

The general fault in our fault lines

Kai **Ruggeri** (Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, NY, USA; Policy Research Group, Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, UK)

Bojana **Večkalov** (Department of Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, North Holland, The Netherlands)

Lana **Bojanić** (Department of Psychology and Mental Health, University of Manchester, Manchester, UK)

Thomas L. **Andersen** (PPR Svendborg, Svendborg Kommune, Svendborg, Denmark)

Sarah **Ashcroft-Jones** (Department of Experimental Psychology, University of Oxford, Oxford, Oxfordshire, United Kingdom)

Nélida **Ayacaxli** (School of General Studies, Columbia University, New York, NY, USA)

Paula **Barea Arroyo** (Faculty of Psychology, University of Seville, Seville, Andalusia, Spain)

Mari Louise **Berge** (Eötvös Loránd University, Budapest, Hungary)

Ludvig D. **Bjørndal** (Department of Psychology, University of Oslo, Oslo, Norway)

Aslı **Bursalioğlu** (Department of Psychology, Loyola University Chicago, Chicago IL, USA)

Vanessa **Bühler** (Department of Occupational, Economic and Social Psychology, Faculty of Psychology, University of Vienna)

Martin **Čadek** (Carnegie School of Sport, Leeds Beckett University)

Melis **Çetincelik** (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands)

Georgia **Clay** (Department of Social Psychology, Institute of Work, Organisational and Social Psychology, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany)

Anna **Cortijos-Bernabeu** (Institute of Psychology, Faculty of Social and Political Sciences, University of Lausanne, Lausanne, Vaud, Switzerland)

Kaja **Damjanović** (Department of Psychology, Laboratory for experimental psychology, University of Belgrade, Belgrade, Serbia)

Tatianna M. **Dugue** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)

Maya **Esberg** (Barnard College, Columbia University, New York, NY, USA)

Celia **Esteban-Serna** (Division of Psychology & Language Sciences, University College London, London, UK)

Ezra N. **Felder** (Department of Psychology, School of General Studies, Columbia University, New York, NY, USA)

Maja **Friedemann** (Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom)

Darianna I. **Frontera Villanueva** (Department of Psychology, Faculty of Arts and Sciences, University of Puerto Rico, Mayagüez, Puerto Rico, USA)

Patricia **Gale** (Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia)

Eduardo **Garcia-Garzon** (School of Education and Health Sciences, Universidad Camilo José Cela, Villanueva de la Cañada, Madrid, Spain)

Sandra J. **Geiger** (Department of Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, North Holland, The Netherlands)

Leya **George** (UCL Interaction Centre, Faculty of Brain Sciences, University College London, London, UK)

Aleksandra **Gracheva** (Political Humanities, Euro-Asia Program, Paris Institute of Political Studies [SciencesPo], Paris, France; Columbia School of General Studies, Columbia University, New York, NY, USA)

Anastasia **Gracheva** (Department of Political Science, Columbia School of General Studies, Columbia University, New York, NY, USA; Political Humanities, Euro-Asia Program, Paris Institute of Political Studies [SciencesPo], Paris, France)

Marquis **Guillory** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)

Marlene **Hecht** (Department of Psychology, Faculty of Life Sciences, Humboldt University of Berlin, Berlin, Germany)

Katharina **Herte** (Social, Health and Organisational Psychology, Utrecht, Utrecht University, The Netherlands)

Barbora **Hubená** (unaffiliated)

William **Ingalls** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)

Lea **Jakob** (National Institute of Mental Health, Klecany, Czech Republic; Third Faculty of Medicine, Charles University, Prague, Czech Republic)

Margo **Janssens** (Department of Organization Studies, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, North Brabant, The Netherlands)

Hannes **Jarke** (Policy Research Group, Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, UK)

Ondřej **Kácha** (Department of Psychology, University of Cambridge, Cambridge, United Kingdom)

Kalina **Nikolova Kalinova** (unaffiliated)

Ralitsa **Karakasheva** (unaffiliated)

Peggh R. **Khorrami** (Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA; Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA)

Žan **Lep** (Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia)

Samuel **Lins** (Department of Psychology, Faculty of Psychology and Educational Sciences, University of Porto, Porto, Portugal)

Ingvild S. **Lofthus** (Department of Psychology, University of Oslo, Oslo, Norway)

Salomé **Mamede** (Department of Psychology, Faculty of Psychology and Educational Sciences, University of Porto, Porto, Portugal)

Silvana **Mareva** (Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK)

Mafalda F. **Mascarenhas** (Instituto de Ciências Sociais e Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal)

Lucy **McGill** (Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands)

Sara **Morales-Izquierdo** (Department of Psychology, University of Warwick, Coventry, United Kingdom)

Bettina **Moltrecht** (Evidence-Based Practice Unit, Anna Freud National Centre & University College London, London, UK)

Tasja S. **Mueller** (Department of Social and Organizational Psychology, University of Groningen, Groningen, the Netherlands)

Marzia **Musetti** (Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy)
 Joakim **Nelsson** (Department of Psychology, Faculty of Social Sciences, Uppsala University, Uppsala, Sweden)
 Thiago **Otto** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)
 Alessandro F. **Paul** (Department of Social, Economic, and Organizational Psychology, Leiden University, Leiden, The Netherlands)
 Irena **Pavlović** (Laboratory for experimental psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia)
 Marija B. **Petrović** (Department of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia)
 Dora **Popović** (Department of Psychology, Faculty of Croatian Studies, University of Zagreb, Zagreb, Croatia)
 Gerhard M. **Prinz** (Bezirkskrankenhaus Straubing, Straubing, Germany)
 Josip **Razum** (Ivo Pilar Institute of Social Sciences, Zagreb, Croatia)
 Iyaylo **Sakelariiev** (unaffiliated)
 Vivian **Samuels** (Department of Psychology, School of General Studies, Columbia University, New York, NY, USA)
 Inés **Sanguino** (Department of Social Policy and Intervention, University of Oxford, Oxford, Oxfordshire, United Kingdom)
 Nicolas **Say** (Department of Psychology, Faculty of Arts, Charles University, Prague, Czech Republic)
 Jakob **Schuck** (Department of Psychology, Faculty of Science, University of Tübingen, Tübingen, Baden-Württemberg, Germany)
 Irem **Soysal** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)
 Anna Louise **Todsen** (Department of Psychology and Neuroscience, Faculty of Arts, University of St Andrews, St Andrews, UK)
 Markus R. **Tünte** (Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Austria)
 Milica **Vdovic** (Department of psychology, Faculty of Media and Communications, Singidunum University, Belgrade, Serbia)
 Jáchym **Vintr** (Department of Psychology, Faculty of Arts, Charles University, Prague, Czech Republic)
 Maja **Vovko** (Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia)
 Marek A. **Vranka** (Charles University, Prague, Czech Republic)
 Lisa **Wagner** (Department of Psychology, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland)
 Lauren **Wilkins** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)
 Manou **Willems** (unaffiliated)
 Elizabeth **Wisdom** (Department of Psychology, School of General Studies, Columbia University, New York, NY, USA)
 Aleksandra **Yosifova** (Department of Cognitive Science and Psychology, School of Graduate Studies, New Bulgarian University, Sofia, Sofia City, Bulgaria)
 Sandy **Zeng** (Department of Psychology, Columbia College, Columbia University, New York, NY, USA)
 Mahmoud A. **Ahmed** (Department of Psychology, St Olaf College, Northfield, MN, USA)
 Twinkle **Dwarkanath** (Columbia University, New York, NY, USA)
 Mina **Cikara** (Department of Psychology, Harvard University, Cambridge, MA, USA)
 Jeffrey **Lees** (Harvard Business School, Harvard University, Cambridge, MA, USA; Department of Economics, Clemson University, Clemson, SC, USA);
 Tomas **Folke** (Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, NY, USA)

Abstract

A pervading global narrative suggests that political polarisation is increasing in the US and around the world. Beliefs in increased polarisation impact individual and group behaviours regardless of whether they are accurate or not. One driver of polarisation are beliefs about how members of the out-group perceive us, known as group meta-perceptions. A 2020 study by Lees and Cikara in US samples suggests that not only are out-group meta-perceptions highly inaccurate, but informing people of this inaccuracy reduces negative beliefs about the out-group. Given the importance of these findings for understanding and mitigating polarisation, it is essential to test to what extent they generalise to other countries. We assess that generalisability by replicating two of the original experiments in 10,207 participants from 26 countries in the first experiment and 10 in the second. We do this by studying local group divisions, which we refer to as fault lines. In line with our hypotheses, results show that the pattern found in the US broadly generalises, with greater heterogeneity explained by specific policies rather than between-country differences. The replication of a simple disclosure intervention in the second experiment yielded a modest reduction in negative motive attributions to the out-group, similar to the original study. These findings indicate first that inaccurate and negative group meta-perceptions are exhibited in a large number of countries, not only the US, and that informing individuals of their misperceptions can yield positive benefits for intergroup relations. The generalisability of these findings highlights a robust phenomenon with major implications for political discourse worldwide.

Main

Beliefs about polarisation have significant social consequences, whether accurate or not (Mendez, 2007; Moore-Berg et al., 2020; O'Brien, 2017; Westfall et al., 2015). 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 (Nyhan & Reifler, 2018). This can result in significant distortions in numerous behaviours, including health, voting, and financial choices, each of which has major consequences on population well-being (Nyhan, 2016).

There is growing interest in the origins of polarisation across populations as well as its features and impacts across communities and society (Shi, Teplitskiy, & Duede, 2019; Shapiro & Taddy, 2015). This interest is global and includes both scientific research as well as general public interest (Luguri & Napier, 2013; Schaeffer, 2020). As polarisation seemingly permeates a growing number of personal and public domains, there is an increasing urgency to understand it and its effects more deeply (Duro, 2014; Wojcieszak, 2009). Thus, it is vital to empirically evaluate the extent of polarisation on community and global levels.

However, with increasing interest in polarisation itself, broadening evidence indicates that inaccuracies in perceptions of how the out-group feels about the in-group can be harmful (Lees & Cikara, 2020). The origin of these “meta-perceptions” may be rooted in negative stereotypes that individuals feel have been applied to them, often incorrectly (Vorauer et al., 1998). This results in an inaccurate perception of differences in beliefs and attitudes between groups (Flynn et al., 2017), which can have negative results for individuals (Finchilescu, 2010; Klein & Azzi, 2001). On a population level, such misperceptions can even result in overstated reactions that exacerbate existing biases (Lau et al., 2016). Furthermore, there are indications that overly negative group meta-perceptions in a global context may lead to more aggressive foreign policy (O'Brien, 2017; Hermann & Keller, 2004; Kteily et al., 2016). Therefore,

interventions that can mitigate such negative perceptions might make compromises more feasible and attractive, both internationally and domestically.

To investigate roots and moderators of polarisation across groups, Lees and Cikara (2020) ran a series of experiments with US participants identifying as Republican or Democrat. In what they refer to as “group meta-perception” (GMP), participants estimated how their own out-group perceived the behavior 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 either their own perception of the out-group’s political action (“individual rating”), their estimate of fellow in-group member’s 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 polarisation both in terms of 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 if 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). 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 (Ensari & Miller, 2002). The greatest reduction in bias was found for those with the most inaccurate pre-intervention meta-perceptions, indicating potential value specifically toward addressing false polarisation narratives, by encouraging accuracy through greater disclosure about beliefs rather than emphasising perceived disagreements.

Most work on group meta-perceptions has been done in US samples. For example, Moore-Berg et al. (2020) 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 polarisation narrative in a representative US sample by presenting a robust argument that true differences are often overestimated, yet the US is far from the only country with alleged increases in polarisation (Carothers & Donahue, 2019). Additionally, though it is relatively simple to conceptualise polarisation in a two-party system, such as the United States, in-group/out-group divisions are not limited to political affiliations.

Testing the generalisability of Lees and Cikara’s findings specifically would therefore have major implications for empirically quantifying the global accuracy of group meta-perceptions, given the purported increases in polarisation. If the results were to generalise, 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 polarisation may be overstated in the US, 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 mobilise a global network of researchers to test the generalisability and replicability of group meta-perceptions in 26 countries. The purpose is to determine the extent to which study conclusions made by Lees and Cikara (2020) apply 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 generalisability of a recent finding with strong policy implications. That this relates to polarisation and inaccurate group perceptions - critical topics in science and society - makes it even more meaningful to attempt in a highly powered, robust, multi-country study.

Importantly, 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, a major 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 (Ruggeri et al., 2020). This observation highlights replication as a crucial tool for examining the generalisability of behavioral effects (Klein et al., 2018). In the context of investigating out-group meta-perceptions cross-culturally, the potential differences—not only in the presence (or lack thereof) 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 polarisation creates inaccurate out-group meta-perceptions. Additionally, it will indicate whether and to what degree inaccurate out-group meta-perceptions can be mitigated.

METHODS

This study followed the published and pre-registered protocols of Lees and Cikara (2020) 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 will be explained later, 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 generalisability 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 US, which would suggest that polarisation is less extreme than what is commonly assumed. As such, we emphasise generalisability over replication for this body of work.

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 generalisability of the original findings in the US, we did not aim to include or exclude any specific countries. Instead, we focused on countries that members of the research network were familiar with, though we did aim for including at least one country from every continent. Our pre-registered sample size targets, based on bootstrapped simulations of the original study results, 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 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 group meta-perception (GMP) ratings (unacceptable, dislike, 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 (e.g., 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 (i.e., group distinction and classification creating the intergroup boundary) are provided in Supplement A. Supplement B 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 (i.e., how do you perceive this out-group action, called “actual perceptions” by Lees & Cikara), in-group meta-perception (i.e., how does your group [Democrats/Republicans] perceive this out-group action?), and out-group meta-perception (i.e., 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 if 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 provided in Table 1. 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 US (strongly associated with Republican party members) or legalisation of recreational drugs in European countries (strongly associated with members of liberal policy parties).

Table 1. Example issue development from original study to country adaptations.

	Issue	Fault line
Original version	Redrawing congressional maps for voting districts	Democrat/Republican
Adaptation example	Changing the number of seats in Parliament	Ruling coalition/Opposition coalition

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 an American 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 operationalised. 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 generalised 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 (i.e., Israeli and Arab) or geopolitical alignments (i.e., pro-Europe and pro-Russia).

Procedure

The procedures for both experiments were equivalent to what is reported in Lees and Cikara (2020), with the exception that we removed the Hypocrisy condition in 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 have been as closely aligned to the original language as possible to 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 (i.e., the greatest mean discrepancy between first-order perceptions and out-group meta-perceptions). We focused on these scenarios because the hypothesised 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 (randomised order) and provided three ratings (dislike, oppose, unacceptable; randomised order) for each, the phrasing of which depended on condition (first-order perception vs. in-group meta-perception vs. out-group meta-perception). After all scenarios were rated, participants provided basic demographic information (age, gender, education level, income). Experiment 4 took between 4 and 7 minutes 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 randomised) 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 minutes to complete Experiment 6.

For this study, ratings were given on a 1 to 100 scale in line with the original study. While such granularity may be unnecessary and lack substantive differences in some ranges (e.g., there is no meaningful distance for disliking something at a 41 compared to a 46) and it may generate noise around mean results, we felt it was important to maintain consistency of measurement. This was also important as GMPs run a likelihood of ceiling effects, which would limit granularity (i.e., important variability) if

reduced to a Likert-type scale of 1-5, 1-7, or even 1-10. Additionally, this scale is also useful at 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 are 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.

Research questions and hypotheses

Prior to this study, we had six primary research questions, three related to each experiment. These focused mainly on aggregate comparisons between first-order and both group (meta-)perceptions. In this section, we list the hypotheses associated with each question.

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 aggregated in-group meta-perceptions would be significantly more negative than first-order perceptions. Finally, we expected 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 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 in 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 section on Experiment 6 Modelling Strategy). 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

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 that were excluded based on not identifying with either group.

We did not rely on student sample pools for this study. Less than half of the sample were actively involved in any form of education (43.9% in Experiment 4 and 40.3% in Experiment 6), and only a third were full-time students (33.6% in Experiment 4 and 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 to most of our target populations (In Experiment 4, 55.7% had completed at least one university degree, compared to 51.6% for Experiment 6). Some participants reported ages that were implausibly high, to avoid these biasing analyses we capped the age variable at 75. 53 observations from Experiment 4 and 38 observations in Experiment 6 were capped in this manner. For a demographic overview of each country sample see Table 2. For a full demographic breakdown see Supplement C.

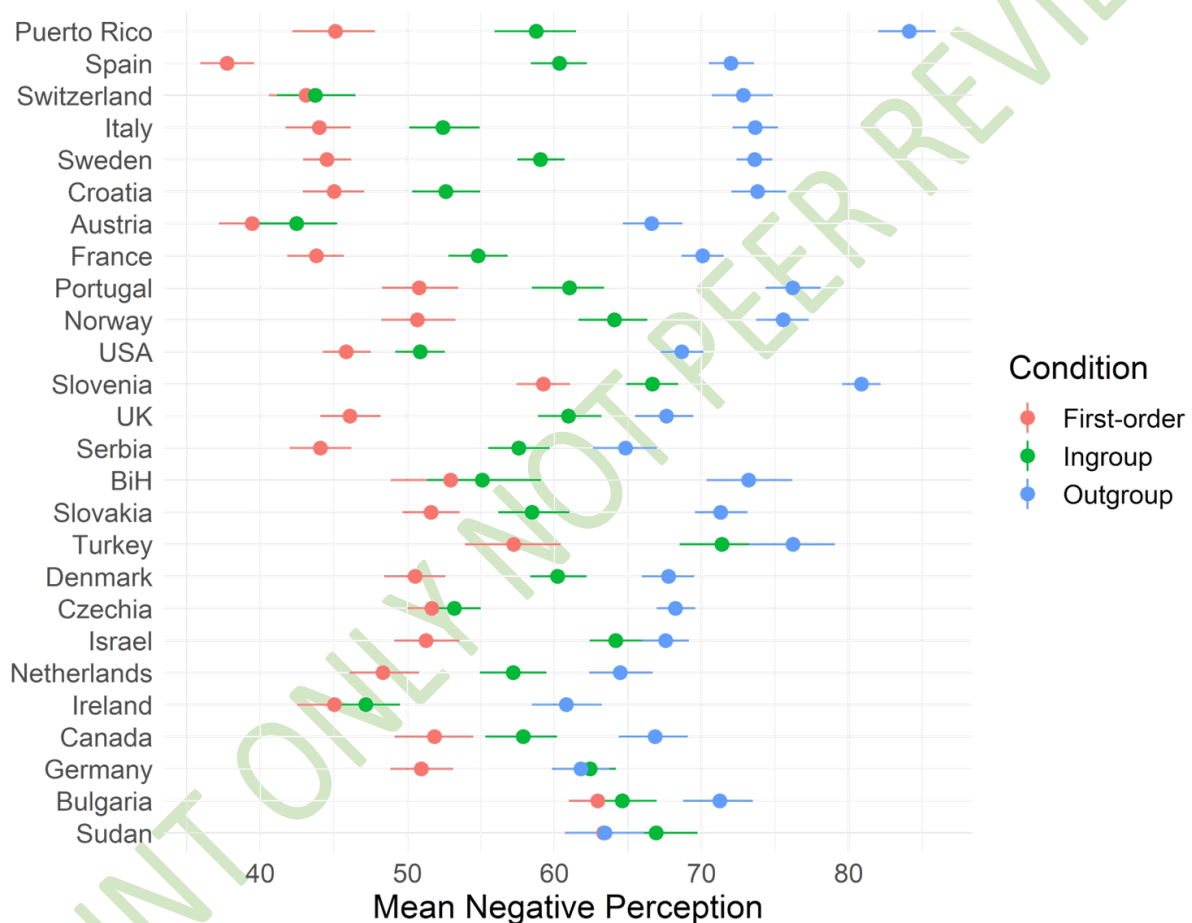
Table 2. Demographic Overview

Country	N	Experiment 4			N	Experiment 6		
		Mean Age	% Female	% University degree or above		Mean Age	% Female	% University degree or above
ALL	5406	34.5	43.5	56.8	4801	35.4	41.3	52.4
Austria	158	37.6	40.5	48.4				
Bosnia	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				
Czechia	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				
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				
UK	213	34.0	40.4	69.4				
USA	366	24.4	56.9	41.6	619	34.1	51.9	58.2

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 (See Figure 2).

Figure 2: Negative perceptions are consistently over-estimated for the out-group, for every country but Sudan.



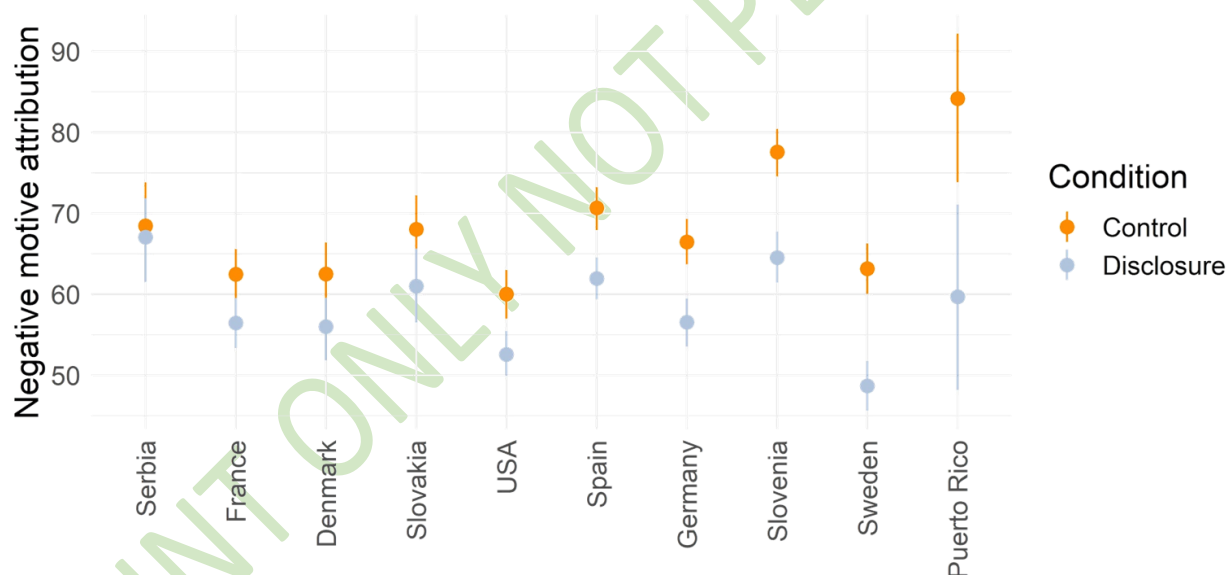
Negative perceptions by country and condition, collapsed across scenarios and rating types. Dots signify means, error bars signify bootstrapped 95% confidence intervals. Countries are ordered based on the difference in mean between the first-order ratings and out-group meta-perception ratings (from largest to smallest).

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 Discussion). 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 to the two other categories).

Additionally, there is no obvious pattern in the ordering of effect sizes. In other words, the variation in effect sizes seem 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, see Figure 3. The disclosure intervention is consistently associated with lower negative motive attributions than the control group, suggesting that the effect of the intervention generalises outside the US. Again we see no clear pattern in effect sizes. Crucially, sample size seems to be unrelated to effect size. To formally evaluate generalisability, 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 have included the pre-registered figure on negative motive attributions based on fault line groups in the supplement for completeness (see Supplement D).

Figure 3: 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.



Mean negative motive attribution of each condition by country. 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.

Pre-registered analyses

Experiment 4: Are meta-perceptions of the out-group's attitudes toward the in-group significantly more negative than the out-group's first-order perceptions?

In our pre-registration, we specified a model that closely reflected the approach of Lees and Cikara, extrapolated to a multi-country context. This model included fixed effects 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. We intended to fit this model using glmmTMB (Brooks et al., 2017), 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 (Bürkner, 2018) 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 DF, centered on 0 and with standard deviations of 10. For the beta-coefficients, we used normal distributions centered on 0, and with a standard deviation of 10. For the dispersion parameter we used a gamma prior with k and theta both set to 0.01.

We ran 4 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 closely matched the estimates from the pre-registered glmmTMB model, see Table 3. 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 has 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 operate directly on the samples drawn from the posterior distributions. For example, if we want to know if the in-group meta-perceptions are more severe than the first-order perceptions globally, we check what proportion of the posterior samples of the fixed-effect in-group parameter is greater than 0.

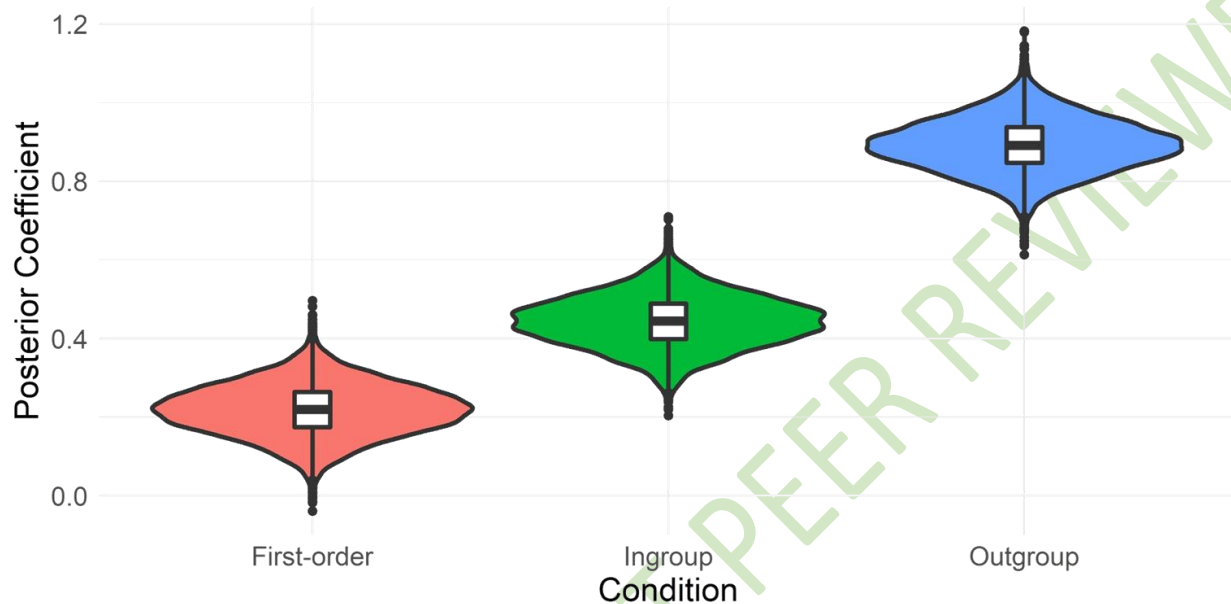
Table 3. Comparing parameter estimates between the preregistered restricted maximum likelihood model and the final, fully Bayesian model.

	Intercept	In-group parameter	Out-group parameter
REML Estimate	21.6	22.5	67.3
Bayesian Posterior Mean	21.9	22.5	67.3

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 show very little overlap (see Figure 1). Both in-group and out-group meta-perceptions are associated with more severe judgements than first-order perceptions ($\text{proportion}_{\text{in-group}} = 1$, $\text{proportion}_{\text{out-group}} = 1$). This means that our sample tends to overestimate how strongly their in-group and their out-group will react to actions taken by the opposition. Furthermore, out-group meta-perceptions are more severe than in-group meta-perceptions ($\text{proportion} = 1$). In other words, participants expect that their out-

group will react more strongly to policies proposed by their in-group than their in-group will react to policies proposed by their out-group. These results all align with what was reported by Lees and Cikara.

Figure 4: Posterior distributions for the population-level mean effects of the different conditions.



Posterior distributions of the marginal effects of the different conditions when accounting for variations in scenario, fault line group, and country.

Despite these robust effects at the population level, it is important to note that the model also estimates substantial between-country variation (Posterior $SD_{in-group} = 0.17$ [95% Credible Intervals = 0.10-0.25], Posterior $SD_{out-group} = 0.30$ [95% Credible Intervals = 0.21-0.41]).

Experiment 6: Does being informed of the true average first-order perception of the out-group reduce negative motive attribution toward the out-group?

For Experiment 6, we pre-registered 4 different models of increasing complexity. As before, 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, we predict obstruction ratings from the experimental conditions, with random intercepts and experimental conditions by country, and random intercepts by fault line groups. This model captures whether the intervention impacts obstruction ratings. In the second model, we add fixed effects for how inaccurate participant meta-perceptions were prior to the intervention, and an inaccuracy-intervention interaction term. This model measures whether belief inaccuracy predicts obstruction ratings, and whether the intervention impacts 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 adds some additional flexibility by allowing the effect of belief inaccuracy to vary between countries, and the final model allows 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 3 models on AIC (see Table 4).

Table 4. Experiment 6 model comparison

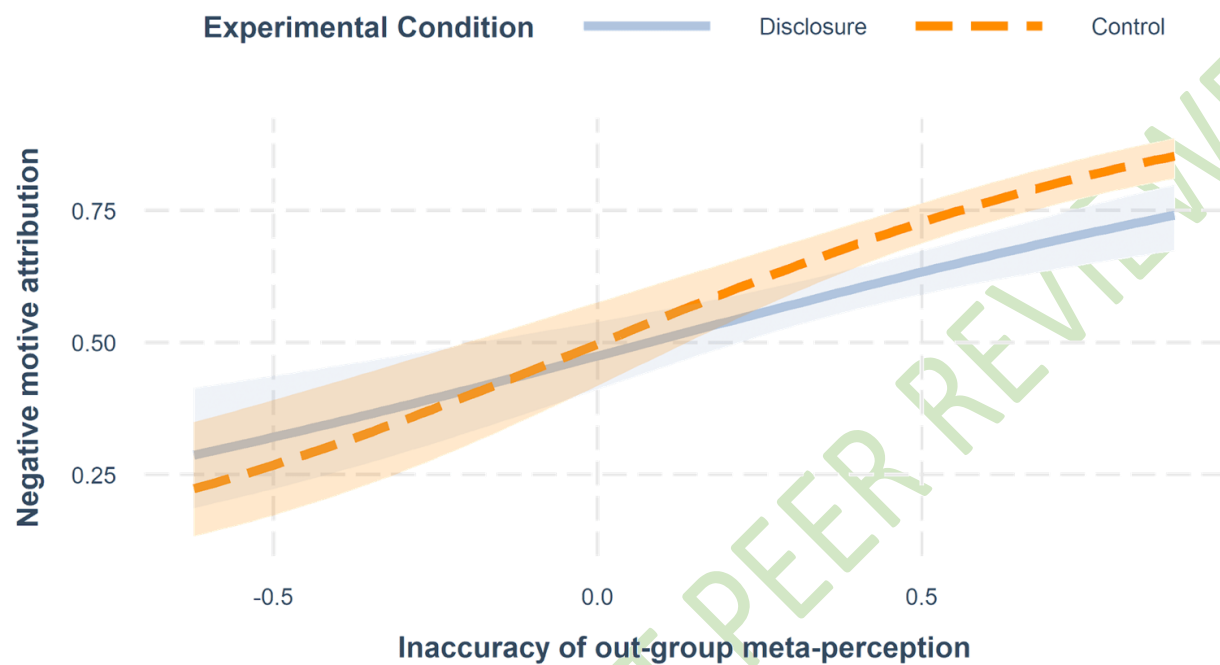
Model	Model Dfs	AIC
1	9	-4170.2
2	11	-5052.8
3	17	-5096.8

The third model provided the best fit to the data. For a summary of the fixed-effects of the winning model, see Table 5. Since the winning model included an interaction term between inaccuracy and the intervention, our pre-registered replication criterion was whether there was a significant (alpha threshold of .01) reduction in the belief inaccuracy on obstruction ratings after the disclosure intervention. This criterion is met, as the relevant p-value is less than .0001. In line with Lees and Cikara, we find that people who hold inaccurate beliefs about the negative opinions of the out-group give significantly higher obstruction ratings, but that this effect is mitigated when the disclosure is presented, see Figure 3. Though this interaction effect is convincingly significant at the population level there is considerable variation between countries ($SD = 0.36$) and fault line groups ($SD = 0.49$).

Table 5. Experiment 6 fixed-effects coefficients

Coefficient	Estimate (SE)	Z-value	p-value
Disclosure condition	-0.08(0.08)	-1.00	0.32
Meta-perception inaccuracy	2.48(0.19)	13.04	<0.0001
Disclosure x Inaccuracy	-0.69(0.14)	-5.05	<0.0001

Figure 5: Interaction effect between belief inaccuracy and conditions.



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.

Exploratory analyses

Because our interest was primarily in generalisability of the Lees and Cikara conclusions, 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 (i.e., are meta-perceptions 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 to what extent 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 US and thus could not inform a comparative component. In this exploratory analysis, we find three qualitatively distinct patterns of fault line group perceptions that repeat in multiple countries (see Figure 6).

In most countries, the perception of both fault line groups mirror each other, such that (for example) the Right's perception of itself closely matches the Left's perception of itself, and the Right's perception of the out-group matches 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 have similar first-order perceptions, and out-group meta-perceptions, but one fault line group perceives that their in-group has 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.

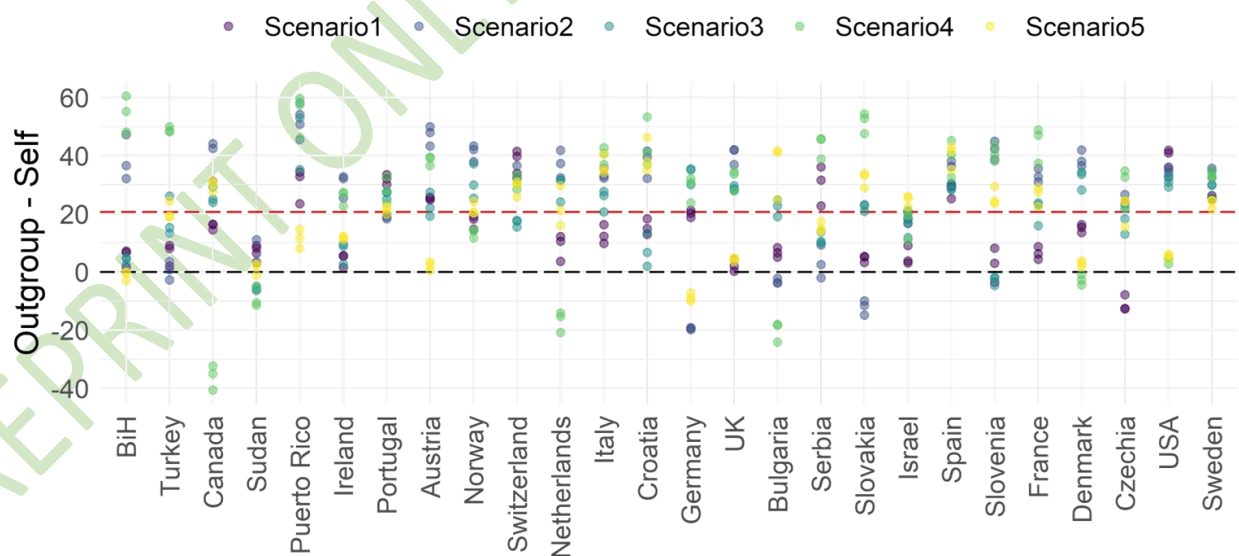
Mean negative perceptions delineated by fault line group, averaged across scenarios. Note that the colour coding refers to which group being evaluated. For example, for Austria, "Right Self" and "Right In-

group” both capture how the right-wing 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).

Sudan is showcasing this final pattern, with the important distinction that one fault line group actually *underestimates* how negatively the out-group would feel. Closer inspection reveals that the Sudanese results might depend on the specifics of the scenarios, as one fault line group finds 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 find to be particularly disagreeable (see Supplement E). We explain possible reasons for this in the Discussion, stressing the need for caution about over-interpretation.

Finally, to test the generalisability 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 (see Figure 7). 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 to 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 Figure 7 is the absence of a clear relationship between sample size and effect, further indicating the robustness and generalisability of the conclusion from the original study.

Figure 7. Reported effects are in the expected direction in 89% of the ratings.



The y-axis shows the difference between out-group meta-perceptions and first-order perceptions. The x-axis shows the 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 line indicates the null-hypothesis that there is no difference between first-order judgements and out-group meta-perceptions. The red line indicates the difference originally reported by Lees and Cikara. See Supplement A for scenario by country.

Discussion

The purpose of this study was to assess the generalisability of recent, important findings on inaccuracy in out-group meta-perceptions across myriad policy actions and political identities. We tested it by running a near-direct replication of the method from the original study in 26 countries, including the United States, where the original study was conducted. To do this, 5,406 participants viewed five scenarios in Experiment 4, providing three ratings for each (a total of 15) from either their own perception, their in-group meta-perception, or their out-group meta-perception. We also replicated the Experiment 6 disclosure intervention from the original study with an additional 4,801 participants from ten countries, using only the issue with the most extreme difference between first-order and out-group meta-perceptions. For both experiments, we found that results from the original study broadly generalise, and conclude the initial findings are valid across a wide range of settings.

(Not) only in America: Generalisability across countries and fault lines

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 similar effect sizes to the original study in Experiment 4. We also expected a majority of those countries would in turn show significant effects (likely 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 generalisability 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 emphasise the generalisability 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 the introduction.

To prevent individual scenarios from exaggerating or mitigating general patterns, the 15 ratings per country were aggregated to compare differences between the three conditions for every country. We find results from the original study largely generalise across 25 of the 26 countries included in this study, the exception being Sudan. We discuss potential reasons for this deviation later.

In-group meta-perceptions

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.

Effects of a disclosure intervention

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 10 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 generalisable at the country level.

Because we focused on a single scenario per country, we cannot distinguish between generalisability across issues and generalisability 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 greatest for individuals and issues where misperceptions are largest.

Deviations from pre-registered plan

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 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 faultline 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 skepticism 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 (see Supplement F).

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 those

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.

WEIRDing out

While it is tempting to interpret the outlier results of Sudan as a meaningful insight of a non-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 faultline 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.

Samples and effects

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, under-powered 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 have 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 less than 30 complete responses), so those have been omitted.

Involving original researchers in the replication

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. Recognising 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 (Lees & Cikara) 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.

Limitations

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 identities, and also resulted in the exclusion of a large number of potential participants that chose “other” for their political affiliation. It 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 to highlight that the findings should only be applied to groups that have clear identification parameters that explicitly stand in juxtaposition to another out-group. Shifting to a scalar classification system (even in the US, where the largest group identification tends to be “moderate”) would certainly be a welcomed 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 generalisations to out-groups that could be viewed as being in direct and polar opposition with an in-group.

Conclusion

We find the results of Lees and Cikara broadly generalise 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 generalisable. 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 generalisability, 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 outgroups. 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 tend 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, it should reduce unfortunate misperceptions that groups are irreparably divided to extremes. Reducing that belief has the potential to increase social cohesion and well-being of populations around the world.

Acknowledgements

The research team would like to thank Nazli Altinok, Lizzie N. Jenudi, Yemen Kheir, Dennis Fyhr, Matthew K. Knock, Buthaina Elnaiem, Aldirderi Abdalla, Marya Abd El Hadi, Joud Haidar, Dragana Tomić, Julieta Azevedo, and Ismaël Benslimane. We would also like to thank the Global Scholars Program, Undergraduate Global Engagement, and the Provost’s Office at Columbia University. Specifically, we also thank Meg Booth, Fay Ju, and Caroline Marvin.

References

- Alt, J. E., & Lassen, D. D. (2006). Transparency, Political Polarization, and Political Budget Cycles in OECD Countries. *American Journal of Political Science*, 50(3), 530–550. <https://doi.org/10.1111/j.1540-5907.2006.00200.x>
- Antonczyk, D., DeLeire, T., & Fitzenberger, B. (2018). Polarization and Rising Wage Inequality: Comparing the U.S. and Germany. *Econometrics*, 6(2), 20. <https://doi.org/10.3390/econometrics6020020>
- Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). GlmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2), 378–400.
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395–411.
- Carothers, T., & O'Donohue, A. (Eds.). (2019). *Democracies divided: The global challenge of political polarization*. Brookings Institution Press.
- Duro, J., & Teixidó-Figueras, J. (2014). World polarization in carbon emissions, potential conflict and groups: An updated revision. *Energy Policy*, 74, 425–432. <https://doi.org/10.1016/j.enpol.2014.08.020>
- Ensari, N., & Miller, N. (2002). The out-group must not be so bad after all: the effects of disclosure, typicality, and salience on intergroup bias. *Journal of Personality and Social Psychology*, 83(2), 313.
- Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38, 127–150. Retrieved March 5, 2020, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12394>
- Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, 28(3), 497–516. <https://doi.org/10.1002/jcpy.1047>
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel hierarchical models*. New York, NY, USA: Cambridge University Press.
- Herrmann, R. K., & Keller, J. W. (2004). Beliefs, values, and strategic choice: U. S. Leaders' decisions to engage, contain, and use force in an era of globalization. *The Journal of Politics*, 66(2), 557–580. <https://doi.org/10.1111/j.1468-2508.2004.00164.x>
- Ho, D., & Chau, A. (2009). Interpersonal perceptions and meta-perceptions of relationship closeness, satisfaction and popularity: A relational and directional analysis. *Asian Journal of Social Psychology*, 12, 173–184. <https://doi.org/10.1111/j.1467-839X.2009.01283.x>
- Kteily, N., Hodson, G. & Bruneau, E. (2016) They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, 110, 343–370.
- Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4(3), 279–286. <https://doi.org/10.1038/s41562-019-0766-4>

Luguri, J. B., & Napier, J. L. (2013). Of two minds: The interactive effect of construal level and identity on political polarization. *Journal of Experimental Social Psychology*, 49(6), 972–977.

<https://doi.org/10.1016/j.jesp.2013.06.002>

Méndez, E., Gómez, Á., Tropp, L. R. (2007). When meta-perceptions are affected by intergroup processes. *International Journal of Psychology & Psychological Therapy*, 7, 237–250. Retrieved from

<http://www.ijpsy.com/volumen7/num2/167.html>

Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*.

O'Brien, T. C., Leidner, B., & Tropp, L. R. (2018). Are they for us or against us? How intergroup meta-perceptions shape foreign policy attitudes. *Group Processes & Intergroup Relations*, 21(6), 941–961.

<https://doi.org/10.1177/1368430216684645061>

Pew Research Center. (2020). Far more Americans see 'very strong' partisan conflicts now than in the last individual two presidential election years. *Pew Research Center*. Retrieved March 5, 2020, from

<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/>

The roles of information deficits and identity threat in the prevalence of misperceptions: *Journal of Elections, Public Opinion and Parties*: Vol 29, No 2. (n.d.). Retrieved March 5, 2020, from

<https://www.tandfonline.com/doi/abs/10.1080/17457289.2018.1465>

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., ... & Gibson, S. P. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 4, 622–633.

Shapiro, J. M., & Taddy, N. M. (2015). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech. <https://www.gsb.stanford.edu/sites/gsb/files/politext.pdf>

Shi, F., Teplitskiy, M., Duende, E. et al. (2019). The wisdom of polarized crowds. *Nature Human Behaviour*, 3, 329–336). <https://doi.org/10.1038/s41562-019-0541-6>

Vorauer, J. D., Main, K. J., & O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, 75(4), 917–937. <https://doi.org/10.1037//0022-3514.75.4.917>

Wagenmakers, E. J., Verhagen, J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., ... & Morey, R. D. (2015). A power fallacy. *Behavior Research Methods*, 47(4), 913–917. <https://doi.org/10.3758/s13428-014-0517-4>

Washington Post-University of Maryland Poll (2020, May 5-10). Accessed on June 9, 2020 at https://capc.umd.edu/sites/capc.umd.edu/files/may_5-10_post-u.md_.national_poll_trend.pdf

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 10(2), 145–158. <https://doi.org/10.1177/1745691615569849>

Wojcieszak, M., & Price, V. (2009). What underlies the false consensus effect? How personal opinion and disagreement affect perception of public opinion. *International Journal of Public Opinion Research*, 21(1), 25-46. doi:<http://dx.doi.org.ezproxy.cul.columbia.edu/10.1093/ijpor/edp001>

PREPRINT ONLY NOT PEER REVIEWED