Computers in Human Behavior 139 (2023) 107537

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Unintended effects of open data policy in online behavioral research: An experimental investigation of participants' privacy concerns and research validity

Bingjie Liu [a], Lewen Wei [b,*]

[a] Department of Communication Studies, California State University, Los Angeles, USA
[b] Gamification Group, Faculty of Information Technology and Communication Sciences, Tampere University, Finland

ARTICLE INFO

Keywords:
Open science
Open data
Privacy management
Self-disclosure
Research validity

ABSTRACT

Research outlets are increasingly adopting the Open Data policy—requiring or encouraging researchers to release data publicly. However, public data sharing in the digital age may threaten participants' privacy and thereby discourage participants to disclose. We tested this possibility and explored solutions with a between-subjects online experiment. Participants from Amazon's Mechanical Turk ($N = 294$) were randomly assigned to one of five conditions with data sharing policies varying on the level of publicness or not mentioned. Participants in the public-access condition reported greater privacy concerns and fewer unethical conducts than those in the private condition. Public data sharing also indirectly decreased participants' amount of open-ended disclosure via privacy concerns. When asked, participants reported privacy violation as their primary concern regarding data sharing. Findings suggest sharing data with researchers' gatekeeping, rather than indiscriminately public, may be a solution that better serves the goals of Open Science—ethical and valid science.

Over the last decade, with a few large-scale reproducibility projects failing to replicate many canonical findings in both natural and social sciences (e.g., Open Science Collaboration, 2015), the science community has been increasingly recognizing, concerned about, and striving to fix this "replication crisis" (Baker, 2016). Open science, the idea of increasing transparency in research and publication processes (Nosek et al., 2015), has emerged as a promising solution and gained popularity in the past few years across several fields, including both natural and social sciences. Among the practices recommended in the Open Science agenda, one practice is to publicly share research data by means such as uploading data sets onto public repositories (e.g., Dienlin et al., 2021; Nosek et al., 2015). Compared with the traditional model of data storage (i.e., keeping data with the research team), the benefits of sharing data publicly and freely, or open data (OD), seem obvious in that it allows others to independently verify the reported results, helps reviewers check for potential errors and unethical data manipulations (Munafò et al., 2017), and helps both the scientific community and the public at large better understand the results and generate cumulative knowledge (LeBel et al., 2018).

There are, however, potential downsides of OD if it is illy operationalized, which may lead to unintended consequences contradicting its goals of more valid and ethical science (Freiling et al., 2021). In most behavioral research (e.g., psychology and communication), human participants are the owners and contributors of data and may not appreciate the level of publicness of their data that OD requires, especially in this digital age when techniques that can re-identify data prevail. As such, OD, although advocated with researchers' good intentions, may impose unexpected threats to participants' welfare, raise privacy concerns among the participants, and reduce or even distort their natural responses during research participation, thereby compromising the validity of the research findings. To ensure ethical and valid research, it is imperative to empirically test the effects of OD on participants and find solutions if such downsides exist (Cummings & Day, 2019; Freiling et al., 2021).

So far, little empirical research has examined the consequences of OD on participants' perceptions and behaviors and research validity (Cummings & Day, 2019; Freiling et al., 2021). We aim to fill this gap by (1) investigating the effects of OD policy on participants' privacy concerns and their disclosure and (2) exploring potential solutions to the unintended effects, if any, with the ultimate goal of informing ethical

* Corresponding author. Mailing address: Gamification Group, Faculty of Information Technology and Communication Sciences, Tampere University, Kansler-inrinne 1, 33100, Tampere, Finland.
E-mail addresses: bliu3@calstatela.edu (B. Liu), lewen.wei@tuni.fi (L. Wei).

https://doi.org/10.1016/j.chb.2022.107537
Received 11 June 2022; Received in revised form 9 October 2022; Accepted 15 October 2022
Available online 22 October 2022

and valid research practices.

1. Literature review

1.1. Open data: ideal, operationalization, and risks

Open data or public data sharing is one of the practices advocated in the Open Science movement. "In simplest terms, data sharing is the practice of making empirical data from research studies *freely available* and *without qualifications*" (Bowman & Spence, 2020, p. 710). So far, many research outlets have already adopted such a practice by requiring researchers to publicly share data, such as *Science* and the *Public Library of Science*. Although not all social science journals are mandating public data sharing by the time of this writing, major professional associations of pertinent fields such as International Communication Association (ICA) and American Psychological Association (APA) require all authors to include a Data Availability Statement in their articles, describing and providing means to access the data or explaining why they are not sharing the data (Dienlin et al., 2021). Technically speaking, authors can choose from several ways to share their data (or not) with different levels of publicness, ranging from "Data are available in a public repository" to "Data available on request" (Oxford Journals, 2021); but public data sharing is the most recommended practice, providing that no ethical codes are violated (e.g., Journal of Communication, 2022; Psychological Science, 2022).

The intention underlying OD is admittedly benign—to reduce problematic research practices and to enhance collaboration (Dienlin et al., 2021). In the ideal, best-scenario case of OD, research data are responsibly shared and ethically reused so that it facilitates valid and efficient knowledge production without incurring loss to any stakeholders, including both the researchers and participants. In reality, however, there is a lack of guidelines on how to handle data ethically and responsibly (Fox et al., 2021), and not many researchers have been trained on the specialized skills and tools needed for ethical data sharing (Joel et al., 2018; Meyer, 2018). As such, even for well-intentioned researchers, unprofessional and unregulated OD practices might inflict risks on the participants, an essential player of most social science research whose views have been largely neglected in the Open Science movement (Cummings & Day, 2019; Cummings et al., 2015; Fox et al., 2021). Below we discuss the potential risks of OD for human participants and the potential influence of OD on participants' psychology and research validity.

1.1.1. The risk of re-identification in the digital age

Most harms associated with data sharing are incurred by the post hoc re-identification of the data, which can result from the researchers' lack of skills in de-identifying the data, the nature of the data, and third parties who are incentivized to snoop the data or act upon the participants. One false belief is that as long as participants' identifiers are removed, information is protected. But in fact, re-identification is still possible for such seemingly anonymous data with the current techniques, such as scripts that scrape online data and algorithms that look for patterns within and across data sets (e.g., Sweeney et al., 2018). Variables that do not appear to allow personal identification may be linked across multiple data sets to re-identify a data set (Bishop & Gray, 2017; Rocher et al., 2019). This could particularly put data in behavioral research such as psychology at risk, where all data relating to one participant are conventionally kept together without permutation. There have already been several cases of presumably anonymous data being re-identified by third parties (Zimmer, 2010).

In addition, certain types of data and populations are easier to (re) identify. For example, non-independent data that involve linked individuals, such as dyadic, family, network, and organizational data, are easier to re-identify, especially for the related others such as partners of the participants who might want to snoop the data (e.g., Joel et al., 2018). Individuals of small or defined populations such as undergraduate communication students (Joel et al., 2018) and underrepresented, marginalized individuals are also easier to identify within and across data sets (Sweeney et al., 2018).

When there is sensitive, private information in the data, re-identification may induce severe negative consequences for participants. But even when the information appeared to be innocent at the time of data collection, re-identification may bring about unforeseeable and uncontrollable threats and harms from third parties. For example, data of people's political attitudes might appear innocent when the research was conducted; but when under politically volatile conditions, authorities in a dictatorial regime may identify the participants with archived social media data years later (Pearce, 2020). In the worst-case scenario, the data might be abused by third parties for malicious purposes of harming participants, such as for mass surveillance (Rocher et al., 2019) and discriminatory profiling (Bishop & Gray, 2017).

1.2. The effects of open data policy on participants' privacy concerns

The abovementioned risks associated with OD are salient in behavioral research in fields such as communication and psychology, because the data collection often involves soliciting participants' disclosure of their private information, defined as information that potentially makes them vulnerable (Petronio, 2002). For example, communication and psychology researchers often need participants to report their demographic information, attitudes, beliefs, and behaviors. Sometimes they ask participants to generate texts or speeches (e.g., conversing with a partner, describing a hurtful conflict, etc.) that they later examine to discern participants' psychological attributes and communicative patterns. Thus, the validity of such behavioral research depends on the quantity and quality of the information that participants provide.

According to the theory of Communication Privacy Management (CPM), individuals consider themselves as the owners of their private information (Petronio, 2002). Even when they have already shared their information with others (e.g., the researchers), they consider the recipients merely as guardians of that information, and they still have the ultimate right to control the flow of their information such as deciding with whom and how their information can be shared. Any violation of such an expectation might frustrate the information owners, compromise trust, and damage the relationships between the two parties (Petronio, 2002).

When participants share their private information with researchers in a research study, according to CPM, they may consider themselves as the owners of the private information and the researchers as only guardians of that information. The implementation of OD, however, could negatively violate participants' expectancy of their privacy and cause privacy concerns among them. Under the OD policy, researchers need to evaluate the risk of disclosing certain data (Joel et al., 2018), decide what to keep in the published data set, proactively make the data publicly accessible, after which the responsibility of protecting participants is delegated to unspecified third parties. In this chain of decisions on the information flow that researchers make on behalf of participants, any miscalibration may result in the information flowing or being used in ways that are inconsistent with the participants' expectations and may even harm the participants. As the scope of data sharing broadens, the probabilities of privacy violation and re-identification increase. Although perfect privacy protection is impossible, and this point is often explicitly conveyed to participants in standard research practice, OD might raise additional privacy concerns as compared with the traditional, private model where data are kept with the research team.

Some preliminary evidence suggests that participants do hold privacy concerns regarding OD. When asked about their opinions on data sharing, participants reported concerns related to control and ownership of data, the responsibility and reliability of the third parties with whom the data are shared, information sensitivity, and being identified by governments, corporations, and other entities using the data (Albornoz, 2018; Cummings et al., 2015; Joly et al., 2016). In Bottesini and Vazire

(2019), among the 128 participants who were asked about their thoughts on data sharing and storage, about 50% spontaneously mentioned data security. Typical responses include "I think it should be stored securely, privacy protect [ed] and shared only with researchers"; "Data should be stored on a secure computer and only used for the purposes of this survey. It should be destroyed when the survey results are completed." Hence, we predict that the implementation of OD—sharing data indiscriminately with the public—will induce greater privacy concerns among participants as compared with when the data are retained privately by the research team.

H1. Open data will induce greater privacy concerns than the private condition.

1.3. The effects of open data policy on participants' disclosure

To have participants voluntarily disclose information, it is important for researchers to ensure participants' privacy and make them feel comfortable and safe enough to disclose honestly (Campbell et al., 2019). As Freiling et al. (2021) have argued, because OD allows more people to access the data, it may induce more risks for participants, which may in turn discourage participants to disclose information to the researchers. Should participants decline to disclose or respond with less or untruthful information due to OD, research validity is at stake. Although past research has not directly tested this hypothesis, theories and research in privacy management and disclosure decisions suggest it is likely.

The central determinant in many models of decision-making about disclosure is risk, that is, any potential harm to self and related others (Afifi & Steuber, 2009; Fisher, 1986; Greene, 2009; Petronio, 2002). Risk is determined by two factors: the sensitivity of the divulged information and who may have access to that information (John et al., 2011). Sensitive topics are "topics that participants may feel uncomfortable discussing," which typically "include taboo topics, topics associated with shame or guilt, and topics that generally reside in the private spheres of our lives" (Noland, 2012, p. 3). Communication and psychology researchers who are interested in people's attitudes, beliefs, and behaviors in various life domains sometimes unavoidably need to collect data on such sensitive topics. In such cases, participants' decisions on whether and how much to disclose should depend on who may access their data and how much risk is involved.

Research has found that one's disclosure depends on the relationship between the discloser and the recipient—the more one trusts the recipient, the more they are willing to disclose (Bansal et al., 2016; Joinson et al., 2010). With OD implemented, however, the data are accessible to unspecified third parties, and there is no guarantee on whether they would identify the data and what they will do with the data. In fact, Cummings et al. (2015) found that participants did not trust third parties. Hence, when OD is implemented, participants may become less willing to disclose their information in a research study when knowing that their data will be shared with unspecified and potentially untrustworthy third parties.

Participants may reduce their disclosure via various methods. We examine three common forms of nondisclosure that may influence research validity in a typical questionnaire-based study where sensitive information is requested from participants.

1.3.1. Effects of open data on participation

The most radical form of nondisclosure in research is non-participation or dropping out (Cummings et al., 2015). As Freiling et al. (2021) predicted, a statement of OD in the consent form such as "we may share your data publicly online" may lead many to refuse to participate. If people who say "no" are systematically different from those who agree to participate, researchers will end up with unrepresentative data and draw biased conclusions. That said, when motivated by incentives (e.g., payment, extra credits), participants may still decide to stay in the study despite the OD policy (Liu & Sundar, 2018), and they may manage their privacy via other means to be discussed below. Therefore, we raise the following research questions to explore whether and to what extent OD may cause non-participation in the current study, and in general, less willingness to participate in a research study.

RQ1. Compared with private data policy, does open data policy influence a) the rate of withdrawal in this study and b) one's willingness to participant in a general study?

1.3.2. Effects of open data on disclosure in answering close-ended questions

In addition to withdrawing, participants can reduce their disclosure by a) choosing "I prefer not to answer" (termed as *active nondisclosure*) when such a choice is provided, b) leaving the questions unanswered (termed as *passive nondisclosure*), or c) disclosing untruthful information (John et al., 2011; Joinson et al., 2010). All these strategies compromise research validity. In the cases of active and passive nondisclosure, researchers will end up having missing data, that is, less information. The third strategy (i.e., lying) is even more destructive because the untruthful responses might lead researchers to reach wrong conclusions.

Unfortunately, previous research suggests that participants are likely to adopt these strategies when they have privacy concerns. In Joinson et al. (2010), participants were asked questions on sensitive topics (e.g., number of different sexual partners), and they had three options: submitting the default option (i.e., passive nondisclosure), disclosing the information, or choosing "I prefer not to say" (i.e., active nondisclosure). Results show that participants' frequency of active nondisclosure was positively related with their privacy concerns (Study 1); participants also disclosed less under weak privacy protection (e.g., stimuli mentioning data sharing, web-servers' data tracking, etc.) as compared with strong privacy protection (e.g., stimuli mentioning no identifiers to be collected, no data sharing, secure data storage, etc.), when the information requestor was not trusted (Study 2).

Similarly, in John et al. (2011), researchers asked participants whether they had done a series of sensitive, socially disapproved conducts (e.g., "Have you ever cheated while in a relationship?") when participants could respond to each question with "Yes" or "No." Because people are unlikely to admit to unethical behaviors when they worry about privacy violation and its consequences such as identification and incrimination, the proportion of questions answered affirmatively (i.e., affirmative admission rate, AAR) indicates the amount of truthful disclosure. That is, the higher AAR to such socially disapproved behaviors, the more likely that the responses are being truthful. John et al. (2011) found that participants had higher AAR on the website with a design that could suppress privacy concerns (Experiment 2 and 3); in other words, greater privacy concerns were associated with less truthful disclosure.

Because OD may increase risks of re-identification and induce more privacy concerns among participants, we anticipate that participants' nondisclosure will be higher under the OD policy. As we are interested in both the amount and the truthfulness of participants' disclosure, we combined the paradigms in John et al. (2011) and Joinson et al. (2010). In the questionnaire of the current study, we included a battery of sensitive questions asking participants whether they had engaged in certain behaviors that are generally deemed unethical or socially disapproved, where they could respond by choosing from "Yes," "No," and "I prefer not to answer," or just leave the question(s) unanswered. Specifically, we examined participants' (non)disclosure in terms of a) the rate of active nondisclosure, b) the rate of overall nondisclosure (both active and passive nondisclosure), and c) AAR, the rate of admission to conducting these socially disapproved, unethical behaviors. In light of previous research, we predict that participants' nondisclosure will be higher with OD implemented.

H2. In response to questions about one's socially disapproved conducts, the open data policy will induce a) a higher rate of active nondisclosure, b) a higher rate of overall nondisclosure, and c) a lower

AAR, as compared with the private condition.

With nondisclosure conceptualized as participants' means of privacy management, we further hypothesize that privacy concerns mediate the effects of OD on nondisclosure.

H3. Privacy concerns mediate the effects of open data policy on a) the rate of active nondisclosure, b) the rate of overall nondisclosure, and c) AAR, relative to the private condition.

1.3.3. Effects of open data on open-ended disclosure

In addition to using close-ended self-report measures, behavioral researchers often collect participants' open-ended responses to understand their psychological experiences and behavioral patterns (e.g., McLaren & Solomon, 2010; Pennebaker, 1997). With other things being equal, the more participants disclose, the more researchers can learn from the data. Thus, researchers often strive to solicit more disclosure from participants with various strategies such as using modes that induce less social presence in data collection (e.g., using web-based questionnaire, see a meta-analysis Weisband & Kiesler, 1996) or letting the experimenters disclose about themselves first in order to elicit reciprocal disclosure (Joinson, 2001). Should participants become more concerned about their privacy due to the OD policy, they may disclose less of their personal experience, which, again, compromises the validity of research findings based on such reserved, abridged disclosure.

In the current study, we asked participants three open-ended questions regarding their personal experiences that social psychology and communication researchers often study: hurtful conflicts (e.g., McLaren & Solomon, 2010), affectionate communication (e.g., Floyd & Custer, 2020), and relational turbulence (e.g., Solomon et al., 2016), which are also topics that have been found as intimate and private in previous studies (e.g., Moon, 2000; Shaffer & Tomarelli, 1989). Specifically, we expect that the OD policy will decrease participants' disclosure amount, which is traditionally indexed by word count of participants' responses (Collins & Miller, 1994; Joinson, 2001; Moon, 2000), because of their heightened privacy concerns in the condition of OD policy.

H4. The open data policy will decrease the word count of participants' open-ended disclosure of their personal experience, as compared with the private condition.

H5. Privacy concerns mediate the effect of open data policy on the word count of participants' open-ended disclosure relative to the private condition.

1.4. Exploring solutions to the unintended effects of OD

If OD raises additional privacy concerns among the participants and thereby discourages their candid, rich disclosure, then it becomes counterproductive toward its original goal—more ethical and valid research. Given this potential downside of OD, what strategies can be employed to both enhance transparency in research practice and keep participants assured? Previous research suggests that researchers seem to be an exempted category that participants trust in general, and that compared with sharing data with anyone indiscriminately (i.e., OD in its extreme form), participants are more tolerant of their data being shared with other researchers (e.g., Gilmore et al., 2018; Joel et al., 2018; Lee et al., 2019). As such, having researchers as gatekeepers in the data sharing process or limiting the scope of sharing within the research community may be a middle ground where the original aims of OD can be reached without compromising research ethics and validity.

We therefore explored two alternative OD practices that afford descending levels of data publicness as potential solutions: 1) sharing data upon reasonable requests made to the researchers and 2) sharing data upon requests by university-affiliated researchers only. If they do not raise additional privacy concerns or discourage disclosure, they can serve as better means to open and valid research. We raise the following research questions to explore the effects of these two more restrictive variants of the OD policy relative to private data storage.

RQ2. Will sharing data upon requests by anyone induce a) greater privacy concerns, b) a higher rate of withdrawal, c) a higher rate of active nondisclosure, d) a higher rate of overall nondisclosure, e) a lower AAR, and f) fewer words, as compared with the private condition?

RQ3. Will sharing data only with researchers induce a) greater privacy concerns, b) a higher rate of withdrawal, c) a higher rate of active nondisclosure, d) a higher rate of overall nondisclosure, e) a lower AAR, and f) fewer words, as compared with the private condition?

2. Method

2.1. Overview

We conducted a five-condition between-subjects online experiment with participants ($N = 412$) recruited from *Amazon's Mechanical Turk* (MTurk) via *CloudResearch* (formerly *TurkPrime*, Litman et al., 2017). Participants were invited to answer a questionnaire purported to be about "lifestyle and media use habits" on *Qualtrics*. They first read a consent form with standard language about privacy and confidentiality, stating, "Reasonable efforts will be made to keep the personal information in your research record private. However, absolute confidentiality cannot be guaranteed." All consented to participate and were then randomized to one of the five conditions with different policies of data sharing or with data policy not mentioned. The questionnaire began with nonsensitive, lead-in questions about media use, followed by 33 questions asking about one's previous engagement in socially disapproved behaviors, three open-ended questions about one's personal experience, a manipulation check, and the measure of privacy concerns. Participants were then debriefed about the true purpose of the study and asked for their willingness to participate in any study under OD. Each participant received $1.5 for their participation.

2.2. Participants

To ensure data quality, we restricted the link to this study only available to workers who had completed at least 100 HITs (i.e., tasks on Mturk), had an approval rate above 95%, and passed the quality check of CloudResearch. In total, 412 participants participated in the study, but only those who passed the attention check and manipulation check remained in the main data analysis. The final data set included 294 participants between the ages of 18 and 73 years ($M = 41.55$, $SD = 12.86$); the majority (60.2%) held Bachelor's degrees or above; 144 participants self-identified as male (48.98%), 143 as female (48.64%), three as "other," and four preferred not to disclose their gender. The majority self-identified as White ($n = 211$, 71.77%), followed by Asian ($n = 32$, 10.88%), Black ($n = 21$, 7.14%), multiracial ($n = 19$, 6.46%), Hispanic ($n = 5$, 1.70%), and native American ($n = 1$, 0.34%); five participants preferred not to disclose their race or ethnicity (1.70%).

We performed a post hoc power analysis using G*Power (Faul et al., 2007) based on our analysis strategies (more details in the Results section below) to check the achieved statistical power given our sample size. With analysis of covariance as the target statistical test, we specified a medium effect size $f = 0.25$, significance level $\alpha = 0.05$, total sample size = 294, numerator $df = 4$, number of groups = 5, and number of covariates = 1, which returned an achieved statistical power of .94. We then considered our sample size as sufficient to detect the hypothesized effects.

2.3. Manipulation

We created four conditions of data policies with various levels of data publicness. We also included a baseline condition in which data sharing was not explicitly mentioned to account for the mere effect of mentioning data policy. The exact wordings of the four data policies are

presented in Table 1.

Because previous research found participants spending little time reading the consent form (Douglas et al., 2021) and disregarding the data sharing policies in it (Cummings et al., 2015), we expected that if we merely mentioned the data sharing policy in the consent form, participants might not notice it. To test this speculation, we conducted a pilot test with a separate group of participants ($N = 43$) recruited from Mturk. We found participants, on average, only spent 33.48 s ($Mdn = 11.79$, $SD = 57.70$) reading the informed consent form of 929 words. Given the average speed of silent reading—238 English words in non-technical material per minute (see a meta-analysis, Brysbaert, 2019)—participants should spend about 3.90 min (i.e., 234 s) on reading the consent form, which is much longer than the time our participants spent. In other words, participants in the pilot test did not read the consent form carefully or fully. Therefore, instead of stating the data sharing policy in the consent form, we placed the statement in red font on the top of each page of the questionnaire, so that participants could see and acknowledge the data policy when participating in our study.

2.4. Measures

Time Spent on Consent Form. The time participants spent reading the consent form was recorded with the timer embedded in the page of the consent form.

Active Nondisclosure. Following previous research (Joinson et al., 2008, 2010), we measured active nondisclosure as the rate of choosing "I prefer not to answer" in response to the 33 sensitive questions (see Supplementary Materials) on whether one had engaged in socially disapproved behaviors adapted from previous research (John et al., 2011), such as "Have you had more than five sexual partners?" ($M = 2.69\%$, $SD = 12.56\%$).

Overall Nondisclosure. Following previous research (Joinson et al., 2010), we also recorded passive nondisclosure, that is, one's rate of unanswered questions among the 33 sensitive questions. The overall nondisclosure was computed by summing one's active and passive nondisclosure ($M = 2.84\%$, $SD = 12.67\%$).

Rate of Affirmative Answers (AAR). Following John et al. (2011), we measured AAR as the proportion of sensitive questions answered affirmatively by a participant (i.e., choosing "Yes") among the 33 sensitive questions ($M = 22.63\%$, $SD = 17.31\%$).

Amount of Open-Ended Disclosure. Participants saw three open-ended questions (see Supplementary Materials). The first question asked them to recall and describe an episode of hurtful conflict they had experienced (Pickard et al., 2018); the second asked them to describe a person they love and appreciate (Pickard et al., 2018); and the third asked them to recall their experience of relational uncertainty and the communication they had about it with their partners. Word count was computed by summing the number of words in participants' responses to all three questions ($M = 123.64$, $SD = 67.94$).

Table 1
Data policy stimuli on each page of questionnaire on qualtrics.

Conditions	Instructions on each page
Baseline	None.
Public access in repository	"Please note: Your input will be shared in a public repository accessible to any other individuals and entities (e.g., researchers, companies, and governments)."
Public on request from anyone	"Please note: Your input will be shared on reasonable requests from other individuals and entities (e.g., researchers, companies, and governments)."
Public on request from researchers only	"Please note: Your input will be shared on reasonable requests from other university-affiliated researchers."
Private	"Please note: Your input will not be shared beyond our research team."

Manipulation Check. Participants (except those in the baseline condition) were asked to recognize the data policy they received among five options, 1) "Data will be shared in a public repository accessible to any other researchers, companies, governments, and other individuals and entities"; 2) "Data will be shared on reasonable requests from other researchers, companies, governments, and other individuals and entities"; 3) "Data will be shared on reasonable requests from other university-affiliated researchers"; 4) "Data will not be shared beyond our research team"; and 5) "I did not see such information mentioned."

Privacy Concerns. Privacy concerns were measured by five items adapted from previous research (John et al., 2011; Joinson et al., 2010). Participants were asked to rate the extent to which they were concerned about "incriminating myself," "whether my answers would truly be private," "who might have access to my answers," "whether the survey was truly anonymous," and "whether my answers would truly be confidential" on 7-point scales from 1 ("Not at all") to 7 ("A great deal") (Cronbach's $\alpha = 0.93$, $M = 4.06$, $SD = 1.97$).

Perceived Sensitivity. To validate the overall sensitivity of the questionnaire as perceived by the participants, we measured perceived sensitivity by two items, "On average, how sensitive are the questions in this survey to you?" and "On average, how intrusive are the questions in this survey to you?" on 7-point semantic differential scales, respectively with "Not sensitive at all—Very sensitive" and "Not intrusive at all—Very intrusive." We created an index by averaging participants' ratings (Cronbach's $\alpha = 0.81$, $M = 5.37$, $SD = 1.35$).

Willingness to Participate Under OD. To explore participants' general attitudes toward OD, after the debrief, we asked participants an open-ended question "Could you please tell us why Open Data policy (i.e., sharing research data to the public) might or might not influence your decision to participate in a study?" Two of the authors first went through all answers to develop a preliminary codebook with two variables, respectively (1) valence of attitudes toward OD and (2) explanations. Then they coded 31 responses (10.54% of all responses) individually, resolved disagreements, and revised the codebook. They then coded another 30 responses with acceptable intercoder reliability (Krippendorff's $\alpha = 0.90$ for valence; Krippendorff's $\alpha = 0.89$ for reason). Then they resolved the disagreements and split the rest responses to complete the coding process. The final codebook can be found in the Supplementary Materials.

3. Results

3.1. Preliminary analysis and data cleaning

Prior to data cleaning, we performed preliminary analyses with the full data set ($N = 412$) to answer research questions regarding the effects of data sharing policies on dropout rate (RQ1a, 2b, and 3b) and to test whether the manipulation affected data quality, so as to determine whether we would have to handle differential attrition in the main analysis.

3.1.1. Withdrawal

In total, 73 participants (17.72%) dropped out of the study in the process. In response to RQ1a, 2b, and 3b on whether different OD policies influenced participants' decisions to leave or stay in the study, we compared the dropout rate across the five conditions with a Chi-square analysis and found no statistically significant difference, χ^2 ($df = 4$) = 0.82, Cramer's $V = 0.05$, $p = .94$.

3.1.2. Attention check

In the middle of the 33 sensitive questions, we inserted an item instructing "Please choose 'No' here" as an attention check. In total, 17 participants (4.13%) failed to follow the instruction. Chi-square test revealed no statistically significant differences across the five conditions on the rate of passing this attention check, χ^2 ($df = 4$) = 0.40, Cramer's

$V = 0.03$, $p = .98$.

3.1.3. Manipulation Check

In total, 46 participants (11.17%) did not pass the manipulation check. The rates of passing/failing the manipulation check were not significantly different across the four experimental conditions at the 0.05 level (see the Supplementary Materials for the contingency table).

Taken together, the presence and the publicness of data policies did not cause differential attrition or impact the data quality in a prominent manner. We then removed participants who had not completed the whole study, failed the attention check, or failed the manipulation check, and conducted the main analysis with clean, high-quality data ($N = 294$) only.

3.1.4. Perceived sensitivity

With the clean data, overall, participants found questions asked in our study were moderately-to-highly sensitive ($M = 5.37$, $SD = 1.35$). This validates the sensitivity of the overall study to observe the effect of OD on participants' in-study behaviors.

3.2. Main analysis

To test H1, H2, and H4 and to answer RQ2–3, we specified a set of analyses of covariance with data sharing policy as the independent variable and with privacy concerns, the rate of active nondisclosure, the rate of overall nondisclosure, AAR, and the amount of open-ended disclosure (i.e., word count) as the dependent variable separately while controlling for the study version.[1] Results are summarized in Table 2.

H1 was supported (F (4, 288) = 4.66, $p = .001$, partial $\eta^2 = 0.06$) such that participants in the public-access condition ($M = 4.62$, $SE = 0.25$) reported significantly greater privacy concerns than those in the private condition ($M = 3.45$, $SE = 0.24$), Cohen's $d = 0.61$. Participants in the public-upon-request condition ($M = 4.29$, $SE = 0.29$) also reported significantly greater privacy concerns than those in the private condition (RQ2a), Cohen's $d = 0.44$. But privacy concerns in the public-to-researchers condition ($M = 3.55$, $SE = 0.26$) were not significantly different from the private condition (RQ3a).

H2a and H2b were not supported such that the effects of OD were nonsignificant on active nondisclosure, F (4, 288) = 0.56, $p = .69$, partial $\eta^2 = 0.01$, and overall nondisclosure, F (4, 288) = 0.46, $p = .77$, partial $\eta^2 = 0.01$. The effects of the public-upon-request condition (RQ2c and d) and the public-to-researchers condition (RQ3c and d) were also nonsignificant.

H2c was supported (F (4, 288) = 3.01, $p = .02$, partial $\eta^2 = 0.04$) such that AAR in the public-access condition ($M = 16.94\%$, $SE = 2.22\%$) was significantly lower than that in the private condition ($M = 26.95\%$, $SE = 2.12\%$), Cohen's $d = 0.59$. AARs in the public-upon-request condition (RQ2e) and the public-to-researchers condition (RQ3e) were not significantly different from the private condition.

H4 was not supported such that OD did not reduce the word count of open-ended disclosure, F (4, 288) = 0.46, $p = .77$, partial $\eta^2 = 0.01$. The effects of the public-upon-request policy (RQ2f) and the public-to-researchers policy (RQ3f) were also nonsignificant.

To test H3 and H5, we performed mediation analyses using the Model 4 (i.e., simple mediation model) of PROCESS macro (Hayes, 2018) and requested 5000 samples of bootstrapping. We specified data policy as a multicategorical independent variable with the private condition as the reference group, privacy concerns as the mediator, each of the three measures of participants' truthful disclosure and the disclosure breadth as the dependent variable separately, and the study version as the model covariate. Table 3 presents the indirect effects of each data policy (relative to the private condition) via privacy concerns.

H3(a) and H3(b) were not supported because the indirect effects of data policies on the rate of active nondisclosure and overall nondisclosure were not significant. In support of H3(c), we found a significant negative indirect effect of public access of data on AAR via privacy concerns. In addition, we found the public-upon-request condition also had a significant negative indirect effect via privacy concerns on AAR.

H5 was supported, such that compared with the private condition, the public-access condition had a significant negative indirect effect on disclosure amount via privacy concerns. In addition, we also found the public-upon-request condition also had a significant negative indirect effect on disclosure amount via privacy concerns.

To answer RQ1b, we analyzed participants' open-ended responses about their willingness to participate in a study under OD. Only 5.44% ($n = 16$) were willing to participate in an OD study unconditionally, mainly out of altruism—to benefit science and the public ($n = 12$). Another 35.37% ($n = 104$) were unwilling to participate, mainly for reasons of their norm of privacy ($n = 58$, e.g., "I will not participate … I would not have any control over this information") and concerns of risk associated with data sharing such as identification ($n = 25$, e.g., "There is always a chance the data could be used to identify me"); some even mentioned self-defense strategies such as lying and selective disclosure ($n = 15$, e.g., "If I'm worried, I can always give false answers"). The majority mentioned contingent participation (41.16%, $n = 121$); specifically, they would only participate when certain conditions are met such as anonymity guaranteed ($n = 47$), no sensitive questions asked ($n = 43$), high payment ($n = 15$), and if the data are only shared with researchers but not the general public ($n = 8$, e.g., "I believe public will start misusing it, but I am sure the researcher will not do it").

4. Discussion

With open science increasingly advocated and adopted, it is imperative to gain more evidence-driven understanding of quality criteria in research (Freiling et al., 2021). The current study serves as one of the initial efforts in empirically testing the effects of OD on participants' psychology and research validity. Findings reveal concerns for OD while also suggest potential solutions. We first interpret the findings and then offer suggestions for data sharing practices.

First, we predicted that public data sharing would induce greater privacy concerns among participants (H1). Results supported our hypothesis such that participants in the condition of public data sharing, the most recommended OD practice, reported the highest level of privacy concerns, significantly higher than that in the condition of private data storage. To be noted, no personal identifiers were collected in this study. Despite this, public data sharing still increased people's privacy concerns. This could be due to participants' general fear of privacy violation in the digital age and that public data sharing reminded them of the potential risks of being identified and data misused by untrusted third parties.

Second, we examined the effects of OD on one's decision to participate (RQ1a) and whether OD discouraged participants' disclosure on sensitive topics (H2) and open-ended disclosure of personal experiences (H4) due to their privacy concerns (H3, H5). Although some researchers suggest OD might lead to non-participation or dropout (Cummings et al., 2015; Freiling et al., 2021), we did not find evidence for OD driving participants to leave the study or participate carelessly, as no one declined to participate, and we did not find significant differences across

[1] Although not considered in the analysis, we also measured participants' dispositional privacy value. To account for the potential priming effect of this measurement on participants' responses to our core manipulation (publicness of data sharing), we prepared two versions of study with this measure positioned either before or after participants completed the questions on the main outcome variables. We had run analyses to ensure that there was neither statistically significant main effect of the study version nor any interaction effect between data policy and study version on our outcome variables. Therefore, in the main analyses, we combined all data and statistically controlled for the study version.

Table 2
Means and SEs of outcome variables in each condition.

Outcome variable	Publicness of data sharing *M* (*SE*)				
	Private	Public on request from researchers only	Public on request from anyone	Public access in repository	Baseline
Privacy concern	3.45 (0.24)[a]	3.55 (0.26)[a,b]	4.29 (0.29)[b,c]	4.62 (0.25)[c]	4.43 (0.23)[c]
Active nondisclosure (%)	1.94 (1.56)	1.82 (1.70)	3.09 (1.88)	4.74 (1.64)	2.09 (1.51)
Overall nondisclosure (%)	2.21 (1.58)	2.20 (1.72)	3.10 (1.90)	4.74 (1.66)	2.17 (1.52)
Affirmative answers (%)	26.95 (2.12)[a]	23.09 (2.30)[a,b]	25.12 (2.55)[a]	16.94 (2.22)[b]	21.45 (2.04)[a,b]
Word count	132.12 (8.47)	124.98 (9.21)	119.65 (10.19)	116.46 (8.89)	123.32 (8.17)

Note. Within rows, means with no superscript in common differ at $p < .05$.

Table 3
Indirect effects of data policies (vs. Private data) on disclosure via privacy concerns.

Data policy	Dependent variable			
	Active nondisclosure	Overall nondisclosure	Affirmative answers	Word count
Public on request from researchers only	$B = 0.05$, Boot$SE = 0.22$, 95% CI [−0.31, 0.60]	$B = 0.05$, Boot$SE = 0.23$, 95% CI [−0.35, 0.63]	$B = -0.26$, Boot$SE = 0.90$, 95% CI [−2.06, 1.53]	$B = -0.61$, Boot$SE = 2.11$, 95% CI [−5.18, 3.56]
Public on request from anyone	$B = 0.40$, Boot$SE = 0.37$, 95% CI [−0.12, 1.31]	$B = 0.43$, Boot$SE = 0.39$, 95% CI [−0.12, 1.39]	$B = -2.08$, Boot$SE = 0.98$, 95% CI [−4.23, −0.31]	$B = -4.78$, Boot$SE = 2.55$, 95% CI [−10.53, −0.57]
Public access in repository	$B = 0.55$, Boot$SE = 0.47$, 95% CI [−0.17, 1.67]	$B = 0.60$, Boot$SE = 0.49$, 95% CI [−0.14, 1.77]	$B = -2.89$, Boot$SE = 0.99$, 95% CI [−5.03, −1.22]	$B = -6.64$, Boot$SE = 3.07$, 95% CI [−13.48, −1.52]
Baseline	$B = 0.46$, Boot$SE = 0.41$, 95% CI [−0.16, 1.44]	$B = 0.50$, Boot$SE = 0.43$, 95% CI [−0.12, 1.56]	$B = -2.42$, Boot$SE = 0.96$, 95% CI [−4.53, −0.77]	$B = -5.56$, Boot$SE = 2.74$, 95% CI [−11.80, −1.08]

Note. Reference group = private data.

conditions in terms of dropout rate, attention, and recall of the data policy (i.e., the manipulation check) in our preliminary analyses.

But we found evidence of OD discouraging candid disclosure such that the rate of admission to socially disapproved behaviors in the condition of public data sharing was significantly lower than that in the condition or private data storage, and this difference was mediated by privacy concerns. Hence, the lower rate of "yes" in the OD condition should be interpreted as participants' strategy to protect their privacy. This finding is consistent with previous research on privacy concerns and disclosure (John et al., 2011).

To be noted, only the rate of affirmative answers but not nondisclosure (active or overall) was significantly affected by the OD policy. This is inconsistent with previous studies where privacy concerns were found significantly related to nondisclosure (e.g., Joinson et al., 2010). On average, less than 1/33 of all 33 sensitive questions were responded with nondisclosure (including active nondisclosure, that is, choosing "I prefer not to answer"). Given the low rate of such behaviors, the non-significance might be interpreted as a floor effect. This could be due to the characteristics of our sample (i.e., MTurk workers, or Turkers). Previous research found that Turkers are primarily driven by monetary incentives; when psychological discomforts (e.g., cognitive dissonance induced by low payment) were incurred, they tended to remain in the study regardless in order to get the researchers' approval and earn the payment, while using other cognitive and behavioral strategies to regulate their emotions, such as paying less effort or convincing themselves that their effort was valuable to science (Liu & Sundar, 2018). Hence, even though Turkers in the current study had greater privacy concerns under the OD policy, to earn the payment, they might have purposefully insisted on staying in the study and responding to as many questions so that the researchers would not reject them for the apparent nondisclosure and would still pay them. Meanwhile, they might have opted for denial of the socially disapproved behaviors as the strategy to protect themselves from the potential consequences of privacy violation.

We also did not find evidence for the effect of OD on the amount of participants' open-ended disclosure (i.e., word count). But we still found a negative indirect effect of OD policy on disclosure amount via privacy concerns, which suggests that privacy concerns still discouraged disclosure. Again, this finding might be attributable to the competing motivation of Turkers as discussed above, such that participants might have wanted to ensure that their effort would appear to be acceptable by the researchers and thereby maintained a relatively large amount of disclosure across conditions, regardless of their privacy concerns.

Third, we explored two other forms of OD that afford lower levels of publicness as potential solutions to the unintended effects of OD on privacy concerns and disclosure (RQ2–3)—sharing data upon reasonable request by anyone or by university-affiliated researchers only. Results are consistent with previous research that suggests researchers as an exempted category that participants trust in general (e.g., Gilmore et al., 2018; Joel et al., 2018; Lee et al., 2019). Specifically, participants' level of privacy concerns in the public-to-researchers condition did not differ from that in the private condition, whereas in the conditions of public upon request (from anyone) and public access (in a repository), participants reported significantly higher levels of privacy concerns than in the private condition. In addition, with researchers as the gatekeepers deciding what merits "reasonable requests," AARs in both the public-upon-request conditions were no lower than that in the private condition, whereas only the public-access condition had AAR significantly lower than the private condition. Taken together, the findings suggest that data sharing within the research community may neither induce additional privacy concerns nor discourage participants' disclosure of sensitive and personal information, which is desirable to researchers.

Lastly, by asking participants explicitly about their willingness to participate in a study under OD (RQ1b), we found that most participants held concerns toward public data sharing. Although a small subset wanted to help with research and science, many expressed privacy-related worries. This is consistent with the previous findings by Bottesini and Vazire (2019) in which the top-two spontaneous themes in Turkers' opinions expressed toward OD were "data security" and "privacy." Overall, our findings suggest participants have two competing motivations. On the one hand, in line with CPM (Petronio, 2002), people do consider private information as their property and care about their ability to control its flow and consequences. On the other hand, they are also willing to grant researchers certain liberty and discretion in making decisions for the larger good, such as research and science, provided that the sharing does not impose any harm on them. Hence, the sweet spot of data sharing should lie in a middle ground with a moderate level of publicness—public enough to facilitate knowledge production, while private enough for data protection.

It is also worth reporting that participants spent little time reading the consent form. Consistent with previous research (Cummings et al.,

2015; Douglas et al., 2021), in both our pilot test ($M = 33.48$ s, $Mdn = 11.79$, $SD = 57.70$) and the preliminary analysis of the main study ($M = 21.85$ s, $Mdn = 9.66$, $SD = 42.21$), participants spent much shorter time reading the 929-word consent form than the expected reading time of 3.90 min (Brysbaert, 2019). This suggests participants paid very little attention to the consent form or only read a part of it, which has important implications for consent obtaining in research practice.

4.1. Practical implications

Although findings in the current study suggest some downsides of OD, OD may still serve its original purposes—more transparent, ethical, and valid science—if the data can be shared so responsibly that risks are minimized, and participants' privacy concerns are addressed. Based on our findings, we propose the following strategies pertaining to research practices in data sharing and consent obtaining to address the unintended effects of OD on participants' privacy concerns and disclosure.

4.1.1. Strategy No. 1: sharing data within the research community

Results suggest that when the data were only shared with researchers, participants' privacy concerns were not significantly different from that in the private condition; when researchers gate-keep the sharing process, participants' disclosure truthfulness and amount were also not significantly different from those in the private condition. Hence, we recommend only sharing research data with researchers while implementing certain gatekeeping procedures to ensure such restrictive sharing.

There are at least three benefits of doing this. First, with less privacy concern, participants can feel more comfortable disclosing honestly. Second, the gatekeeping procedures allow the authorized researchers to be held accountable if they end up using the data unethically and can help prevent unauthorized third parties from abusing the data for unethical purposes. Third, although sharing data with authorized researchers affords less publicness than sharing indiscriminately, it still allows other researchers to verify and replicate the original findings and perform secondary analysis. In addition, technologies can be used to facilitate the gatekeeping and sharing processes. Therefore, the primary goal of OD—promoting transparency and knowledge distribution—will not be compromised. So far, there are already technologies and data sharing guidelines that support formal gatekeeping in the data sharing process (for more resources, see Gilmore et al., 2018; Joel et al., 2018; Levenstein & Lyle, 2018; Meyer, 2018). For example, *Databrary* (databrary.org), an online data library specialized for storing and sharing video data, can restrict the data as open only to institutionally approved researchers (Gilmore et al., 2018).

4.1.2. Strategy No. 2: privacy protection

Results suggest that participants' primary concern with data sharing is privacy concern, which was also found as the mechanism for the reduced disclosure in terms of both truthfulness and quantity. Hence, one solution could be directly reducing participants' privacy concerns. To this end, researchers could inform participants on how they plan to protect their privacy and to minimize the risks associated with disclosure and identification and discuss the effectiveness of these measures so as to reduce participants' privacy concerns.

4.1.3. Strategy No. 3: obtaining meaningful informed consent

When compared with the common consent gaining process (i.e., keeping all information in one consent form preceding the study), our manipulation of data policies (i.e., data policy highlighted on each page of the questionnaire) appeared to fall short in ecological validity. A more natural method would be including some standard languages in the consent form to inform participants about the data sharing policy, as suggested by some researcher and institutions (ICPSR, 2017; Meyer, 2018). Nevertheless, as found in previous research (e.g., Cummings et al., 2015; Douglas et al., 2021) and replicated in the current study, participants spend extremely short time reading the consent form. Hence, simply including the statement of data policy in the consent form may be ineffective in informing participants, especially in online research.

In light of the observed effects of data policies highlighted on each page in the current study, our approach offers suggestions on how to modify the consent process to ensure that participants are really informed of the nature of the study and the implications of their participation. For example, to obtain meaningful consent, researchers could emphasize the data policy they adopt and its potential risks (e.g., privacy violation, identification) and benefits (e.g., transparent science), such as using larger fonts, more conspicuous colors, etc., to alert participants about this recent change in the research community, so that they can then make informed decisions on whether to participate.

4.2. Limitations and future directions

The current study has several limitations which should be addressed in future work. First, we only used a convenience sample recruited from MTurk, which can be different from other populations on related characteristics such as digital literacy, financial need, and motivations. Therefore, our findings might not generalize to other populations. For example, tech-savvy participants might be more aware of the techniques of re-identification, more concerned about the risk of data sharing via digital platforms, and therefore more sensitive to OD. Participants recruited from other sources and who are less concerned about getting monetary rewards, such as studies broadcasted on social media platforms, may just vote with their feet and simply not participate upon seeing the OD policy or withdraw in the middle when they see sensitive questions. Future research should test the effects of OD with different populations and consider their motivations and vulnerability to data sharing.

Second, sensitive questions were asked in this study and the effects of OD on disclosure seemed to only apply to the sensitive questions. For studies that do not touch on such sensitive topics, OD might have less impact on participants' privacy concerns and disclosure, if at all. Future research may test the effects of OD on other measures and determinants of research validity for both sensitive and nonsensitive topics. In addition, we only analyzed the amount of open-ended disclosure (indicated by word count). Future research then could consider exploring other dimensions, such as the depth of disclosure, to yield a richer understanding of OD effects in online behavioral research.

Third, we examined the effects of OD from the perspective of privacy management, but there might be other operative psychological mechanisms underlying the observed effects of OD. For example, OD exposes participants' data to more people and may lead participants to feel being observed by others. The social presence of the third parties might trigger processes such as self-monitoring, impression management, and social facilitation, which may have different implications for research validity depending on the nature of the study. Future research may explore those alternative mechanisms to better understand the implications of OD for participants' psychology and research validity.

5. Conclusion

In response to behavioral researchers' call for more evidence-driven considerations on the implementation of OD (e.g., Cummings & Day, 2019; Freiling et al., 2021), findings in the current study reveal some downsides of OD, such that even when the risk of identification is not apparent, OD may still induce additional privacy concerns among participants and distort their responses to sensitive questions, which threatens research validity and ethics. We propose sharing data within the research community as a potential solution, as opposed to the extreme form of OD—sharing data to the public indiscriminately. In addition, we call for more responsible actions in data sharing and more meaningful conversations with participants about the risks and benefits

of data sharing at the stage of gaining their informed consent.

Credit Author Statement

Bingjie Liu: Conceptualization, Methodology, Formal analysis, Writing – original draft preparation. Lewen Wei: Conceptualization, Methodology, Data collection, Formal analysis, Writing – original draft preparation.

Funding

Data collection was funded with the first author's faculty development fund at California State University, Los Angeles.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2022.107537.

References

Afifi, T., & Steuber, K. (2009). The revelation risk model (RRM): Factors that predict the revelation of secrets and the strategies used to reveal them. *Communication Monographs, 76*(2), 144–176. https://doi.org/10.1080/03637750902828412

Albornoz, D. A. (2018). *Reimagining open science through a feminist lens*. https://medium.com/@denalbz/reimagining-open-science-through-a-feminist-lens-546f3d10fa65.

Baker, Monya (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*, 452–454. https://doi.org/10.1038/533452a

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management, 53*(1), 1–21. https://doi.org/10.1016/j.im.2015.08.001

Bishop, L., & Gray, D. (2017). Ethical challenges of publishing and sharing social media research data. In K. Woodfield (Ed.), *The ethics of online research (advances in research ethics and integrity* (Vol. 2, pp. 159–187). Emerald Publishing Limited.

Bottesini, J., & Vazire, S. (2019). Do participants want their data to be shared?. https://osf.io/8x5d9/.

Bowman, N. D., & Spence, P. R. (2020). Challenges and best practices associated with sharing research materials and research data for communication scholars. *Communication Studies, 71*(4), 708–716. https://doi.org/10.1080/10510974.2020.1799488

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language, 109*, Article 104047. https://doi.org/10.1016/j.jml.2019.104047

Campbell, R., Goodman-Williams, R., & Javorka, M. (2019). A trauma-informed approach to sexual violence research ethics and open science. *Journal of Interpersonal Violence, 34*(23–24), 4765–4793. https://doi.org/10.1177/0886260519871530

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

Cummings, J. A., & Day, T. E. (2019). But what do participants want? Comment on the "data sharing in psychology" special section (2018). *American Psychologist, 74*(2), 245–247. https://doi.org/10.1037/amp0000408

Cummings, J. A., Zagrodney, J. M., & Day, T. E. (2015). Impact of open data policies on consent to participate in human subjects research: Discrepancies between participant action and reported concerns. *PLoS One, 10*(5), Article e0125208. https://doi.org/10.1371/journal.pone.0125208

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., … De Vreese, C. (2021). An agenda for open science in communication. *Journal of Communication, 71*(1), 1–26. https://doi.org/10.1093/joc/jqz052

Douglas, B. D., McGorray, E. L., & Ewell, P. J. (2021). Some researchers wear yellow pants, but even fewer participants read consent forms: Exploring and improving consent form reading in human subjects research. *Psychological Methods, 26*(1), 61–68. https://doi.org/10.1037/met0000267

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fisher, D. V. (1986). Decision-making and self-disclosure. *Journal of Social and Personal Relationships, 3*(3), 323–336. https://doi.org/10.1177/0265407586033005

Floyd, K., & Custer, B. E. (2020). *Affection exchange theory*. Oxford Research Encyclopedia of Communication. https://doi.org/10.1093/acrefore/9780190228613.013.937

Fox, J., Pearce, K. E., Massanari, A. L., Riles, J. M., Szulc, Ł., Ranjit, Y. S., … L Gonzales, A. (2021). Open science, closed doors? Countering marginalization through an agenda for ethical, inclusive research in communication. *Journal of Communication*. https://doi.org/10.1093/joc/jqab029

Freiling, I., Krause, N. M., Scheufele, D. A., & Chen, K. (2021). The science of open (communication) science: Toward an evidence-driven understanding of quality criteria in communication research. *Journal of Communication*. https://doi.org/10.1093/joc/jqab032

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science, 1*, 121–130. https://doi.org/10.1177/2515245917746500

Greene, K. (2009). An integrated model of health disclosure decision-making. In T. D. Afifi, & W. A. Afifi (Eds.), *Uncertainty, information regulation in interpersonal contexts: Theories and applications* (pp. 226–253). Routledge.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

ICPSR. (2017). *Recommended informed consent language for data sharing*. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confientiality/conf-language.html.

Joel, S., Eastwick, P. W., & Finkel, E. J. (2018). Open sharing of data on close relationships and other sensitive social psychological topics: Challenges, tools, and future directions. *Advances in Methods and Practices in Psychological Science, 1*, 86–94. https://doi.org/10.1177/2515245917744281

John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of Consumer Research, 37*(5), 858–873. https://doi.org/10.1086/656423

Joinson, A. N. (2001). Knowing me, knowing you: Reciprocal self-disclosure in Internet-based surveys. *CyberPsychology and Behavior, 4*(5), 587–591. https://doi.org/10.1089/109493101753235179

Joinson, A. N., Paine, C., Buchanan, T., & Reips, U. D. (2008). Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior, 24*(5), 2158–2171. https://doi.org/10.1016/j.chb.2007.10.005

Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human-Computer Interaction, 25*(1), 1–24. https://doi.org/10.1080/07370020903586662

Joly, Y., Dyke, S. O., Knoppers, B. M., & Pastinen, T. (2016). Are data sharing and privacy protection mutually exclusive? *Cell, 167*(5), 1150–1154. https://doi.org/10.1016/j.cell.2016.11.004

Journal of Communication. (2022). Author guidelines https://academic.oup.com/joc/pages/General_Instructions#Research%20Data%20Policy.

Journals, O. (2021). *Research data policy*. https://academic.oup.com/journals/pages/authors/preparing_your_manuscript/research-data-policy#data2.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389–402. https://doi.org/10.1177/2515245918787489

Lee, S. S. J., Cho, M. K., Kraft, S. A., Varsava, N., Gillespie, K., Ormond, K. E., Wilfond, B. S., & Magnus, D. (2019). I don't want to be Henrietta Lacks": Diverse patient perspectives on donating biospecimens for precision medicine research. *Genetics in Medicine, 21*(1), 107–113. https://doi.org/10.1038/s41436-018-0032-6

Levenstein, M. C., & Lyle, J. A. (2018). Data: Sharing is caring. *Advances in Methods and Practices in Psychological Science, 1*, 95–103. https://doi.org/10.1177/2515245918758319

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Liu, B., & Sundar, S. S. (2018). Microworkers as research participants: Does underpaying Turkers lead to cognitive dissonance? *Computers in Human Behavior, 88*, 61–69. https://doi.org/10.1016/j.chb.2018.06.017

McLaren, R. M., & Solomon, D. H. (2010). Appraisal and distancing responses to hurtful messages II: A diary study of dating partners and friends. *Communication Research Reports, 27*(3), 193–206. https://doi.org/10.1080/08824096.2010.496325

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science, 1*, 131–141. https://doi.org/10.1177/2515245917747656

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26*(4), 323–339. https://doi.org/10.1086/209566

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., … Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Noland, C. M. (2012). Institutional barriers to research on sensitive topics: Case of sex communication research among university students. *Journal of Research Practice, 8*(1). Article M2 http://jrp.icaap.org/index.php/jrp/article/view/332/2621.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716

Pearce, J. (2020). Political regime and the reproduction of violence and criminality in Latin America: An interdisciplinary conversation. *Latin American Research Review, 55* (4), 859–868. https://doi.org/10.25222/larr.1288

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8*(3), 162–166. https://doi.org/10.1111/j.1467-9280.1997.tb00403.x

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. SUNY Press.

Pickard, M. D., Wilson, D., & Roster, C. A. (2018). Development and application of a self-report measure for assessing sensitive information disclosures across multiple modes. *Behavior Research Methods, 50*(4), 1734–1748. https://doi.org/10.3758/s13428-017-0953-z

Psychological Science. (2022). *Psychological Science submission guidelines*. https://www.psychologicalscience.org/publications/psychological_science/ps-submissions#TRAN.

Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*(1), 1–9. https://doi.org/10.1038/s41467-019-10933-3

Shaffer, D. R., & Tomarelli, M. M. (1989). When public and private self-foci clash: Self-consciousness and self-disclosure reciprocity during the acquaintance process. *Journal of Personality and Social Psychology, 56*(5), 765–776. https://doi.org/10.1037/0022-3514.56.5.765

Solomon, D. H., Knobloch, L. K., Theiss, J. A., & McLaren, R. M. (2016). Relational turbulence theory: Explaining variation in subjective experiences and communication within romantic relationships. *Human Communication Research, 42* (4), 507–532. https://doi.org/10.1111/hcre.12091

Sweeney, L., von Loewenfeldt, M., & Perry, M. (2018). *Saying it's anonymous doesn't make it so: Re-Identifications of "anonymized" law school data*. Technology Science. article 2018111301 https://techscience.org/a/2018111301/.

Weisband, S., & Kiesler, S. (1996). Self-disclosure on computer forms: Meta-analysis and implications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3–10). April https://doi.org/10.1145/238386.238387.

Zimmer, M. (2010). But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology, 12*(4), 313–325. https://doi.org/10.1007/s10676-010-9227-5