participants or raters.

dependent variable (see online appendix G1). This association is all the more noteworthy, as we observed it relying on a single message in an artificial setting without real stakes. We know from Studies 1–3 that political hostility is a relatively rare phenomenon: even the more hostile respondents appear to keep their calm in most of their interactions. Consequently, it is not surprising that we do not see huge differences between people at various levels of the scale.

Do people on average write comments that are more hostile than the original post despite instructions to match tone? No. In Study 7, we subtract the crowdsourced mean hostility of the given target post from the mean hostility of the reply to measure the average difference in hostility between the target post and the comment written to it.[11] Positive values indicate comments more hostile than the target, whereas negative values denote less hostile comments. Next, we perform a simple one-sample *t*-test to investigate whether the average distance is statistically significant from 0. We find that comments are on average 11 percentage points *less* hostile than the target ($\Delta M = -0.11$, $t(1639) = -25.6$, $p < 0.001$). This indicates that the average respondent is unlikely to suffer from a perception bias. Still, about a quarter of our sample did write comments more hostile than the target. Next, we investigate whether individual differences correlate with comment hostility in this task.

Do people higher in status-driven risk taking write more hostile comments despite instructions to match tone? No. We regress the response hostility variable on SDRT allowing for varying intercepts for target hostility (8 levels) and target position (2 levels). There is no evidence that varying slopes for the personality variables improves model fit ($p = 0.51$ and $p = 0.94$, respectively), and they are omitted from the model. In line with the instructions, more hostile target posts received more hostile responses. However, we do *not* find that high status-drive respondents are more likely to overreact in the task. In other words, while the initial validation tests showed a relationship between response hostility and status drive, the instruction to match the tone of the target Facebook post completely washes away this relationship.

STUDY 8: TESTING THE CONNECTIVITY HYPOTHESIS

Studies 1–7 indicate that, for the most part, people behave and process information in similar ways offline and online. The question is, then, what the cause of the hostility gap is. We have raised an alternative to the mismatch hypothesis, focusing not on the psychological effects of online environments but on the actual affordances of online discussion environments. Notably, online discussion environments are highly connected, public, and permanent. Consequently, the actions of just a few hostile individuals will be significantly more visible online than offline.

Predictions

A primary observable implication of the connectivity hypothesis is that people are exposed to particularly large numbers of hostile actions against third parties online (vs. offline), whereas the hostility gap against friends and the self is smaller. In other words, people are exposed to hostile interactions involving strangers online that are hidden from view offline.

Methods and Materials

We collected data from the samples employed in Studies 2 (Denmark, $N = 1{,}041$) and 3 (USA, $N = 998$) to test this notion. Specifically, we asked respondents how often they witness attacks against self, friends, and strangers. As before, we repeated these questions both for online and online discussions. We estimate the perceived frequency of witnessing an attack on (a) the target of attack (self, friends, or strangers), (b) the environment (offline or online), and (c) their interactions with OLS models. We report standard errors clustered at the level of individuals.

Results

Respondents reported witnessing more hostility against each of these parties online than offline. However, as Figure 6 demonstrates, the gap between the two environments is largest for witnessing attacks on strangers (USA: β online = 0.06, SE = 0.01, $p < 0.001$; Denmark: β = 0.21, SE = 0.01, $p < 0.001$) and much smaller for friends (USA: β online × friends = −0.03, SE = 0.01, $p < 0.001$; Denmark: β online × friends = −0.14, SE = 0.01, $p < 0.001$) and the self (USA: β online × self = −0.04, SE = 0.01, $p < 0.001$; Denmark: β online × self = −0.17, SE = 0.01, $p < 0.001$).

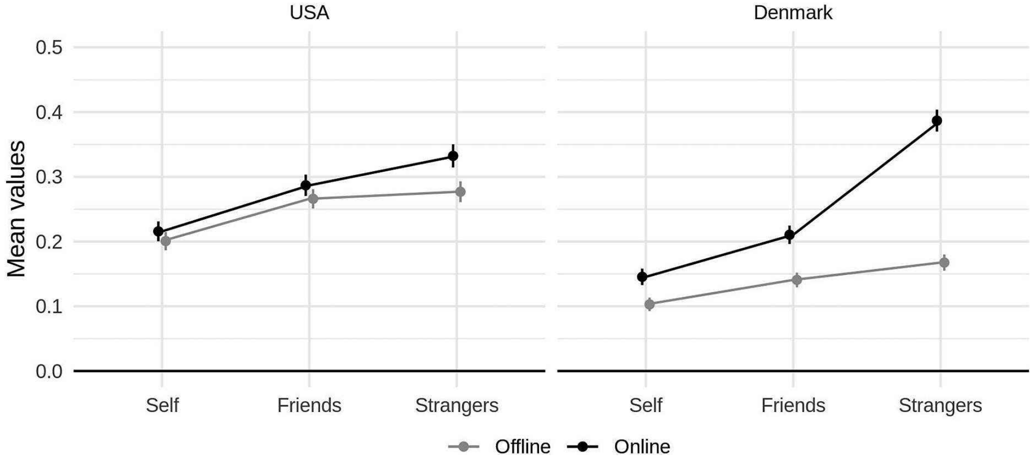
These additional analyses are consistent with the connectivity hypothesis, which entails that the perception that online discussions are more hostile than offline discussions is simply because people witness a much larger number of discussions online as they browse through their feeds on social media.[12]

Given this, they perceive—without any bias—a much larger number of encounters where the discussants are hostile with each other or at the expense of "absent"

[11] Here, we rely on crowd-sourced hostility ratings from Study 6 participants, who rated the target post before the matching task was introduced. We replicate these results with Study 7 participants' ratings in online appendix G. A slightly different scale was employed in the pretesting of the target posts, so the two ratings are not directly comparable (but the results replicate if we rely on those ratings to calculate the difference scores, nonetheless).

[12] Our talking about politics variables from Studies 1–4 offers additional evidence that, indeed, these experiences come from passively witnessing hostility rather than participating in discussions that contain hostility (against the respondent or others). Even if we compare specifically political discussions with partners who disagree with the respondent, we find that, if anything, people talk less about politics with disagreeable partners online than offline (S1 US $\Delta M = -0.04$, $p < 0.001$; S2 DK $\Delta M = -0.05$, $p < 0.001$; S4 US $\Delta M = 0.01$, n.s.).

FIGURE 6. Average Exposure to Attacks against Various Parties Online and Offline



third parties or groups. To put it bluntly, people might also be faced with significant offline hostility if they were able to monitor all the private chats at parties, bars, and dinner tables about, for example, minority groups. Thus, despite common concerns about the negative effects of online echo chambers, perceptions of online hostility may be exacerbated by the publicity and fluidity of these discussion environments (Eady et al. 2019; Gentzkow and Shapiro 2011; Settle 2018).

CONCLUSION AND GENERAL DISCUSSION

In this article, we documented that online political discussions seem more hostile than offline discussions and investigated the reasons why such hostility gap exists. In particular, we provided a comprehensive test of the mismatch hypothesis, positing that the hostility gap reflects psychological changes induced by mismatches between the features of online environments and human psychology. Overall, however, we found little evidence that mismatch-induced processes underlie the hostility gap. We found that people are not more hostile online than offline, that hostile individuals do not preferentially select into online (vs. offline) political discussions, and that people do not overperceive hostility in online messages. We did find some evidence for another selection effect: nonhostile individuals select out from all, both hostile and nonhostile, online political discussions. Thus, despite the use of study designs with high power, the present data do not support the claim that online environments produce radical psychological changes in people.

Our ambition with the present endeavor was to initiate research on online political hostility, as more and more political interactions occur online. To this end, we took a sweeping approach, built an overarching framework for understanding online political hostility, and provided a range of initial tests. Our work highlights important fruitful avenues for future research. First, future studies should assess whether mismatches could propel hostility in specific environments, platforms, or situations, even if these mismatches do not generate hostility in all online environments. Second, all our studies were conducted online, so it is important for future research to assess the mismatch hypothesis by using behavioral data from offline discussions. Contrasting online versus offline communications directly in a laboratory setting could yield important new insights on the similarities and differences between these environments. Third, there is mounting evidence that, at least in the USA, online discussions are sometimes hijacked by provocateurs such as employees of Russia's infamous Internet Research Agency. While recent research implies that the amount of content generated by these actors is trivial compared with the volume of social media discussions (Bail et al. 2020), the activities of such actors may nonetheless contribute to instilling hostility online, even among people who are not predisposed to be hostile offline.

Most importantly, however, our findings suggest that future research could fruitfully invest in developing and testing the main alternative to the mismatch hypothesis, the connectivity hypothesis. Thus, our findings suggest that the feeling that online interactions are much more hostile than offline interactions emerges because hostile individuals—especially those high in status-driven risk taking—have a significantly larger reach online; they can more easily identify targets, and their behavior is more broadly visible.

We recommend future research that examines both the technological and psychological sides of the connectivity hypothesis. In terms of technology, we do not know which features are the most responsible for this connectivity: is it recommendation algorithms, the permanence of online communication, the salience of cross-cutting networks, or something else? In terms of psychology, further research is required on the motivations of those who are hostile. Is it possible that many hostile online messages are not meant to reach members of the out-group (e.g., an offensive joke meant only for close, like-minded friends) or are hostile individuals directly motivated to engage their political opponents? Status-driven motives could underlie both behaviors, and more research is needed to understand their relative importance.

Further research on the relative importance of situational and personality factors is also required. On one hand, most of our respondents are rarely hostile. Consistent with the suggestion of Cheng et al. (2017), situational triggers likely play a large role in determining whether and when these respondents become furious. On the other hand, the evidence against the mismatch hypothesis implies that online environments do not contain more anger-inducing situations. "Anyone can become a troll" online (as suggested by Cheng et al. 2017), just as anyone can get angry offline, but as our findings demonstrate, those high in status-driven risk taking are much more likely to do so.[13] Our results indicate that at least among these people, aggression is not an accident triggered by unfortunate circumstances but a strategy they employ to get what they want, including a feeling of status and dominance in online networks.

In summary, our research suggests that people do not engage in online political hostility by accident. Online political hostility reflects status-driven individuals' deliberate intentions to participate in political discussions and offend others in both online and offline contexts. In large online discussion networks, the actions of these individuals are highly visible, especially compared with more private offline settings. These results imply that policies against hostility should seek to reduce the connectivity of hostile individuals, for example, by decreasing the visibility of the content they produce or increasing the possibility of enforcing legal actions against illegal behavior (Haidt and Rose-Stockwell 2019). Of course, these policies need to be finely balanced in order to avoid curtailing the freedom of expression. Furthermore, our findings show that norms of civility are somewhat weaker online than offline and continued exposure to hostile messages may increase this gap, potentially propelling more hostility through a vicious cycle (Brady et al. 2021). Consequently, our research is consistent with prior findings that interventions strengthening norms or highlighting norm violations can reduce hostility (Matias 2019; Siegel and Badaan 2020). At the same time, it is relevant to highlight the danger of encouraging users to set and to police norms of civility, as this may ignite novel forms of hostility about appropriateness and could undermine goals of improving the general tone of online discussions. From this perspective, public discourse should inform the setting of norms, but third-party referees—such as platform policies or trained moderators—would be mainly responsible for promoting and enforcing them in specific discussion networks.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0003055421000885.

DATA AVAILABILITY STATEMENT

Replication and reproduction materials can be found at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/8I6NOT.

ACKNOWLEDGMENTS

This research has benefitted from discussions with Vin Arceneaux, Matt Levendusky, Mark Van Vugt, John Tooby, and members of the Research on Online Political Hostility (ROHP) group, among many others. We are grateful for constructive comments to workshop attendees at the Political Behavior Section of Aarhus University, at the NYU-SMAPP Lab, at the NYU Social Justice Lab, at the Hertie School, and to conference audiences at APSA 2019, HBES 2019, and ROPH 2020.

FUNDING STATEMENT

This study was funded by grant AUFF-E-2016-9-22 from the Aarhus University Research Foundation and grant CF18-1108 from the Carlsberg Foundation, both awarded to Michael Bang Petersen.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research.

REFERENCES

Aguinis, Herman, and Kyle J. Bradley. 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4): 351–71.

Ashton, Michael C., Kibeom Lee, Julie A. Pozzebon, Beth A. Visser, and Narnia C. Worth. 2010. "Status-Driven Risk Taking and the Major Dimensions of Personality." *Journal of Research in Personality* 44 (6): 734–37.

Baek, Young Min, Magdalena Wojcieszak, and Michael X. Delli Carpini. 2012. "Online versus Face-to-Face Deliberation: Who? Why? What? with What Effects?" *New Media & Society* 14 (3): 363–83.

Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and

[13] In fact, an exploratory analysis shows that many high-SDRT respondents report being hostile as often as they talk about politics (see OA, Table D9), leaving less room for situational factors.

Alexander Volfovsky. 2020. “Assessing the Russian Internet Research Agency’s Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017.” *Proceedings of the National Academy of Sciences* 117 (1): 243 LP—50.

Baltes, Boris B., Marcus W. Dickson, Michael P. Sherman, Cara C. Bauer, and Jacqueline S. LaGanke. 2002. “Computer-Mediated Communication and Group Decision Making: A Meta-Analysis.” *Organizational Behavior and Human Decision Processes* 87 (1): 156–79.

Bartusevičius, Henrikas, Florian van Leeuwen, and Michael Bang Petersen. 2020. “Dominance-Driven Autocratic Political Orientations Predict Political Violence in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) and Non-WEIRD Samples.” *Psychological Science* 31 (12): 1511–30.

Bisbee, James, and Jennifer M. Larson. 2017. “Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity.” *American Political Science Review* 111 (3): 502–21.

Bohnet, Iris, and Bruno S. Frey. 1999. “The Sound of Silence in Prisoner’s Dilemma and Dictator Games.” *Journal of Economic Behavior and Organization* 38: 43–57.

Bor, Alexander, and Michael Bang Petersen. 2021. “Replication Data for: The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis.” Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/8I6NOT.

Brady, William J., M. J. Crockett, and Jay J. Van Bavel. 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, William J., Killian McLoughlin, Tuan Nguyen Doan, and Molly Crockett. 2021. “How Social Learning Amplifies Moral Outrage Expression in Online Social Networks.” *PsyArXiv*. https://doi.org/10.31234/osf.io/gf7t5.

Buckels, Erin E., Paul D. Trapnell, and Delroy L. Paulhus. 2014. “Trolls Just Want to Have Fun.” *Personality and Individual Differences* 67 (September): 97–102.

Chaplin, William F., Oliver P. John, and Lewis R. Goldberg. 1988. “Conceptions of States and Traits: Dimensional Attributes with Ideals as Prototypes.” *Journal of Personality and Social Psychology* 54 (4): 541–57.

Cheng, Joey T., Jessica L. Tracy, Tom Foulsham, Alan Kingstone, and Joseph Henrich. 2013. “Two Ways to the Top: Evidence That Dominance and Prestige Are Distinct yet Viable Avenues to Social Rank and Influence.” *Journal of Personality and Social Psychology* 104 (1): 103–25.

Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. “Anyone Can Become a Troll.” In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, chairs. Charlotte P. Lee and Steve Poltrock, 1217–30. New York: Association for Computing Machinery Press. https://doi.org/10.1145/2998181.2998213.

Coe, Kevin, Kate Kenski, and Stephen A. Rains. 2014. “Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments.” *Journal of Communication* 64 (4): 658–79.

Dahlberg, Lincoln. 2001. “The Internet and Democratic Discourse: Exploring the Prospects of Online Deliberative Forums Extending the Public Sphere.” *Information, Communication & Society* 4 (4): 615–33.

Diamond, Pamela M., and Philip R. Magaletta. 2006. “The Short-Form Buss-Perry Aggression Questionnaire (BPAQ-SF): A Validation Study with Federal Offenders.” *Assessment* 13 (3): 227–40.

Duggan, Maeve. 2017. “Online Harassment 2017.” *Pew Research*. July 11. https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/.

Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker. 2019. “How Many People Live in Political Bubbles on Social Media? Evidence from Linked Survey and Twitter Data.” *SAGE Open* 9 (1). https://doi.org/10.1177/2158244019832705.

Epley, Nicholas, and David Dunning. 2000. “Feeling ‘Holier than Thou’: Are Self-Serving Assessments Produced by Errors in Self- or Social Prediction?” *Journal of Personality and Social Psychology* 79 (6): 861–75.

Eveland, William P., and Myiah Hutchens Hively. 2009. “Political Discussion Frequency, Network Size, and ‘Heterogeneity’ of Discussion as Predictors of Political Knowledge and Participation.” *Journal of Communication* 59 (2): 205–24.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. “Ideological Segregation Online and Offline.” *Quarterly Journal of Economics* 126 (4): 1799–839.

Gerber, Alan S., and Todd Rogers. 2009. “Descriptive Social Norms and Motivation to Vote: Everybody’s Voting and so Should You.” *Journal of Politics* 71 (1): 178–91.

Gibson, Rachel, and Marta Cantijoch. 2013. “Conceptualizing and Measuring Participation in the Age of the Internet: Is Online Political Engagement Really Different to Offline?” *Journal of Politics* 75 (3): 701–16.

Gorrell, Genevieve, Mark Greenwood, Ian Roberts, Diana Maynard, and Kalina Bontcheva. 2018. “Online Abuse of UK MPs in 2015 and 2017: Perpetrators, Targets, and Topics.” Working Paper. http://arxiv.org/abs/1804.01498.

Gratz, Kim L., and Lizabeth Roemer. 2004. “Multidimensional Assessment of Emotion Regulation and Dysregulation.” *Journal of Psychopathology and Behavioral Assessment* 26 (1): 41–54.

Griezel, Lucy, Linda R. Finger, Gawaian H. Bodkin-Andrews, Rhonda G. Craven, and Alexander Seeshing Yeung. 2012. “Uncovering the Structure of and Gender and Developmental Differences in Cyber Bullying.” *Journal of Educational Research* 105 (6): 442–55.

Grossman, Lev. 2010. “Person of the Year 2010: Mark Zuckerberg.” *Time*, December 15.

Guess, Andrew M. 2015. “Measure for Measure: An Experimental Test of Online Political Media Exposure.” *Political Analysis* 23 (1): 59–75.

Guess, Andrew M., Kevin Munger, Jonathan Nagler, and Joshua Tucker. 2019. “How Accurate Are Survey Responses on Social Media and Politics?” *Political Communication* 36 (2): 241–58.

Haidt, Jonathan, and Tobias Rose-Stockwell. 2019. “The Dark Psychology of Social Networks.” *The Atlantic*, December, 6–60.

Hertel, Guido, Susanne Geister, and Udo Konradt. 2005. “Managing Virtual Teams: A Review of Current Empirical Research.” *Human Resource Management Review* 15 (1): 69–95.

Hesse, Bradford W., Carol M. Werner, and Irwin Altman. 1988. “Temporal Aspects of Computer-Mediated Communication.” *Computers in Human Behavior* 4 (2): 147–65.

Hill, R. A., and R. I. M. Dunbar. 2003. “Social Network Size in Humans.” *Human Nature* 14: 53–72.

Holmes, Marcus. 2013. “The Force of Face-to-Face Diplomacy: Mirror Neurons and the Problem of Intentions.” *International Organization* 67 (4): 829–61.

Kalmoe, Nathan P. 2014. “Fueling the Fire: Violent Metaphors, Trait Aggression, and Support for Political Violence.” *Political Communication* 31 (4): 545–63.

Kurzban, Robert. 2001. “The Social Psychophysics of Cooperation: Nonverbal Communication in a Public Goods Game.” *Journal of Nonverbal Behavior* 25: 241–59.

Li, Norman P., Mark van Vugt, and Stephen M. Colarelli. 2018. “The Evolutionary Mismatch Hypothesis: Implications for Psychological Science.” *Current Directions in Psychological Science* 27 (1): 38–44.

Lim, Daniel, Paul Condon, and David De Steno. 2015. “Mindfulness and Compassion: An Examination of Mechanism and Scalability.” *PLoS ONE* 10 (2): 1–8.

Matias, J. Nathan. 2016. “High Impact Questions and Opportunities for Online Harassment Research and Action.” MIT Center for Civic Media, Report. September 7. https://civic.mit.edu/index.html%3Fp=930.html.

Matias, J. Nathan. 2019. “Preventing Harassment and Increasing Group Participation through Social Norms in 2,190 Online Science Discussions.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (20): 9785–89.

Mendes, Kaitlynn, Jessica Ringrose, and Jessalynn Keller. 2018. “#MeToo and the Promise and Pitfalls of Challenging Rape Culture through Digital Feminist Activism.” *European Journal of Women’s Studies* 25 (2): 236–46.

Nelson, Michelle R., and Sharon Shavitt. 2002. "Horizontal and Vertical Individualism and Achievement Values: A Multimethod Examination of Denmark and the United States." *Journal of Cross-Cultural Psychology* 33 (5): 439–58.

Olaniran, Bolanle. 2002. "Computer-Mediated Communication: A Test of the Impact of Social Cues on the Choice of Medium for Resolving Misunderstandings." *Journal of Educational Technology Systems* 31 (2): 205–22.

Petersen, Michael Bang, Andreas Roepstorff, and Søren Serritzlew. 2009. "Social Capital in the Brain?" Chap. 5 in *Handbook of Social Capital: The Troika of Sociology, Political Science and Economics*, eds. Gert Tinggaard Svendsen and Gunnar Lind Haase Svendsen. Cheltenham, UK: Edward Elgar

Petersen, Michael Bang, Mathias Osmundsen, and Alexander Bor. 2020. "Beyond Populism: The Psychology of Status-Seeking and Extreme Political Discontent." Chap. 4 in *The Psychology of Populism*, eds. Joseph Forgas, Bill Crano, and Klaus Fiedler. New York: Routledge.

Rasmussen, Stig H. R., Alexander Bor, Mathias Osmundsen, and Michael Bang Petersen. 2021. "Super-unsupervised Text Classification for Labeling Online Political Hate." PsyArXiv. June 8. doi:10.31234/osf.io/8m5dc.

Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2019. "Auditing Radicalization Pathways on YouTube." Working Paper. http://arxiv.org/abs/1908.08313.

Rowe, Ian. 2015. "Civility 2.0: A Comparative Analysis of Incivility in Online Political Discussion." *Information Communication and Society* 18 (2): 121–38.

Scharlemann, Jörn P. W., Catherine C. Eckel, Alex Kacelnik, and Rick K. Wilson. 2001. "The Value of a Smile: Game Theory with a Human Face." *Journal of Economic Psychology* 22 (5): 617–40.

Settle, Jaime E. 2018. *Frenemies: How Social Media Polarizes America*. Cambridge: Cambridge University Press.

Settle, Jaime E., and Taylor N. Carlson. 2019. "Opting out of Political Discussions." *Political Communication* 36 (3): 476–96.

Siegel, Alexandra A., and Vivienne Badaan. 2020. "# No2Sectarianism: Experimental Approaches to Reducing Sectarian Hate Speech Online." *American Political Science Review* 114 (3): 837–55.

Simpson, Jeffry A., and Jay Belsky. 2008. "Attachment Theory within a Modern Evolutionary Framework." Chap. 5 in *Handbook of Attachment: Theory, Research, and Clinical Application*, 2nd edition, eds. Jude Cassidy and Phillip R. Shaver. New York: The Guilford Press.

Stein, Joel. 2016. "Tyranny of the Mob." *Time* 188 (8): 26–32.

Sydnor, Emily. 2019. *Disrespectful Democracy: The Psychology of Political Incivility*. New York: Columbia University Press.

TrygFonden. 2017. "Ytringsfrihed Og Digital Usikkerhed (Freedom of Expression and Digital Uncertainty)." Copenhagen. http://reader.livedition.dk/trygfonden/306/html5/.

Valenzuela, Sebastian, Yonghwan Kim, and Homero Gil de Zuniga. 2012. "Social Networks That Matter: Exploring the Role of Political Discussion for Online Political Participation." *International Journal of Public Opinion Research* 24 (2): 163–84.

Wilson, Margo, and Martin Daly. 1985. "Competitiveness, Risk Taking, and Violence: The Young Male Syndrome." *Ethology and Sociobiology* 6 (1): 59–73.

Witschge, Tamara. 2004. "Online Deliberation: Possibilities of the Internet for Deliberative Democracy." In *Democracy Online: The Prospects for Political Renewal through the Internet*, ed. Peter M. Shane, 109–22. New York: Routledge.

Wolchover, Natalie. 2012. "*Why Is Everyone on the Internet So Angry?*" Scientific American, July 25. https://www.scientificamerican.com/article/why-is-everyone-on-the-internet-so-angry/.