

Public Attitudes Towards Social Media Field Experiments

Straub, Vincent J.; Burton, Jason W.; Geers, Michael; Lorenz-Spreen, Philipp

Document Version

Final published version

Published in:

Scientific Reports

DOI:

[10.1038/s41598-024-76948-z](https://doi.org/10.1038/s41598-024-76948-z)

[10.1038/s41598-025-89155-1](https://doi.org/10.1038/s41598-025-89155-1)

Publication date:

2024

License

CC BY

Citation for published version (APA):

Straub, V. J., Burton, J. W., Geers, M., & Lorenz-Spreen, P. (2024). Public Attitudes Towards Social Media Field Experiments. *Scientific Reports*, 14(1), Article 26110. <https://doi.org/10.1038/s41598-024-76948-z>, <https://doi.org/10.1038/s41598-025-89155-1>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jul. 2025





OPEN Public attitudes towards social media field experiments

Vincent J. Straub^{1,5,6✉}, Jason W. Burton^{2,3}, Michael Geers^{3,4,7} & Philipp Lorenz-Spreen^{3,8}

The growing use of social media field experiments demands a rethink of current research ethics in computational social science and psychological research. Here, we provide an exploratory empirical account of key user concerns and outline a number of critical discussions that need to take place to protect participants and help researchers to make use of the novel opportunities of digital data collection and field studies. Our primary contention is that we need to elicit public perceptions to devise more up-to-date guidelines for review boards whilst also allowing and encouraging researchers to arrive at more ethical individual study design choices themselves. To ground our discussion in real-world examples of online experiments, we focus on recent social media studies in the field of misinformation, polarization, and hate speech research. We conclude by discussing how we can better strike a balance between meeting ethical guidelines and the concerns of social media users alongside maximizing scientific impact and credibility.

Keywords Ethics, Field experiments, Social media, Public attitudes, Survey

Computational social scientists (CSS) studying online human behavior are increasingly conducting field experiments on social media platforms to overcome the limited ecological validity associated with lab experiments¹. These experiments typically entail administering some treatment to users and analyzing their online behavior before and after the treatment, with or without cooperation of the social media platform². Treatments have involved sending private messages to users³, publicly replying to users' posts⁴, and encouraging users to "like" (subscribe) certain content⁵. In cases where there is platform collaboration, researchers can go further and, for example, alter the algorithmic ranking of content on users' news feeds⁶. The aim of such studies can range from interventions debunking misinformation, reducing hate speech or boosting people's ability to spot manipulative techniques⁷ to a real-time assessment of users' motives for sharing and creating content online⁸.

While social media field experiments have generated exciting results that have appeared in high-impact venues, the innovations they offer also introduce new ethical concerns, such as experimental manipulations in public, networked domains, and sometimes the infeasibility or waiving of informed consent and debriefing for the sake of ecological validity, to name a few. Moreover, their novelty means current ethical guidelines remain underdeveloped for those types of experiments and institutional review boards (IRBs) may be ill-equipped to evaluate associated risks. Not only are these research designs novel, but the social networks themselves are a moving target as their functionality for users can be altered with only small changes⁶. The potential risks of social media studies range from influencing elections⁹ to reducing the willingness to vaccinate¹⁰. Fortunately, researchers can mitigate some of these risks by adopting more reflexive ethical research practices. For instance, debriefing participants can inform them of the true purpose of the study and help alleviate potential misunderstandings or concerns. Although quantitative estimates are difficult to obtain, ¹¹ report that fewer than 30% of papers currently mention any debriefing at all.

Although social media field experiments may typically seem harmless and are likely to be enacted in good faith, the researchers conducting such experiments are currently often left to justify their experimental designs on their own ad hoc, despite potential conflicts with key pillars of research ethics, such as informed consent. Unfortunately, the lack of clear institutional guidance and blurred boundaries for ethical conduct has repeatedly been reflected in public backlash over the past decade. Both older publications, such as the infamous Facebook contagion study¹², which involved researchers exposing users to fewer positive posts to see if they would lead to

¹Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford, Oxford, UK. ²Department of Digitalization, Copenhagen Business School, Frederiksberg, Denmark. ³Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. ⁴Department of Psychology, Humboldt University of Berlin, Berlin, Germany. ⁵Public Policy Programme, Alan Turing Institute, London, UK. ⁶Faculty of Life Sciences, Humboldt University of Berlin, Berlin, Germany. ⁷College of Business, University College Dublin, Dublin, Ireland. ⁸Center Synergy of Systems and Center for Scalable Data Analytics and Artificial Intelligence, Dresden University of Technology, Dresden, Germany. ✉email: vincent.straub@ndph.ox.ac.uk

greater expressions of sadness, as well as more recent studies, like a large study run on LinkedIn that varied the proportion of weak and strong contacts suggested by its “People You May Know” algorithm¹³, have instigated debate and been met with public criticism. Despite passing an institutional ethical review process or involving private sector scientists collaborating with academic scientists who should have obtained institutional review board approval for all projects they conduct (in the case of the Facebook study), the key point is that both studies were ultimately conducted without users’ awareness. As such, they could have affected some people’s livelihoods and well-being, with potential long-term consequences (e.g., on job prospects), in ways we’ll never fully know, raising questions about transparency and research oversight.

We contend that the public backlash against social media field experiments signals a need to revisit and potentially reform computational social science research ethics, much in the same way that public backlash against biomedical experiments inspired the development of The Belmont Report¹⁴ in the 1970s^{15–17}. Of course, the purpose of research ethics is not to avoid public backlash *per se* and we do not suggest that the research community should be beholden to public opinion alone. However, taking proactive steps to protect (or enhance) the public’s trust in science seem necessary to ensure that important, timely discoveries are well-received when communicated, as demonstrated by the real-world harms that have been induced by anti-science attitudes and lacking trust in vaccine and climate science domains^{18,19}. As a first proactive step towards possible research ethics reform, we propose that researchers should engage with the perspective of social media users’ themselves. The incorporation of users’ perspectives could help devise guidelines that effectively protect research participants from harm, especially vulnerable populations (e.g., by shining a light on the actual experiences of participants, rather than experiences inferred by researchers). Moreover, engaging with the public and incorporating the perspectives of affected communities can increase transparency and accountability of research ethics guidelines. Given that it is in the interest of the academic research community to be trusted by the public, openly engaging with and responding to public opinion could serve to foster that trust.

In the present work, we aim to initiate consideration of social media users’ preferences and attitudes towards social media field experiments, focusing on research in domains of misinformation, polarization, and hate speech studies, before considering how these can contribute to the development of revamped guidelines for ethical experiments more broadly. We first unpack the state of affairs by spelling out the problems with existing research practices in more depth. Then we present the results of our survey of public attitudes towards dominant ethical practices, providing a first look at what we can learn when we ask the users’ perspective. Informed by this analysis, which includes public perceptions of high-profile social media field experiments, we then suggest that researchers need to: (1) collect more data on user attitudes towards evolving study designs; (2) better disseminate the values of research impact and scientific reproducibility; and (3) be more transparent about the fact that they are running social media field experiments in the first place. While we focus on academic studies, many of the observations we make also hold true for the online experiments constantly run by private companies (e.g., A/B testing), which often suffer from even less regulatory ethical oversight²⁰.

Pitfalls of current practice

The rapid development in what is technically possible when it comes to designing social media experiments with high ecological validity means new challenging ethical questions are raised continuously. Is it ethical for researchers to place subjects into minimal risk studies without informed consent? What characteristics make an intervention on social media unethical? When does affecting the content users see on social media constitute a form of harm? We are naturally not the first to recognize this or comment on the worrying and persistent lack of institutional attention paid to ethical concerns and public perceptions when it comes to designing online experiments²¹. However, we believe that the potential of current ethical governance procedures to not be fit for meeting the changing nature of research using social media is too urgent an issue to not warrant further investigation, especially from the standpoint of CSS research, a field that can greatly benefit from this new approach to research—if used appropriately.

The reason why current practices appear to not suffice in adequately dealing with social media research, something which has been regularly discussed by research ethicists over the last few years^{22–24}, are manifold but ultimately relate to the fact that as technology continues to drive forward, researchers and institutional processes like IRBs can struggle to keep up. As already discussed in a special issue on ethical issues in social media research in 2020²³, while academic debates about ethical practice seek to catch up with new technological realities and guidelines are suggested by experts and experienced researchers, IRBs and novice researchers can remain unaware of these guidelines and feel uncertain about how to conduct research in this field. This sentiment, of researchers themselves feeling that current procedures are not enough and their demand for additional support, training, and ethics guidance^{25,26} is arguably cause enough to question current procedures.

Given the nascent state of CSS research ethics, prior discussions of how to design ethical social media field experiments have tended to focus on the structural changes that need to take place, such as the establishment of new guidelines, which would benefit the field as a whole. One reoccurring call to action is the suggestion to create new institutions that mirror the state of play in other fields like medical ethics, that is, by creating new committees and codes of practice that can aid researchers in better making the distinction between what constitutes the public and private domain, among other things. Far less attention has, however, been paid to the ethics of particular interventions as they relate to the typical online field experiment. This is likely in part a reflection of the high number of manipulations that researchers can implement when using social media sites, and the difficulty of providing strict ethical guidance for each. To help ground our discussion in real-world examples of online experiments, we focus herein on recent studies in the field of misinformation, polarization, and hate speech research as a case in point^{3–5,27}. These usually involve research designs that require users to think about a concept like accuracy and track how an intervention affects real-world sharing of misinformation,

or entail survey respondents ranking false content on the likelihood that they believe it to be true and whether would share it²⁸.

Beyond ethics as a box-checking exercise

The first dominant approach to ethical review that arguably epitomizes the current rigid process and stifles progress in researcher engagement is the treatment of ethics as an exercise separate from the research process. Specifically, the tendency of ethics to be viewed by universities and research organizations as something that has to be instituted as an administrative ritual prior to research wherein certain ethics review forms may take precedence over critical judgment, collaboration and the process of ethical inquiry in essence can become what social researchers and critical ethics scholars have previously described as a box-checking exercise^{29–32}, given that standard ethics forms required by IRBs consist of batteries of questions, with many being in checklist form. Typically, such forms involve at minimum answering questions addressing: The merit of the project, its scientific rigor, anticipated outcomes, researcher qualifications, consent, reimbursement and risk management considerations, participant recruitment (including vulnerable populations), and data handling. Given these forms are not continuously updated in line with changing methodological possibilities, prescribed standards may quickly become outdated or, as in the context of misinformation, polarization, and hate speech research, forms and guidelines that have been developed primarily for traditional laboratory-style experiments in artificial settings may be blindly adapted to social media field experiments. To be clear, we are not suggesting that researchers do not think about the additional or unprecedented ethical implications of their research at all prior to engagement with IRBs and institutional procedures, nor are we questioning the utility of forms as a way to guide researchers' critical judgment. Instead, we simply contend that dominant practices of ethical review primarily as an administrative process can stifle engagement if ethics committees do not promote a culture of respectful mutual learning in favor of more rigid top-down processes often epitomized by the use of such forms³³.

Albeit more speculative, we consider a further danger of standardizing ethical reflection and review as a box-checking exercise to be that it can give the impression that ethical review is a binary process wherein a study that has received IRB approval will not lead to unethical practices or public backlash. Yet, as the exact grounds for why a particular experimental setup was approved are seldom revealed, existing studies may receive ethical approval but could still involve questionable design practices—and no in-depth discussion of why these are justified beyond the claim that they increase ecological validity. While this may sound like a bold claim, it simply follows from the fact that many social science studies with IRB approval recognize that such boards, particularly ones from a different culture to the study participant pool, do not carry out comprehensive ethical assessments.

As discussed in detail by³⁴, this concern is underscored by the recognition that standard ethical review processes often overlook the nuanced ethical considerations that arise from the specific context of a study. The ethical standards applied by IRBs may not fully consider cultural or contextual factors that could make certain research practices ethically problematic. The idea that ethical judgments are context-dependent means that practices deemed acceptable in one setting may be inappropriate in another, leading to potential ethical oversights when reviews do not account for these differences. As a result, even research that passes through ethical review may include practices that, while approved, could be ethically contentious or even harmful if the review process fails to adapt to the unique ethical challenges posed by diverse research contexts.

In the context of misinformation, polarization, and hate speech research, it has even been shown experimentally that the entrenched assumption that ethical considerations are met as long as the authors of a study can say that the study has been approved by an IRB may result in interventions (e.g., publicly replying to posts with a fact-check without debriefing users) that increase misinformation sharing and toxicity³⁵. This is despite the fact that a careful and thorough debriefing procedure has been shown to help researchers safely and ethically conduct research on these topics¹¹.

Avoiding the rigid commitment to principles

A second approach to ethical review that may enable questionable experimental setups can be argued to result from the fact that many informal recommended best practices are overly concerned with being concept-driven—that is, focused primarily on meeting ethical standards by operationalizing particular principles, such as privacy and autonomy. The roots of this approach can be traced back to the late 1990s and a line of scholarship that sought to refine core concepts relating to research ethics in light of increasing digitization³⁶. The issue with this approach (or more accurately, the logic of conceptualization), however, is that it can result in presumptive assumptions, limiting nuanced and contingency-based ethical decision-making. That is, it privileges past practice and historical precedent as the primary guide to ethical reasoning³⁷. In other words, by relying first on particular concepts to frame ethics it assumes that these concepts, and historical reasoning, best capture the dilemmas a researcher may face now. Yet, by prioritizing these concepts and assigning them to particular actions or regulations, this approach devalues focusing instead first on the actual and potential consequences of the contemporary actions researchers must take, especially when methods rapidly evolve. An example in the case of misinformation research is taking risk to mean the risk that users would otherwise be exposed to or assuming that consent is given due to the fact that users have agreed to a platform's terms of service. In reality, we increasingly know that risk on social media is not experienced the same by all users (and researchers) and so depending on the specific intervention, the consequences can vary dramatically³⁸. Yet, the heterogeneity of consequences posed by field studies is not accounted for in inflexible concept-driven models of ethical review, which do not encourage researchers to view ethics as a process that can be re-contextualized as needed on a case-by-case basis³⁷.

The continued over-reliance on single concepts and little focus on their intersectionality means ethical reasoning can remain a reified process that is resistant or hesitant to modification based upon changes in context,

public opinion or technical possibilities. Perhaps the greatest problem with this exclusive reliance on concepts in contrast to contextual details is that it can lead both to misdefining the vulnerability of participants and mistaking predetermined notions with actual experiences of harm³⁷. Moreover, it disregards altogether the fact that public opinion can rapidly change when it comes to issues of privacy, vulnerability, anonymity, and harm. Whilst we acknowledge that completely avoiding the potential of harm in online experimentation is difficult (and perhaps impossible), it is perhaps worthwhile remembering that the medical ethics axiom to ‘first do no harm’ has inspired other fields to increase discussions on defining harm and how to recognize it, especially in relation to vulnerable participants, helping to better prepare researchers for when ethical dilemmas do arise³⁹.

To take the notion of ‘vulnerability’ as a case in point can help elucidate the above points further. This principle, which is fundamental to ethical research involving human participants, encapsulates the idea that potential risks to subjects must be kept to a minimum, and vulnerable individuals need even greater safeguards, such as pregnant women, incarcerated individuals, and minors. The dominant, concept-driven, categorical use of vulnerability suggests that there is a hierarchy, with certain populations being labeled as vulnerable. As a result, it disregards the degree of vulnerability within these groups, and does not ask researchers to consider and identify the situations in which individuals might be considered vulnerable. Alternative proposals suggest researchers should instead consider this contextual, ‘layered’ intersectional conceptualization of vulnerability^{40–42}. Whilst the difficulty of operationalizing this view is often used as a way to justify existing practices, the problems and inconsistencies with the categorical approach are arguably becoming too big to ignore.

In a recent analysis,⁴³ find that, in 355 official documents governing social/behavioral human subjects research across 107 countries, there are 68 distinct vulnerability categories. Their analysis reveals that there are significant differences across different regions and medical classifications are often given too much weight, while important factors like displacement are overlooked. Additionally, there is likely a great deal of diversity within and between these groups. We welcome their and prior suggestions²¹ to focus not only on groups of subjects but the characteristics of research studies and the potential they have to induce or exacerbate vulnerability. Whilst in need-of-testing,⁴³’s TAPIR framework, which outlines the implications of researcher decisions about Topics, steps to Appraise vulnerability, strategies for vulnerability Protections, considerations for Implementation fidelity, and commitments to ongoing Reflection, seems a viable step in the right direction.

Public attitudes of social media field experiments

Given the issues with the current ethics process, we contend that one sensible step is to more deeply consider the perspectives of those directly affected by social media experiments: The users. For one, users’ attitudes are significant because they can help shape ethical research principles (see⁴⁴ for a notable example regarding public attitudes toward content moderation). Additionally, public acceptance of social media research is crucial for legitimizing and enhancing its effectiveness. Of course, user preferences and attitudes are only one aspect of the broader picture, which must be complemented by established ethical principles to ensure that studies are conducted as ethically as possible. Nevertheless, the research community stands to benefit from understanding what kinds of research practices may be likely to elicit negative public responses, and to understand from which subgroups such responses might come. For example, a researcher designing an experiment with practices that are likely to elicit a negative response could benefit from reflection on that practice and its implementation. If a researcher decides to proceed with a practice that is likely to elicit a negative response, then proactive science communication measures could be taken to, for example, communicate to the public the cost-benefit analyses that have been done throughout the ethical approval process.

To provide more specific guidance to researchers engaged in social media field experimentation, we conducted a pre-registered analysis¹ of social media users’ perceptions of ethical practices as they relate to the existing high-profile social media studies discussed above. Here we present our findings and provide interpretations of key results.

Our study involved running an online survey questionnaire for which we recruited 500 US-based participants aged 18–78 years ($M = 41.7$, $SD = 13.6$, 41.5% female) via Prolific that were regular users of at least one of Facebook, X (formerly Twitter), or Reddit. Participants (final $N = 499$; one participant did not conclude the study) proceeded through three main survey blocks corresponding to three main topics of investigation (Fig. 1). In the first block we explored participants’ awareness of social media field experiments. We find that while users are generally aware that social media data are sometimes used by academic researchers (84.8% were at least “slightly aware”), users are less aware of which interactional treatments researchers have administered. For instance, 66.3% of participants indicated that they were aware that researchers had “created fake accounts (‘bots’)”, 51.8% thought they had conducted experiments involving “privately messaging users”, and 42.9% were aware that researchers had “publicly posted on users’ profiles” (Fig. 2A). Moreover, we find that while users are generally aware that social media data is sometimes used by academic researchers, they are less aware of which interactional treatments researchers have administered. This observation is perhaps unsurprising given the novelty of social media field experiments, yet it underscores the need to consider when specific manipulations should be adopted and how they could be better regulated in part to avoid potential abuses and subsequent reputational damage to the research community.

¹The pre-registration for our study is available on the Open Science Framework at: <https://osf.io/swcvq/>.

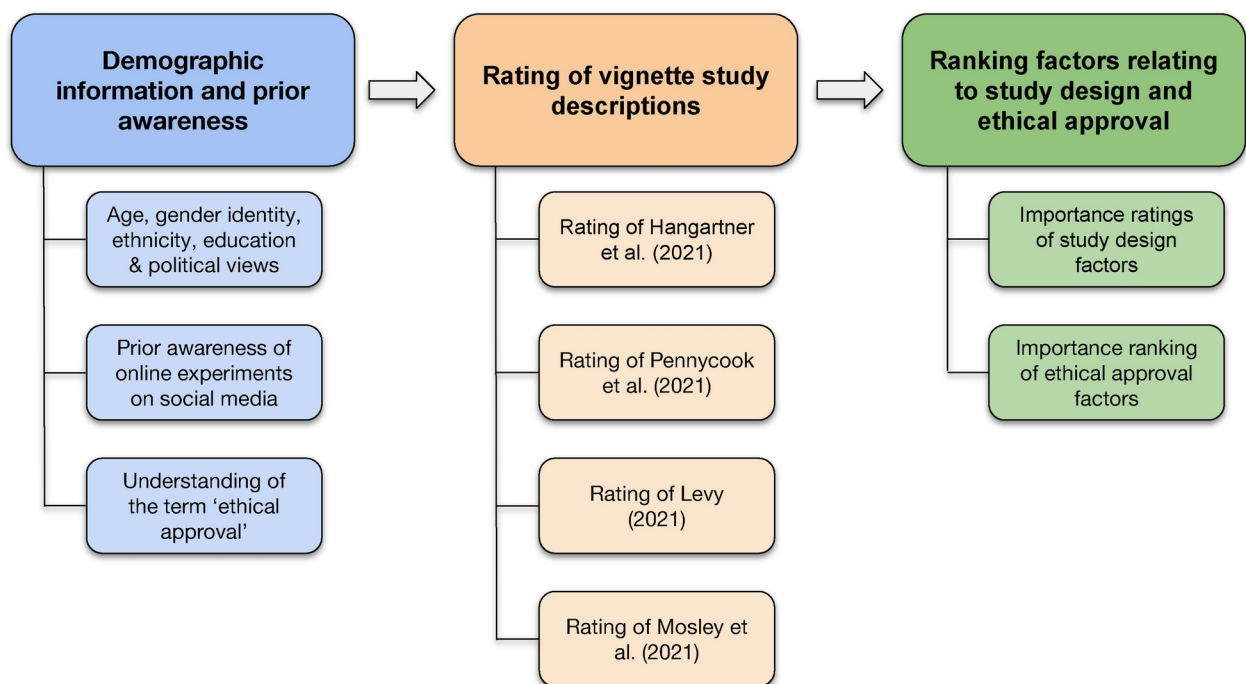


Fig. 1. Schematic of study design. The survey questionnaire involved three distinct blocks, including a rating of four vignette study descriptions (in a randomized order) described in the main text. See Supplementary Material Sect. 2 for details.

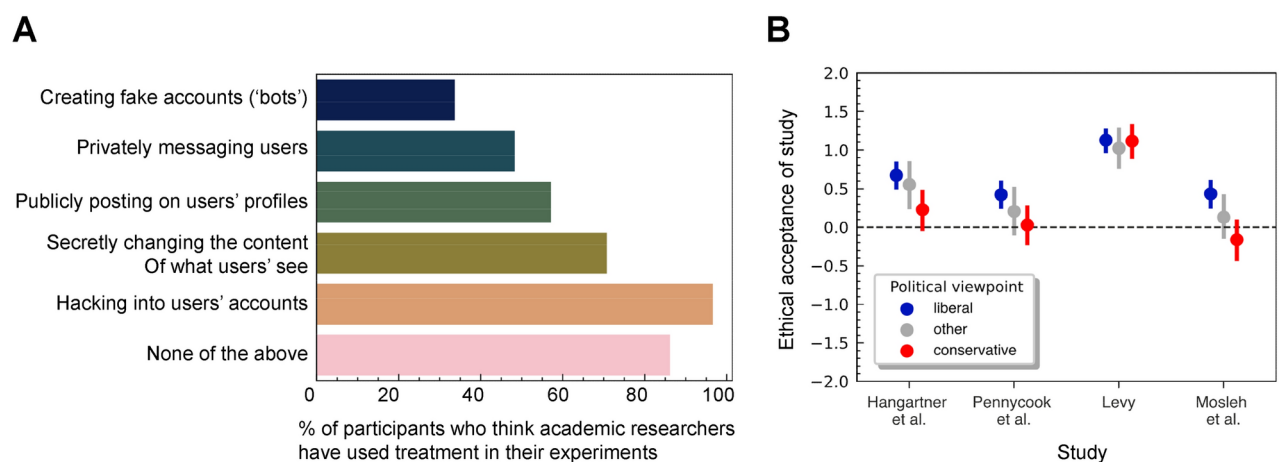


Fig. 2. Participant awareness and ratings of vignette study descriptions. **(A)** Participant awareness of treatments administered by researchers. Participants ($N = 499$) were asked “which of the following ways of interacting on social media do you think academics have used in their experiment? (Select all that apply)”. **(B)** Ethical acceptability of four published social media field experiments^{3–5,27}. Participants were provided with a brief description of the experiment (approved by the respective authors) and asked “How ethically unacceptable versus acceptable do you find the described study?” using a 5-point rating scale ($-2 =$ completely unacceptable, $2 =$ completely acceptable). Points represent means and vertical bars represent standard error. Data is split out by participants’ political viewpoint, such that “liberal” participants ($N = 276$) indicated that they are either slightly or very liberal, “conservative” participants ($N = 131$) indicated that they are either slightly or very conservative, and “other” participants ($N = 92$) indicated that they are neutral or they preferred not to say.

In the second survey block we asked whether participants perceived actual studies' procedures to be ethically acceptable—specifically⁴, Study 7 of^{3,5}, and²⁷². Our findings suggest that participants generally viewed the procedures of past studies to be somewhat ethically acceptable. However, We find that, while participants generally viewed the procedures of past studies to be ethically acceptable, this depended in part on personal user attributes, such as political ideology. For instance, for three of the four studies, users who self-identified as politically conservative tended to view the studies as less positive than those who self-identified as liberal (Fig. 2B). In an unregistered exploratory analysis, we fit a linear mixed-effects model with perceived acceptability (five-point Likert scale responses; - 2: completely unacceptable to 2: completely acceptable) as the dependent variable, political identity (liberal, conservative, or other) as a fixed factor, and random intercepts by participant and stimulus (study). This model revealed that liberals rated the ethical acceptability of experimental designs significantly higher than conservatives ($\beta = 0.37$, $SE = 0.10$, $t(496) = 3.63$, $p < 0.001$), while the difference between neutrals and conservatives was not statistically significant ($\beta = 0.21$, $SE = 0.13$, $t(496) = 1.56$, $p = 0.12$). Note, however, that this kind of hypothesis testing was not the primary objective of our study, hence the limited number of stimuli used and imbalanced sample of each political identity (276 participants indicated that they are either slightly or very liberal, 131 indicated that they are either slightly or very conservative, 89 indicated they are either neutral of neither liberal nor conservative, and 3 preferred not to say). This observation could be reflective of anti-academic attitudes being more prevalent among conservatives⁴⁵. However, further hypothesis-driven research seems needed to replicate our observations and distinguish which specific elements of a study design are influencing differences among participants' acceptability ratings.

Our analysis also serves to illustrate another important consideration often forgotten in more top-down, traditional approaches to ethical review: Participants are seldom a homogeneous group. Being aware of this reminds researchers of the classical ethical principal to serve and take more care of protecting subjects the greater their vulnerability. As highlighted in prior work⁴⁶, this may include paying particular attention to minorities, LGBT individuals, and specific communities; importantly, this also applies to the research process, as awareness and public scholarship present unequal challenges for researchers that echo the differences in user expectations³⁸. Our results similarly suggest that if researchers continue running experiments like those considered in our survey, they might contribute to the polarization of liberal versus conservatives attitudes towards science, as conservatives feel more targeted by academic researchers⁴⁷. Yet, to fully understand the effects of various personal user attributes on ethical perceptions of research, more work is clearly needed.

An additional observation we make is that participants place a high importance on the type of intervention (treatment) implemented in experiments (i.e., private messaging vs. publicly replying to posts) but care less about its impact (i.e., whether it changed users' behavior) and how it is administered (e.g., manually by a researcher vs. via an automated account). Moreover, in exploring whether participants judge certain experimental design features to be more important than others in their views of ethical acceptability, we find that users are most concerned about whether study participants have been asked for consent and whether study details have been disclosed. Finally, in the third survey block (importance of study design factors), we explored whether participants judged key experimental design features ($n = 9$) identified in prior literature to be more important than others in their views of ethical acceptability. Here, we find that the primary concern is whether participants have been asked for consent and study details have been disclosed to them after completion of the study, with 38.3% of participants ranking 'Participant awareness of study' as extremely important, compared to 21.2% for the 'Type of data' and 11.2% for the 'Length of data collection', for instance (see Fig. 3). Notably, two of the three factors which the smallest percentage of participants ranked as extremely important, 'Number of users studied' (16.08%) and 'Length of data collection', relate to design features where online and laboratory field experiments differ in practice but not in theory (in contrast to, say, 'Type of data' that can be collected, which is a key difference between the two types of design), suggesting that what matters to the public may broadly be the same between the two types of experimental design.

Discussion: accounting for public perceptions and promoting researcher judgment

Taken together, our results provide empirical support for the need to revamp guidelines for ethically sound social media field experiments. In line with previous surveys⁴⁸, our results affirm that participant awareness is a highly valued ethical principle and—in contrast to earlier commentaries⁴⁹—is considered more important than research purpose and impact. Overall, our findings ultimately lead us to conclude that we need to pay far greater attention to public attitudes than we currently do, especially that users wish to be informed about the use of social media data in scientific studies. While studies on public attitudes towards online field experiments are still sparse, past research and our results further show that these can also be highly contextual, depending on factors such as how the research is conducted or disseminated, who is conducting it, and what the study is about⁴⁸. Research domain, content type, purpose of data use, and awareness of data collection can, for instance, all impact respondents' comfort⁵⁰. As such, developing new ethical guidelines will require both greater awareness of public perceptions at a population level, and an appreciation of how these may vary across different demographic groups and social media user bases as a result of factors like the specific type of experimental intervention used. Only when IRBs and individual researchers develop a firmer understanding of both of these topics can the CSS community as a whole convincingly claim to be conducting research that is fully in the public interest and

²These studies were selected due to their publication in high-profile venues and because their experimental designs involve a range of different treatments (e.g., private messages, public replies, browser extensions, use of "bot" accounts, etc.). The standardized-vignette study descriptions presented to participants in our survey were approved by the respective authors for accuracy (see Supplementary Material Sect. 2 for details).

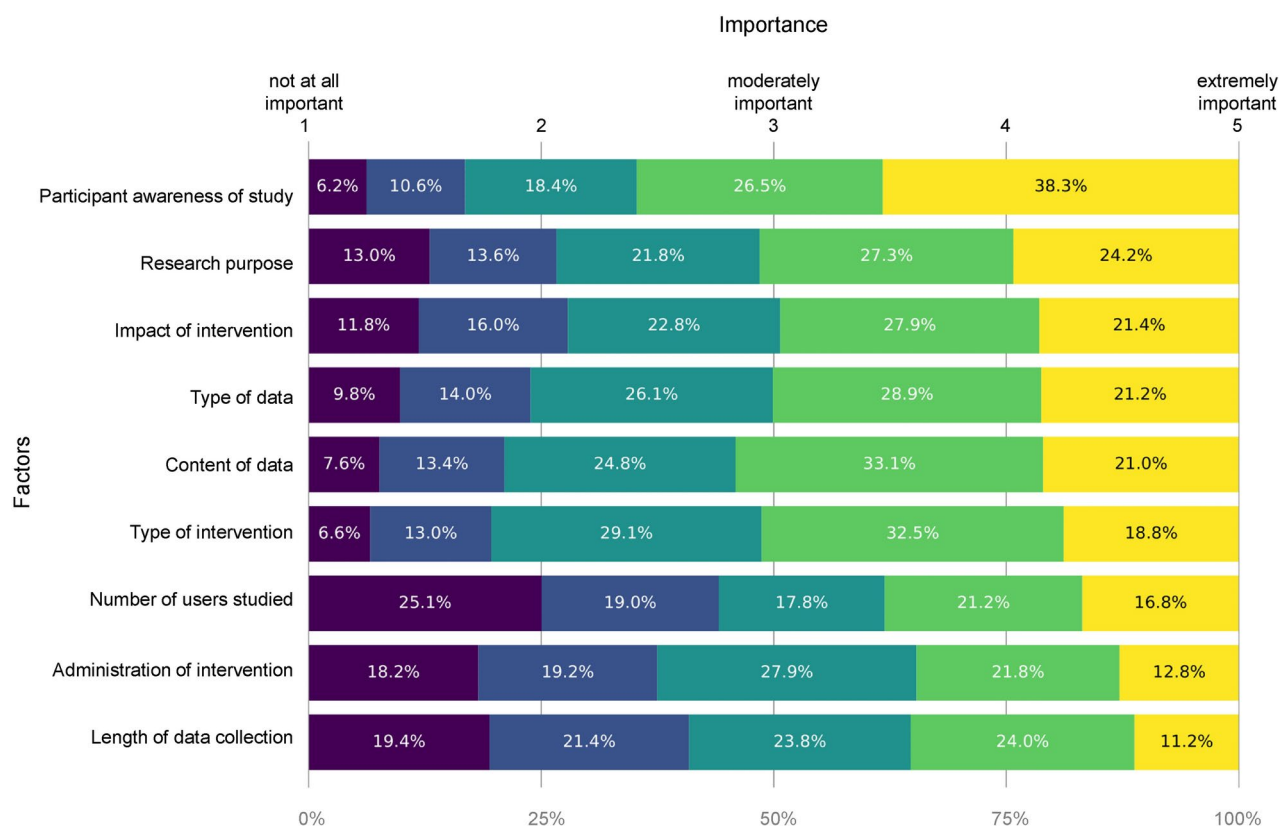


Fig. 3. Participant perceptions of the importance of different study design factors. Distribution of participant responses to the question “How important are the following factors for determining your level of concern with academic research conducted on social media, such as online experiments?” using a 5-point rating scale (1 = not at all important, 5 = extremely important). The factors ($n = 9$) represent key study design choices identified in prior literature and are depicted in descending order of importance, as measured by the percentage of participants who ranked the factor as ‘extremely important’ (ranging from 38.3% for the top-ranked factor, ‘Participant awareness of study’, to 11.2% for the bottom-ranked factor, ‘Length of data collection’).

strengthen trust in science. This is especially relevant at a time when there is an urgent need to understand the increasing online informational risks users face⁵¹.

With the above in mind, a final point that we wish to stress is the need for researchers themselves to develop a more informed understanding and judgment of the implications of their study designs. Naturally, all researchers are expected to examine and understand the ethical challenges in their studies’ design, regardless of the field. Yet, new possibilities in social media field experiments mean computational social scientists need to be particularly well-attuned to the changing benefits and risks of experimenting on platforms like X (formerly Twitter), Facebook or Reddit. For example, ongoing work on (mis)information sharing goes beyond the popular approach to link users’ survey responses (e.g., demographics or political views) with their digital footprints⁵², but additionally elicits users’ behavioral motives “in the moment”⁵³. Given the dynamic nature of social media, a developed sense of critical judgment on the part of researchers is arguably becoming only more important. Relatedly, given the changing relationship the public holds with regards to social media, the other important ingredient that is arguably missing from the current landscape of research ethics is a greater ongoing researcher awareness of public attitudes to current practices and specific interventions.

Although there is a growing recognition that the relationship between IRBs, researchers, and the public should be one of continuing cooperative dialogue³⁷, we contend that this combination of (1) more informed researcher judgment, and (2) more engagement with public sentiment is what is needed to ensure future experiments are more ethically sound and publicly acceptable. To be clear, we are not suggesting that current practices (i.e., the tendency to leave ethical review to IRBs) are dropped wholesale; rather, we wish to stress how they can be augmented through the adoption of these additional practices. In practice, this means researchers engaging in deeper, bottom-up ethical reflection when it comes to the day-to-day practices of designing a study, regardless of the discipline, country and context they work in³⁷. Ethical reflection is here taken to entail spending more time personally grappling with specific questions related to the experiment setup at hand, before and after study completion, such as: To what extent can the current intervention be justified? How may it exacerbate the vulnerability of participants? Are there alternative methodologies that have not been considered? As described by⁴³ and their TAPIR framework, researchers can do so by considering how the topic of study might introduce

new vulnerabilities, appraise how an intervention affects individuals in different country contexts, devise and implement relevant protections to account for any new vulnerabilities, and reflect on or ideally verify that no new vulnerabilities were introduced to appropriate steps were taken to address them.

In spending more time wrestling with questions about the ethical nature of their studies, alongside consulting IRB guidelines, researchers should be encouraged to embrace ethical pluralism and cross-cultural awareness, allowing them to iterate over design choices while accepting the possibility of multiple, ethically legitimate judgment calls—similar to public attitudes towards field experiments. Importantly, this call for further training in research ethics is not meant as a way to prevent the design of future studies like the ones we use in our survey here, regardless of how ethical they are. Rather, it is a low-cost and easy to implement practical preventative measure that arguably not only has inherent value as a pedagogical practice but ultimately benefits research in the long run, by fostering collaboration and ensuring researchers are aware of and conduct research on changing social and ethical concerns, among other factors⁵⁴. Moreover, given the potential of large language models to transform CSS⁵⁵, including in experimentation, and usher in further ethical quandaries, training in research ethics can perhaps also be seen as a necessary preventative measure to foster greater awareness of the potential perils of AI-assisted experimentation in the next generation of CSS students; something that is already regularly discussed in the field of computer science and cybersecurity⁵⁶ and has been readily observed in the integration of AI in social media⁵⁷.

Concluding thoughts

Taken together, our discussion and the results of our survey lead us to emphasize that a number of critical discussions and decisions still need to happen to ensure researchers conducting social media field experiments do so in a way that is ethically sound and publicly acceptable. Whilst Kinder-Kurlanda and Zimmer conclude that, “there is no ‘ethical research’ of the internet in the social age ... we can really only strive for ‘ethically-informed’ research practices”⁵⁸, we believe there is a way forward. A first step in the right direction is further data collection on user attitudes, both to get a more comprehensive picture of differences amongst different subpopulations and to paint a clearer landscape of the public acceptability of novel interventions. Secondly, the research community as a whole arguably needs to better disseminate the values of research impact and scientific reproducibility, as achieving greater ecological validity will naturally mean that studies are more meaningful—allowing any findings that can be translated to policies to ultimately benefit the public more in the long term. Third, research organizations and researchers arguably need to begin being more proactively transparent about the fact that they are running social media field experiments, as a sizable proportion of the public appears to remain unaware of this and some appear to consistently prefer to be excluded from field experiments altogether.

Importantly, the considerations and recommendations that we discuss here are not limited to those conducting misinformation, polarization, and hate speech research. Our primary objective is to guide all researchers wishing to conduct field experiments on social media in choosing a research design that is the result of considered ethical reflection, abides by existing ethical review standards, and is in tune with public sentiment. In the long term, ensuring studies are aligned with public sentiment will ensure trust and funding in CSS is maintained, benefiting both the public and researchers. Thus, explicitly eliciting public perceptions of study designs will both help protect participants and aid researchers in maximizing the scientific impact and credibility of their work.

Data availability

This study and all experimental protocols were reviewed and approved by the Max Planck Institute for Human Development in accordance with the procedures laid down by the Max Planck Society for ethical approval of all research involving human participants. Informed consent was obtained from all subjects (see Supplementary Material Sect. 1 for details) and all methods were carried out in accordance with relevant guidelines and regulations. Our main research questions, survey design, and analyses were pre-registered before the collection of the data (AsPredicted: #103940). The datasets generated and analyzed during the current study are available in a Open Science Framework repository: <https://osf.io/swcvq/>. Data and code are also available on GitHub at the following URL: <https://github.com/vincejstraub/article-digex-survey>. Participants were recruited using the Prolific platform. The survey responses were recorded using Qualtrics software and analyzed in the R statistical language and Python.

Received: 27 March 2024; Accepted: 17 October 2024

Published online: 30 October 2024

References

- Mosleh, M., Pennycook, G. & Rand, D. G. Field experiments on social media. *Curr. Dir. Psychol. Sci.* **31**, 69–75 (2022).
- Aridor, G., Jiménez-Durán, R., Levy, R. & Song, L. Experiments on social media. In *Handbook of Experimental Methods in the Social Sciences* (Edward Elgar Publishing, Cheltenham, 2024).
- Pennycook, G. et al. Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).
- Hangartner, D. et al. Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proc. Natl. Acad. Sci.* **118**, e2116310118 (2021).
- Levy, R. Social media, news consumption, and polarization: Evidence from a field experiment. *Am. Econ. Rev.* **111**, 831–870 (2021).
- Guess, A. M. et al. How do social media feed algorithms affect attitudes and behavior in an election campaign?. *Science* **381**, 398–404 (2023).
- Kozyreva, A. et al. Toolbox of individual-level interventions against online misinformation. *Nat. Human Behav.* **8**, 1–9 (2024).
- Lorenz-Spreen, P. et al. Real-time assessment of motives for sharing and creating content among highly active Twitter users. *OSF* (2024).

9. Tucker, J. A. et al. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: A review of the scientific literature* (March 19, 2018) (2018).
10. Kata, A. Anti-vaccine activists, web 2.0, and the postmodern paradigm-an overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine* **30**, 3778–3789 (2012).
11. Greene, C. M. & Murphy, G. Debriefing works: Successful retraction of misinformation following a fake news study. *PLoS ONE* **18**, e0280295 (2023).
12. Hunter, D. & Evans, N. Facebook emotional contagion experiment controversy. *Res. Ethics* **12**, 2–3 (2016).
13. Singer, N. LinkedIn ran social experiments on 20 million users over five years. *The New York Times* <https://www.nytimes.com/2022/09/24/business/linkedin-social-experiments.html> (2022).
14. The belmont report (1979). https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.
15. Heller, J. Syphilis victims in us study went untreated for 40 years. *N.Y. Times* **1**, 8 (1972).
16. Reverby, S. M. *Examining Tuskegee: The Infamous Syphilis Study and Its Legacy* (University of North Carolina Press, Chapel Hill, 2009).
17. Schupmann, W. & Moreno, J. D. Belmont in context. *Perspect. Biol. Med.* **63**, 220–239 (2020).
18. Oreskes, N. What is the social responsibility of climate scientists?. *Daedalus* **149**, 33–45 (2020).
19. Hotez, P. J. Anti-science kills: From soviet embrace of pseudoscience to accelerated attacks on us biomedicine. *PLoS Biol.* **19**, e3001068 (2021).
20. Boyd, D. Untangling research and practice: What facebook s emotional contagion study teaches us. *Res. Ethics* **12**, 4–13. <https://doi.org/10.1177/1747016115583379> (2016).
21. McDermott, R. & Hatemi, P. K. Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms. *Proc. Natl. Acad. Sci.* **117**, 30014–30021 (2020).
22. Samuel, G. et al. Is it time to re-evaluate the ethics governance of social media research?. *J. Empir. Res. Hum. Res. Ethics* **13**, 452–454 (2018).
23. Samuel, G. & Buchanan, E. *Ethical Issues in Social Media Research*, (Guest editorial, 2020).
24. Al-Zaman, M. S., Khemka, A., Zhang, A. & Rockwell, G. The defining characteristics of ethics papers on social media research: A systematic review of the literature. *J. Acad. Ethics* **22**, 163–189 (2024).
25. Hokke, S. et al. Ethical considerations in using social media to engage research participants: Perspectives of Australian researchers and ethics committee members. *J. Empir. Res. Hum. Res. Ethics* **15**, 12–27 (2020).
26. Lazer, D. M. et al. Computational social science: Obstacles and opportunities. *Science* **369**, 1060–1062 (2020).
27. Mosleh, M., Martel, C., Eckles, D. & Rand, D. Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a twitter field experiment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13 (2021).
28. Guay, B., Berinsky, A. J., Pennycook, G. & Rand, D. How to think about whether misinformation interventions work. *Nat. Human Behav* **7**(8), 1–3 (2023).
29. Whelan, A. Ethics are admin: Australian human research ethics review forms as (un) ethical actors. *Soc. Media Soc.* **4**, 2056305118768815 (2018).
30. Markham, A. & Buchanan, E. Internet research: Ethical concerns. In *International Encyclopedia of Social and Behavioral Science* 606–613 (Elsevier, 2015).
31. Macintyre, M. Problems with ethics committees. *Aust. J. Anthropol.* **25**, 381–383 (2014).
32. Faria, R. Being ethical in research. In *Qualitative Research in Criminology Cutting-Edge Methods* 229–240 (Springer, 2022).
33. Brown, C., Spiro, J. & Quinton, S. The role of research ethics committees: Friend or foe in educational research? An exploratory study. *Br. Edu. Res. J.* **46**, 747–769 (2020).
34. Asiedu, E., Karlan, D., Lambon-Quayefio, M. & Udry, C. A call for structured ethics appendices in social science papers. *Proc. Natl. Acad. Sci.* **118**, e2024570118 (2021).
35. Clifford, S. & Sullivan, B. W. Do survey questions spread conspiracy beliefs?. *J. Exper. Polit. Sci.* **10**, 1–11 (2023).
36. Nissenbaum, H. *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press, Stanford, 2009).
37. Markham, A. N. Afterword: Ethics as impact-moving from error-avoidance and concept-driven models to a future-oriented approach. *Soc. Media Soc.* **4**, 2056305118784504 (2018).
38. Massanari, A. L. Rethinking research ethics, power, and the risk of visibility in the era of the alt-right gaze. *Soc. Media Soc.* **4**, 2056305118768302. <https://doi.org/10.1177/2056305118768302> (2018).
39. Buchanan, D. & Warwick, I. First do no harm: Using ‘ethical triage’ to minimise causing harm when undertaking educational research among vulnerable participants. *J. Furth. High. Educ.* **45**, 1090–1103 (2021).
40. Gordon, B. G. Vulnerability in research: Basic ethical concepts and general approach to review. *Ochsner J.* **20**, 34–38 (2020).
41. Luna, F. ‘vulnerability’, an interesting concept for public health: The case of older persons. *Public Health Ethics* **7**, 180–194 (2014).
42. Luna, F. Identifying and evaluating layers of vulnerability-a way forward. *Dev. World Bioeth.* **19**, 86–95 (2019).
43. Findley, M. G., Ghosn, F. & Lowe, S. J. Vulnerability in research ethics: A call for assessing vulnerability and implementing protections. *Proc. Natl. Acad. Sci.* **121**, e2322821121 (2024).
44. Kozyreva, A. et al. Resolving content moderation dilemmas between free speech and harmful misinformation. *Proc. Natl. Acad. Sci.* **120**, e2210666120 (2023).
45. Oreskes, N. & Conway, E. M. From anti-government to anti-science: Why conservatives have turned against science. *Daedalus* **151**, 98–123 (2022).
46. Franzke, A., Bechmann, A., Ess, C. & Zimmer, M. (eds.) *Internet Research: Ethical Guidelines 3.0*, vol. 3 (AoIR (The International Association of Internet Researchers), 2020).
47. Mann, M. & Schleifer, C. Love the science, hate the scientists: Conservative identity protects belief in science and undermines trust in scientists. *Soc. Forces* **99**, 305–332 (2020).
48. Fiesler, C. & Proferes, N. Participant perceptions of Twitter research ethics. *Soc. Media Soc.* **4**, 1–14 (2018).
49. Kleinsman, J. & Buckley, S. Facebook study: A little bit unethical but worth it?. *J. Bioeth. Inq.* **12**, 179–182 (2015).
50. Gilbert, S., Vitak, J. & Shilton, K. Measuring Americans’ comfort with research uses of their social media data. *Soc. Media Soc.* **7**, 20563051211033824 (2021).
51. Aslett, K. et al. Online searches to evaluate misinformation can increase its perceived veracity. *Nature* **625**, 548–556 (2023).
52. Guess, A., Nagler, J. & Tucker, J. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Sci. Adv.* **5**, eaau4586 (2019).
53. Geers, M. Linking lab and field research. *Nat. Rev. Psychol.* **2**, 458–458 (2023).
54. Glasziou, P., Scott, A. M., Chalmers, I., Kolstoe, S. E. & Davies, H. T. Improving research ethics review and governance can improve human health. *J. R. Soc. Med.* **114**, 556–562 (2021).
55. Ziems, C. et al. Can large language models transform computational social science?. *Comput. Linguist.* **50**, 237–291 (2024).
56. Cruz, T. & Simões, P. Fostering cybersecurity awareness among computing science undergraduate students: Motivating by example. In *20th European Conference on Cyber Warfare and Security*, vol. 72 (2020).
57. Saheb, T., Sidaoui, M. & Schmarzo, B. Convergence of artificial intelligence with social media: A bibliometric & qualitative analysis. *Telemat. Inform. Rep.* **14**, 100146 (2024).
58. Kinder-Kurlanda, K. & Zimmer, M. *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. (Peter Lang Verlag, 2019).

Acknowledgements

This research was assisted by an SSRC/Summer Institutes in Computational Social Science (SICCS) Research Grant. The funder had no role in the decision to publish or the preparation of this manuscript. We thank the organizers and participants of SICSS London 2021, SICSS London 2022, and the International Conference for Computational Social Science 2023 for helpful discussions and feedback. We would especially like to thank Joshua Becker and Anastasia Kozyreva for their constructive comments. We would also like to thank David Rand, Dominik Hagnarter, and Ro  e Levy for providing us with context about the field experiments they conducted, which we refer to herein. Michael Geers is a fellow of the International Max Planck Research School on the Life Course in Berlin and supported by the Joachim Herz Foundation. Vincent J. Straub is supported by ESSGN (HORIZON - MSCA - DN - 2021 (101073237)) and the Leverhulme Trust Large Centre Grant LCDS (RC - 2018 - 003).

Author contributions

V.J.S, J.B., and M.G. designed the study. V.J.S. secured the funding and collected the data; V.J.S. and J.B. analyzed the data, created the visualizations, and wrote the first draft. M.G. and P.L.S contributed to writing and editing, and all read and approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-76948-z>.

Correspondence and requests for materials should be addressed to V.J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

  The Author(s) 2024