scientific **data**

OPEN

ARTICLE

Check for updates

Estimating social bias in data sharing behaviours: an open science experiment

Claudia Acciai [1 ✉], Jesper W. Schneider [2] & Mathias W. Nielsen [1 ✉]

Open data sharing is critical for scientific progress. Yet, many authors refrain from sharing scientific data, even when they have promised to do so. Through a preregistered, randomized audit experiment (N = 1,634), we tested possible ethnic, gender and status-related bias in scientists' data-sharing willingness. 814 (54%) authors of papers where data were indicated to be 'available upon request' responded to our data requests, and 226 (14%) either shared or indicated willingness to share all or some data. While our preregistered hypotheses regarding bias in data-sharing willingness were not confirmed, we observed systematically lower response rates for data requests made by putatively Chinese treatments compared to putatively Anglo-Saxon treatments. Further analysis indicated a theoretically plausible heterogeneity in the causal effect of ethnicity on data-sharing. In interaction analyses, we found indications of lower responsiveness and data-sharing willingness towards male but not female data requestors with Chinese names. These disparities, which likely arise from stereotypic beliefs about male Chinese requestors' trustworthiness and deservingness, impede scientific progress by preventing the free circulation of knowledge.

Introduction

Scientific discoveries are made through cumulative and collective efforts, ideally based on full and open communication. For science to work, published claims must be subject to organized skepticism[1]. Yet, science's ethos of rigorous, structured scrutiny is contingent on data sharing. Lack of data prevents results from being reexamined with new techniques, and samples from being pooled for meta-analysis[2,3]. This ultimately hinders the cumulative knowledge-building that drives scientific progress[4].

Open data improves the credibility of scientific claims, and while journal editors increasingly acknowledge the importance of disclosing data[5,6], many authors refrain from sharing their data, even when they have promised to do so[7–13].

Previous research has focused on the supply-side determinants of data-sharing. Surveys find that scientists' decisions to share data depend on (i) contextual factors such as journal requirements, funding incentives and disciplinary competition, (ii) individual factors such as perceived risks of data misuse, lost publication opportunities, and efforts associated with making data available, and (iii) demographic factors such as experience level, tenure-status and gender[7,14–19].

Much less attention is given to the demand-side issues of data sharing. Ideally, everyone, irrespective of background, should be able to contribute to science[1]. As such, data access should not differ depending on who is asking for the data. Yet, research indicates persistent gender, ethnic and status-related bias in science[20–27] that likely also affects data-sharing practices.

Social bias in data-sharing may arise from scientists' stereotypic beliefs about data requestors. According to status characteristics theory, nationality, ethnicity, gender and institution prestige are diffuse cues that, when salient, may influence scientists' impressions of requestors' trustworthiness, competence or deservingness[28,29]. Such status cues are more likely to guide people's judgments in ambiguous situations, where information is scarce[30,31]. Further, status cues may be critical for data sharing, as knowledge transfer is generally more likely in high-trust situations, here including the potential data sharer's trust in the requestor's competences and intentions[32,33].

[1]Department of Sociology, University of Copenhagen, Øster Farimagsgade 5, 1353, Copenhagen, Denmark. [2]Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Bartholins Allé 7, 8000, Aarhus C, Denmark. ✉e-mail: cla@soc.ku.dk; mwn@soc.ku.dk

We conducted a pre-registered (https://osf.io/378qd), randomized audit experiment to examine possible ethnic, gender and status bias in data sharing. We requested data from authors of papers published in widely recognized journals that commit to the FAIR data principles (*the Proceedings of the National Academy of Sciences* [PNAS] and *Nature*-portfolio outlets), where data were indicated to be available upon request. We varied the identity of the fictitious data requestor on four factors: (i) country of residence (China vs. United States [US]), (ii) institution prestige (high status vs. lower status university), (iii) ethnicity (putatively Chinese vs. putatively Anglo-Saxon), and (iv) gender (masculine-coded vs. feminine-coded name).

Motivated by evidence of gender, ethnic and status bias in trust games[34,35], in correspondence tests of employers, legislators, educators and citizens[27,36–39], and in field- and survey experiments of peer evaluation[22,40,41], we hypothesized that scientists would be less willing to share data, when a requestor (i) was from China (compared to the US); (ii) was affiliated with a lower-status university (compared to a higher status university); (iii) had a Chinese-sounding name (compared to a typical Anglo-Saxon name); and (iv) had a feminine-coded name (compared to a masculine-coded name).

In addition to gender and institution status, which have previously been covered in correspondence tests of scientists[42,43], we were interested in the specific disadvantages facing researchers with Chinese names and university affiliations. China is currently the world's largest producer of scientific outputs[44] and Chinese expatriates by far outnumber any other group of foreign graduate students at US universities[45,46]. Considering these figures, a study of the possible bias facing Chinese nationals and descendants in a globalized science system seems timely.

Materials and Methods

Sampling.   Our data collection included four steps, summarized in Figure S1. The experimental population consisted of authors of scientific papers published in PNAS and Nature-portfolio journals (between 2017 and 2021), where data were indicated to be 'available upon request'. We queried journal websites to identify all peer-reviewed papers that included the text-string "available upon request". This resulted in an initial sample of 6,798 papers. C.A. and M.W.N. manually checked and coded the data availability section of each paper to identify cases where data were unambiguously made available upon request (Table S1). If the primary author contact listed in a data statement occurred multiple times in our sample (due to >1 publications), we only included the author's most recent publication. We removed all retracted papers and sources with corrections. We matched the authors listed as primary data contacts in the papers with bibliographic metadata from Clarivate's Web of Science (WoS) to retrieve up-to-date information on emails and affiliations (Figure S2). Our final sample consists of 1,826 author-paper pairs. Due to bounced emails and authors withdrawing from the study after receiving a debriefing statement (in total 78 authors decided to withdraw), our analysis sample is further reduced to 1,634 author-paper pairs. According to our registered power analysis, this sample size should be sufficient for detecting a small effect size of Cohen's $f^2 = 0.02$, with $\alpha = 0.01$ and a power of 0.95.

Procedures.   Efforts to measure ethnic, gender and status bias are complicated by observer effects and issues of social desirability[47]. Audit studies can mitigate such biases by allowing the experimenter to estimate participants' responses to the treatment in a realistic setting[47–49].

Our unpaired audit study randomized participants across twelve treatment conditions (Figure S3). Given our four treatments, a 4×4 factorial design would be the typical set-up for our study. Yet, to keep the treatments realistic, we decided not to include a putatively Anglo-Saxon (male or female) treatment associated with a (higher or lower status) Chinese University (during the COVID-19 pandemic, the number of international students in China decreased enormously). Thus, we adopted 12, instead of 16, treatment conditions. Previous research indicates that data-sharing practices differ depending on the author's gender and disciplinary field[11,13,50]. Hence, we block-randomized the sample population according to scientific field and gender. Inspired by previous audits on discrimination in science[42,43], our data-sharing requests were emailed from fictitious "about-to-become" PhD students. We created Gmail accounts for all of the gender-ethnicity combinations (in total 4 email addresses). In the emails, the fictitious data requestors asked the participants to share data related to a specific publication (Figure S4). If participants did not respond to our initial email, we sent out a follow-up request after one week. Data collection was completed two weeks after the follow-up request. All data were collected in April and May 2022. Email correspondences were managed through YAMM (https://yamm.com/), a private email service provider (google sheet add-on for Gmail and Google Workspace users). This tool allowed us to track the email delivery metrics and provided information on whether an email had been received (or bounced) and opened (or left unopened).

Our sampling and analysis plan was preregistered at the Open Science Framework. We have followed all steps presented in the registered plan, with one minor deviation. In the preregistration we planned to run two linear probability models with cluster-robust standard errors at the field level, while in the results section, we report the outcomes of these models without the cluster-robust standard errors. Cluster robust standard errors proved unsuitable given the low number of clusters and because randomized treatments were assigned at the individual level as opposed to the group-level[51].

Manipulations.   Our treatment conditions varied on the following four factors: gender, ethnicity, country of residence and institutional affiliation. We used first and last names to signal the fictitious requestors' gender and ethnicity (Figure S4). Emails from putatively Chinese treatments included a masculine- or feminine-coded Anglo-Saxon middle name in a parenthesis to signal the requestor's gender (e.g., 'Yadan (Cecilia) Xing'). This is a widely used name-practice among transnational Chinese students[52]. We created four email addresses for our treatments. All accounts were opened and used for at least 90 days before the experiment started. Warm-up activities of email domains (e.g., sending small volumes of emails and slowly increasing) helped build a positive

sender reputation for the accounts. Such standard email activity from the newly created account can reduce the risk of an account being labeled as a fake, hence reducing the email's likelihood of being filtered as spam. To select relevant names, we relied on a historical dataset of Olympic athletes[53]. We limited our focus to Chinese and American athletes participating in the Olympic Games between 1932 and 2016. For each country treatment (US and China), we randomly selected separate first and last names until finding appropriate combinations. For the putatively Anglo-Saxon requestor conditions, we used the following names: Jeffrey Killion and Hilary Witmer. For the putatively Chinese requestor conditions, we used the following names: 张嘉实 (Jia Shi Zhang) and 邢雅丹 (Yadan Xing). We used the R package "rethnicity"[54] to ensure that the selected Anglo-Saxon first names were typical Anglo-Saxon names and that the selected Chinese first names were well-known Asian names. In addition, we manually verified the rethnicity estimates by looking up the first names on Linkedin. To ensure that participants perceived the Chinese names as Chinese, we wrote the signature names (at the bottom of each email) in both English letters and Chinese characters ('Yadan (Cecilia) Xing|邢雅丹' and 'Jiashi (Wilson) Zhang|张嘉实') (see also Figure S4). According to the US Census of Frequently Occurring Surnames[55], 96% of individuals with the surname "Witmer" identified as white, 87% of individuals with the surname Killion identified as white, 97% of individuals with the surname Xing identified as Asian, while 98% with the surname Zhang identified as Asian. Previous research suggests that some race and ethnicity-based manipulations also include class primes[56]. In our design, we followed Block *et al.*[36] and attempted to reduce the potential effect of socio-economic status by holding occupation constant. Specifically, all participants received an email from an 'about to become' Ph.D. student at a Chinese or US research institution.

Emails from putatively Chinese treatments (located in the US or China) that targeted authors outside of China were written in English. Emails from Chinese nationals that targeted authors located in China were written in Mandarin. We used four international university rankings to identify appropriate university affiliations that varied in institution status (high status [Carnegie Mellon and Zhejiang University] vs. lower status [Baylor University and Chongqing University]) (SI *Appendix*, Table S2). This manipulation also signaled the fictitious requestors' country of residence (China or the US). Appropriate university affiliations were identified using a combination of different university rankings (Times Higher Education, Shanghai national and international rankings, QS ranking, and the PP-top 10% indicator from the Leiden Ranking). To select high status affiliations, we focused on universities that scored consistently high across the rankings. When selecting lower-status affiliations, we focused on universities that scored consistently low across rankings. We decided to exclude top-ranked (top 10%) and bottom-ranked universities (bottom 10%) to lower the likelihood that participants would discern the purpose of the experiment. In addition, we restricted our selection to universities with multiple faculties and active Ph.D. programs within each faculty.

Measures. We measured the experimental treatments using four dichotomous variables: country (US = 0, China = 1), ethnicity (Anglo-Saxon name = 0, Chinese name = 1), institution status (high status university = 0, lower status university = 1), and gender (masculine coded name = 0 and feminine coded name = 1). In addition, all statistical models included five dichotomous variables to adjust for scientific field and publication outlet (Table S3).

Our preregistered dependent variable is a binary measure of data-sharing willingness. This measure is based on a systematic coding of participants' email responses (respondents did not share data nor indicate willingness to share data = 0, respondents shared data or indicated willingness to share data = 1). Two authors systematically coded all responses. The codebook (Table S4) was first tested by coding 10% of the sample. The pilot phase was repeated, and the codebook was further adjusted, until coding reliability measures reached a satisfactory level (Kappa coefficient ≥ 0.8). In the manual coding, the coders were blinded to the treatments.

As a second outcome (not registered), we measured whether participants responded to our data requests or not (non-response = 0, response = 1). This outcome measure is widely used in previous email-based correspondence tests[57]. As opposed to data-sharing willingness, our measure of email responses did not involve any coding of textual content and can thus be viewed as a more objective indicator of bias in data sharing.

In accordance with our registered analysis plan, we measured and reported the dependent variables in two ways. In one case, we excluded all unopened emails from the analysis. In the results section, we refer to this sample as the sample of "opened emails". In the other case, we included unopened emails and coded them as indicating unwillingness to share data or non-responses. In the results section, we refer to this sample as the "full sample". Given that all participants in the sample of opened emails have been directly exposed to a treatment, we would expect the treatment effects to be larger for this sample. In contrast, given that some participants in the full sample were not directly exposed to a treatment, we would expect the noise-to-signal ratio to be larger and the treatment effects to be smaller for this sample. Thus, while the sample of opened emails gets us closer to the direct effect of the treatments, the full sample gives us a better sense of the real-world disadvantages associated with a given treatment. Due to a minor error in our email management, responses to one of the treatment emails were associated with an alias during the first wave of data collection. Specifically, participant responses to emails sent from Yadan Xing were forwarded to a different alias address with a similar manipulation condition. Only two participants noticed this issue in their responses, but nonetheless positively engaged with our request. The rest of the respondents exposed to this error either responded to the alias email while addressing their message to Yadan (e.g., "Dear Yadan") or responded directly to the correct email manipulation. Based on this evidence, we find it reasonable to assume that the vast majority of respondents exposed to the error either did not notice the issue or perceived it to be irrelevant. This mistake was corrected in the follow-up email addressed to all recipients who did not respond to the first email. Table S5 presents the response rates across treatments for the first and the second wave. The treatment including the alias mistake (Yadan Xing) had the highest response rate of the four treatment emails after wave one. Note also that our main findings concerning a bias in responsiveness
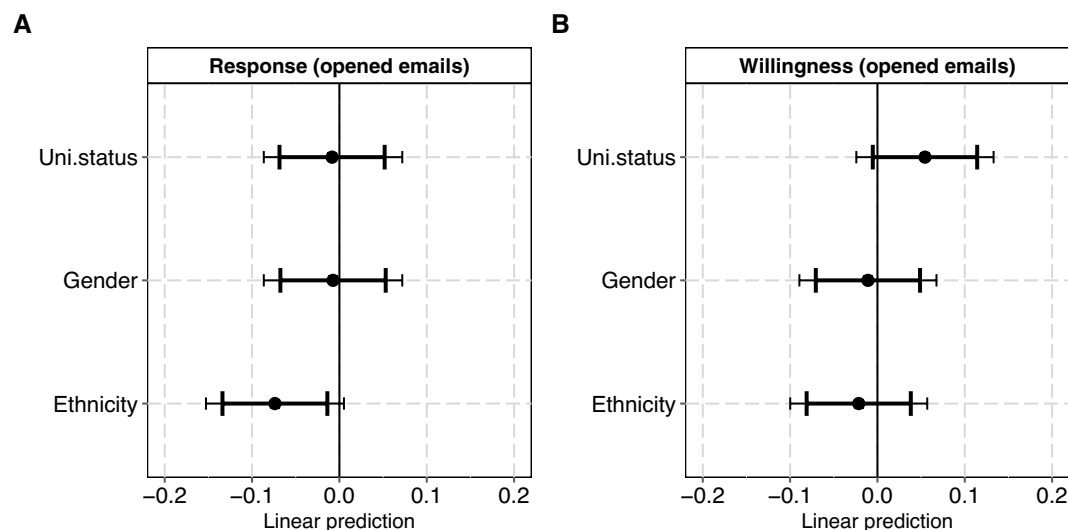
Fig. 1 Plots of coefficients from two linear probability models with "Response" (Panel A) and "Willingness" (Panel B) as outcomes. The sample (N: 770) covers all participants (that have opened the treatment email[s]) exposed to treatments with US affiliations. The panels plot the fixed coefficients for the three main predictors University status (high = 0, low = 1), Gender (masculine-coded name = 0 feminine-coded name = 1) and Ethnicity (typical Anglo-Saxon name = 0, Chinese-sounding name = 1). Error bars represent 95% and 99%. Both models adjust for scientific field and publisher. For model specifications, see Supplementary Tables S6, S7.

and data-sharing willingness (Figs. 1,4) pertain to the male Chinese treatment (Jiashi Zhang) and not the female Chinese treatment (Yadan Xing).

Statistical analyses. Given that no participants in our study were exposed to the putatively Anglo-Saxon treatments located in China, we estimated two groups of linear probability models. In one group, we estimated the direct effects of gender, ethnicity and institution prestige on data sharing and email responses among participants exposed to treatments affiliated with US universities (Fig. 1 and S5). In another group, we estimated the direct effects of gender, country-location, and institution prestige on data-sharing willingness among putatively Chinese treatments located in the US and China (Fig. 2 and S6). Despite the unidirectional nature of our hypotheses, we report all outcomes with two-directional 95% and 99% confidence intervals.

All analyses were conducted in R[58], and we used the estimator package[59] to perform the linear probability models.

Ethics. Since the social biases examined in our study may operate subconsciously, a pre-experimental consent process could damage the validity of the experiment. For this reason, we decided to operate with ex-post consent and information disclosure. Our study was approved by the Ethics Review Board at the Department of Sociology, University of Copenhagen (UCPH-SAMF-SOC-2022-03). At the end of the experiment, a debriefing email was sent to all participants (non-respondents as well as respondents). In this debriefing, we explained the general purpose of the study, its experimental manipulations, the potential risks and benefits to participants, and the principles of data management and anonymization. Moreover, we informed participants about their right to withdraw from the study without penalty. In total, 78 authors (5%) decided to withdraw from our study after receiving the debriefing statement.

Results

Eight-hundred-and-eighty-four scientists responded to our data requests, and 226 either shared or indicated willingness to share all or some of their data. This corresponds to 54% (884 of 1634) and 14% (226 of 1634) of the full sample, and 75% (884 of 1179) and 20% (226 of 1179) of the sample of opened emails.

In Fig. 1, we estimate how institution prestige, gender and ethnicity influence participants' responsiveness and data-sharing willingness in the sub-sample (that opened emails) exposed to treatments with US affiliations. As shown in the figure (Panel A), neither university status nor gender affected participants' likelihood of responding to data requests from treatments with US affiliations. Yet, treatments with Chinese-sounding names were 7 percentage points less likely to receive a response than treatments with putatively Anglo-Saxon names ($\beta = -0.07$, 95% CI: -0.13:-0.01; 99% CI: -0.15:0.01). This corresponds to an odds ratio of 0.66, or 34% lower odds of obtaining a response for the treatments with Chinese-sounding names compared to putatively Anglo-Saxon treatments (Table S16). Results are similar, though associated with larger uncertainties, when estimates are based on the full sample as opposed to the sample of opened emails (Figure S5).

A

B
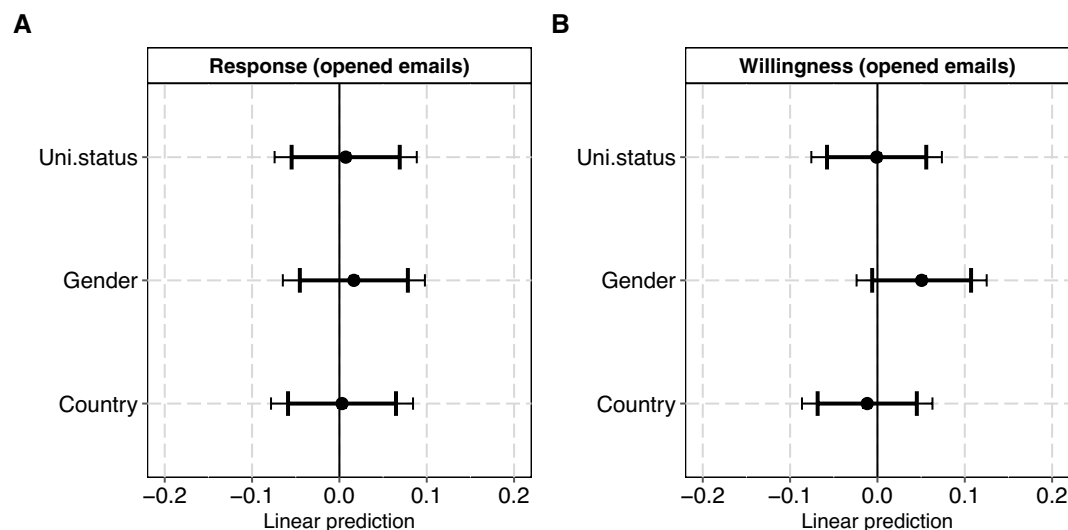
Fig. 2 Plots of coefficients from two linear probability models with "Response" (Panel A) and "Willingness" (Panel B) as outcomes. The sample (N: 802) covers all participants (that have opened the treatment email[s]) exposed to treatments with Chinese-sounding names located in the US and China. The panels plot the fixed coefficients for the three main predictors University status (high = 0, low = 1), Gender (masculine-coded name = 0 feminine-coded name = 1) and Country (US university = 0, Chinese university = 1). Error bars represent 95% and 99%. Both models adjust for scientific field and publisher. For model specifications, see Supplementary Tables S10, S11.
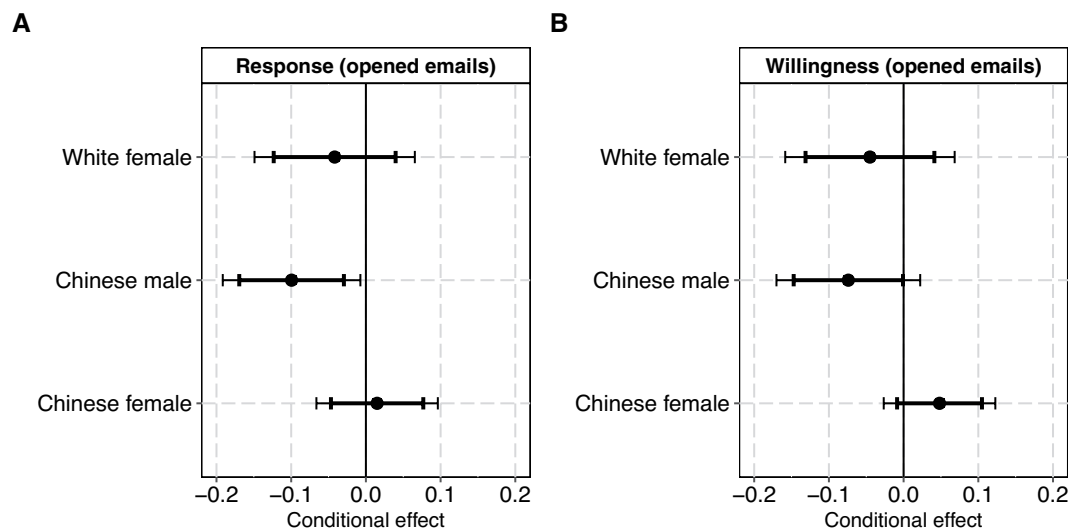
A
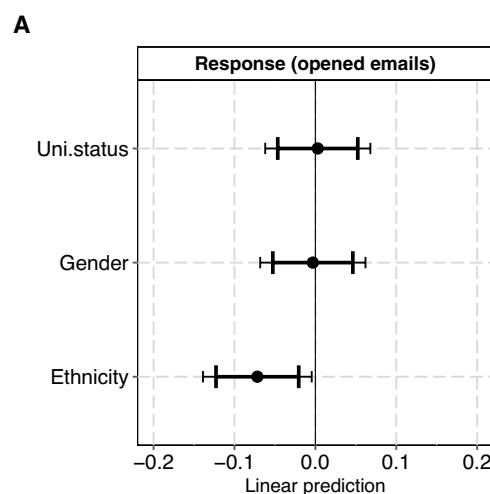
Fig. 3 Plots of coefficients from a linear probability model with "Response" as outcome. The sample (N: 1179) covers all participants that have opened the treatment email(s). The panel plots the fixed coefficients for the three main predictors University status (high = 0, low = 1), Gender (masculine-coded name = 0 feminine-coded name = 1) and Ethnicity (Typical Anglo-Saxon name = 0, Chinese-sounding name). Error bars represent 95% and 95%. The model adjusts for scientific field and publisher. For model specifications, see Supplementary Table S14.

Our pre-registered analysis of ethnic, gender and status-related bias in scientists' willingness to share data with treatments located in the US proved inconclusive in both the full sample (Figure S5, Panel B) and in the sample of individuals that opened emails (Fig. 1, Panel B). In conflict with one of our hypotheses, participants seemed slightly more willing to share data, when the request came from a lower status US institution (Baylor University) compared to a higher status US institution (Carnegie Mellon), but the confidence intervals for this effect all include zero ($\beta = 0.05$, 95% CI: $-0.01:0.11$; 99% CI: $-0.02:0.13$).

In Fig. 2, we estimate how university status, gender and country matter for participants' responsiveness and data-sharing willingness in the sub-sample exposed to treatments with Chinese-sounding names, located in China vs. the US. As shown in Fig. 2 (Panel A), we do not find any association between the treatment conditions and participants' responsiveness to data-requests in this subsample. All coefficients are close to zero, indicating

Fig. 4 Conditional coefficients derived from interactions between Ethnicity and Gender in two linear probability models with Response (Panel A) and Willingness (Panel B) as outcomes. The sample (N: 1179) covers all participants who have opened the treatment email(s). The panels plot conditional coefficients of the ethnicity variable by treatment gender. The reference group is the putatively Anglo-Saxon male treatment. Errors bars represent 95% and 99% confidence intervals. For model specifications, see Supplementary Tables S22, S23.

no discernible effects. Similarly, for data-sharing willingness (Fig. 2, Panel B), results are weak and inconclusive. The effects of university status and country on data-sharing willingness are close to zero, albeit women, not men (contrasting our pre-registered hypothesis), are met with slightly higher data sharing willingness in this sub-sample. Yet, the confidence intervals for this gender effect include zero ($\beta = 0.05$, 95% CI: -0.01:0.11; 99% CI: -0.02:0.13). Again, the results are similar, when models are based on the full sample as opposed to the sample of opened emails (Figure S6).

In Fig. 3, we explore if the ethnicity bias indicated in Fig. 1 (Panel A) is specific to the US treatments with Chinese-sounding names, or whether it affects putatively Chinese data requestors more generally. In this analysis, which covers both the US-located and China-located treatments, we obtain results comparable to those reported in Fig. 1 (Panel A). The effects for university status and gender remain inconclusive, but treatments with Chinese-sounding names have a 7 percentage points lower likelihood of receiving a response than treatments with typically Anglo-Saxon names ($\beta = -0.07$, 95% CI: -0.12:-0.02; 99% CI: -0.14:-0.00). This corresponds to an odds ratio of 0.67, or 33% lower odds of response for putatively Chinese treatments compared to putatively Anglo-Saxon treatments (Table S20). This estimated effect is smaller and associated with larger uncertainties when the analysis is based on the full sample compared to the sample of opened emails (Figure S7).

Given the indication in Fig. 2 (Panel B) that data-sharing willingness is lower for men with Chinese names compared to women with Chinese names, we also estimated the conditional effects of ethnicity on data-sharing behaviors for male and female treatments. To secure the highest possible statistical power, we ran interaction-analyses on the combined samples of US-located and China-located treatments. Figure 4 plots the conditional coefficients from four interactions between Ethnicity and Gender in the sample of authors that opened emails. Male treatments with Chinese-sounding names (compared to male treatments with putatively Anglo-Saxon names) face consistent disadvantages both with respect to responsiveness ($\beta = -0.10$, 95% CI: -0.17:-0.03; 99% CI: -0.19:-0.01) and willingness ($\beta = -0.07$ 95% CI: -0.15:-0.00; 99% CI: -0.17:0.02) when requesting data, while this is not the case for female treatments with Chinese-sounding names. We obtain comparable results when the conditional coefficients are estimated based on the full sample (Figure S8).

Discussion

Previous research on social bias in data sharing is scarce. Several studies document low response rates and low data-sharing willingness among scientists who agreed to make their data available upon request[7,9,11,12,15]. Yet, demand-side determinants of data sharing remain largely unexamined.

In a small-scale field experiment (N = 200), Krawczyk and Reuben[60] tested whether economists' willingness to share supplementary materials differed depending on the prestige of the requestor's university and position level (Columbia University vs. University of Warsaw) and found negligible effects.

In our audit experiment, which draws on a larger multi-disciplinary sample of participants, none of the four preregistered hypotheses predicting national, ethnic, gender and institutional prestige bias in data-sharing willingness were supported.

Yet, tests for differences in scientists' responsiveness to data requests (our most objective measure of disparities in data-sharing) indicated lower response rates for Chinese treatments compared to Anglo-Saxon treatments. This may indicate that ethnic bias is more likely to occur at the initial stage of a (potential) data exchange

when scientists make rapid and unreflective judgments on whether to engage with a requestor or not. Indeed, previous research into ethnic bias in pro-sociality also emphasizes the role of implicit attitudes (activated quickly and spontaneously) in discriminatory behaviors[61,62].

Contrary to our expectations, scientists exposed to US-located treatments seemed slightly less willing to share data with requestors from prestigious universities compared to requestors from lower status universities (although the 95% confidence bounds spanned zero for this estimate). One possible explanation for this may be that the perceived career risks associated with data sharing (in terms of lost publication opportunities and low-ered competitive advantages), on average, are higher, when requests are made from prestigious US universities compared to lower-status US universities or Chinese universities. Indeed, previous research finds that scientists' data-sharing willingness tends to be lower when perceived competition is high[7,50,63,64]. Such risks and concerns could potentially be reduced through the use of data licensing on public repositories like OSF.

Importantly, the negative prestige effect was only salient for scientists exposed to treatments from US univer-sities. This may be because the participants in our sample, which primarily reside in Europe or North America (Table S26), are more familiar with the prestige hierarchy of US institutions and less knowledgeable about the relative standing of Chinese institutions.

Also contrary to our predictions, Chinese treatments with feminine-coded names were met with slightly higher data-sharing willingness than Chinese treatments with masculine-coded names (although the 95% confidence bounds spanned zero for this estimate). Importantly, this finding reflects an underlying pattern of male-specific ethnic discrimination. Conditional effects derived from interaction analyses suggested a clear bias in responsiveness and data-sharing willingness against male Chinese treatments, while results for female Chinese treatments were inconclusive.

Given the "double-burden hypothesis" in intersectional theory[65], which states that minority women are most likely to face discrimination, these findings may seem counter-intuitive. Nevertheless, they have some bearing in previous studies on trust and discrimination. Indeed, evidence suggests that women are stereotypically seen as more trustworthy than men[66], and studies on helping behavior also indicate a greater taste for helping women than men in various social situations[67–70]. Further, while field experiments have generally neglected intersec-tional perspectives on ethnic and gender discrimination[71], studies that *do* cover this aspect typically find that minority males are subject to larger ethnic penalties than minority females in job-markets, housing markets, and sharing economies[72–79]. Building on research on gender and nationality stereotypes[80], Arai and colleagues theo-rize that when stereotypes against specific ethnic groups are negative, they are more likely to disadvantage men than women, because it is men who are primarily presumed to embody these stereotypes[72]. Additional research is required to determine the root causes of the observed bias towards Chinese men. However, we hypothesize that it likely arises from stereotypic beliefs about the group's trustworthiness and deservingness in data exchange relationships. Such beliefs may have been particularly salient during 2022, when we collected our data, due to recurring discussions about China's alleged intellectual property theft in the US and Europe[81], but also in the wake of COVID-19, where prejudice and discriminatory intent against Asians aggravated[82,83].

While our field-experiment has clear advantages over survey-based approaches to measuring data-sharing behaviors, it is not without limitations. Our data requests were sent from Gmail accounts, and this may have increased the likelihood that emails ended up in the recipients' spam filters. Further, some recipients may have found our data requests more suspicious than they would have, if the same emails were sent via institutional accounts. We have attempted to account for this issue by tracking opened and unopened emails throughout the study-period and by reporting results for samples that include and exclude unopened emails (SI *Appendix*).

Compared to previous correspondence studies, where data-requests were made from institutional email accounts, the response rate in our study is quite high. Participants ignored 49% of the requests made by Tedersoo and colleagues[11] concerning data for recent papers in *Nature* and *Science*; while 86% of the data requests made by Gabelica and colleagues to authors publishing in BioMed Central journals were also ignored[12]. In comparison 54% of scientists in our sample responded to the data requests. This suggests that the drawbacks of using Gmail compared to an institutional account have been small.

Another limitation of our study design concerns the generic nature of our data request, which may have increased the level of suspicion among some recipients. After the study was completed, a few authors approached us with concerns that the email request lacked detail and was not tailored to the specific practices of their dis-cipline, which made them hesitant to respond. While we acknowledge this critique, experiments like ours will always be subject to trade-offs between ecological validity and treatment bias. In this case, we decided to keep the emails generic to hold constant all other components than our manipulations.

A fourth limitation concerns our sampling strategy. Because we only targeted authors of papers in Nature portfolio journals and PNAS, our results are limited to scientists publishing in these journals. In the future, researchers should examine whether data sharing behaviors differ for authors publishing in journals that are less committed to open science and the FAIR data-sharing principles. Finally, given that none of our four preregis-tered hypotheses were directly confirmed, our results concerning gender-specific ethnic discrimination in data sharing can only be seen as suggestive.

Despite these limitations, our paper offers important new insights on scientific data-sharing practices in sci-ence. Compared to unregistered experiments, our preregistered analysis has the advantage of providing a clear record of what ideas our study was designed to evaluate, how we planned to examine them, and how our most notable finding of ethnic bias in data-sharing relate to these ideas[84]. Put differently, the preregistered analysis plan has limited our degrees of freedom as researchers and thereby increased the validity and reliability of our study.

Our paper has important implications for open science policies. Despite clearly indicating intent to make their data available upon request, only around half of the targeted authors responded to our data requests, and only 14% indicated willingness to share all, or some, of their data. While some participants may have had good

reasons not to share, this behavior conflicts with the FAIR principles adopted by PNAS and Nature portfolio journals, hence demonstrating the drawbacks of enabling researchers to make data available upon request. Our study further complicates this issue by exposing potential inequalities in who can benefit from data-sharing, when disclosure decisions are left to the discretion of individual scientists[7]. In accordance with previous work, our study shows that data requests often require more than trivial efforts from the side of the requestor. These efforts could be reduced if funders and publishers required authors to release all relevant data, whenever possible[11,12].

Unfortunately, the reality is that most journals do not incentivize data sharing. In a review of editorial policies for 318 biomedical journals[5], only 12% explicitly required data sharing as a condition for publication, 9% required data sharing without stating it as a condition for publication, while around one third of the journal sample did not mention data-sharing at all. Under such conditions, we expect that "data availability upon request" will remain a widespread practice in many disciplines.

Importantly, disclosure is sometimes also challenged by practical issues (e.g., data size and propriety rights) or ethical issues (e.g., sensitive information on human subjects), and publishers could do more to help mitigate these challenges. From our experience, it seems that many authors that cannot share their data for practical or ethical reasons currently opt to indicate data availability upon request to circumvent a journal's data-sharing requirements. Assistance from the side of publishers in providing the necessary storage space or easy-to-use methods for making synthetic datasets for sensitive populations, could help mitigate these problems[85].

In summary, our field experiment extends on research about scientists' compliance with open data principles by indicating that sharing behaviors may differ depending on who is asking for the data. These disparities, which likely arise from stereotypic beliefs about specific requestors' trustworthiness and deservingness, hamper the core principle of mutual accountability in science and impede scientific progress, by preventing the free circulation of knowledge.

Data availability
Data available at this link: osf.io/kzrc4[86]

Code availability
Code can be accessed at this link: osf.io/kzrc4

References
1. Merton, R. K. The normative structure of science. in *In Norman W. Storer (Ed.), The sociology of science: Theoretical and empirical investigations*. 267–278 (The University of Chicago Press., 1942).
2. Nosek, B. A. & Bar-Anan, Y. Scientific Utopia: I. Opening Scientific Communication. *Psychol. Inq.* **23**, 217–243 (2012).
3. Murray-Rust, P. Open Data in Science. *Nat. Preced.* https://doi.org/10.1038/npre.2008.1526.1 (2008).
4. Bird, A. What Is Scientific Progress? *Noûs* **41**, 64–89 (2007).
5. Vasilevsky, N. A., Minnier, J., Haendel, M. A. & Champieux, R. E. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* **5**, e3208 (2017).
6. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
7. Andreoli-Versbach, P. & Mueller-Langer, F. Open access to data: An ideal professed but not practised. *Res. Policy* **43** (2014).
8. Federer, L. M. *et al.* Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLoS One* **13**, e0194768 (2018).
9. Savage, C. J. & Vickers, A. J. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* **4** (2009).
10. Roche, D. G. *et al.* Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution. *Proc. R. Soc. B Biol. Sci.* **289**, 20212780 (2022).
11. Tedersoo, L. *et al.* Data sharing practices and data availability upon request differ across scientific disciplines. *Sci. Data 2021 81* **8**, 1–11 (2021).
12. Gabelica, M., Bojčić, R. & Puljak, L. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *J. Clin. Epidemiol.* **150**, 33–41 (2022).
13. Tenopir, C. *et al.* Data sharing by scientists: Practices and perceptions. *PLoS One* **6**, (2011).
14. Feigenbaum, S. & Levy, D. M. The market for (ir)reproducible econometrics. *Soc. Epistemol.* **7**, 215–232 (1993).
15. Campbell, H. A., Micheli-Campbell, M. A. & Udyawer, V. Early Career Researchers Embrace Data Sharing. *Trends Ecol. Evol.* **34**, 95–98 (2019).
16. Tenopir, C., Christian, L., Allard, S. & Borycz, J. Research Data Sharing: Practices and Attitudes of Geophysicists. *Earth Sp. Sci.* **5**, 891–902 (2018).
17. Stieglitz, S. *et al.* When are researchers willing to share their data? – Impacts of values and uncertainty on open data in academia. *PLoS One* **15**, e0234172 (2020).
18. Houtkoop, B. L. *et al.* Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Adv. Methods Pract. Psychol. Sci.* **1**, 70–85 (2018).
19. Linek, S. B., Fecher, B., Friesike, S. & Hebing, M. Data sharing as social dilemma: Influence of the researcher's personality. *PLoS One* **12**, e0183216 (2017).
20. Weisshaar, K. Publish and Perish? An Assessment of Gender Gaps in Promotion to Tenure in Academia. *Soc. Forces* **96**, 529–560 (2017).
21. Ross, J. S. *et al.* Effect of blinded peer review on abstract acceptance. *J. Am. Med. Assoc.* **295**, 1675–1680 (2006).
22. Tomkins, A., Zhang, M. & Heavlin, W. D. Reviewer bias in single- versus double-blind peer review. *Proc. Natl. Acad. Sci. USA* **114**, 12708–12713 (2017).
23. Krawczyk, M. & Smyk, M. Author's gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. *Eur. Econ. Rev.* **90**, 326–335 (2016).
24. Card, D., DellaVigna, S., Funk, P. & Iriberri, N. Are Referees and Editors in Economics Gender Neutral?*. *Q. J. Econ.* **135**, 269–327 (2020).
25. Peng, H., Teplitskiy, M. & Jurgens, D. Author Mentions in Science News Reveal Wide-Spread Ethnic Bias. *ArXiv Prepr.* **abs/2009.0**, (2020).

26. Peng, H., Lakhani, K. & Teplitskiy, M. Acceptance in Top Journals Shows Large Disparities across Name-inferred Ethnicities. *SocArXiv* https://doi.org/10.31235/osf.io/mjbxg (2021).
27. Milkman, K. L., Akinola, M. & Chugh, D. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *J. Appl. Psychol.* **100**, 1678–1712 (2015).
28. Ridgeway, C. L. Why Status Matters for Inequality. *Am. Sociol. Rev.* **79**, 1–16 (2013).
29. Berger, J., Cohen, B. P. & Zelditch, M. Status Characteristics and Social Interaction. *Am. Sociol. Rev.* **37**, 241–255 (1972).
30. Correll, S. J., Weisshaar, K. R., Wynn, A. T. & Wehner, J. D. Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment. *Am. Sociol. Rev.* **85**, 1022–1050 (2020).
31. Melamed, D. & Savage, S. V. Status, Numbers and Influence. *Soc. Forces* **91**, 1085–1104 (2013).
32. Hsu, M.-H & Chang, C.-M. Examining interpersonal trust as a facilitator and uncertainty as an inhibitor of intra-organisational knowledge sharing. *Inf. Syst. J.* **24**, 119–142 (2014).
33. Rutten, W., Blaas-Franken, J. & Martin, H. The impact of (low) trust on knowledge sharing. *J. Knowl. Manag.* **20**, 199–214 (2016).
34. Fershtman, C. & Gneezy, U. Discrimination in a Segmented Society: An Experimental Approach*. *Q. J. Econ.* **116**, 351–377 (2001).
35. Cettolin, E. & Suetens, S. Return on Trust is Lower for Immigrants. *Econ. J.* **129**, 1992–2009 (2019).
36. Block, R., Crabtree, C., Holbein, J. B. & Monson, J. Q. Are Americans less likely to reply to emails from Black people relative to White people? *Proc. Natl. Acad. Sci. USA* **118**, (2021).
37. Booth, A. L., Leigh, A. & Varganova, E. Does Ethnic Discrimination Vary Across Minority Groups? Evidence from a Field Experiment*. *Oxf. Bull. Econ. Stat.* **74**, 547–573 (2012).
38. Baert, S. Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005 BT - Audit Studies: Behind the Scenes with Theory, Method, and Nuance. in (ed. Gaddis, S. M.) 63–77. https://doi.org/10.1007/978-3-319-71153-9_3 (Springer International Publishing, 2018).
39. Gaddis, S. M. & Ghoshal, R. Searching for a Roommate: A Correspondence Audit Examining Racial/Ethnic and Immigrant Discrimination among Millennials. *Socius* **6**, 2378023120972287 (2020).
40. Ross, J. S. *et al.* Effect of Blinded Peer Review on Abstract Acceptance. *JAMA* **295**, 1675–1680 (2006).
41. Harris, M. *et al.* Explicit bias toward high-income- country research: A randomized, blinded, crossover experiment of English clinicians. *Health Aff.* **36**, 1997–2004 (2017).
42. Milkman, K. L., Akinola, M. & Chugh, D. Temporal Distance and Discrimination: An Audit Study in Academia. *Psychol. Sci.* **23**, 710–717 (2012).
43. Gerhards, J., Hans, S. & Drewski, D. Global inequality in the academic system: effects of national and university symbolic capital on international academic mobility. *High. Educ.* **76**, 669–685 (2018).
44. Tollefson, J. China declared world's largest producer of scientific articles. *Nature* **553**, 390–391 (2018).
45. Brumfiel, G. Chinese students in the US: Taking a stand. *Nature* **438**, 278–280 (2005).
46. Bartlett, T., & Fischer, K. The China Conundrum. *The New York Times. (Retrieved October 2022)* (2011).
47. Pager, D. & Quillian, L. Walking the Talk? What Employers Say Versus What They Do: Am. *Sociol.* **70**, 355–380 (2005).
48. Riach, P. A. & Rich, J. Field Experiments of Discrimination in the Market Place*. *Econ. J.* **112**, F480–F518 (2002).
49. Gaddis, S. M. Understanding the "How" and "Why" Aspects of Racial-Ethnic Discrimination: A Multimethod Approach to Audit Studies: Sociology of Race and Ethnicity **5**, 443–455 (2019).
50. Thursby, J. G., Haeussler, C., Thursby, M. C. & Jiang, L. Prepublication disclosure of scientific results: Norms, competition, and commercial orientation. *Sci. Adv.* **4**, eaar2133 (2022).
51. Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. When should you adjust standard errors for clustering? (*No. w24003*). *Natl. Bur. Econ. Res.* (2022).
52. Diao, W. Between Ethnic and English Names: Name Choice for Transnational Chinese Students in a US Academic Community. *J. Int. Students* **4**, 205–222 (2014).
53. Griffin, R. 120 years of Olympic history: athletes and results. (2022).
54. Xie, F. rethnicity: An R package for predicting ethnicity from names. *SoftwareX* **17**, 100965 (2022).
55. U.S. Census Bureau: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html (2022).
56. Gaddis, S. M. Signaling class: An experiment examining social class perceptions from names used in correspondence audit studies. Preprint at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350739 (2019).
57. Crabtree, C. An Introduction to Conducting Email Audit Studies BT - Audit Studies: Behind the Scenes with Theory, Method, and Nuance. in (ed. Gaddis, S. M.) Ch. 6 (Springer International Publishing, 2018).
58. R Core Team. *R: A language and environment for statistical computing.* (2016).
59. Blair, G., Cooper, J., Coppock, A., Humphreys, M. & Sonnet, L. Estimatr: Fast estimators for design-based inference. *R Packag. version* (2019).
60. Krawczyk, M. & Reuben, E. (Un)Available upon Request: Field Experiment on Researchers' Willingness to Share Supplementary Materials. *Account. Res.* **19**, (2012).
61. Bhati, A. Does Implicit Color Bias Reduce Giving? Learnings from Fundraising Survey Using Implicit Association Test (IAT). *Volunt. Int. J. Volunt. Nonprofit Organ.* **32**, 340–350 (2021).
62. Stepanikova, I., Triplett, J & Simpson, B. Implicit racial bias and prosocial behavior. *Soc. Sci. Res.* **40**, 1186–1195 (2011).
63. Vogeli, C. *et al.* Data Withholding and the Next Generation of Scientists: Results of a National Survey. *Acad. Med.* **81**, (2006).
64. Kim, Y. & Zhang, P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Libr. Inf. Sci. Res.* **37**, 189–200 (2015).
65. Browne, I. & Misra, J. The Intersection of Gender and Race in the Labor Market. *Annu. Rev. Sociol.* **29**, 487–513 (2003).
66. Fiske, S. T., Cuddy, A. J. C., Glick, P. & Xu, J. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* **82**, 878–902 (2002).
67. Eagly, A. H. & Crowley, M. Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychol. Bull.* **100**, 283–308 (1986).
68. Dufwenberg, M. & Muren, A. Generosity, anonymity, gender. *J. Econ. Behav. Organ.* **61**, 42–49 (2006).
69. Weber, M., Koehler, C. & Schnauber-Stockmann, A. Why Should I Help You? Man Up! Bystanders' Gender Stereotypic Perceptions of a Cyberbullying Incident. *Deviant Behav.* **40**, 585–601 (2019).
70. Erlandsson, A. *et al.* Moral preferences in helping dilemmas expressed by matching and forced choice. *Judgm. Decis. Mak.* **15**, 452–475 (2020).
71. Bursell, M. The Multiple Burdens of Foreign-Named Men—Evidence from a Field Experiment on Gendered Ethnic Hiring Discrimination in Sweden. *Eur. Sociol. Rev.* **30**, 399–409 (2014).
72. Arai, M., Bursell, M. & Nekby, L. The Reverse Gender Gap in Ethnic Discrimination: Employer Stereotypes of Men and Women with Arabic Names. *Int. Migr. Rev.* **50**, 385–412 (2016).
73. Carol, S., Eich, D., Keller, M., Steiner, F. & Storz, K. Who can ride along? Discrimination in a German carpooling market. *Popul. Space Place* **25**, e2249 (2019).
74. Dahl, M. & Krog, N. Experimental Evidence of Discrimination in the Labour Market: Intersections between Ethnicity, Gender, and Socio-Economic Status. *Eur. Sociol. Rev.* **34**, 402–417 (2018).
75. Flage, A. Ethnic and gender discrimination in the rental housing market: Evidence from a meta-analysis of correspondence tests, 2006–2017. *J. Hous. Econ.* **41**, 251–273 (2018).

76. Midtbøen, A. H. Discrimination of the Second Generation: Evidence from a Field Experiment in Norway. *J. Int. Migr. Integr.* **17**, 253–272 (2016).
77. Simonovits, B., Shvets, I. & Taylor, H. Discrimination in the sharing economy: evidence from a Hungarian field experiment. *Corvinus J. Sociol. Soc. Policy* **9**, 55–79 (2018).
78. Sidanius, J. & Pratto, F. *Social dominance: An intergroup theory of social hierarchy and oppression. Social dominance: An intergroup theory of social hierarchy and oppression.* (Cambridge University Press, 1999).
79. Ert, E., Fleischer, A. & Magen, N. Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tour. Manag.* **55**, 62–73 (2016).
80. Eagly, A. H. & Kite, M. E. Are stereotypes of nationalities applied to both women and men? *J. Pers. Soc. Psychol.* **53**, 451–462 (1987).
81. Guo, E., Aloe, J., & Hao, K. The US crackdown on Chinese economic espionage is a mess. We have the data to show it. *MIT Technology Review* (2021).
82. Lu, Y., Kaushal, N., Huang, X. & Gaddis, S. M. Priming COVID-19 salience increases prejudice and discriminatory intent against Asians and Hispanics. *Proc. Natl. Acad. Sci.* **118**, e2105125118 (2021).
83. Cao, A., Lindo, J. M. & Zhong, J. *Can Social Media Rhetoric Incite Hate Incidents? Evidence from Trump's" Chinese Virus" Tweets.* (2022).
84. Ryan, T. J. & Krupnikov, Y. Split Feelings: Understanding Implicit and Explicit Political Persuasion. *Am. Polit. Sci. Rev.* **115**, 1424–1441 (2021).
85. Quintana, D. S. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *Elife* **9**, e53275 (2020).
86. Acciai, C., Jesper, W. S. & Mathias, W. N. Estimating social bias in data sharing behaviours: an open science experiment. *Open Science Framework* https://doi.org/10.17605/OSF.IO/PJC9G (2023).

Acknowledgements

Author contributions

C.A. and M.W.N. designed research with input from J.W.S.; C.A. and M.W.N. performed research; J.W.S. contributed supplementary data. C.A. and M.W.N. analyzed data; C.A. and M.W.N. and wrote the paper with input from J.W.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-023-02129-8.

Correspondence and requests for materials should be addressed to C.A. or M.W.N.

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.