The Tip of the Iceberg: How the Social Media Production-Consumption Gap Distorts Public Opinion for Citizens and Researchers

Lisa Oswald[1*] William Small Schulz[2] Ralph Hertwig[1] David Lazer[3] Sebastian Stier[4,5]

October 22, 2025

[1] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin
[2] Social Media Lab and Human-Centered AI Institute, Stanford University
[3] NULab for Digital Humanities and Computational Social Science, Northeastern University
[4] Department Computational Social Science, GESIS – Leibniz Institute for the Social Sciences, Cologne
[5] School of Social Sciences, University of Mannheim

Abstract

The production–consumption gap on social media is a consistent finding across time, platforms, and cultural contexts: A small minority of highly active users produce the majority of online political content, while the majority of users consume content passively and remain largely silent. Online content thus reveals only the tip of an iceberg, from which citizens and scholars alike are apt to draw incorrect inferences regarding the submerged mass of public opinion. This has substantive as well as methodological consequences for social media research, which must be taken into account when designing studies to describe and understand how social media use relates to content exposure, public opinion, and political behavior, and when designing and testing pro-democratic interventions.

Online participation: The tip of the iceberg

Despite enormous interest in how social media affects politics, research findings have been mixed [63], possibly due to the inherent challenges of studying complex and rapidly-evolving systems [72]. Yet one finding persists reliably across time frames, platforms, and political cultures: a gap between the producers and consumers of online content [21, 73, 76]. By way of illustration, Figure 1 presents findings from the major election survey in Germany, showing that most online political content is produced by a small minority of highly active users ("power users" [17]), while the majority remain silent, passive consumers of this content ("lurkers" [77, 102]). In the U.S., most users say they "never" (37%) or "rarely" (34%) share their political views online [35]; likewise, only 9.5% of Germans reported posting about politics on social media in the week leading up to the 2021 national election [37]. Recent evidence from YouTube shows that just 2% of commenters produce 50% of comments under political videos [73]. Though evidence of the narrow origins of online content is widespread and longstanding, scholars and citizens have yet to fully confront social media's failure to fulfill its promise of truly democratizing public discourse.

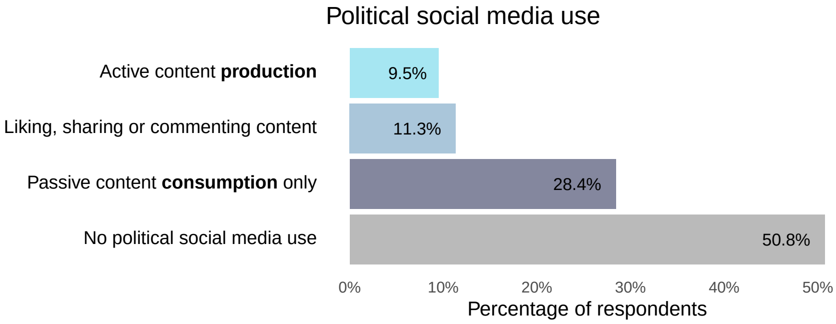

Figure 1: *Data*: German Longitudinal Election Study (N = 10,143 [37]). Replication materials are available on OSF.

We argue that the online production–consumption gap has three important potential consequences that we must address: (1) Citizens develop a distorted perception of public opinion; (2) Researchers may develop a skewed understanding of public opinion and phenomena in digital media, stemming from a set of methodological challenges; (3) As analyses based on this biased data are communicated to the public, the distortion is amplified. Furthermore, if politicians and media outlets wrongly interpret social media content as a window into public opinion, the effects of the production–consumption gap may distort the perceptions of non-social-media users, and even the positions taken by legislators and other public policymakers. Finally, as large language models (LLMs) are trained on text corpora including social media content, distortions of public opinion are translated into generative AI models and reproduced, for example, in interactions with AI Chatbots.

This makes the production–consumption gap particularly problematic in the context of *political* behavior, as the consequences are relevant for all political decisions–including elections–that impact every citizen, even those who choose to remain silent. Although selection effects occur in other domains, such as consumer choices to write reviews on platforms like Amazon, the potential harm in the political realm is likely much greater. Indeed, while the production-consumption gap echoes other fat tailed distributions inherent to complex systems [21], there is evidence for a qualitative difference between producers and consumers of political content online. While the consumption of political content is predicted by classic factors such as political interest [65, 84], its production is predicted by additional factors, such as partisan ideological strength [55][1] – which suggests that the production-consumption gap could have a polarizing influence on political discourse.

More generally, because content producers self-select into their role, they are often systematically different from other users, and the content they produce is often highly unrepresentative of the broader user population. The most politically-outspoken users represent the proverbial *tip of the iceberg* (see Figure 2): a very visible part of public opinion that may mislead observers as to the contours of the whole object, the variance of opinion distributions, and the majority's leanings. This production–consumption gap has been found for political posting on Twitter [7, 49], misinformation sharing [8, 28, 41], and toxic commenting [55, 73, 82], which are widely considered to be among the most consequential online political behaviors.

However, while these patterns of (non-)participation have been evident for some time, science has been slow to appreciate their important implications for research on online platforms and their users. In this article, we identify several substantive and methodological implications that

[1] This pattern is also observed among politicians [30].

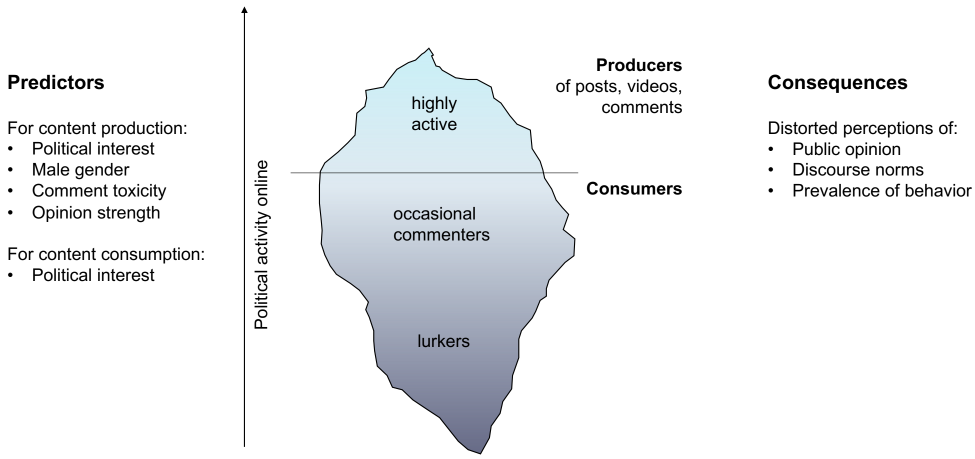

Figure 2: Predictors and consequences of participation inequality in political discourse on social media identified in the literature.

threaten the validity of certain social media research designs. We recommend best practices for study design and analysis to mitigate these threats. Further, we discuss the implications for ordinary users, and how the misrepresentativeness of online discourse might contribute to attitudinal polarization. We conclude by discussing how to apply these recommendations to address various open questions in future social media research.
While the iceberg metaphor suggests a simplified binary split between visible producers and passive consumers, the hierarchy of engagement is more fine grained and differs between platforms depending on the specific affordances. Usually, at the very tip of the iceberg, we have content production (in the form of writing posts and comments), followed by content evaluation (sharing content or emotive reactions such as "likes"). Both forms of active engagement are visible to everyone—users, platforms and researchers. Below this, we have content consumption and attention which scientists can only access via web-tracking data but which is an important internal metric for platforms. Finally, those disengaged with political content are invisible to anyone as they don't leave digital traces and could only be inferred in comparison to, e.g., census data.

Implications for social media research methods

Until the recent advent of the so-called "post-API age," [34] social media data constituted a low-hanging fruit for computational social scientists, providing unprecedented access to vast amounts of user-generated content, which scholars have put to use in many different study designs. Yet it has become increasingly clear that certain kinds of such analyses are influenced by the production-consumption gap.
First among these is the use of social media content as a proxy for public opinion. Although some analyses of social media data have found that social media discourse tracks changes in public opinion over time, reflects differences between geographic regions, or has "predicted" specific election outcomes [1, 9, 10, 11, 22, 23, 27, 29, 32, 36, 38, 52, 54, 79, 80, 105, 112], online content is not a reliable measure of public opinion (for a critique and guidance, see [98]). Traditional methods of measuring public opinion, such as polls and surveys, are designed to ensure that the barriers to opinion expression are minimal, and that every individual's perspective is given equal

weight—reflecting democratic values. Estimating public opinion on the basis of social media content abandons this principle in the interest of more convenient data collection, giving more weight to the views of highly vocal and potentially unrepresentative users. This can distort evidence about what the broader population really thinks. To phrase it in terms of the iceberg metaphor: Most behavior is below the water, and what is above the water is not reflective of what is below the water. First, there is a descriptive discrepancy that researchers tend to focus on what is above the water. Second, there can be an inferential discrepancy when researchers wrongly generalize from above the water to below—while likely less frequent than the descriptive discrepancy, these mistaken inferences are most problematic.

The fundamental challenge that social media scholars face is that the availability of evidence diverges drastically for different types of users and different modes of data collection. Only people who produce online content, such as posts and comments, leave visible digital traces that can be studied. In exploiting the relative abundance of such data in recent years, most research has focused on these content producers—the tip of the iceberg (Figure 2). Yet this kind of expressive behavioral data provides few insights into the underlying attitudes, motives, or identities of users—a limitation that has not prevented researchers from making inferences about these constructs, and potentially drawing incomplete or misleading conclusions. Survey data, in contrast, provides robust indicators of complex latent constructs that people can validly self-report, such as attitudes and identities [3], and allows more transparent sampling strategies. This is not to say that surveys are generally preferable to social media data. Surveys often fail, for example, to accurately capture online *behavior*: Self-reports tend to have limited validity when it comes to digital behavior, media consumption, and online content production [87]. One important behavioral dimension that is largely missing from *both* social media data and survey data, meanwhile, is information about *exposure* to online content – an ironic lacuna, given the tendency of social-scientific theories to focus on social media's role in exposing users to certain types of content. Passive web-tracking data can provide insights into the online behavior of the silent majority, including how they consume news media [101], although this method, too, comes with its own costs and frictions, as we discuss below.

Overall, these differences in the signals provided by various data sources result in a scientific knowledge gap, where research reveals different things about different populations. While this discrepancy is not inherently problematic, failing to acknowledge that studies are based on a skewed subset of the population can lead to incorrect conclusions. In the following subsections, we discuss several specific methodological challenges and provide recommendations to help mitigate them (see Table 1 for summary).

User-centered data collection for user-centered research questions

For some years now, there has been widespread concern that social media users may be negatively affected by the content they encounter online—especially when it comes to politics [35, 63]. Consequently, considerable effort has been invested in collecting and analyzing the political content shared on social media, and the dynamics by which it spreads through online networks [109]. For example, Barberá et al. [6] analyzed the ideological alignment between authors and retweeters of political content on Twitter to assess the extent to which online discourse on various topics took place within echo chambers, where content is predominantly shared by like-minded users. This pathbreaking work shed light on patterns of content-sharing among the small group of users who authored and retweeted political content. However, it cannot gauge the extent to which *typical users* (largely political content consumers, see Fig. 1) are also in echo chambers—that is, the extent to which they encounter like-minded content more often than content that challenges their views. Furthermore, the concept of online echo chambers relies heavily on understanding

Methodological challenges	Recommendations
Overrepresentation of active users in social media data (tip of the iceberg)	• Adopt user-centered data collection
Covariance of social media activity with online study compliance	• Reduce friction in study participation • Examine and report differential study compliance and attrition
Selection effects and collider bias	• Clarify theoretical and empirical estimand • Select sampling approach that fits the research objective
Limited generalizability of intervention findings	• Identify and recruit the intended target population • Examine treatment effect heterogeneity • Do not exclude participants based on failed manipulation checks, learn about realistic intervention take-up instead

Table 1: Implications for social media research

the content that people are *exposed to* [86, 103], a measure that is missing from most data sources that are accessible to researchers. More broadly, it is inadvisable to assume that a content-centered analysis is representative of user-level experiences, attitudes, or behaviors.

User-centered data collection methods [16] can help by bridging the discrepancy between those whose data is supplied by platforms and those who can be recruited into surveys. User-centered sampling in social media research involves recruiting and surveying participants before collecting digital trace data in order to avoid biasing the sample toward users who are disproportionately likely to generate digital traces. It covers various types of data collection, including "donated data" that participants provide voluntarily for research purposes. For example, users can download data packages from social media platforms [12] or install research software that tracks their browsing [24] and smartphone activity [91]. Working directly with social media platforms [39, 44, 78] can allow researchers to observe users who actively consume content without engaging (silent users) from those who are completely inactive but industry collaborations naturally challenge research independence [110]. Including a survey component is almost always beneficial for ensuring the representativeness of the sample and directly assessing participants' internal states, such as political attitudes [101]. By combining surveys with digital behavioral data, and making the sample's composition an integral part of the study design, researchers can better understand how different types of users—both consumers and producers—engage in online political discourse.

User-centered data collection methods adhere to foundational principles of survey methodology, such as carefully considering sampling and measurement issues. For example, following the total error framework [57, 98] helps ensure that researchers are aware of, and can mitigate, potential sources of sampling and measurement error that might distort findings. Moreover, in addition to providing technical solutions for data-access challenges in the post-API era [34, 45] data donations and other custom data collection techniques also uphold high ethical standards, including informed consent.

Given the reality of budget constraints, however, user-centered data collection is not always possible, and there is no one-size-fits-all sampling approach for social media research. For example, random sampling methods may be appropriate for some research questions, but might suffer from high non-compliance or dropout rates—an issue we outline below. In contrast, recruitment via social media advertisements [74] might attract a sample with a higher concentration of active, content-producing users, which could be advantageous or detrimental depending on the research question. Some social media phenomena—such as the emergence of cultural niches, minority political movements, or political extremism—thrive precisely because physical barriers to communication have been reduced for a small number of highly active people. Studying these phenomena may require targeted sampling strategies, including oversampling of populations that show a specific online behavior [71, 81], to ensure sufficient statistical power to analyze rare yet consequential kinds of content. In many such cases, it may be appropriate to use a sampling frame that focuses on active users [18, 83, 93]. However, even research focusing on niche populations can benefit from user-centered approaches [92]. As in all research, the best sample selection approach is the one that best fits the research question at hand.

Differential study compliance and attrition

Study non-compliance and differential attrition have long complicated social science research, and apply specifically to social media research. Those who frequently post content on a specific platform are more likely to comply with a study conducted on that platform [99]. Dropout rates are likely to co-vary with how active participants are online, especially in studies that involve complex data collection procedures that require participants to, for example, install research software, manually upload data, or interact with specific platforms. Furthermore, both frequent posting on social media and online study participation—with some even becoming "professional survey takers" [48]—depend on participants spending time on digital devices.

Accordingly, researchers need to systematically examine study compliance and differential attrition between user groups. Because active producers may be more willing to participate in online panels or accept survey invitations (e.g., to express their political opinion), methodological choices can inadvertently favor certain user groups. The choice of approach should therefore be informed by the research question: Studies aiming to understand typical user behaviors and exposure effects may require sampling strategies that capture a broader spectrum of users, while investigations into the drivers of content production should concentrate on high-output users.

Regarding questions of causal inference, it is essential to have a precise definition of the theoretical estimand—that is, what researchers want to know and *about whom*. Furthermore, it is crucial to be aware of potential collider bias that can arise in the empirical estimand if the study is inadvertently limited to a subpopulation, such as content producers [64]. As noted by Lundberg [64, p. 544], "when selection limits us to observing only a slice of the world, we can get counterintuitive results. Issues of sample selection may grow in importance as sociology explores new data sources."

Aligning the data collection approach with the research objective and seeking to minimize friction in study participation are practical steps toward reducing the risk of sampling bias (i.e., collider bias). Substantial friction can otherwise introduce sampling bias through low opt-in rates and high subsequent dropout patterns.

Considerations for intervention research

Similar considerations apply when designing and empirically testing experimental interventions on social media platforms. It is crucial to carefully consider which segments of the user population are intended to be affected by the treatment, and whether that aligns with the sample that can actually be recruited. Researchers must determine whether the intervention is aimed primarily at content producers and sharers (e.g., to change posting or sharing behavior) or at consumers (e.g., to change how content is consumed and evaluated) [88]. This focus becomes especially relevant when only a subset of users may be receptive to measures like media literacy prompts [56]—a typical treatment that is prone to self-selection [59]. Thus, identifying and understanding the "target" and "take-up" populations is vital. In social media research, the target audience often maps clearly onto the tip of the iceberg of online political engagement. The take-up audience—those who are likely to adopt and engage with the intervention—may be even more narrowly defined. Their receptiveness could be influenced by factors such as their attitudes towards the scientists who developed the intervention [2, 96].

It is, therefore, important to explore treatment effect heterogeneity between the groups, ideally incorporating such analyses into preregistration plans [15, 20] or employing a sequential testing approach. For instance, an intervention could initially be tested on the defined target sample to test its efficacy, subsequently scaled up to a population-representative sample to monitor for unintended side effects that might occur when the intervention is applied more widely, and finally tested in a field setting to assess how it interacts with the existing information environment (see also discussion by IJzerman et al. [50]). In other words, a population-representative sample is not necessarily the best option to test the effectiveness of a "pro-democratic" or "pro-social" intervention—the best sample to use depends strongly on the research question and the characteristics of target and take-up populations concerned. To use a medical analogy, no researcher would test the effectiveness of a cancer treatment in a population-representative sample that consists mostly of healthy people, as the relevant population is those affected by cancer.

In the context of testing interventions, researchers can reconsider the definitions of their theoretical estimands [64]—clearly specifying the target population of interest over which they will aggregate their unit-specific results. Examining patterns of selection and compliance—understanding the take-up population—further helps to clarify potential bias in the empirical estimand.

Intervention research usually seeks to inform policy decisions. However, the common practice of excluding participants who fail manipulation checks can make the study's findings meaningless for policy making—particularly when there is limited overlap between the target population and those who actually engage with the intervention. Manipulation checks are postexperimental measures widely used to verify whether a treatment was actually delivered to a participant. Suppose you want to reduce online hate speech with an empathy-based intervention [82]. You may be interested in exploring the theoretical potential of empathy as upper benchmark of what is possible—conditional on whether participants reported to have experienced empathy with the target of their comment which would be reflected by passing a manipulation check ("complier average treatment effect", CATE). However, this is very different from understanding the likely effects of an intervention rolled out in a population where not everyone will be reached by the treatment ("intention to treat effect", ITT)—despite seeing a prompt to empathize, *not* empathizing with a target. Estimating effects in the subpopulation of study compliers can hugely inflate the effects of an intervention [106]. For instance, if a media literacy intervention is tested in a population-representative sample that only contains people who were motivated enough to engage with the treatment, the results might suggest strong effectiveness. However, if the actual target group is unlikely to accept media literacy training, the intervention's true impact could be minimal. Successfully recruiting target populations is already a significant challenge—such as when an intervention aims to increase trust in science among a low-trust population. However, if a study only analyzes data from the even narrower population of those who fully comply with the treatment, the findings may be entirely detached from reality.

While facing these challenges with individual-level behavioral interventions, acknowledging the production-consumption gap also highlights the potential impact of system-level interventions such as (A) content moderation policies with a focus on high impact accounts [104] and (B) algorithmic interventions affecting the visibility of different types of content [89].

Implications for substantive research

The skewed production of online discourse also has implications for attitudinal polarization in society more broadly—a phenomenon of central interest for social scientists [31, 40, 47, 51, 60, 70, 95]. This is because the excessive visibility of a highly active minority at the tip of the iceberg can not only mislead social scientists, but also deceive social media users themselves: Users seeing the disproportionate contribution of a highly vocal minority may believe that these—often more extreme—views are more common or widely accepted than they actually are. This visibility can lead users to develop a distorted mental representation of social norms and inaccurate perceptions about others' views [94].

A question that may arise is whether this is any different from pre-online times. Participation in political discussions has never been representative of the population at large—those most vocal in town hall meetings or political discussions among the neighborhood also tended to be unusually politically interested and predominantly male [53]. However certain characteristics of digital media, such as potentially much greater audiences for those who voice opinions, the removal of spatial barriers for otherwise isolated individuals holding extreme views, and intransparent algorithmic curation of user feeds with tendencies to fuel toxic content suggest a sharpening skew of participation in public discourse.

The tip of the iceberg and perceptions of public opinion

At a psychological level, viewing an unrepresentative sample of opinion is likely to engender misperceptions about the true distribution of opinion in broader society. This is due in part to "metacognitive myopia," the tendency to correctly process sampled information but neglect its underlying sources and contexts, thereby overestimating the generalizability of the sample to the population [33]. Another cognitive factor that may reinforce these distortions is the "feature-positive effect," which makes it (much) easier for people to process events or information that they can see (positive features) than those that cannot be seen (negative features). For example, on social media, people pay attention to the visible content, but overlook the absence of views from the silent majority, and therefore discount them [75].

The skewing of perceptions of public opinion through the nominal overrepresentation of content from a highly active minority is amplified by the "overperception" of emotionality. Social media users overperceive the levels of moral outrage expressed in posts, which in turn inflates beliefs about intergroup hostility [14]. Metaphors like a "prism" [4, 55] and a "funhouse mirror" [94] have been used to describe this distorting effect, capturing how social media warps perceptions, giving an exaggerated version of reality. In the following, we focus on the nominal overrepresentation and visibility of content—represented by the tip of the iceberg metaphor.

The phenomenon that people hold misperceptions of how the majority of a group thinks or behaves has previously been captured under the term "pluralistic ignorance" [68]. More specifically, perceptions of public opinion can be skewed in a magnitude that turns a majority opinion into a perceived minority opinion, creating a "false social reality" [100, p. 2]. For example, a strong form of pluralistic ignorance has recently been found in the context of climate change where 80 to 90 percent of Americans strongly underestimate the prevalence of support for climate change mitigation policies [100].

In the context of online political behavior, these distortions specifically affect the perceived prevalence of certain political attitudes—highlighting those held by a minority of highly active users, and detracting from those of a largely silent majority [46, 69]. This skew also extends to misconceptions about the proportion of people who contribute to toxic content online, giving the impression that negative views are more widespread than they actually are [58].

When perceptions of public opinion, including discourse norms and mainstream beliefs, become distorted, they shape individual attitudes and behavior through social learning [13]. For example, if someone repeatedly encounters messages that suggest a particular political stance is widely accepted, they may begin to adopt that stance themselves, even if it is not actually as common as it appears. Research indicates that social norms—especially descriptive norms, which are the beliefs about what others are doing based on observed behavior—exert a robust influence on attitudes and actions across various areas [26, 67]. In fact, aligning one's behavior with what is seen as the group norm is a socially adaptive response [25, 111]. So, if hostile and polarizing behavior is over-represented online, it may induce more users to engage in such behavior.

Implications for mass polarization

Political polarization, as a specific case of a shift in public opinion, can from this perspective be understood as an alignment process between individual opinion and *perceived* public opinion, catalyzed through mechanisms of social learning. This proposed mechanism differs from the idea that people self-select into "echo chambers" [103] or "filter bubbles" [86] that reinforce their existing views through *content exposure*. While the existence of homophilic social networks online—but also offline—is undisputed, exposure to diverse content is also common in online settings [42]. The proposed solution to the polarizing effects of echo chambers was to expose people to opposing views. However, empirical support for this solution is weak, and some studies have even suggested that exposure to opposing viewpoints reinforced the majority group and increased polarization [4, 83]. More generally, empirical research has shown mixed effects of content exposure—the primary element in theories of political polarization—often not translating into significant changes in downstream political attitudes [43, 44]. However, strong and sustained changes in exposure do shift attitudes [19, 89, 90].

The alternative, implied by our understanding of the iceberg phenomenon, is a mechanism of alignment between personal and perceived public opinion, and focuses on the perception process that informs social norm learning. It suggests that this alignment process—which may or may not result in public opinion shifts, such as polarization—results from how people perceive the distribution of opinions in society. However, when using social media, one can exclusively observe expressions of support or rejection from the minority of producers.

This alternative mechanism helps to bridge discrepancies between correlational and causal evidence on digital media use and political polarization. While correlational evidence shows consistent associations [63], causal evidence of the polarizing effect of social media through content exposure is mixed [19, 43, 44, 90]. However, the reversed path—people with more extreme political views being more active on social media—is fully in line with findings on the tip of the iceberg phenomenon. While this may be interpreted as good news from the perspective of platforms who may not want to take responsibility to be the root cause of polarization, we argue that this hyper-visibility of extreme political views may be the core problem with digital media and politics.

Beneath the tip of the iceberg: research opportunities beyond the visible

Rethinking designs for interventions

Interventions aimed at counteracting the presumed effects of online echo chambers, by exposing people to cross-cutting content, have not been consistently effective in reducing polarization [5].

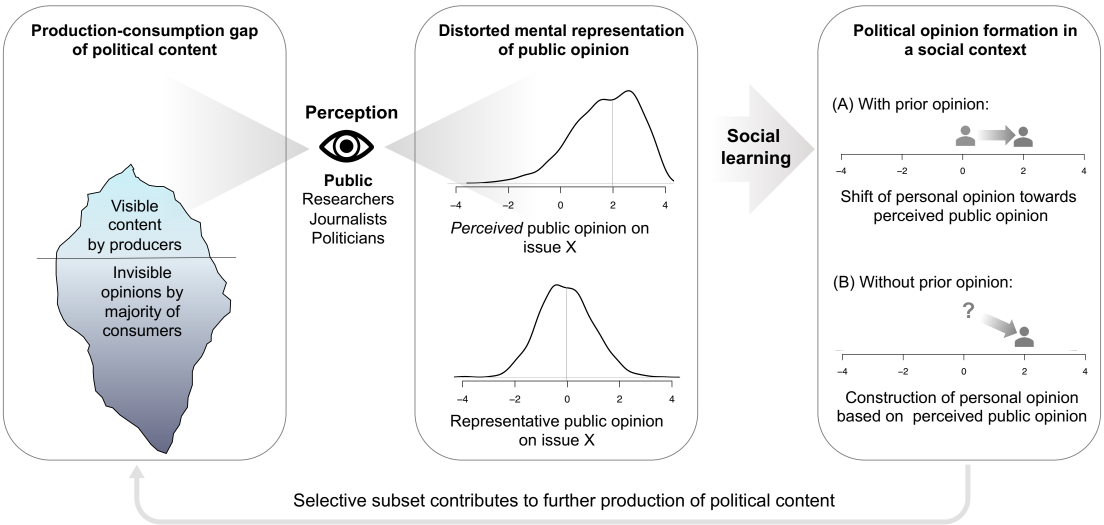

Figure 3: The production–consumption gap on social media and its consequences for public opinion perceptions and opinion formation. A small minority of highly active producers generates visible online political content (tip of the iceberg), while the silent majority of consumers remains submerged and underrepresented (left). This selective visibility distorts the perceived distribution of public opinion among citizens, researchers, journalists, and policymakers (center). In turn, these distorted perceptions shape individual political opinions through social learning—if prior opinions on an issue exist, personal opinions can shift towards perceived majority public opinions, if no prior opinion exists for novel issues, personal opinions can be constructed based on perceived majority public opinions (right). A selective subset, potentially those with stronger opinions, contribute to further political content production on social media (feedback loop).

Given the inconclusive evidence, there is a need to rethink the mechanisms behind political polarization and shifts in discourse norms.

Following our line of reasoning, we instead highlight the potential of placing the production–consumption gap at the core of intervention design, using meta-cognitive techniques to recalibrate public perceptions of social norms and the distribution of opinions [58]. A first step of such interventions is to convey an understanding of the tip of the iceberg phenomenon—that social media creates perceptual distortions that translate into inaccurate mental representations of public opinion. Even though the literature on metacognitive myopia suggests that it is very difficult to fix such distorted mental representations, even if one is aware of the nature of the distortion [33], understanding this distortion is a necessary first step to enable people to meta-cognitively correct for it. This approach finds a parallel in consensus messaging interventions in climate change communication [61, 97, 107], which aim to correct the false balance created by media reports giving disproportionate attention to dissenting views on anthropogenic climate change. Another way to increase the saliency of distortions induced by the production–consumption gap may be to provide social media users with cues that show base rates and passive behavior such as view counts, in addition to like and share counts [62].

Additionally, there is potential in exploring interventions that directly address second-order effects, such as the disparities between what individuals believe others think and actual majority policy preferences. One of the most effective treatments identified in a U.S. mega study [108] involved presenting survey data to correct inflated perceptions of outpartisan support for undemocratic practices, thereby reducing such support. This suggests that the provision of accurate and corrective information can mitigate the polarizing effects of unrepresentative content creation on social media. Interventions could take the form of simple reminders that social media content is unrepresentative, or more sophisticated contextual annotations that indicate how "fringe" a given piece of content is relative to the general population. This approach aims to prevent the normalization of extreme views by ensuring that users are aware that these views are not representative of the broader population.

While these interventions focus on perceptions of the consumers—the demand side of political social media content—the nature of the production-consumption gap highlights the potential impact supply-side interventions in the form of content moderation policies with a focus on high impact accounts [104] and algorithmic interventions affecting the visibility of different types of content [89].

Open substantive research questions

Beyond developing interventions, research is needed to answer several important and unresolved empirical questions. For example, why do some users speak up while others remain silent? Evidence suggests that those who create political content online are more likely to be male, have a strong interest in politics, and hold more extreme ideological views [55, 85]. The association between political interest and online political content creation may be inevitable, but the gender imbalance is troubling, and the absence of moderate voices online is particularly concerning in relation to polarization and merits further study. Some moderates may stay silent simply due to lack of interest in politics—this may not be problematic from a normative standpoint. More concerning is the possibility that moderates are choosing not to speak up due to structural factors, which, if so, may only be addressed by means of structural changes [85].

One important structural factor is a platform's use of content-ranking algorithms. They determine which content appears most prominently in users' feeds, and typically seek to maximize user engagement. However, the content that is most engaging may not be particularly representative of the views of the broader user population; on the contrary, the most attention-grabbing content my tend to be that which is unusual and striking [66] – features that may promote the visibility of fringe views in social media feeds, and may systematically spotlight users who hold such views, enabling them to achieve greater status and reach on platforms, and encouraging their continued creation of fringe political content. Future research should investigate how these algorithms influence participation equilibria and consider interventions to ensure a more balanced representation of voices, mitigating the polarization effect.

Conclusion

The production–consumption gap in online political engagement can be illustrated using the tip of the iceberg metaphor: a few highly visible users actively produce content above the surface, while the largely silent majority passively consume content below the surface. This gap distorts perceptions of public opinion for social media users, journalists, politicians, and researchers alike. It has methodological implications for both the internal validity of social media research—particularly in terms of how social media use relates to political constructs—and its external validity, particularly when designing "pro-democratic" interventions to mitigate certain phenomena.

We identify two areas of methodological challenges in social media research: (1) misleading inferences about "public opinion" and the prevalence of specific behaviors due to the selective study of content producers in social media data (2) limits to the generalizability of treatment effects from "pro-democratic" interventions that were tested on populations that differ from the intended target populations. Linking back to the influential work of Lundberg et al. [64], the first challenge regards the empirical estimand whereas the second regards the theoretical estimand.

We propose several solutions to address these challenges. First, we call for a shift towards user-centered data collection approaches that combine digital behavioral data with survey data and implement a clearly defined sampling strategy, as well as to reduce friction in study designs and systematically examine study compliance and attrition. Second, we emphasize the importance of clearly identifying the target and likely take-up populations of an intervention, examining treatment effect heterogeneity and transparently aligning sampling procedures with the research objective.

Finally, we transfer our observations from the process of scientific knowledge generation to the process of public opinion formation. The proposed perception-and-learning (Fig. 3) process offers an alternative perspective on how social media use relates to phenomena such as opinion and discourse shifts. Placing the production-consumption gap central, allows us to move beyond classic theories of individual-level content exposure but takes into account both: the social process of public opinion formation as well as the structural realities of political content in online environments. The assumed distortions of public opinion perceptions that inform social learning processes are entirely plausible on a cognitive level and in line with extant psychological literature. We encourage empirical researchers to further examine this process through observational, experimental, and theoretical work.

Data availability

Data and code relating to the project are available at OSF:
https://osf.io/hz36b/?view_only=d775647510df4a50a547502d19d08e57

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Susannah Goss for editing this manuscript. L.O. acknowledges funding from the European Commission (Horizon 2020 grant 101094752 SoMe4Dem). R.H. received support from the Volkswagen Foundation (large grant 'Reclaiming individual autonomy and democratic discourse online: How to rebalance human and algorithmic decision making') and from the European Union's Horizon 2020 grant 101094752 (SoMe4Dem).

References

1 Adams-Cohen, N. J. (2020). Policy Change and Public Opinion: Measuring Shifting Political Sentiment With Social Media Data. *American Politics Research*, *48*(5), 612–621. https://doi.org/10.1177/1532673X20920263

2 Agley, J., & Xiao, Y. (2023). Low trust in science may foster belief in misinformation by aligning scientifically supported and unsupported statements [Publisher: SAGE Publications Ltd]. *Perspectives in Public Health*, *143*(4), 199–201. https://doi.org/10.1177/17579139221136722

3 Arslan, R. C., Brümmer, M., Dohmen, T., Drewelies, J., Hertwig, R., & Wagner, G. G. (2020). How people know their risk preference [Publisher: Nature Publishing Group]. *Scientific Reports*, *10*(1), 15365. https://doi.org/10.1038/s41598-020-72077-5

4 Bail, C. A. (2021). *Breaking the social media prism: How to make our platforms less polarizing* (1st). Princeton University Press.

5 Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

6 Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

7 Barberá, P., & Rivero, G. (2015). Understanding the Political Representativeness of Twitter Users [Publisher: SAGE Publications Inc]. *Social Science Computer Review*, *33*(6), 712–729. https://doi.org/10.1177/0894439314558836

8 Baribi-Bartov, S., Swire-Thompson, B., & Grinberg, N. (2024). Supersharers of fake news on Twitter. *Science*, *384*(6699), 979–982. https://doi.org/10.1126/science.adl4435

9 Barnett, G. A., Xu, W. W., Chu, J., Jiang, K., Huh, C., Park, J. Y., & Park, H. W. (2017). Measuring international relations in social media conversations. *Government Information Quarterly*, *34*(1), 37–44. https://doi.org/10.1016/j.giq.2016.12.004

10 Beauchamp, N. (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. *American Journal of Political Science*, *61*(2), 490–503. https://doi.org/10.1111/ajps.12274

11 Bermingham, A., & Smeaton, A. F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results.

12 Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388–423. https://doi.org/https://doi.org/10.5117/CCR2022.2.002.BOES
13 Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, *7*(33), eabe5641. https://doi.org/10.1126/sciadv.abe5641
14 Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., & Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *7*(6), 917–927. https://doi.org/10.1038/s41562-023-01582-0
15 Brashier, N. M. (2024). Fighting misinformation among the most vulnerable users [Place: Netherlands Publisher: Elsevier Science]. *Current Opinion in Psychology*, *57*. https://doi.org/10.1016/j.copsyc.2024.101813
16 Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2023). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research [Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2022.2097015]. *Information, Communication & Society*, *26*(14), 2649–2668. https://doi.org/10.1080/1369118X.2022.2097015
17 Bright, J., Bermudez, S., Pilet, J.-B., & Soubiran, T. (2019). Power users in online democracy: Their origins and impact. *Information, communication & society*, 1–16.
18 Bright, J., Marchal, N., Ganesh, B., & Rudinac, S. (2022). How Do Individuals in a Radical Echo Chamber React to Opposing Views? Evidence from a Content Analysis of Stormfront. *Human Communication Research*, *48*(1), 116–145. https://doi.org/10.1093/hcr/hqab020
19 Broockman, D. E., & Kalla, J. L. (2025). Consuming Cross-Cutting Media Causes Learning and Moderates Attitudes: A Field Experiment with Fox News Viewers [Publisher: The University of Chicago PressChicago, IL]. *The Journal of Politics*. https://doi.org/10.1086/730725
20 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3
21 Carron-Arthur, B., Cunningham, J. A., & Griffiths, K. M. (2014). Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. *Internet Interventions*, *1*(4), 165–168. https://doi.org/10.1016/j.invent.2014.09.003
22 Chen, S.-N., Al-Hamdi, R., Tan, Y.-K., Kasiwi, A. N., & Nurmandi, A. (2020). Measuring Public Opinion with Social Media Use in Local Government of Asian Cities. In J. C. Hung, N. Y. Yen, & J.-W. Chang (Eds.), *Frontier Computing* (pp. 399–406, Vol. 551). Springer Singapore. https://doi.org/10.1007/978-981-15-3250-4_48
23 Chen, Y., Sack, H., & Alam, M. (2022). Analyzing social media for measuring public attitudes toward controversies and their driving factors: A case study of migration. *Social Network Analysis and Mining*, *12*(1), 135. https://doi.org/10.1007/s13278-022-00915-7
24 Christner, C., Urman, A., Adam, S., & Maier, M. (2022). Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations [Publisher: Routledge _eprint: https://doi.org/10.1080/19312458.2021.1907841]. *Communication Methods and Measures*, *16*(2), 79–95. https://doi.org/10.1080/19312458.2021.1907841
25 Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

26 Cialdini, R. B., & Jacobson, R. P. (2021). Influences of social norms on climate change-related behaviors. *Current Opinion in Behavioral Sciences*, *42*, 1–8. https://doi.org/10.1016/j.cobeha.2021.01.005
27 Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll (S. Lehmann, Ed.). *PLOS ONE*, *10*(8), e0136092. https://doi.org/10.1371/journal.pone.0136092
28 DeVerna, M. R., Aiyappa, R., Pacheco, D., Bryden, J., & Menczer, F. (2024). Identifying and characterizing superspreaders of low-credibility content on Twitter [Publisher: Public Library of Science]. *PLOS ONE*, *19*(5), e0302201. https://doi.org/10.1371/journal.pone.0302201
29 DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior (L. M. Martinez, Ed.). *PLoS ONE*, *8*(11), e79449. https://doi.org/10.1371/journal.pone.0079449
30 Eady, G., Bonneau, R., Tucker, J. A., & Nagler, J. (2024). News sharing on social media: Mapping the ideology of news media, politicians, and the mass public. *Political Analysis*. https://doi.org/10.1017/pan.2024.19
31 Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F. K., Bruns, A., O'Connor-Farfan, K. S., & Lubicz-Zaorski, C. (2024). Destructive polarization in digital communication contexts: A critical review and conceptual framework [Publisher: Routledge]. *Information Communication & Society*.
32 Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A. (2018). Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. *Business & Society*, *57*(1), 60–97. https://doi.org/10.1177/0007650316683926
33 Fiedler, K., Prager, J., & McCaughey, L. (2023). Metacognitive Myopia: A Major Obstacle on the Way to Rationality [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *32*(1), 49–56. https://doi.org/10.1177/09637214221126906
34 Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, *35*(4), 665–668. https://doi.org/10.1080/10584609.2018.1477506
35 Gallup. (2022, March). *Media and Democracy: Unpacking America's Complex Views on the Digital Public Square* (tech. rep.). Washington DC.
36 Gao, Y., Chen, Y., Mu, L., Gong, S., Zhang, P., & Liu, Y. (2022). Measuring urban sentiments from social media data: A dual-polarity metric approach. *Journal of Geographical Systems*, *24*(2), 199–221. https://doi.org/10.1007/s10109-021-00369-z
37 GLES. (2022). Gles panel 2021, welle 19. https://doi.org/10.4232/1.13920
38 Gong, Z., Cai, T., Thill, J.-C., Hale, S., & Graham, M. (2020). Measuring relative opinion from location-based social media: A case study of the 2016 U.S. presidential election (J. Zhao, Ed.). *PLOS ONE*, *15*(5), e0233660. https://doi.org/10.1371/journal.pone.0233660
39 González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., ... Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, *381*(6656), 392–398. https://doi.org/10.1126/science.ade7138
40 Green, J., Mccabe, S., Shugars, S., Chwe, H., Horgan, L., Cao, S., & Lazer, D. (2025). Curation Bubbles. *American Political Science Review*, 1–19. https://doi.org/10.1017/S0003055424000984
41 Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

42 Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*, ajps.12589. https://doi.org/10.1111/ajps.12589
43 Guess, A. M., Barberá, P., Munzert, S., & Yang, J. (2021). The consequences of online partisan media. *Proceedings of the National Academy of Sciences*, *118*(14). https://doi.org/10.1073/pnas.2013464118
44 Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? [Publisher: American Association for the Advancement of Science]. *Science*, *381*(6656), 398–404. https://doi.org/10.1126/science.abp9364
45 Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research [Place: United Kingdom Publisher: Taylor & Francis]. *Information, Communication & Society*, *22*(11), 1567–1581. https://doi.org/10.1080/1369118X.2019.1627386
46 Heltzel, G., & Laurin, K. (2024). Why Twitter Sometimes Rewards What Most People Disapprove of: The Case of Cross-Party Political Relations. *Psychological Science*, *35*(9), 976–994. https://doi.org/10.1177/09567976241258149
47 Hobolt, S. B., Lawall, K., & Tilley, J. (2023). The Polarizing Effect of Partisan Echo Chambers [Publisher: Cambridge University Press]. *American Political Science Review*, 1–16. https://doi.org/10.1017/S0003055423001211
48 Hohenberg, B. C. v., Ventura, T., Nagler, J., Menchen-Trevino, E., & Wojcieszak, M. (2024, September). Survey Professionalism: New Evidence from Web Browsing Data. https://doi.org/10.31219/osf.io/jxdc8
49 Hughes, A. (2019). A small group of prolific users account for a majority of political tweets sent by U.S. adults.
50 IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *4*(11), 1092–1094. https://doi.org/10.1038/s41562-020-00990-w
51 Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States [_eprint: https://doi.org/10.1146/annurev-polisci-051117-073034]. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034
52 Ji, X., Chun, S. A., Wei, Z., & Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining*, *5*(1), 13. https://doi.org/10.1007/s13278-015-0253-5
53 Karpowitz, C. F., & Mendelberg, T. (2014, August). *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton Univers. Press.
54 Kim, D., Chung, C. J., & Eom, K. (2022). Measuring Online Public Opinion for Decision Making: Application of Deep Learning on Political Context. *Sustainability*, *14*(7), 4113. https://doi.org/10.3390/su14074113
55 Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity [Publisher: Oxford University Press]. *Journal of Communication*, *71*(6), 922–946.
56 Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., Barzilai, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L. K., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., ... Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation [Publisher:

Nature Publishing Group]. *Nature Human Behaviour*, *8*(6), 1044–1052. https://doi.org/10.1038/s41562-024-01881-0
57 Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century [Publisher: Nature Publishing Group]. *Nature*, *595*(7866), 189–196. https://doi.org/10.1038/s41586-021-03660-7
58 Lee, A. Y., Neumann, E., Hancock, J., & Zaki, J. (2024, October). People Overestimate How Many Social Media Users Post Harmful Content. https://doi.org/10.31234/osf.io/f8y62
59 Leeper, T. J. (2017). How Does Treatment Self-Selection Affect Inferences About Political Communication? *Journal of Experimental Political Science*, *4*(1), 21–33. https://doi.org/10.1017/XPS.2017.1
60 Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, *80*(S1), 392–410. https://doi.org/10.1093/poq/nfw005
61 Lewandowsky, S., Cook, J., Fay, N., & Gignac, G. E. (2019). Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & cognition*, *47*(8), 1445–1456.
62 Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *4*(11), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7
63 Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide.
64 Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory [Publisher: SAGE Publications Inc]. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187
65 Mangold, F., Schoch, D., & Stier, S. (2024). Ideological self-selection in online news exposure: Evidence from Europe and the US [Publisher: American Association for the Advancement of Science]. *Science Advances*, *10*(37), eadg9287. https://doi.org/10.1126/sciadv.adg9287
66 McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). Misinformation exploits outrage to spread online [Publisher: American Association for the Advancement of Science]. *Science*, *386*(6725), 991–996. https://doi.org/10.1126/science.adl2829
67 Melnyk, V., Carrillat, F. A., & Melnyk, V. (2022). The Influence of Social Norms on Consumer Behavior: A Meta-Analysis [Publisher: SAGE Publications Inc]. *Journal of Marketing*, *86*(3), 98–120. https://doi.org/10.1177/00222429211029199
68 Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity [Place: US Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *53*(2), 298–305. https://doi.org/10.1037/0022-3514.53.2.298
69 Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *117*(26), 14864–14872. https://doi.org/10.1073/pnas.2001263117
70 Mukerjee, S., & Shen, C. (2025). The Antecedents and Manifestations of Political Polarization in Visual Media: Key Questions and Future Directions [Publisher: Routledge _eprint: https://doi.org/10.1080/10584609.2024.2446348]. *Political Communication*, *42*(1), 208–214. https://doi.org/10.1080/10584609.2024.2446348

71 Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, *39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5
72 Munger, K. (2019). Knowledge Decays: Temporal Validity and Social Science in a Changing World. Retrieved December 9, 2024, from https://www.semanticscholar.org/paper/Knowledge-Decays%3A-Temporal-Validity-and-Social-in-a-Munger/d52e0fbeaf7926fb72ce22a51361ab74bf8daa3
73 Munger, K., Hindman, M., Yalcin, O., Phillips, J., & Bisbee, J. (2025). Pressing Play on Politics: Quantitative Description of YouTube. *Journal of Quantitative Description: Digital Media*, *5*. https://doi.org/10.51685/jqd.2025.006
74 Neundorf, A., & Öztürk, A. (2023). How to improve representativeness and cost-effectiveness in samples recruited through meta: A comparison of advertisement tools [Publisher: Public Library of Science]. *PLOS ONE*, *18*(2), e0281243. https://doi.org/10.1371/journal.pone.0281243
75 Newman, J., Wolff, W., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 630–650. https://doi.org/10.1037/0278-7393.6.5.630
76 Nielsen, J. (2006). Participation Inequality: The 90-9-1 Rule for Social Features. *Nielsen Norman Group*. http://www.nngroup.com/articles/participation-inequality/.
77 Nonnecke, B., & Preece, J. (1999). Shedding light on lurkers in online communities. *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities, Edinburgh*, *123128*.
78 Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., . . . Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing [Publisher: Nature Publishing Group]. *Nature*, 1–8. https://doi.org/10.1038/s41586-023-06297-w
79 O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 122–129. https://doi.org/10.1609/icwsm.v4i1.14031
80 Oh, G., Zhang, Y., & Greenleaf, R. G. (2022). Measuring Geographic Sentiment toward Police Using Social Media Data. *American Journal of Criminal Justice*, *47*(5), 924–940. https://doi.org/10.1007/s12103-021-09614-z
81 Oswald, L. (2022). *Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users*.
82 Oswald, L. (2023, January). Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users. https://doi.org/10.31235/osf.io/evdxy
83 Oswald, L., & Bright, J. (2022). How Do Climate Change Skeptics Engage with Opposing Views Online? Evidence from a Major Climate Change Skeptic Forum on Reddit [Publisher: Routledge _eprint: https://doi.org/10.1080/17524032.2022.2071314]. *Environmental Communication*, *16*(6), 805–821. https://doi.org/10.1080/17524032.2022.2071314
84 Oswald, L., Munzert, S., Barbera, P., Guess, A., & Yang, J. (2022, October). Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data. https://doi.org/10.31235/osf.io/yfmzh
85 Oswald, L., Schulz, W., & Lorenz-Spreen, P. (2025, April). A Collective Field Experiment Disentangling Participation in Online Political Discussions. https://doi.org/10.31235/osf.io/p2jaq_v1
86 Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

87 Parry, D. A., Davidson, B. I., Sewall, C. J., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 1–13.
88 Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2
89 Piccardi, T., Saveski, M., Jia, C., Hancock, J. T., Tsai, J. L., & Bernstein, M. (2024, November). Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity [arXiv:2411.14652]. https://doi.org/10.48550/arXiv.2411.14652
90 Rathje, S., Pretus, C., He, J., Harjani, T., Roozenbeek, J., Gray, K., Linden, S. v. d., & Van Bavel, J. (2025, August). Unfollowing hyperpartisan social media influencers durably reduces out-party animosity. https://doi.org/10.31234/osf.io/acbwg_v2
91 Reeves, B., Robinson, T., & Ram, N. (2020). Time for the human screenome project. *Nature*, *577*(7790), 314–317.
92 Ribeiro, M., Calais, P., Santos, Y., Almeida, V., & Jr, W. M. (2018). Characterizing and Detecting Hateful Users on Twitter [Number: 1]. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1). https://doi.org/10.1609/icwsm.v12i1.15057
93 Ribeiro, M. H., Blackburn, J., Bradlyn, B., Cristofaro, E. D., Stringhini, G., Long, S., Greenberg, S., & Zannettou, S. (2021). The Evolution of the Manosphere across the Web. *Proceedings of the International AAAI Conference on Web and Social Media*, *15*, 196–207. https://doi.org/10.1609/icwsm.v15i1.18053
94 Robertson, C. E., del Rosario, K. S., & Van Bavel, J. J. (2024). Inside the funhouse mirror factory: How social media distorts perceptions of norms. *Current Opinion in Psychology*, *60*, 101918. https://doi.org/10.1016/j.copsyc.2024.101918
95 Rollwage, M., Zmigrod, L., de-Wit, L., Dolan, R. J., & Fleming, S. M. (2019). What Underlies Political Polarization? A Manifesto for Computational Political Psychology. *Trends in Cognitive Sciences*, *23*(10), 820–822. https://doi.org/10.1016/j.tics.2019.07.006
96 Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., Sultan, M., & Linden, S. v. d. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking [Publisher: Cambridge University Press]. *Judgment and Decision Making*, *17*(3), 547–573. https://doi.org/10.1017/S1930297500003570
97 Said, N., Frauhammer, L. T., & Huff, M. (2023). Consensus messaging in climate change communication: Metacognition as moderator variable in the gateway belief model. *Journal of Environmental Psychology*, *91*, 102128. https://doi.org/10.1016/j.jenvp.2023.102128
98 Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly*, *85*(S1), 399–422.
99 Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S., & Weiß, B. (2022). Linking surveys and digital trace data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *S2*(185), S387–S407. https://doi.org/https://doi.org/10.1111/rssa.12954
100 Sparkman, G., Geiger, N., & Weber, E. U. (2022). Americans experience a false social reality by underestimating popular climate policy support by nearly half [Publisher: Nature Publishing Group]. *Nature Communications*, *13*(1), 4779. https://doi.org/10.1038/s41467-022-32412-y

101 Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, *38*(5), 503–516.
102 Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, *38*, 110–117. https://doi.org/10.1016/j.chb.2014.05.022
103 Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.
104 Thomas, D. R., & Wahedi, L. A. (2023). Disrupting hate: The effect of deplatforming hate organizations on their online audience [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *120*(24), e2214080120. https://doi.org/10.1073/pnas.2214080120
105 Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*, *29*(4), 402–418. https://doi.org/10.1177/0894439310386557
106 Varaine, S. (2023). How Dropping Subjects Who Failed Manipulation Checks Can Bias Your Results: An Illustrative Case. *Journal of Experimental Political Science*, *10*(2), 299–305. https://doi.org/10.1017/XPS.2022.28
107 Većkalov, B., Geiger, S. J., Bartoš, F., White, M. P., Rutjens, B. T., van Harreveld, F., Stablum, F., Akın, B., Aldoh, A., Bai, J., Berglund, F., Bratina Zimic, A., Broyles, M., Catania, A., Chen, A., Chorzępa, M., Farahat, E., Götz, J., Hoter-Ishay, B., … van der Linden, S. (2024). A 27-country test of communicating the scientific consensus on climate change [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *8*(10), 1892–1905. https://doi.org/10.1038/s41562-024-01928-2
108 Voelkel, J. G., Stagnaro, M. N., Chu, J. Y., Pink, S. L., Mernyk, J. S., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., Allen, L. G., Allis, L. V., Baleria, G., Ballantyne, N., Van Bavel, J. J., Blunden, H., Braley, A., Bryan, C. J., Celniker, J. B., Cikara, M., … Willer, R. (2024). Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity [Publisher: American Association for the Advancement of Science]. *Science*, *386*(6719), eadh4764. https://doi.org/10.1126/science.adh4764
109 Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
110 Wagner, M. W. (2023). Independence by permission [Publisher: American Association for the Advancement of Science]. *Science*, *381*(6656), 388–391. https://doi.org/10.1126/science.adi2430
111 West, E. A., & Iyengar, S. (2022). Partisanship as a Social Identity: Implications for Polarization. *Political Behavior*, *44*(2), 807–838. https://doi.org/10.1007/s11109-020-09637-y
112 Zeng, J., Zhang, S., Wu, C., & Xie, J. (2007). Predictive Model for Internet Public Opinion. *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, 7–11. https://doi.org/10.1109/FSKD.2007.453