nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01092-x

The general fault in our fault lines

Kai Ruggeri[1,2] ✉, Bojana Većkalov[3], Lana Bojanić[4], Thomas L. Andersen[5], Sarah Ashcroft-Jones[6], Nélida Ayacaxli[7], Paula Barea-Arroyo[8], Mari Louise Berge[9], Ludvig D. Bjørndal[10], Aslı Bursalıoğlu[11], Vanessa Bühler[12,13], Martin Čadek[14], Melis Çetinçelik[15], Georgia Clay[16], Anna Cortijos-Bernabeu[17], Kaja Damnjanović[18], Tatianna M. Dugue[19], Maya Esberg[20], Celia Esteban-Serna[21], Ezra N. Felder[22], Maja Friedemann[6], Darianna I. Frontera-Villanueva[23], Patricia Gale[24], Eduardo Garcia-Garzon[25], Sandra J. Geiger[3], Leya George[26], Allegra Girardello[27], Aleksandra Gracheva[28,29], Anastasia Gracheva[7,28], Marquis Guillory[19], Marlene Hecht[30], Katharina Herte[31], Barbora Hubená[32,69], William Ingalls[19], Lea Jakob[32,33], Margo Janssens[34], Hannes Jarke[2], Ondřej Kácha[35], Kalina Nikolova Kalinova[32,69], Ralitsa Karakasheva[32,69], Peggah R. Khorrami[36,37], Žan Lep[38], Samuel Lins[39], Ingvild S. Lofthus[10], Salomé Mamede[39], Silvana Mareva[40], Mafalda F. Mascarenhas[41], Lucy McGill[42], Sara Morales-Izquierdo[43], Bettina Moltrecht[44], Tasja S. Mueller[45], Marzia Musetti[46], Joakim Nelsson[47], Thiago Otto[19], Alessandro F. Paul[48], Irena Pavlović[18], Marija B. Petrović[49], Dora Popović[50], Gerhard M. Prinz[51], Josip Razum[52], Ivaylo Sakelariev[32,69], Vivian Samuels[22], Inés Sanguino[53], Nicolas Say[54], Jakob Schuck[55], Irem Soysal[19], Anna Louise Todsen[56], Markus R. Tünte[57], Milica Vdovic[58], Jáchym Vintr[59], Maja Vovko[60], Marek A. Vranka[61], Lisa Wagner[62], Lauren Wilkins[19], Manou Willems[32,69], Elizabeth Wisdom[22], Aleksandra Yosifova[63], Sandy Zeng[19], Mahmoud A. Ahmed[64], Twinkle Dwarkanath[65], Mina Cikara[66], Jeffrey Lees[67,68] and Tomas Folke[1,2]

Pervading global narratives suggest that political polarization is increasing, yet the accuracy of such group meta-perceptions has been drawn into question. A recent US study suggests that these beliefs are inaccurate and drive polarized beliefs about out-groups. However, it also found that informing people of inaccuracies reduces those negative beliefs. In this work, we explore whether these results generalize to other countries. To achieve this, we replicate two of the original experiments with 10,207 participants across 26 countries. We focus on local group divisions, which we refer to as fault lines. We find broad generalizability for both inaccurate meta-perceptions and reduced negative motive attribution through a simple disclosure intervention. We conclude that inaccurate and negative group meta-perceptions are exhibited in myriad contexts and that informing individuals of their misperceptions can yield positive benefits for intergroup relations. Such generalizability highlights a robust phenomenon with implications for political discourse worldwide.

Beliefs about polarization have significant social consequences, whether accurate or not[1–4]. They also complicate the study of social issues as reported attitudes might be impacted by inaccurate out-group perceptions and thus make conflicts around a specific issue or policy appear more severe than they actually are[5]. This can result in significant distortions in numerous behaviours, including health, voting and financial choices, each of which has consequences on population wellbeing[6].

There is growing interest in the origins of polarization across populations as well as its features and impacts across communities and society[7,8]. This interest is global and includes both scientific research as well as general public interest[9,10]. As polarization seemingly permeates a growing number of personal and public domains, there is some sense of renewed urgency to understand it and its effects more deeply[11,12]. This includes extending study to understand the extent of polarization on community and global levels.

However, with increasing interest in polarization itself, broadening evidence indicates that inaccuracies in perceptions of how the out-group feels about the in-group can be harmful[13]. The origin of these 'meta-perceptions' may be rooted in negative stereotypes that individuals feel have been applied to them, often incorrectly[14]. This results in an inaccurate perception of differences in beliefs and attitudes between groups[15], which can have negative results for individuals[16,17]. On a population level, such misperceptions can even result in overstated reactions that exacerbate existing biases[18].

To investigate roots and moderators of polarization across groups, Lees and Cikara[13] ran a series of experiments with US participants identifying as Republican or Democrat. In what they refer

A full list of affiliations appears at the end of the paper.

to as 'group meta-perception' (GMP), participants in experiment 4 estimated how their own out-group perceived the behaviour of their in-group across five scenarios involving political actions (which we will refer to as 'issues'), in which the behaviour of one group would potentially disadvantage the other. In their fourth experiment, for each scenario, participants rated their own perception of the out-group's political action ('individual rating'), their estimate of fellow in-group members' rating of out-group actions (in-group meta-perception) or their estimate of the out-group's rating of their own group's political actions (out-group meta-perception). When comparing perceptions between these three conditions, they found that individuals strongly overestimated the negativity of their out-group's reaction, and even overestimated negative perceptions among their in-group. In other words, the authors found that the participants overestimated perceived polarization in terms of both their own group and the out-group, which they refer to as inaccurate group meta-perceptions.

In experiment 6 of their study, Lees and Cikara tested whether disclosing the true opinions of the out-group would mitigate negative out-group attributions. After participants made their GMP judgements, they were presented with their own quantified GMP ratings as well as the true average individual ratings reported by out-group members (for example, Republicans would be shown what they had estimated for Democrats, along with actual Democrat ratings taken from experiment 4). Following this disclosure, participants were presented with an additional rating asking about how much the out-group was obstructing a particular issue. In these scenarios, participants made significantly less negative out-group attributions than those who did not receive the intervention, which aligns with prior work on the potential for disclosure effects to reduce negative out-group biases[19]. The greatest reduction in bias was found for those with the most inaccurate pre-intervention meta-perceptions, indicating potential value specifically toward addressing false polarization narratives, by encouraging accuracy through greater disclosure about beliefs rather than emphasizing perceived disagreements.

Most work on group meta-perceptions has been done in US samples. For example, Moore-Berg et al.[2] found that especially partisan Americans have hostile beliefs about what the opposite group believes, which were not actually in line with the beliefs held by that group. Lees and Cikara scientifically challenge the polarization narrative in a representative US sample by presenting a robust argument that true differences are often overestimated, yet the United States is far from the only country with alleged increases in polarization[20]. Additionally, though it is relatively simple to conceptualize polarization in a two-party system, such as the United States, in-group/out-group divisions are not limited to political affiliations.

Inaccurate group meta-perceptions (and polarization narratives generally) are not an anomaly limited to US political groups. Rather, such perceptions are very much a global issue and have been studied in a number of contexts, not merely focused on political group affiliation[21,22]. Testing the generalizability of Lees and Cikara's findings specifically would therefore have implications for empirically quantifying the global accuracy of group meta-perceptions, given the purported increases in polarization. If the results were to generalize, it would mean current narratives about group divisions are likely overstated on a global scale. Contrarily, if the results were to not hold consistently, it would suggest that media reporting on polarization may be overstated in the United States, but not elsewhere. As such, an international replication of the original study will help illuminate social and political divisions due to incorrect group meta-perceptions and provide insight for domestic and international policy-making.

In this study, we mobilize a global network of researchers to test the generalizability and replicability of group meta-perceptions in 26 countries. The purpose is to determine the extent to which inaccurate group meta-perceptions generalize to a large number of locations and policy actions. Specifically, the primary aim is to assess the consistency of overestimated negative meta-perceptions (in-group and out-group, though primarily out-group) in a number of countries and languages, as well as whether disclosing such biases will mitigate negative attributions toward out-groups. In this way, we seek to provide a meaningful new approach to incorporate elements from both direct (to the extent possible and/or necessary) and conceptual replication, and to test the generalizability of a recent finding with strong policy implications. That this relates to polarization and inaccurate group perceptions, which are critical topics in science and society, makes it even more meaningful to attempt in a highly powered, robust, multi-country study.

The outcomes of replication studies are not necessarily purely binary in terms of reflecting either success or failure to replicate among variables of interest. Instead, they are more likely to exist in a range of insights. For example, an international replication study designed to critically evaluate prospect theory in 19 countries concluded a generally successful replication, but also highlighted an attenuation of original effect sizes in 77% of replicated sub-questions[23]. This observation highlights replication as a crucial tool for examining the generalizability of behavioural effects[24]. In the context of investigating out-group meta-perceptions cross-culturally, the potential differences (not only in the presence or lack of outwardly perceived divisiveness but also in effect sizes and success of meta-perceptual modification) hold promise for elucidating nuances that may not exist in one single population. In other words, knowing effect sizes cross-culturally can help us understand how strongly inaccurate group meta-perceptions exist from place to place and how easily these misperceptions can be attenuated. To this effect, replication of inaccurate group meta-perception will elucidate potential differences in effect size, allowing us to determine whether and the extent to which perceived polarization creates inaccurate out-group meta-perceptions. Additionally, it will indicate whether and to what degree inaccurate out-group meta-perceptions can be mitigated.

Before this study, we had six primary research questions (see pre-registration, https://osf.io/aj6xd/, and data simulations, https://osf.io/s2r9v/), three related to each experiment. These focused mainly on aggregate comparisons between first-order and both group (meta-)perceptions. For experiment 4, we expected out-group meta-perceptions would be significantly more negative than the first-order perceptions. We expected this finding in a majority of countries for aggregated ratings for a majority of scenarios. We also expected that aggregated in-group meta-perceptions would be significantly more negative than first-order perceptions. Finally, we expected that out-group meta-perceptions would be significantly more negative than in-group meta-perceptions (more often than not, within and between countries).

For experiment 6, we expected that negative motive attribution would be significantly less severe in the disclosure intervention condition for the global effect estimate (we also report country-specific estimates and variance estimates of between-country variation). We expected negative motive attributions to be negatively correlated with meta-perception accuracy, in line with the results reported by Lees and Cikara. However, we anticipated that the modelling of these effects might be too complex (require too many degrees of freedom) to allow for firm conclusions, which we addressed in the pre-registration (see the "Experiment 6 modelling strategy" section at https://osf.io/aj6xd/). Though it was impossible to estimate a specific range, we expected considerable variability in intervention effects, ranging from no change to effects larger than in the original study.

Results

Analyses of the data collected focused on highlighting two things. The first was whether group meta-perceptions were generally

inaccurate across a diverse set of countries, languages and political systems. The second was to determine whether the intervention that had shown some effect at reducing inaccuracies in the United States could be effective in other locations and contexts using the most extreme examples from a subset of countries.

Participants. The final sample consisted of 10,207 total participants, with 5,406 (26 countries) in experiment 4 and 4,801 (10 countries, subsample of the 26) in experiment 6. Country sample sizes ranged from 71 (Bosnia) to 368 (Sweden) in experiment 4 and 53 (Puerto Rico) to 675 (Slovenia) in experiment 6. These sample totals do not include the large number of participants who were excluded based on not identifying with either group.

As described in the "Recruitment" section, a multi-layered approach to data collection was implemented for all locations. No student sample pools were used, and less than half of the sample was actively involved in any form of education (43.9% in experiment 4, 40.3% in experiment 6), with fewer full-time students (33.6% in experiment 4, 29.8% in experiment 6). The sample was roughly gender balanced (experiment 4: 41.9% female, 3.5% non-binary or prefer not to say; experiment 6: 41.3% female, 2.5% non-binary or prefer not to say). Our sample was highly educated compared with most of our target populations (in experiment 4, 55.7% had completed at least one university degree; 51.6% for experiment 6). Demographic overviews of each country sample are presented in Table 1, with full details in Supplementary Tables 1 and 2.

Overview. To provide a broad understanding of the primary study findings, here we present a descriptive overview of the results. In the next section, we systematically present the results of the pre-registered models.

For experiment 4, we calculated the mean of the negative perceptions collapsed over the three rating types and five scenarios to give a crude overview of the differences between the intervention groups in the different countries (Fig. 1). The total sample size by condition was 2,019 for the first-order ratings, 1,602 for the in-group ratings and 1,785 for the out-group ratings.

From this overview, in all countries but one, there appears to be a reliable difference between first-order judgements and out-group meta-perceptions, with Sudan as a notable outlier (see Supplementary Table 6 for inferential statistics; see Discussion for potential reasons why this is the case). There is considerably greater between-country variation with regards to in-group meta-perceptions; in some countries, these are very close to out-group meta-perceptions, whereas they are close to first-order perceptions in others (and in some countries, they are almost equidistant from the two other categories). Additionally, there is no obvious pattern in the ordering of effect sizes. In other words, the variation in effect sizes seems to be equally driven by variations in first-order judgements and meta-perceptions, and there is no obvious pattern with regards to which countries show larger effects and which countries show smaller effects. The absence of a clear pattern on these speaks to the general reliability and robustness of the conclusion of the original study.

We plotted a similar overview for experiment 6. Here we plot mean negative motive attributions as a function of experimental condition for each country (Fig. 2). The disclosure intervention is consistently associated with lower negative motive attributions than the control group, suggesting that the effect of the intervention generalizes outside the United States (Supplementary Table 7). Again, we see no clear pattern in effect sizes. Crucially, sample size seems to be unrelated to effect size. To formally evaluate generalizability, we ran hierarchical beta regressions, in line with our pre-registration. Because we had fewer countries and smaller sample sizes than planned, we do not want to over-interpret differences between fault-line groups. However, we include the pre-registered figure on negative motive attributions based on fault-line groups in the supplement for completeness (Supplementary Fig. 1). The total sample size by condition was 2,417 for the control condition and 2,384 for the disclosure condition. All *P* values reported in these results are two tailed. Our outcome variables in the pre-registered analyses were beta distributed. We ran assumption tests for multicollinearity, heteroscedasticity and non-normally distributed residuals/random effects, because these are the key assumptions of beta regressions[25]. Our analyses met all test assumptions. Equality of variances was not assessed for the *t* tests because we used the Welch approximation of degrees of freedom, which adjusts for unequal variances.

Pre-registered analyses. *Experiment 4.* The fourth experiment from the original study aimed to test whether meta-perceptions of the out-group's attitudes toward the in-group are significantly more negative than the out-group's first-order perceptions. In our pre-registration (https://osf.io/aj6xd/; https://osf.io/s2r9v/), we specified a fixed-effects model for the different experimental conditions, with random effects at the country level for intercepts and experimental conditions and random intercepts for out-group contrasts, participants and scenarios. Those models closely reflected the approach of the original study methods of Lees and Cikara, extrapolated to a multi-country context. We intended to fit this model using glmmTMB[26], the same R package used by the original authors. However, this model caused convergence warnings due to the complexity of the clustering of the data. Because the full model is theoretically justified, we were reluctant to simplify it. Instead, we used brms[27] to run a Bayesian version of the model with regularizing priors to ensure that the parameters were sufficiently constrained. For intercept terms, we used *t* distributions with 1 degree of freedom (df), centred on 0 and with standard deviations of 10. For the beta coefficients, we used normal distributions centred on 0, and with a standard deviation of 10. For the dispersion parameter, we used a gamma prior with k and θ both set to 0.01. These priors were chosen to be minimally informative and serve only as a weak constraint to ensure convergence. Making these priors more diffuse or more precise by an order of magnitude did not meaningfully change our result (Supplementary Table 5).

We ran four chains with 1,000 warm-up samples and 2,000 post-warm-up iterations, leading to 8,000 samples in total. Chains converged with Rhat values of 1 for all fixed effects, but with Rhat values up to 1.02 for the correlations between random effects.

The posterior means of the fixed effects in the Bayesian model (intercept 0.22, in-group parameter 0.23, out-group parameter 0.67) closely matched the estimates from the pre-registered glmmTMB model (intercept 0.22, in-group parameter 0.23, out-group parameter 0.67). Because of this close match between the pre-registered restricted maximum-likelihood model and the Bayesian model, it is unlikely that the shift in modelling framework impacted our conclusions. However, this transition to a fully Bayesian framework made our pre-registered decision criteria irrelevant. Rather than reporting *P* values for the fixed effects, we operated directly on the samples drawn from the posterior distributions. For example, if we wanted to know whether the in-group meta-perceptions are more severe than the first-order perceptions globally, we checked what proportion of the posterior samples of the fixed-effect in-group parameter were greater than 0.

To avoid confusion, we refer to these comparisons as proportions and reserve '*P*' for the frequentist *P* value. With that in mind, the population-level posterior distributions between the different conditions showed very little overlap (Fig. 3). Both in-group meta-perceptions ($n_{\text{observations}}=24{,}030$, $n_{\text{participants}}=1{,}602$, $n_{\text{countries}}=26$, posterior mean 0.44, 95% credible intervals [0.31–0.57]) and out-group meta-perceptions ($n_{\text{observations}}=26{,}773$, $n_{\text{participants}}=1{,}785$, $n_{\text{countries}}=26$, posterior mean 0.89, 95% credible intervals [0.76–1.03]) were associated with more severe judgements

Table 1 | Demographic overview

	Experiment 4				Experiment 6			
Country	***N***	**Mean age (years)**	**% female**	**% university degree or above**	***N***	**Mean age (years)**	**% female**	**% university degree or above**
All	5,406	34.5	43.5	56.8	4,801	35.4	41.3	52.4
Austria	158	37.6	40.5	48.4				
Bosnia and Herzegovina	71	33.8	51.5	60.0				
Bulgaria	218	35.1	40.2	73.8				
Canada	124	35.7	45.9	54.9				
Croatia	209	33.2	45.3	9.7				
Czech Republic	307	36.5	48.5	62.5				
Denmark	291	42.7	43.1	44.7	361	34.4	42.1	36.4
France	277	30.0	39.6	76.9	532	28.2	39.7	65.7
Germany	211	32.9	44.9	65.2	637	33.2	37.8	53.2
Ireland	148	31.0	46.5	57.2				
Israel	251	30.0	52.3	51.2				
Italy	181	30.6	53.9	58.7				
The Netherlands	165	32.1	41.4	55.2				
Norway	163	35.4	29.7	62.5				
Portugal	156	37.4	54.3	81.6				
Puerto Rico	135	44.9	56.1	80.6	52	49.4	69.2	87.5
Serbia	236	34.0	48.5	48.7	267	37.2	40.4	41.6
Slovakia	247	35.2	27.7	52.7	353	36.8	30.0	47.7
Slovenia	262	42.4	48.4	69.3	675	43.2	43.7	43.9
Spain	256	26.6	30.3	49.6	630	30.3	40.6	56.0
Sudan	133	41.2	32.1	57.3				
Sweden	368	38.4	38.0	56.0	675	39.2	37.8	52.9
Switzerland	163	40.8	35.0	51.3				
Turkey	97	29.4	39.6	69.1				
United Kingdom	213	34.0	40.4	69.4				
United States	366	24.4	56.9	41.6	619	34.1	51.9	58.2

than first-order perceptions ($n_{observations}=30{,}284$, $n_{participants}=2{,}019$, $n_{countries}=26$, posterior mean 0.22, 95% credible intervals [0.08–0.35]) ($proportion_{in\text{-}group}=1$, $proportion_{out\text{-}group}=1$). This means that our sample tended to overestimate how strongly their in-group and their out-group would react to actions taken by the opposition. Furthermore, out-group meta-perceptions were more severe than in-group meta-perceptions (proportion = 1). In other words, participants expected that their out-group would react more strongly to policies proposed by their in-group than their in-group would react to policies proposed by their out-group. These results all align with what was reported by Lees and Cikara.

Despite these robust effects at the population level, the model also estimated substantial between-country variation (posterior $SD_{in\text{-}group}$ 0.17, 95% credible Intervals [0.10–0.25]; posterior $SD_{out\text{-}group}$ 0.30, 95% credible Intervals [0.21–0.41]).

Experiment 6. The sixth experiment from the original study tested whether being informed of the true average first-order perception of the out-group reduces negative motive attribution toward the out-group. For experiment 6, we pre-registered four different models of increasing complexity. As above, these models are designed to closely match the analysis strategy of Lees and Cikara, while allowing for variation between countries and fault-line groups. In the simplest model (df 9; Akaike information criterion (AIC) −4,170.2, deviance −4,168.0), we predicted obstruction ratings from the experimental conditions, with random intercepts and experimental conditions by country, and random intercepts by fault-line groups. This model captured whether the intervention impacts obstruction ratings. In the second model (df 11; AIC −5,052.8, deviance −5,054.0), we added fixed effects for how inaccurate participant meta-perceptions were before the intervention, and an inaccuracy–intervention interaction term. This model measured whether belief inaccuracy predicts obstruction ratings and whether the intervention impacted this relationship. Belief inaccuracy was captured by the mean difference between the three out-group meta-perceptions and the first-order attitudes of the out-group, in line with Lees and Cikara's original study. The third model (df 17; AIC −5,096.8, deviance −5,110.0) added some additional flexibility by allowing the effect of belief inaccuracy to vary between countries, and the final model allowed the interaction term to vary by country as well. In our pre-registration, we wrote that we would compare these models via AIC but ignore any models that were too complex to converge. We would then focus on the lowest AIC model. Model 4 failed to converge, so we compared the remaining three models on AIC.

The third model provided the best fit to the data. For a summary of the fixed effects of the winning model, see Table 2. Since the winning model included an interaction term between inaccuracy and the intervention, our pre-registered replication criterion was whether there was a significant reduction in the belief inaccuracy on obstruction ratings after the disclosure intervention (alpha

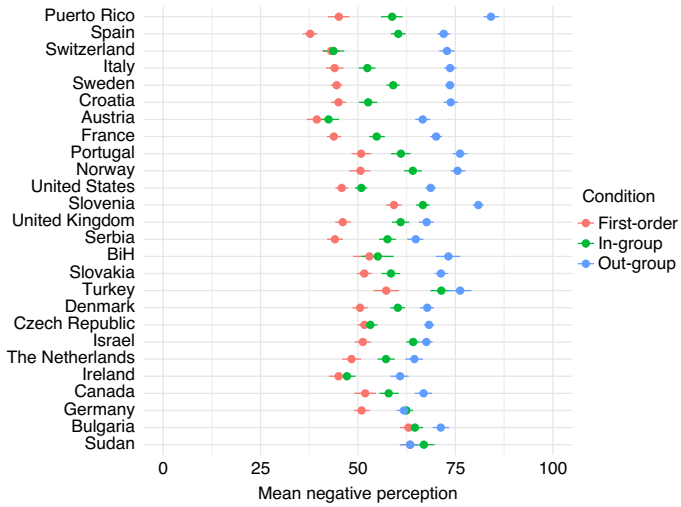

Fig. 1 | Negative perceptions by country and condition, collapsed across scenarios and rating types. Dots signify means; error bars signify bootstrapped 95% confidence intervals. Countries are ordered based on the difference in mean between the first-order ratings and out-group meta-perception ratings (from largest to smallest). Sample sizes ranged from 71 in Bosnia and Herzegovina to 368 in Sweden. For a list of country-specific sample sizes see Table 1. For unpooled country-specific significance tests between the first-order ratings and out-group ratings, see Supplementary Table 6. Negative perceptions are consistently overestimated for the out-group in almost every country ($n = 5{,}406$).

threshold of 0.01). This criterion was met, as the relevant P value is less than 0.0001. Our results show that people who hold inaccurate beliefs about the negative opinions of the out-group gave significantly higher obstruction ratings, but that this effect was mitigated when the disclosure was presented (Table 2 and Fig. 4). Though this interaction effect is convincingly significant at the population level, there is considerable variation between countries (s.d. 0.36) and fault-line groups (s.d. 0.49).

Exploratory analyses. Because our interest was primarily in the generalizability of inaccurate group meta-perceptions, we also wanted to explore two additional concepts *sans doctrina*. The first was to provide some indication as to the reliability of the findings within and between countries and fault-line groups (that is, whether meta-perceptions are group dependent). The second was to look at the robustness of the initial findings based on sample sizes when attempting to replicate.

A question of theoretical interest is the extent to which the group meta-perception effects vary between fault-line groups in different countries. We had no a priori expectations with regards to this, as the original study focused exclusively on the United States and thus could not inform a comparative component. In this exploratory analysis, we found three qualitatively distinct patterns of fault-line group perceptions that repeat in multiple countries (Fig. 5).

In most countries, the perception of both fault-line groups mirrored each other, such that (for example) the Right's perception of itself closely matched the Left's perception of itself, and the Right's perception of the out-group matched the Left's perception of the out-group. Austria, Slovenia and Sweden all show this pattern. The second pattern is a diamond shape, such that both fault-line groups had similar first-order perceptions and out-group meta-perceptions, but one fault-line group perceived that their in-group had more negative views than the other. The Netherlands and Bosnia and Herzegovina (BiH) both show this pattern. Finally, in some countries, one fault-line group has more severe first-order judgements and self-perceptions than the others. The United States and Switzerland both show this pattern.

Sudan showcases this final pattern, with the distinction that one fault-line group actually underestimated how negatively the out-group would feel. Closer inspection revealed that the Sudanese results might depend on the specifics of the scenarios, as one fault-line group found four of the scenarios significantly more disagreeable than the other (in the fifth scenario, both groups find it equally disagreeable). This is in contrast to most other countries, where either both groups find each scenario equally disagreeable, or they are roughly balanced in how many scenarios they found to be particularly disagreeable (Supplementary Fig. 2). We explain possible reasons for this in the Discussion, stressing the need for caution about over-interpretation.

Finally, to test the generalizability of the out-group meta-perception bias, we computed the mean difference between the out-group meta-perception and first-order perception for each rating type and scenario (Fig. 5). Eighty-nine percent of values were in the expected direction of the bias. When values were not in the expected direction, this was typically driven by the scenario, as typically all ratings from that scenario showed that effect. To assess how these results compared with those originally reported by Lees and Cikara, we computed the difference between the average out-group rating and the average first-order rating from experiment 4 in their study. Fifty-seven percent of ratings showed a greater effect than that originally reported by Lees and Cikara, suggesting that their original effect is close to the global expected value across countries as there are roughly the same number of observations above their original finding as below. Also visible in Fig. 6 is the absence of a clear relationship between sample size and effect, further indicating the robustness and generalizability of the conclusion from the original study.

Discussion

This study demonstrates that inaccuracies in group meta-perceptions are a widely generalizable phenomenon on an international scale, such that individuals who identify with a defined group by and large overestimate the negativity of views held by their respective out-group. We find this effect clearly in 25 out of 26 countries in a sample of 5,406 participants in the first experiment, which demonstrated the general pattern, and 4,801 in the second experiment, which demonstrated the potential for an intervention to reduce those inaccuracies in the scenarios that had initially elicited the most extreme meta-perceptions. The wide generalizability of the findings demonstrates that the effect is not limited to countries where most published behavioural research has been conducted. Additionally, it should encourage social scientists to use highly powered study designs in original testing, as the study replicated here has produced broadly consistent findings in spite of a number of differences in instrument, sample and implementation.

In preparing this study, we did not expect replication for all countries with the same effect sizes, but that a majority of countries would show effect sizes similar to the original study in experiment 4. We also expected that a majority of those countries would in turn show significant effects (likely with varied sizes) for experiment 6. Within that matrix, we concluded original findings as having replicated results based on the criteria described in the pre-registered analyses, and the generalizability would be described on a spectrum of zero replications to universal replication. With this general guide, we aimed to present patterns across all locations and issues to emphasize the generalizability aspect rather than to focus on individual settings. In line with this, we focus the discussion on the four primary research questions and related hypotheses as presented in Main.

To prevent individual scenarios from exaggerating or mitigating general patterns, the 15 ratings per country were aggregated to

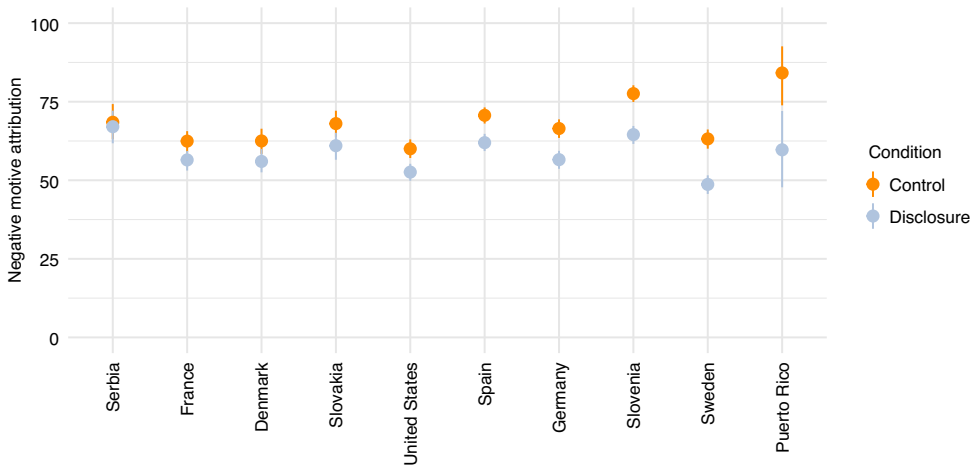

Fig. 2 | Mean negative motive attribution of each condition by country. Dots signify means; error bars are 95% bootstrapped confidence intervals. Countries are ordered based on the mean difference between the experimental conditions, from smallest to largest. Sample sizes range from 52 in Puerto Rico to 675 in Slovenia. For unpooled country-specific significance tests, see Supplementary Table 7. The disclosure intervention is consistently associated with lower negative motive attributions than the control group, though some of these differences are within the margin of error ($n = 4{,}801$).

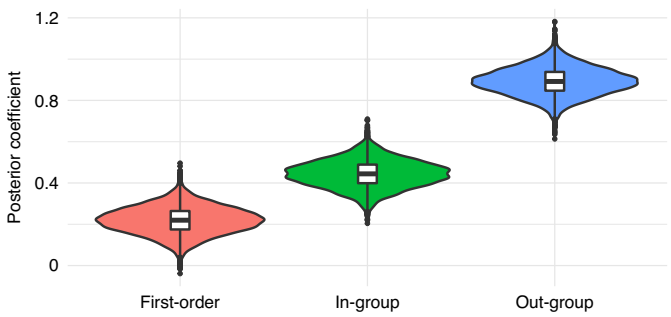

Fig. 3 | Posterior distributions for the population-level mean effects of the different conditions ($n = 5{,}406$). Posterior distributions of the marginal effects of the different conditions when accounting for variations in scenario, fault-line group and country. Violin plots show the full posterior distributions consisting of 8,000 samples combined from four Monte Carlo Markov chains. In the boxplots, the centrelines signify medians, and the lower and upper hinges signify the first and third quartile, respectively. The whiskers extend from the hinge to at most 1.5. times the interquartile range; points signify observations outside of this range. The first-order condition contains *2,019* observations, the second-order condition contains *1,602* observations and the out-group condition contains *1,785* observations.

compare differences between the three conditions for every country. We find that results from the original study largely generalize across 25 of the 26 countries included in this study, the exception being Sudan. We discuss potential reasons for this deviation later.

Though not the focus of this replication, we also studied differences between in-group meta-perceptions and first-order perceptions. In line with the original results, we found that, in aggregate, in-group meta-perceptions were reliably more negative than first-order perceptions but also more positive than out-group meta-perceptions. In other words, at the aggregate level, we tend to overestimate the negativity of our in-group, but not to the same extent that we overestimate the negativity of the out-group. However, relative to out-group meta-perceptions, this finding was much noisier at the country level.

There is no clear pattern for these in terms of proximity to first-order perceptions or to out-group perceptions, even when considering sample size. This may be better studied on a country-by-country basis in terms of the fault lines and variability (or lack thereof) within groups. However, similarity of in-group identities and perceptions was not a focus of this study, so we suggest this be studied by others interested more specifically in political group affiliation as we do not want to casually speculate on potential explanations.

For experiment 6, we find reduced negative motive attributions after being informed of the first-order perception. The intervention appears to work by reducing negative motive attributions in those participants with most inaccurate beliefs, as found in the original study. This finding applies to nine of the ten countries included (Serbia being the exception). Given that the effect was present for even the smallest sample group (in spite of anticipating a large sample necessary to detect a small effect), its effectiveness also appears to be generalizable at the country level.

Because we focused on a single scenario per country, we cannot distinguish between generalizability across issues and generalizability across locations; we can only conclude that, across both these dimensions, we tend to find a difference between the conditions in our sample. However, due to limited observations at the country/scenario level, we do not wish to speculate as to which characteristics produce a stronger or weaker effect.

We do find that the disclosure intervention has the largest effect on negative motive attributions for those participants with the most inaccurate and negative out-group meta-perceptions. At its base, this finding supports the common-sense assumption that inaccurate meta-perceptions around an issue are necessary for the disclosure intervention to be effective. It further implies that effects may be largest for individuals and issues where misperceptions are largest. However, one limitation is that we cannot fully rule out the possibility of participants wanting to appear less politically polarized upon learning they held inaccurate beliefs. This is worth noting given that we have theorized that the effect of the intervention is due to a causal relationship between updated GMP beliefs and out-group motive attributions. Future work should consider ways of mitigating this methodological artefact, or measuring post-intervention belief updating directly.

In terms of core methodology related to the instrument, procedure and data collection, there were no substantive deviations from the pre-registered research plan. Instead, all deviations related to

Table 2 | Experiment 6 fixed-effects coefficients

Coefficient	Estimate (s.e.)	Odds ratios	OR 95% CI	*Z* value	*P* value
Intercept	−0.18 (0.13)	0.83	[0.65, 1.07]	−1.44	0.15
Disclosure condition	−0.08 (0.08)	0.92	[0.78, 1.08]	−1.00	0.32
Meta-perception inaccuracy	2.48 (0.19)	11.98	[8.25, 17.40]	13.04	<0.001
Disclosure × inaccuracy	−0.69 (0.14)	0.50	[0.38, 0.66]	−5.05	<0.001

$n_{participants}=4{,}801$, $n_{countries}=10$

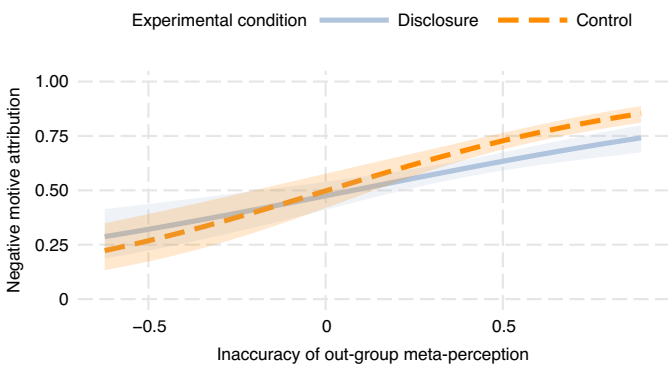

Fig. 4 | Interaction effect between belief inaccuracy and conditions ($n=4{,}801$). Beta regression lines showing how inaccurate meta-perceptions are associated with negative motive attributions for each of the conditions, while statistically controlling for nationality and fault line. Shaded areas show 95% analytic confidence intervals for the fixed effects. For statistics for these results, see Table 2.

the sample size deficits in many of the countries involved. These resulted in fewer countries being considered for experiment 6, and imbalanced sample sizes for fault-line groups due to the lack of involvement of specific political groups. There is some indication that challenges in obtaining target sample sizes stemmed directly from COVID-19 impacts on social study more generally, though it is difficult to assess this impact directly.

Our target sample size was 240 per country for experiment 4 and 600 per country for experiment 6. These targets were met for nine out of 26 countries for experiment 4 and five out of ten countries for experiment 6. It became evident early on that the original targets were unlikely to be achieved for many countries. In some cases, this was due to the usual scepticism faced when collecting social surveys with political themes (responses ranged from strong rebukes, concerns about ulterior motives and exclusionary practices based on the requirement to be in one of two groups to participate). Participation was typically much lower for conservative groups (whether named parties or general description). To ensure power was sufficient, paid samples using the Prolific platform were used in five out of 26 countries in experiment 4 and seven out of ten countries for experiment 6 (Supplementary Tables 3 and 4).

To ensure sufficient power in both experiments, we created an ad hoc rule that inclusion in experiment 6 required at least 50 participants in both groups by the end of the second full week of data collection, and 60 participants in both groups before calculating the average ratings. Some countries passed these thresholds only later, and thus were not included in experiment 6. None of these steps conflicted with the methodological approach, but only reacted to challenges in collecting data during COVID-19.

While it is tempting to interpret the outlier results for Sudan as a meaningful insight of a non-Western, educated, industrialized, rich and democratic (WEIRD) population, we caution against accepting this interpretation uncritically. Notably, the approach to data collection in Sudan differed from all other locations in that roughly one-third were recruited through a voluntary non-profit group that presented the survey to constituents. This introduced some systemic bias into the sampling as participants who were surveyed at the same time were likely more similar than two random persons from the target population. Furthermore, due to translation issues and time constraints, a last-minute change to the wording of the fault line changed the group classifications from specific political parties to general ideological groups. This was adjusted without review from the research team, but as we did not want to exclude any data for purposes of transparency, we have included it along with all other countries.

Along with Sudan, we also chose to include the samples from Turkey and Bosnia and Herzegovina for experiment 4, as well as from Puerto Rico for experiment 6. While these were clearly small, underpowered sample sizes that would not have held much weight reported individually, there were two primary reasons we included them. First, we felt it was best to be fully transparent, and not to exclude marginal samples ad hoc. Second, as is evident in several figures, there is no obvious link between effect sizes and sample sizes. Our initial hope had been to have samples over 240 for each experiment 4 country and over 600 for each experiment 6 country. Such samples, had they been achieved, might have allowed for further analyses with demographic variables. Even so, there is no indication that the smaller samples impacted the main conclusions of our study as there was no indication that sample sizes predicted replication probability and confidence intervals were only modestly larger for most countries with smaller samples.

The study was also attempted in India and Australia, but it was not possible to generate even borderline sample sizes sufficient for inclusion (at the closing of data collection, each survey had fewer than 30 complete responses), so those have been omitted.

Along with the critical scientific importance of this topic, we also placed considerable emphasis on establishing a standard for large-scale replications involving the original research team. Recognizing the potential value of that involvement while also acknowledging the potential for conflicts of interest, the collaboration essentially allocated the bulk of all processes to the new research team, with regular feedback sought from the original team. However, all final methodological decisions were left to the replication team, and the original researchers[13] were only allowed to comment and advise. In this way, we leveraged the unique and intimate knowledge from the original researchers to maintain integrity and fidelity to the study being replicated (which also aided in methodological efficiency and rigour), but did not allocate any decision-making responsibilities to them. We strongly recommend this approach for similar attempts at replication for future research teams.

The study method has an evident limitation in treating all countries as having a binary group identification system, which is not representative or indicative in many cases. Particularly in Europe, where many parliamentary systems require coalitions to form governments, this created some challenges in establishing the group

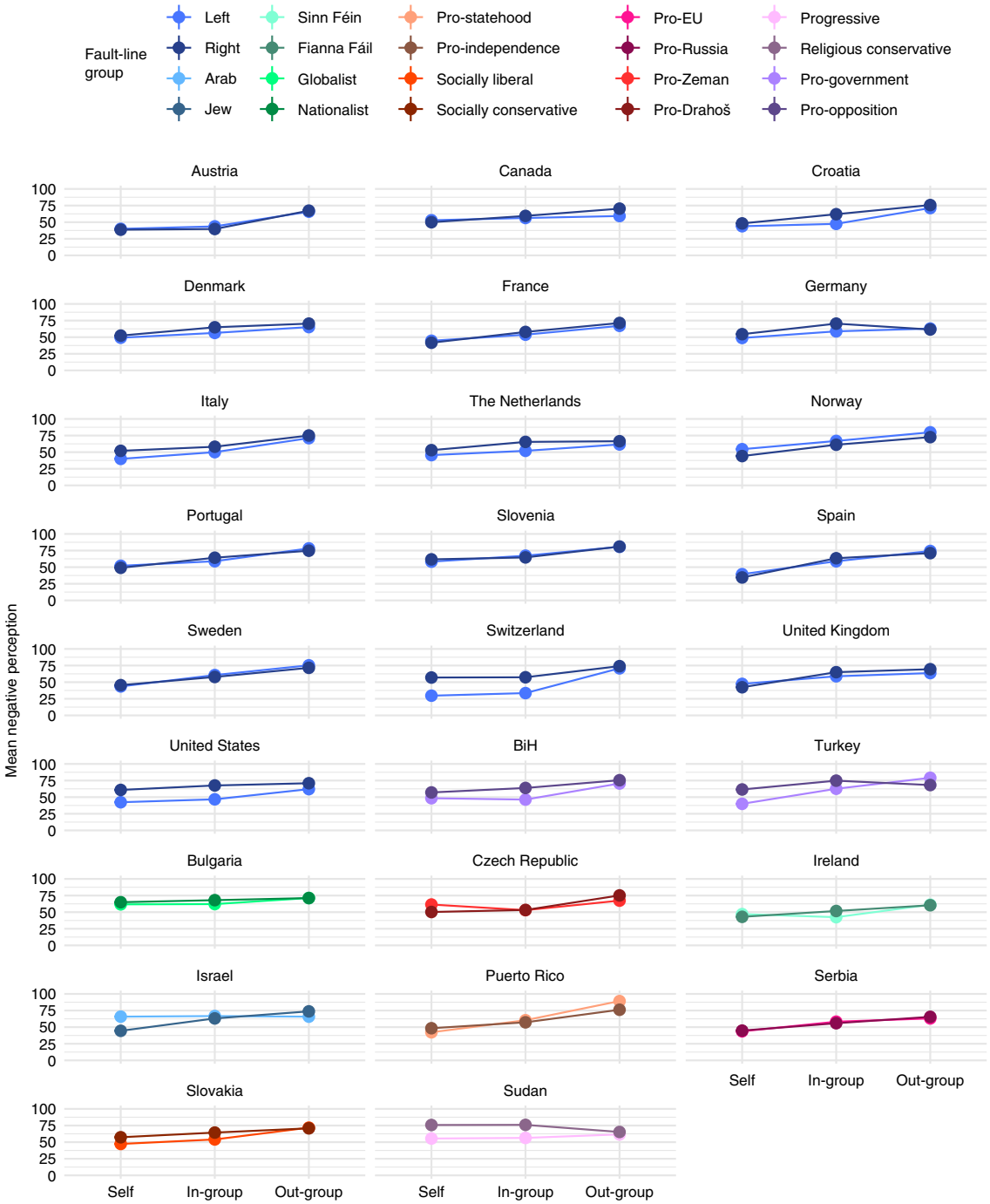

Fig. 5 | Differences between the conditions depending on how each fault-line group is perceived ($n = 5{,}406$). Mean negative perceptions delineated by fault-line group, averaged across scenarios. Dots signify means; error bars signify bootstrapped 95% confidence intervals (but confidence intervals are often so narrow that they are hidden behind the dots). Note that the colour coding refers to which group is being evaluated. For example, for Austria, 'Right Self' and 'Right In-group' both capture how the Austrian Right perceives itself, whereas 'Right Out-group' refers to how the Austrian Left perceives the Austrian Right. Though each context is unique, where fault lines are similar, we have classified them with common colours and put in clusters (to the extent possible).

identities, and also resulted in the exclusion of a large number of potential participants that chose 'other' for their political affiliation. While including those who did not identify with either group might have elicited meaningful data, our interest was specifically on how groups from opposite (alleged or legitimate) viewpoints perceive the other. Including those who did not identify with either group would have required a fundamental shift in the design of the studies. What those unaligned with certain groups think that members of other groups believe about them is a worthwhile research question and an exciting avenue of future research on intergroup perceptions. However, it lies beyond the scope of the current investigation.

Limiting to two groups also resulted in various statements made both via email and on social media in response to the survey instrument about potential participants feeling excluded from potentially relevant discourse, which is obviously not the intent of the research team. In the end, to avoid exacerbating this limitation, it is critical

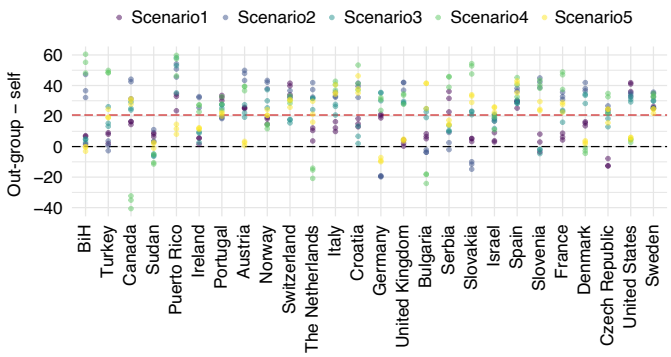

Fig. 6 | The difference between out-group meta-perceptions and first-order perceptions for countries ordered by sample size. Each point indicates a mean difference for a specific rating type and scenario (resulting in 15 points per country). The black dashed line indicates the null hypothesis that there is no difference between first-order judgements and out-group meta-perceptions. The red dashed line indicates the difference originally reported by Lees and Cikara. See Supplementary Methods 1.1 for scenario by country. Reported effects are in the expected direction in 89% of the ratings ($n = 5{,}406$).

to highlight that the findings should only be applied to groups that have clear identification parameters that explicitly stand in juxtaposition to another group, even if not specifically fitting under a single party or ideology. Shifting to a scalar classification system (even in the United States, where the largest group identification tends to be 'moderate') would certainly be a welcome study, but would require a substantial shift in approach. We therefore caution against interpreting our findings as applicable to all out-group meta-perceptions in the political context, but instead limit generalizations to out-groups that could be viewed as being in direct and polar opposition with an in-group.

Conclusion

We find that the results of Lees and Cikara broadly generalize across 26 countries by replicating two of their experiments on group meta-perceptions with over 10,000 participants. Methodologically, the most critical insight is that the results of a highly powered original study were widely replicated in samples of varying sizes. In sum, we conclude that, the greater the power of the original study, the more likely its conclusions are to be widely generalizable. While this may not be a surprising conclusion, it should provide another compelling reason to invest in high power when studying novel phenomena. In other words, if you want your findings to replicate or make claims about their generalizability, power them appropriately the first time.

In terms of the specific topic of interest, we also conclude that individuals are likely to overestimate how negatively political out-groups perceive policy actions initiated by their side. Disclosing the true perceptions of the out-group has an additional, meaningful impact on reducing (inaccurate) negative out-group motive attributions. This is not to say that there are no group differences between political groups toward specific actions or in their perceptions of out-groups. We also do not claim that individuals from different political identities will suddenly converge on a common belief set by learning about each other. Instead, it demonstrates that, even where differences may exist, the magnitude of these differences tends to be exaggerated. While these insights should not and will not impact beliefs about a given policy issue or affiliation to a particular political identity, they should reduce unfortunate misperceptions that groups are irreparably divided to extremes. Reducing that belief has the potential to increase social cohesion and wellbeing of populations around the world.

Methods

Ethics. Ethical approval was given by the Columbia University Irving Medical Center to cover all aspects of data collection. All team members responsible for collecting data outside the United States produced a validation letter on appropriateness of language and content. All data were collected via Qualtrics using a Columbia University Irving Medical Center account, therefore a single institutional review board application was used. Only the lead and senior author had access to full data. All participants gave informed consent at the start of the survey. Where appropriate, an extra layer of consent was included for countries covered by General Data Protection Regulation. Most participants did not receive any payment for completing the study. Where Prolific was necessary (Supplementary Figs. 3 and 4), participants received approximately US$0.80 for completing experiment 4 ($n = 270$) and US$0.25 ($n = 2{,}403$) for experiment 6, though precise values varied based on local currency.

Methods overview. This study followed the published and pre-registered protocols of Lees and Cikara[13] for experiments 4 and 6, with adjustments only to facilitate a multi-country, multi-lingual sample. To the extent possible, identical scenarios, procedures and participant experiences were implemented, with modifications only made to match local relevance and item understanding following translation. As explained below, we also added new scenarios to the US version. Direct replication was not possible for issues in multiple countries, so we outline the full approach to conceptual replication while maintaining the methodological integrity and fidelity of the original study. This study serves as both a partial attempt to directly replicate the original study as well as to assess the generalizability of findings on a more global scale. In other words, we focus on whether we find inaccurate group meta-perceptions in other countries as well as in the United States, which would suggest that polarization is less extreme than is commonly assumed. As such, we emphasize generalizability over replication for this body of work.

Replication choice. While past work has examined a variety of instances of polarization, the study of Lees and Cikara[13] was chosen for several reasons. First, the original effect was observed in a stratified, representative sample and both the effect (experiment 4) and intervention (experiment 6) were well powered and pre-registered and pursued a generalizable design and analysis. They utilized multiple generic scenarios and model random effects, making the design itself well suited for extension to different intergroup contexts. To the best of the authors' knowledge, no existing work on inaccurate perceptions and polarization shares all these qualities, and many studies on polarization are so anchored to the US political context that 'replicating' them internationally would effectively mean designing entirely new studies. Lees and Cikara provided a unique and robust framework for scaffolding research on polarization, and an intervention that reduces it, across the globe.

Second, the work of Lees and Cikara[13] builds upon a growing body of research which identifies inaccurate second-order beliefs (that is, meta-perceptions) as a unique vector by which misperceived polarization leads to actual polarization via mutual reinforcement[2,19,28,29]. This growing literature suggests that interventions designed to target inaccurate meta-perceptions may be particularly effective at reducing inaccurate perceptions of polarization and their negative effects due to the reputationally relevant nature of second-order beliefs (that is, that they are about 'me' or 'us'). People may be relatively more receptive to feedback that "'they" like "us" more than I thought' than they are to fact-checking interventions which seek to change individuals' own ideological worldviews[30]. This meant that an international replication of Lees and Cikara[13] would itself contribute to this growing view among scholars regarding the role of second-order beliefs in driving polarization.

Consortium and country selection. A consortium of approximately 80 behavioural scientists based in 26 countries carried out this study. With the exception of the explicit aim of testing the generalizability of the original findings in the United States, we did not aim to include or exclude any specific countries. Instead, we focused on countries with which members of the research network were familiar, though we did aim to include at least one country from each continent. Our pre-registered (https://osf.io/aj6xd/; 23 July 2020) sample size targets, based on bootstrapped simulations of the original study results (https://osf.io/s2r9v/), were 240 participants per country for experiment 4 and 600 participants per country for experiment 6. As data collection started, it quickly became clear that we would struggle to reach these targets. Because the intervention in experiment 6 involves showing participants results from experiment 4, we created an ad hoc inclusion rule for experiment 6 by requiring at least 50 participants in both groups by the end of the second full week of data collection. These issues are covered more fully in the Discussion.

Instrument construction. For each country, five scenarios were developed with three GMP ratings (unacceptable, dislike and oppose) for replicating experiment 4, and a fourth rating (obstruction) of out-group motives for experiment 6. For operational reasons, we distinguish 'issue' and 'scenario', and define 'fault lines' uniquely by country. 'Issues' refer to the specific policy being proposed. 'Scenarios'

are the full statements involving explicit mentions of in-groups and out-groups. The distinction in classifying those discrete groups (for example, Democrats and Republicans) is referred to as 'fault lines'. Each scenario has three ratings asking the extent to which a proposed action is unacceptable, disliked and opposed. In experiment 6, a fourth rating asks participants the extent to which they believe the out-group would like to block the proposed action. Though modifications exist to fit each language and country, or to ensure general alignment across countries, all aspects are directly in line with the original study. Each country issue and fault line (that is, group distinction and classification creating the intergroup boundary) is provided in Supplementary Methods 1.1. Supplementary Methods 1.2 includes notes where individual countries had unique deviations from the original study.

Issue design. Lees and Cikara presented participants five political issues within scenarios that alternated between first-order perceptions (that is, how do you perceive this out-group action, called 'actual perceptions' in the original study), in-group meta-perception (that is, how does your group [Democrats/Republicans] perceive this out-group action?) and out-group meta-perception (that is, how does the other group [Republicans/Democrats] perceive this in-group action?). The five issues they used involved changing the name of a highway from the name of someone from one group to the other, changing the committee that draws district lines for voting, requiring governors to disclose tax returns, banning anonymous political donations and allowing the legislature to appoint judges in groups rather than individually.

To an extent, these issues were generally suitable for reuse in other countries. However, this was not universal within issues or countries, as some issues did not apply, even with some adjustment. To resolve this, while maintaining cohesion that would avoid biasing results by having participants rate dissimilar issues between countries, all collaborators were instructed to select issues using a common set of guidelines. First, all collaborators were instructed to assess whether the original issues could be directly recycled within the country. If so, those issues remained. Issues that required only minor adjustments, such as shifting 'state' to indicate municipal or other non-national entities, were considered as directly replicated.

For issues that already existed in some way, such as existing bans on anonymous donations, individual country research teams were asked to consider the reversal of those laws. Where this did not fit, teams were asked to design issues that fit the general level of importance. This involved, for example, renaming a plaza or public building instead of a highway, as the level of divisiveness would be about the same and could be relevant to either group along the fault line. Conversely, changing marriage equality laws in place of the highway issue would not be permitted. An example of these transitions is (see also Supplementary Methods 1.1):Original issue: Redrawing congressional maps for voting districts (fault line: Democrat/Republican)

Adaptation: Changing the number of seats in parliament (ruling coalition/opposition coalition)

Comprehensively divisive issues that were clearly associated with one group but not the other were prohibited. This would include issues such as abortion bans in the United States (strongly associated with Republican party members) or legalization of recreational drugs in European countries (strongly associated with members of liberal policy parties).

As the original version of the study is extremely recent and was a highly powered trial with all relevant safeguards in place such as pre-registration and publicly available data, there was limited value in a comprehensive direct replication in a US sample. Instead, three of the new issues developed to replace original issues that did not translate were used:

1. Income-based parking tickets (rather than paying a fixed penalty, fines will be based on a percentage of monthly income)
2. State governments take control over refugee intake levels
3. State governments offer tax breaks for manufacturing companies that create local jobs

Banning anonymous donations and replacing names of highways were the two most commonly used original issues in other countries. Those were retained as a baseline for comparison with the original study. However, minimal divergences from the original study issues were anticipated (and differences would largely be attributed to sampling adjustments).

Fault lines. As with the issues, the scenarios required country-by-country precision on how in-groups and out-groups are operationalized (see Supplementary Methods 1.2 for full details). Since most of the countries involved in this global iteration do not have strictly or even predominantly two-party systems, each country team proposed a clear fault line where participants would have a common understanding of in-group and out-group without requiring explanation within the study. In some cases, this generalized to liberal–conservative ideologies (often in the form of a coalition, but not necessarily a specific party). In other locations, fault lines were based on ethnic groups (that is, Israeli and Arab) or geopolitical alignments (that is, pro-Europe and pro-Russia).

Limiting to only two groups may inadvertently produce participants who have distinct psychological mechanisms, which would potentially predispose them to certain intergroup perceptions and potentially have a confounding influence on ratings. We assess this in the Exploratory analyses and Discussion.

Recruitment. To ensure a diverse and, to the extent possible, unbiased representation across all of the countries involved, multiple recruitment approaches were taken within all countries. In each location, widespread dissemination and circulation of the survey was carried out over social media (personal account posts as well as targeted group posts), email distribution lists and various discussion forum threads. Additionally, we recruited participants through the comment sections of news sites to increase the likelihood of non-moderate group members. Generally speaking, snowballing methods were not used. We also created a website with a central repository of all versions where individuals could choose which country and language they preferred to complete.

In most cases (Supplementary Tables 3 and 4), sufficient participation was generated through these means for experiment 4. In a small number of countries, participation was somewhat lower, particularly among conservative participants, and the Prolific platform was used to ensure that the analyses would be sufficiently powered. No student sample pools were used, though some universities did assist with circulating the study participation link. All participants reached the survey through a Qualtrics link, being informed only that the study was about political perspectives. This approach is similar to that used in a recent multi-country replication study[24] which yielded a diverse sample.

In Sudan, participants were recruited in the ways described as well as through a non-profit organization that is involved in community service projects (see Discussion for a description of potential bias). In Israel, an existing participant email circulation list was also used for recruiting, but there was no systematic bias for who was engaged through that list.

Procedure. The procedures for both experiments were equivalent to what is reported by Lees and Cikara[13], with the exception that we removed the 'Hypocrisy' condition from experiment 6 because it showed no additional effect above the main 'Truth' condition. We retained 'Obstruction' in experiment 6, but the term 'purposefully obstruct' (original version) created translation issues, thus several countries had to adapt the wording to articulate an intent to block or disrupt the action. Therefore, translations were as closely aligned to the original language as possible to describe an active form of interfering with the policy that is intentional. Additionally, for experiment 6, we presented only one scenario per country, chosen as the one with the greatest inaccuracy in beliefs within country (that is, the greatest mean discrepancy between first-order perceptions and out-group meta-perceptions). We focused on these scenarios because the hypothesized mechanism of the intervention is reducing these discrepancies, so failure to find a significant effect of the intervention for scenarios where a discrepancy was not present would be theoretically uninformative. Additionally, testing five scenarios in each country would require sample sizes that we could not realistically obtain.

Using these instruments, participants in experiment 4 began by choosing a group affiliation, which determined their side of the fault line. They were then randomly assigned between subjects to one of three conditions: first-order perceptions, in-group meta-perceptions or group meta-perceptions. They then read all five scenarios (in randomized order) and provided three ratings (dislike, oppose and unacceptable, in randomized order) for each, the phrasing of which depended on condition (first-order perception versus in-group meta-perception versus out-group meta-perception). After all scenarios were rated, participants provided basic demographic information (age, gender, education level and income). Experiment 4 took between 4 and 7 min on average, but this time was partially exaggerated due to lengthy consent forms required due to the multinational nature of the study.

For experiment 6, participants first provided three ratings (order randomized) for the scenario with the largest difference between out-group meta-perceptions and first-order perceptions in experiment 4. After the ratings, intervention group participants were shown the mean first-order perception of the out-group from experiment 4, and then gave a single additional rating for obstruction, followed by demographic indicators. Control group participants saw their own scores in the intermediate page but were not shown the first-order ratings of the out-group. Participants needed approximately 2 min to complete experiment 6.

For this study, ratings were given on a 1–100 scale, in line with the original study. While such granularity may be unnecessary and lack substantive differences in some ranges (for example, there is no meaningful distance for disliking something at 41 compared with 46) and it may generate noise around mean results, we felt it was best to maintain consistency of measurement. This was also optimal as GMPs run a likelihood of ceiling effects, which would limit granularity (that is, variability of interest) if reduced to a Likert-type scale of 1–5, 1–7 or even 1–10. Additionally, this scale is also useful for providing the necessary sensitivity for the intervention component in experiment 6. As such, it will also absorb or mirror the same noise that would be present in experiment 4.

Translation. Once issues were confirmed for all countries, forward- and back-translations were applied within teams. In this way, all issues were translated to the local language by a native speaker, back-translated by a different, independent reviewer and assessed for consistency with other scenarios. Since there would be variation in issues between countries, it was not necessary that all scenarios be perfectly identical, but we required that they must have been perceived as presented identically in terms of tone in how policy is presented. More

emphasis was placed on how each of the rating items aligned with the original version.

For English-speaking countries, teams were required to have local language checks to ensure phrasing and verbiage was not perceived as inherently American.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data will be publicly available at http://osf.io/fb8wu/ from 9 November 2022. Delayed release has been chosen to avoid the potential for their use or misuse in current election cycles. Data are available upon request from the corresponding authors at any time.

Code availability

All code is available at http://osf.io/fb8wu/.

Received: 5 September 2020; Accepted: 5 March 2021;
Published online: 22 April 2021

References

1. Méndez, E., Gómez, Á. & Tropp, L. R. When meta-perceptions are affected by intergroup processes. *Int. J. Psychol. Psychol. Ther.* **7**, 237–250 (2007).
2. Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B. & Bruneau, E. Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc. Natl Acad. Sci. U. S. A.* **117**, 14864–14872 (2020).
3. O'Brien, T. C., Leidner, B. & Tropp, L. R. Are they for us or against us? How intergroup meta-perceptions shape foreign policy attitudes. *Group Process. Intergroup Relat.* **21**, 941–961 (2018).
4. Westfall, J., Van Boven, L., Chambers, J. R. & Judd, C. M. Perceiving political polarization in the United States: party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspect. Psychol. Sci. A* **10**, 145–158 (2015).
5. Nyhan, B. & Reifler, J. The roles of information deficits and identity threat in the prevalence of misperceptions. *J. Elect. Public Opin. Parties* **29**, 222–244 (2018).
6. Nyhan, B. Facts and myths about misperceptions. *J. Econ. Perspect.* **34**, 220–236 (2020).
7. Shi, F., Teplitskiy, M., Duede, E. & Evans, J. A. The wisdom of polarized crowds. *Nat. Hum. Behav.* **3**, 329–336 (2019).
8. Shapiro, J. M. & Taddy, N. M. Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech. NBER Working Paper 22423 https://www.gsb.stanford.edu/sites/gsb/files/politext.pdf (Stanford Graduate School of Business, 2015).
9. Luguri, J. B. & Napier, J. L. Of two minds: the interactive effect of construal level and identity on political polarization. *J. Exp. Soc. Psychol.* **49**, 972–977 (2013).
10. Schaeffer, K. Far more Americans see 'very strong' partisan conflicts now than in the last two presidential election years. *Pew Research Center.* https://www.pewresearch.org/fact-tank/ 2020/03/04/far-more-americans-see-very- strong-partisan-conflicts-now-than-in- the-last-two-presidential-election-years/ (4 March 2020).
11. Duro, J. & Teixidó-Figueras, J. World polarization in carbon emissions, potential conflict and groups: an updated revision. *Energy Policy* **74**, 425–432 (2014).
12. Wojcieszak, M. & Price, V. What underlies the false consensus effect? How personal opinion and disagreement affect perception of public opinion. *Int. J. Public Opin. Res.* **21**, 25–46 (2009).
13. Lees, J. & Cikara, M. Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat. Hum. Behav.* **4**, 279–286 (2020).
14. Vorauer, J. D., Main, K. J. & O'Connell, G. B. How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *J. Personal. Soc. Psychol.* **75**, 917–937 (1998).
15. Flynn, D. J., Nyhan, B. & Reifler, J. The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. *Polit. Psychol.* **38**, 127–150 (2017).
16. Finchilescu, G. Intergroup anxiety in interracial interaction: the role of prejudice and metastereotypes. *J. Soc. Issues* **66**, 334–351 (2010).
17. Klein, O. & Azzi, A. E. The strategic confirmation of meta-stereotypes: how group members attempt to tailor an out-group's representation of themselves. *Br. J. Soc. Psychol.* **40**, 279–293 (2001).
18. Lau, T., Morewedge, C. K. & Cikara, M. Overcorrection for social-categorization information moderates impact bias in affective forecasting. *Psychol. Sci.* **27**, 1340–1351 (2016).
19. Ensari, N. & Miller, N. The out-group must not be so bad after all: the effects of disclosure, typicality, and salience on intergroup bias. *J. Personal. Soc. Psychol.* **83**, 313–329 (2002).
20. Carothers, T. & O'Donohue, A. *Democracies Divided: The Global Challenge of Political Polarization* (Brookings Institution Press, 2019).
21. Obaidi, M., Kunst, J. R., Kteily, N., Thomsen, L. & Sidanius, J. Living under threat: mutual threat perception drives anti-Muslim and anti-Western hostility in the age of terrorism. *Eur. J. Soc. Psychol.* **48**, 567–584 (2018).
22. Obaidi, M., Thomsen, L. & Bergh, R. "They think we are a threat to their culture": meta-cultural threat fuels willingness and endorsement of extremist violence against the cultural outgroup. *Int. J. Confl. Violence* **12**, 1–13 (2018).
23. Ruggeri, K. et al. Replicating patterns of prospect theory for decision under risk. *Nat. Hum. Behav.* **4**, 622–633 (2020).
24. Klein, R. A. et al. Many Labs 2: investigating variation in replicability across samples and settings. *Adv. Methods Pract. Psychol. Sci.* **1**, 443–490 (2018).
25. Smithson, M. & Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* **11**, 54 (2006).
26. Brooks, M. E. et al. GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* **9**, 378–400 (2017).
27. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411 (2018).
28. Livingstone, A. G., Fernández Rodríguez, L. & Rothers, A. "They just don't understand us": the role of felt understanding in intergroup relations. *J. Pers. Soc. Psychol.* **119**, 633–656 (2020).
29. Lees, J. & Cikara, M. Understanding and combating misperceived polarization. *Phil. Trans. R. Soc. B* **376**, 20200143 (2021).
30. Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. Fact-checking: a meta-analysis of what works and for whom. *Polit. Commun.* **37**, 350–375 (2020).

Acknowledgements

The authors thank N. Altinok, L. N. Jenudi, E. Kheir, D. Fyhr, M. K. Knock, B. Elnaiem, A. Abdalla, M. Abd El Hadi, J. Haidar, D. Tomić, J. Azevedo and I. Benslimane. The authors also thank the Global Scholars Program, Undergraduate Global Engagement and the Provost's Office at Columbia University, which provided indirect funds for the training component preceding the study and partial funding for data collection. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors also thank M. Booth, C. Hart, F. Ju and C. Marvin.

Author contributions

K.R. conceived the study and was responsible for all aspects of design, preparation, management, implementation, analysis, writing and revision. B.V. and L.B. were responsible for initial design and team coordination, as well as part of the country teams. T.L.A., S.A.-J., N.A., P.B.A., M.L.B., L.D.B., A.B., V.B., M.Č., M.Ç., G.C., A.C.-B., K.D., T.M.D., M.E., C.E.-S., E.N.F., M.F., D.I.F.V., P.G., E.G.-G., S.J.G., L.G., A.Gi., Al.Gr., An.Gr., M.G., M.H., K.H., B.H., W.I., L.J., M.J., H.J., O.K., K.N.K., R.K., P.R.K., Ž.L., S.L., I.S.L., S.Mam., S.Mar., M.F.M., L.M., S.M.-I., B.M., T.S.M., M.M., J.N., T.O., A.F.P., I.P., M.B.P., D.P., G.M.P., J.R., I.San., V.S., I.San., N.S., J.S., I.Soy., A.L.T., M.R.T., M.Vd., J.V., M.Vo., M.A.V., L.Wa., L.Wi., M.W., E.W., A.Y., S.Z., M.A.M. and T.D. were each part of country teams responsible for issue design, translation, implementation and local data collection. M.C. and J.L. were senior advisers responsible for the original study and input on all aspects of the study. T.F. was responsible for general design and all analyses.

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/s41562-021-01092-x.

Correspondence and requests for materials should be addressed to K.R.

Peer review information *Nature Human Behaviour* thanks Gillian Finchilescu, Thomas O'Brien, Steven Webster and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

[1]Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, NY, USA. [2]Policy Research Group, Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, UK. [3]Department of Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, the Netherlands. [4]Department of Psychology and Mental Health, University of Manchester, Manchester, UK. [5]PPR Svendborg, Svendborg Kommune, Svendborg, Denmark. [6]Department of Experimental Psychology, University of Oxford, Oxford, UK. [7]Department of Political Science, School of General Studies, Columbia University, New York, NY, USA. [8]Faculty of Psychology, University of Seville, Seville, Spain. [9]Eötvös Loránd University, Budapest, Hungary. [10]Department of Psychology, University of Oslo, Oslo, Norway. [11]Department of Psychology, Loyola University Chicago, Chicago, IL, USA. [12]Department of Occupational, Economic and Social Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria. [13]Centre for Behaviour Change, Faculty of Brain Sciences, University College London, London, UK. [14]Carnegie School of Sport, Leeds Beckett University, Leeds, UK. [15]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [16]Department of Social Psychology, Institute of Work, Organisational and Social Psychology, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany. [17]Institute of Psychology, Faculty of Social and Political Sciences, University of Lausanne, Lausanne, Switzerland. [18]Department of Psychology, Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia. [19]Department of Psychology, Columbia College, Columbia University, New York, NY, USA. [20]Barnard College, Columbia University, New York, NY, USA. [21]Division of Psychology and Language Sciences, University College London, London, UK. [22]Department of Psychology, School of General Studies, Columbia University, New York, NY, USA. [23]Department of Psychology, Faculty of Arts and Sciences, University of Puerto Rico, Mayagüez, Puerto Rico, USA. [24]Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia. [25]School of Education and Health Sciences, Universidad Camilo José Cela, Villanueva de la Cañada, Madrid, Spain. [26]UCL Interaction Centre, Faculty of Brain Sciences, University College London, London, UK. [27]Department of Neuroscience and Psychology, Columbia College, Columbia University, New York, NY, USA. [28]School of General Studies, Columbia University, New York, NY, USA. [29]Political Humanities, Euro-Asia Program, Paris Institute of Political Studies [SciencesPo], Paris, France. [30]Department of Psychology, Faculty of Life Sciences, Humboldt University of Berlin, Berlin, Germany. [31]Social, Health and Organisational Psychology, Utrecht, Utrecht University, Utrecht, the Netherlands. [32]National Institute of Mental Health, Klecany, Czech Republic. [33]Third Faculty of Medicine, Charles University, Prague, Czech Republic. [34]Department of Organization Studies, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, the Netherlands. [35]Department of Psychology, University of Cambridge, Cambridge, UK. [36]Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA. [37]Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. [38]Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia. [39]Department of Psychology, Faculty of Psychology and Educational Sciences, University of Porto, Porto, Portugal. [40]Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK. [41]Instituto de Ciências Sociais e Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal. [42]Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, the Netherlands. [43]Department of Psychology, University of Warwick, Coventry, UK. [44]Evidence-Based Practice Unit, Anna Freud National Centre and University College London, London, UK. [45]Department of Social and Organizational Psychology, University of Groningen, Groningen, the Netherlands. [46]Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy. [47]Department of Psychology, Faculty of Social Sciences, Uppsala University, Uppsala, Sweden. [48]Department of Social, Economic, and Organizational Psychology, Leiden University, Leiden, the Netherlands. [49]Department of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia. [50]Department of Psychology, Faculty of Croatian Studies, University of Zagreb, Zagreb, Croatia. [51]Bezirkskrankenhaus Straubing, Straubing, Germany. [52]Ivo Pilar Institute of Social Sciences, Zagreb, Croatia. [53]Department of Social Policy and Intervention, University of Oxford, Oxford, UK. [54]Faculty of Business Administration, University of Economics, Prague, Czech Republic. [55]Department of Psychology, Faculty of Science, University of Tübingen, Tübingen, Germany. [56]Department of Psychology and Neuroscience, Faculty of Arts, University of St Andrews, St Andrews, UK. [57]Department of Developmental and Educational Psychology, University of Vienna, Vienna, Austria. [58]Department of Psychology, Faculty of Media and Communications, Singidunum University, Belgrade, Serbia. [59]Department of Psychology, Faculty of Arts, Charles University, Prague, Czech Republic. [60]Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia. [61]Charles University, Prague, Czech Republic. [62]Department of Psychology, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland. [63]Department of Cognitive Science and Psychology, School of Graduate Studies, New Bulgarian University, Sofia, Bulgaria. [64]Department of Psychology, St. Olaf College, Northfield, MN, USA. [65]Columbia University, New York, NY, USA. [66]Department of Psychology, Harvard University, Cambridge, MA, USA. [67]Harvard Business School, Harvard University, Cambridge, MA, USA. [68]Department of Economics, Clemson University, Clemson, SC, USA. [69]unaffiliated. ✉e-mail: kai.ruggeri@columbia.edu

nature research

Corresponding author(s): Kai Ruggeri

Last updated by author(s): Feb 14, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected through the online platform Qualtrics XM.
Data analysis	All data were analyzed using R version 3.6.1 for 64bit Windows. Data cleaning and preprocessing were conducted with Tidyverse (1.2.1). Maximum Likelihood beta-regressions were fitted using glmmTMB (1.0.0), Bayesian beta regression were fit using brms (2.10.0). All figures were plotted using ggplot2 (3.3.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data will be publicly available at osf.io/fb8wu/ from November 9, 2022. Delayed release has been chosen to avoid the potential for their use or misuse in current election cycles. Data are available upon request from the corresponding authors at any time.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a quantitative experimental study with two experiments. The first experiment (labeled Experiment 4 to match the study our methods are based on) involved three experimental groups (first-oder perception, ingroup meta-perception and outgroup meta-perception), across two observational group (fault lines, e.g. left vs right political orientation), resulting in 6 between-participant cells per country, with a total of 26 countries. Each participant provided 3 attitude ratings for 5 country-specific scenarios, resulting in 15 DV observations per participant. The second experiment (labeled Experiment 6 to match the study our methods are based on) involved two experimental groups (control or disclosure), across two observational group (fault lines, e.g. left vs right political orientation) resulting in 4 between-participants cells per country, with 10 countries in total. Each participants gave a single obstruction rating, creating a single dv observation per participant.
Research sample	The research sample consisted of residents of the 26 target countries for Experiment 4 or one of the 10 target countries for Experiment 6, who identified with one of the fault line groups of that country. We do not have reliable enough population data to determine whether the samples are representative of the fault line groups but both samples include a wide range of ages and are approximately gender balanced (see Table 2 in the manuscript for country breakdowns).
Sampling strategy	Unpaid participants were recruited through various online forums. Paid participants were recruited through Prolific.co. Sample sizes were determined numerically by resampling with replacement from a previous study using the same design in a single country, thus determining what sample size would be required to reliably detect an effect. See the preregistration for further details (https://osf.io/w62fs/).
Data collection	All data collection was conducted via the Online survey platform Qualtrics.
Timing	Data collection took place between the 30th of July and the 30th of August, 2020.
Data exclusions	Participants were excluded if they terminated the survey before having been assigned an experimental condition, or they reported belonging to neither fault line group in their country, or they failed to give informed consent (including consent to GDPR in European countries), or their Qualtrics ID indicated that they were part of the research team. All other observations were included in the analysis.
Non-participation	9802 participants were excluded from Experiment 4 for the reasons outlined. 3503 Participants were excluded from Experiment 6.
Randomization	Participants were randomly allocated to experimental groups for both experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Our target population is very diverse since we wanted to capture residents of all of 26 countries who identified with with one of the fault line groups in their country of residence. We wanted to have a roughly gender-balanced sample, and wide range of ages, both of which were achieved.
Recruitment	Most recruitment came from informal posts on social media, similar to "If you are interested in the topic of polarization and political beliefs, we strongly suggest you take this survey [link to Survey]". Some countries involved paid samples recruited through prolific.co.
Ethics oversight	All testing was approved by the Columbia University Irving Medical Center IRB. To receive approval for non-US samples, all countries had to produce a letter confirming cultural and linguistic appropriateness of all measures.

Note that full information on the approval of the study protocol must also be provided in the manuscript.