Supporting Information

How Terrorism Does (and Does Not) Affect Citizens' Political Attitudes: A Meta-Analysis

Amélie Godefroidt

American Journal of Political Science

Contents

A Preregistration Motivations and Deviations

A.1 Motivations to Preregister the Meta-Analysis

The review protocol was preregistered before data collection at PROSPERO (#CRD42018106160) and during data collection at Open Science Framework. There were three main reasons to preregister this study. First, like any other type of empirical research, a meta-analysis is susceptible to research practices that might threaten the reliability and replicability of its results. For example, inclusion criteria or search terms can be adjusted after the results are known to accommodate sought-after results, remove studies that lower the overall effect size(s), or reduce evidence of publication bias (Quintana, 2015). Alternatively, one might selectively report only those outcome measures for which a statistically significant overall effect size is found (i.e., selective outcome reporting). Such research practices are quite common, not because of malicious intent but rather because they are caused in part by the incentive structures that underlie academia (Munafò et al., 2017). Hence, to increase the transparency, integrity, and reproducibility of the results reported in this paper, I documented my *a priori* expectations and analysis intentions (Munafò et al., 2017). Second, and related, the preregistration enabled me to highlight that the moderator analyses were conducted in an *exploratory* fashion which distinguishes them from the confirmatory main analyses on the overall effect sizes. A clearer and *a priori* distinction between the hypothesis-testing and -generating parts guarded me against hypothesizing after the results are known ("HARKing") and "posthoc hunting for covariates" that explain the results (Quintana, 2015). Third, preregistered studies are found to publish more negative findings that do not support their hypothesis (Warren, 2018). As a result, the review protocol was also preregister to facilitate publication in case the results were insignificant.

At the same time, it is important to note that review protocols are likely to change over time given that meta-analyses are iterative processes (Quintana, 2015). Indeed, over 20% of meta-analyses are found to make changes to original protocols (Dwan et al., 2011). By having a record of a protocol prior to analysis and by stating deviations from that protocol in the paper or appendix, such changes are made in a more transparent way. To do so, I describe deviations from the preregistration and their justifications in the next section.

A.2 Deviations from the Preregistration

Regarding the hypotheses: The out-group hostility (H1) and conservative shift (H2) hypotheses were preregistered, but the rally-around-the-flag hypothesis was not. An initial descriptive review of the sample (hence, before conducting any inferential tests) revealed that this was an important sub-field in the literature and I therefore added the rally-around-the-flag tests to this article after preregistration. Next, as touched upon above, the overarching hypothesis on the moderators was preregistered an exploratory hypothesis (i.e., "the overall effect size of terrorism on public opinion will be moderated by several theoretical and methodological moderators"). However, although the vast majority of moderators was examined in an exploratory fashion, I did preregister that appraisals of anger would elicit stronger attitudinal shifts (preregistered H3; based on, e.g., Lambert et al., 2019; Marcus et al., 2000; Vasilopoulos et al., 2019) and that the relationship between terrorism and out-group hostility would also be stronger when there is an overlap between the ideology of terrorism and the particular out-group studied (preregistered H4;

based on, e.g., Doosje et al., 1998). While this so-called "guilt-by-association" effect was tested and confirmed in the main paper, there were very few—and mainly recent—effect sizes derived from anger appraisals. Hence, to gain statistical power, emotional appraisals are assessed together in the main paper. Results for the separate negative emotions can be obtained upon request or via the online Replication Files. In general, there is not much evidence in support of an anger effect for outgroup hostility and rally effects, but anger elicits significantly stronger conservative worldviews compared to most other terrorism exposure measures.

Regarding the data collection: The originally preregistered search string was refined as the original one resulted in too much ineligible reports. Particularly the original terms delineating the outcome variables of interest (i.e., "effects" OR "impact" OR "attitudes" OR "opinion") needed a narrower specification (i.e., prejudice OR stereotyp* OR out-group OR attitud* OR authoritarian* OR conservat* OR "public opinion" OR "policy support" OR "political consequences" OR "political tolerance" OR ideolog* OR voting OR vote*). The full final search string is also included below. In addition, the data has been collected by one coder (i.e., the author of the manuscript), and not two coders as preregistered, due to budgetary constraints. About 15% of the data was double checked by that coder, however.

Regarding the data analysis: The option to run three- and not two-level random-effects models was discussed in the PROSPERO preregistration (hence, before data collection) and officially preregistered in the Open Science Framework preregistration (hence, during data collection). I preregistered the original review protocol before I collected any data and, therefore, had no idea yet about the data structure. After the pilot study, it became clear that the sample would entail multiple effect sizes clustered within manuscripts and, therefore, a three-level model was only specified in the Open Science Framework preregistration (see Section B for more information on why a three-level model is desirable in this case). Additionally, how to handle missing data was not specified in the preregistration. By default, the used R functions apply listwise deletion (Cheung, 2015a, 2015b). As a result, the likelihood ratio tests (LRTs) reported in Section C, comparing the deviance of an intercept-only model to that of a model including covariates, could not be conducted when there were missing values on the moderator variable because, in that case, moderated models are not nested within the full-sample intercept-only model. Finally, in the protocol, I explained how I would assess the impact of study quality on the results using a threats-to-validity approach. Yet, while most of the items included in the quality assessment scale (see Section C.2.3) were discussed in the protocol, the 7-item scale itself was not preregistered.

B Additional Information on Data Collection and Analysis

B.1 Selection of Relevant Records

In what follows, I provide more details on the four-fold search strategy used to collect relevant research records. First, several electronic databases were searched for relevant records using a broad search string (last search: 22/12/2019). Records were extracted using databases included in Web of Science (i.e., SSCI, A&HCI, CPCI-SSH, BKCI-SSH), ProQuest (i.e., Criminal Justice Database, Political Science Database, Psychology Database, Social Science Database, Sociology Database, PsycARTICLES) and EBSCO (i.e., Criminal Justice Abstracts, OpenDissertations, Communication & Mass Media Complete, ATLA Religion Database with ATLASerials PLUS). The precise search term was: (prejudice OR stereotyp* OR out-group OR attitud* OR authoritarian* OR conservat* OR "public opinion" OR "policy support" OR "political consequences" OR "political tolerance" OR ideolog* OR voting OR vote*) AND (terror* OR attack* OR "political violence" OR bomb* OR "September 11" OR "9/11" OR "March 11" OR "Charlie Hebdo" OR "Paris attacks" OR "Utoya" OR "Utøya").

Second, a call for additional published or in-press articles, working papers, and unpublished data was put out on Twitter and sent via the e-mail listservs of relevant societies in political sciences (i.e., American Political Science Association, European Political Science Association, European Consortium of political Research, International Society for Political Psychology, and Society of Terrorism Research). I also personally contacted 33 prominent scholars in the field with the same call for additional data. Importantly, I ensured to contact scholars working on non-Islamist (e.g., extreme right) terrorism as my search term predominately generated studies on unspecified or Islamist terrorism. Third, four review articles were screened (i.e., Ayer et al., 2017; Jost et al., 2017; Onraet et al., 2013; Sniderman et al., 2019). These review articles are qualitative in nature or, when quantitative, did not perform a meta-analysis on the specific topic of this article. Last, the reference list of about half of the retrieved manuscripts was screened for additional citations, while the Social Sciences Citation Index was used to check on later citations of influential terrorism-effects manuscripts. The online data file "02-metaanalysis-data.xlsx" documents the number of retrieved manuscripts per search strategy.

This comprehensive strategy resulted in 12,133 possibly relevant records. All these records were imported into Mendeley Reference Manager to check for and delete duplicates (n = 1,742 duplicates). The remaining 10,391 records were screened for inclusion in the meta-analysis based on a list of pre-defined inclusion and exclusion criteria (Table B.1). In the first phase, a broad selection based on the title and abstract was performed, excluding research that clearly did not meet the selection criteria. In this respect, it is worth noting that the original corpus included numerous congressional reports, policy briefings, statements on votes and bills, presidential addresses, and news articles without any statistical information/quantification on the relationship of interest—hence, the large reduction in records in this initial screening phase. Next, the full text of the 947 remaining manuscripts was retrieved and thoroughly screened to ascertain whether they met all inclusion and exclusion criteria. A final of 241 manuscripts were withheld for analysis in this article. Together, the manuscripts cover 326 unique empirical studies and 1,733 effect sizes quantifying the relationship between terrorism exposure and sociopolitical attitudes.

Table B.1: Inclusion and Exclusion Criteria

	Inclusion criteria	Exclusion criteria
Constructs	Studies on the relationship between terrorism[*] (independent variable) and intergroup and/or political attitudes (dependent variable).[†]	Studies on the impact of civil war, repression, genocide, random or non-political violence, or morality salience.
Units	Studies with individuals as the unit of analysis.	Studies with communities, countries, or other aggregated units of analysis. Studies using news articles, vote shares, tweets, etc. as their unit of analysis.
Study	English, French, and Dutch studies. No restriction regarding publication type.	Studies in other languages or published after August 2020.
Designs	Studies reporting a quantitative measure of association and its precision between terrorism and sociopolitical attitudes.	Systematic reviews and meta-analyses, focus groups, in-depth interviews, case reports, editorials, and commentaries.
Statistics	Sufficient information to calculate effect sizes and precision.	Insufficient information to calculate effect sizes and precision.[‡]

[*] Studies examining reactions to violence falling under the academic consensus definition of terrorism were eligible (Schmid, 2011) as well as studies that used terrorism as a label to describe one of their variables of interest. This two-folded strategy allowed us to exhaustively take stock of how scholars in this field of research define and operationalize "terrorism."

[†] A broad operationalization of the outcome of interest was applied: Intergroup attitudes could include affective (e.g., feeling thermometers, out-group target emotions such as hatred/disgust/fear/anxiety or liking/warmth/sympathy/ happiness), cognitive (e.g., beliefs, stereotypes, evaluations), behavioral (e.g., actual, intended, and self-reported actions toward out-group members such as helping, harming, social distance, avoidance) or policy-related (e.g., support for policies commonly associated with prejudice, such as immigration or confirmative action policies) attitudes towards an out-group. Political attitudes could include measures of general political ideology (e.g., authoritarianism, social dominance orientation, general/social/economic conservatism, left-right/liberal-conservatism self-placement), nationalism (e.g., political and/or institutional trust, patriotism, national pride, and identification), political participation (e.g., voting), and support for specific politicians or policies not related to an out-group (e.g., support for President Bush, domestic or foreign counter-terrorism measures).

[‡] When studies did not report enough information, the corresponding author was contacted to obtain the necessary data. If the authors did not reply or supply this information after two reminders, the study was excluded from the meta-analysis. If, however, the (in)significance of the association was stated, a conservative approach was used by setting p-values $> \alpha$ to .90 and p-values $\leqslant \alpha$ to α.

B.2 Additional Explanation of Three-Level Meta-Analytical Models

In this section, I explain why *three*-level random-effects meta-analytic models are used in the main manuscript. In addition, I elaborate on the parameters of interest estimated via these models.

B.2.1 A Fixed-Effect or Random-Effects Model?

The first decision one has to make when performing a meta-analysis is whether a fixed-effect or a random-effects model is more appropriate. A fixed-effect model assumes that all studies along with their effect sizes stem from a single homogeneous population (Borenstein et al., 2010; Harrer et al., 2019). As a result, fixed-effect models provide conditional inferences and results only generalize to studies sharing the *exact same* characteristics as the

studies included in the meta-analysis (Borenstein et al., 2010; Cheung, 2015a). Statistically, this means that fixed-effect models assume that all studies share a common population effect (β_F) and that differences in the observed effect sizes (y_i) are only caused by sampling error (e_i). The univariate fixed-effect model for the i'th study can thus be written as:

$$y_\mathrm{i} = \beta_\mathrm{F} + e_\mathrm{i}, \tag{1}$$

To calculate the overall effect using a fixed-effect model, one simply averages all y_i's with more weight given to studies with higher precision (Cheung, 2015a; Harrer et al., 2019). However, studies in the social sciences often differ in many ways (e.g., differences in country and context, sample characteristics, dependent and independent variables and measures, design, quality, etc.). It is, therefore, not reasonable to assume that studies share the same common effect size. In contrast, real-world data in the social sciences are likely to have variable population parameters and, hence, heterogeneity is expected *a priori*. In this case, a random-effects model is more appropriate since it allows the studies to have their own population effect sizes. Statistically, random-effects models assume that differences in the observed effect sizes can—but do not have to—be caused by differences between the studies themselves (u_i) in addition to sampling variation (e_i). The random-effects model for the i'th study equates:

$$y_\mathrm{i} = \beta_\mathrm{R} + u_\mathrm{i} + e_\mathrm{i}, \tag{2}$$

where β_R denotes the *average* (and not common) population effect size across all studies, u_i the random deviation of the true effect for study i and the average population effect, and e_i the residual due to sampling fluctuation. Thus, random-effects models provide unconditional inferences to be generalized beyond the studies included in the meta-analysis under the assumption that the studies are randomly sampled from the larger super-population of studies (Borenstein et al., 2010; Cheung, 2014, 2015a). Again, the estimated average effect size is obtained via a weighted mean of y_i (i.e., $\hat{\beta}_\mathrm{R} = \frac{\sum \tilde{w}_i y_i}{\sum \tilde{w}_i}$, with $\tilde{w}_i$ = the reciprocal of the precision; Cheung, 2015a, p. 90).

B.2.2 A Two- or Three-Level Random-Effects Model?

Traditionally, if using a random-effects model, scholars have used a two-level random effect model that account for sampling and between-study variation (or, more accurate, between-manuscript variation).[1] An important assumption of these models is that one specific observed effect size does not give information about another observed effect's direction or the size of deviation from the value we would expect based on the meta-analytic model. In other words, most statistical models used in previous meta-analyses assume independent effect sizes. However, this assumption is often violated in practice (Cheung, 2014, 2015a; Van den Noortgate et al., 2013). Effect size dependence can be introduced by either the researchers who conducted the primary studies (e.g., researchers might have compared different treatment groups to the same control group, used multiple measures for the same constructs, and/or questioned the same respondents at various time points) or by the reviewers conducting the meta-analysis (e.g., reviewers might include various effect sizes from the same research teams, authors, or collected within the same country). In this study, included manuscripts report 7.191 effect sizes on average, and effect sizes from within one manuscript are thought to be more similar to each other than effect sizes derived from different manuscripts.

[1] A manuscript denotes a (un)published collection of one or more unique studies and is often the cluster level used in previous meta-analyses.

Various meta-analysts in the social sciences have opted to ignore such dependence by including all effect sizes as if each effect size stemmed from an independent study (e.g., Jost et al., 2017) or to avoid it by either selecting one effect size per study or averaging effect sizes within studies (e.g., Burke et al., 2013; Pettigrew and Tropp, 2006). There are issues with both approaches, however. On the one hand, ignoring overlap in information between effect sizes typically leads to an underestimation of the standard errors (*SE*s) of the parameter estimates and a consequent inflation of the Type I error rates. As a result, researchers may overestimate confidence in their results and incorrectly conclude that there is an effect (i.e., false positive). Selecting or averaging effect sizes, on the other hand, may lead to less precise parameter estimates with larger *SE*s and lower statistical power because informative differences between the outcomes may have been lost. This approach also reduces the research questions that one can address via a meta-analysis because the differences between effect sizes cannot be further explained via moderator analyses (Borenstein et al., 2009; Cheung, 2014; Van den Noortgate et al., 2013).

Hence, it is recommended to apply a third but more complex approach by modelling the dependent effect sizes. To model unknown dependence, a three-level meta-analysis can be used adding another level (or cluster effect) that allows the effect sizes to be correlated within a predefined cluster. This predefined cluster or nested structure depends on the research question and data structure of the meta-analysis (Cheung, 2014, 2015a). In this paper, y_{ij} represents the reported *i*th effect size in the *j*th manuscript given that most manuscripts on terrorism and political attitudes use more than one measure of terrorism and/or attitudes within the same sample. This nesting of participants within effect sizes and effect sizes within manuscripts yields a third potential source of variance, in addition to the *sampling variance* (i.e., differences between observed effect sizes and population effects) and *between-manuscript variance* (i.e., differences between the population effect sizes from different manuscripts). The third source of variance is *between-effect sizes/within-manuscript variance* and denotes the systematic differences between effect sizes stemming from using different measures within the same manuscript. Mathematically, this three-level random-effects meta-analysis with multiple effect sizes nested within manuscripts can be depicted as follows:

$$y_{ij} = \beta_0 + u_{(2)ij} + u_{(3)j} + e_{ij}. \tag{3}$$

This single equation can be split into three equations helping to understand the model:

$$\begin{aligned} \text{Level-1:} \quad & y_{ij} = \lambda_{ij} + e_{ij}, \\ \text{Level-2:} \quad & \lambda_{ij} = f_j + u_{(2)ij}, \\ \text{Level-3:} \quad & f_j = \beta_0 + u_{(3)j}, \end{aligned} \tag{4}$$

where y_{ij} is the *i*th observed effect size within the *j*th manuscript; β_0 is the overall mean effect size across all effect sizes and manuscripts; $u_{(3)j}$ denotes the random deviation of the mean effect in the *j*th manuscript from the overall population effect (β_0); $u_{(2)ij}$ denotes the deviation of the *i*th effect size in the *j*th manuscript from the mean effect in *j*th manuscript (f_j); and e_{ij} is the residual due to sampling fluctuation indicating the deviation of the observed effect size (y_{ij}) from the *true* effect size for *i*th effect size in the *j*th manuscript (λ_{ij}). All three error terms are assumed to be independent and normally distributed with zero mean. The sampling variance for each effect size is not to be estimated anymore in the meta-analysis since it is considered as known (i.e., $\frac{1}{(n-3)}$ for Fisher's Z correlations as

used in this study). In sum, as explained in the main manuscript, the main estimates in this meta-analysis are:

- β_0: The overall mean effect size
- $\text{Var}(u_{(2)\text{ij}}) = \tau^2_{(2)}$: The Level-2 heterogeneity or within-manuscript variance
- $\text{Var}(u_{(3)\text{j}}) = \tau^2_{(3)}$: The Level-3 heterogeneity or between-manuscript variance

B.2.3 Testing the Homogeneity of Effect Sizes and Quantifying the Degree of Heterogeneity

Although it is generally not recommended to decide between a fixed-effect, two-level, or three-level model based on statistical information,[2] one can formally test the homogeneity of effect sizes. In other words, there exists ways to test whether the variance between the observed effect sizes is indeed larger than one would expect based on sampling variance alone. Conventionally, meta-analysts use Cochran's Q statistic to test the null hypothesis of homogeneity of effects sizes. The Q-statistic on the homogeneity of effect sizes is highly significant for all three outcome types ($p < .0001$), so we reject the null hypothesis of no differences between effect sizes across manuscripts. However, this test assumes conditionally independent effect sizes (see Cheung, 2015a, pp.184-186, for more information)—an assumption which is often violated in social science meta-analyses (as explained above). As a result, the Q statistic may not be accurate and likelihood ratio tests (LRTs) may be preferred instead. As it is clear from Equations 1, 2, and 3, the fixed-effect model is nested within the two-level random-effects model which is again nested within the three-level random-effects model. As a result, the difference on the number of parameters is 1 and LRT's can be used to compare models with and without between- and within-manuscript variance component by testing $H_0 : \tau^2_{(3)} = 0$ and $H_0 : \tau^2_{(2)} = 0$. As Table B.2 shows, all test statistics indicate that there exist significant between- and within-manuscript variance in this study. Lastly, all original heterogeneity estimates ($\tau^2_{(2)}$ and $\tau^2_{(3)}$) were also significant (see Table B.3). Taken together, there is much evidence that there are true differences between the effect sizes within a manuscript as well as between those across different manuscripts.

Table B.2: Testing the Homogeneity of Effect Sizes, by Outcome Type

	Outgroup Hostility		Conservative Shift		Rally Effects	
	χ^2(df=1)	p-value	χ^2(df=1)	p-value	χ^2(df=1)	p-value
Testing $H_0 : \tau^2_{(3)} = 0$	403.5183	$< .0001$	253.0271	$< .0001$	133.2251	$< .0001$
Testing $H_0 : \tau^2_{(2)} = 0$	1763.041	$< .0001$	2246.512	$< .0001$	4706.859	$< .0001$

Note: To obtain the LR statistics (i.e., χ^2's), I fitted two constrained models for each outcome type, and compared these models against the main three-level model (Eq. 3) with the `anova()` function. The results indicate that a three-level model is statistically better than a two-level model (first row) and a three-level model better than a model with the between-manuscript variance constrained to 0 (second row). This means that effect sizes between ànd within manuscripts are not merely direct replications of each other and confirms the need for a three-level model.

[2] Instead of relying on statistics, scholars recommend to select a meta-analytic model based on whether a conditional or an unconditional inference is required (Hedges and Vevea, 1998 as cited in Cheung, 2015a). Additionally, three-level models are especially advantageous when there are many different outcome variables, the presence of these variables varies widely across manuscripts, and/or when researchers want to generalize across outcome variables as well (Fernández-Castilla et al., 2018). Given that I aim to generalize the findings from this meta-analysis across both Level-2 (i.e., to different measures used to obtain individual effect sizes) and Level-3 (i.e., to different manuscripts) and given that there are many different effect sizes reported within manuscripts, a three-level model was considered most appropriate.

In addition to testing the homogeneity of effect sizes, one can also quantify the degree of heterogeneity of the effect sizes at both levels by using Higgins' I^2 (see Cheung, 2015a, p.186, for the equations). In this meta-analysis, $I^2_{(2)}$ and $I^2_{(3)}$ represent the proportions of the total variation of the effect sizes that is due to differences within a manuscript (i.e., Level-2 heterogeneity) and across manuscripts (i.e., Level-3 heterogeneity). In other words, these parameters quantify the percentage of variance that is not due to sampling error. As Table B.3 shows, the $I^2_{(2)}$ is above 22.14% and the $I^2_{(3)}$ above 62.85% for all three hypotheses, which signals substantial differences in observed effect sizes both within and between manuscripts (Cheung, 2015a).

Table B.3: Heterogeneity Estimates

	Heterogeneity Variances				Proportion	
Outcome Type	$\tau^2_{(2)}$	LBCI	$\tau^2_{(3)}$	LBCI	$I^2_{(2)}$	$I^2_{(3)}$
Outgroup hostility	.009	[.007; .010]	.029	[.022; .039]	22.14%	75.11%
Conservative shift	.009	[.007; .010]	.017	[.012; .023]	32.75%	63.28%
Rally effects	.009	[.007; .011]	.018	[.011; .029]	31.57%	66.42%

Note: The τ estimates indicate significance heterogeneity both within and between manuscripts. As these heterogeneity variances are hard to interpret, I^2 estimates are used to quantify the proportion of variation explained by Level-2 and Level-3. These estimates confirm that there is substantial variation at both levels.

B.2.4 Adding Within- and Between-Manuscript Moderators

After rejecting the homogeneity and quantifying the heterogeneity of effect sizes, one can try to explain both within- and between-manuscript variation by adding Level-2 and Level-3 moderators (as done in Figure 3 in the main article or in the corresponding Tables C.1-C.2-C.3 below). Specifically, my definition of Level-2 entails that $\tau^2_{(2)}$ indicates the heterogeneity of effect sizes due to different ways of measuring the same underlying constructs (i.e., terrorism exposure as well as outgroup attitudes, conservative ideology, or rally tendencies). Characteristics of the measurements (e.g., features of the outgroup under scrutiny, type of political outcome, type of terror threat) were used as within-manuscript moderators to explore how different measures affect found effect sizes. My definition of Level-3 entails that $\tau^2_{(3)}$ indicates the heterogeneity among the true effect sizes across manuscripts, after controlling for the multiple measures used at Level-2. Manuscript-characteristics (e.g., publication status and year, country of study) can thus be included to explain remaining between-manuscript heterogeneity at Level-3. In short, the random-effects model (Eq. 3) can be extended to the following mixed effects model to explore and explain variation in effect sizes at both Level-2 and Level-3:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_j + u_{(2)ij} + u_{(3)j} + e_{ij}. \tag{5}$$

C Numerical Results and Additional Analyses

C.1 Numerical Results of Meta-Regressions

On the next pages, I provide the numerical results for the moderator analyses as described in the main manuscript. The meta-regressions are based on a series of three-level random-effects models, in which moderators are entered separately (using Eq. 5). As explained in the main manuscript, when assessing moderator effects of categorical variables, as many dummy indicators as there are categories were included in the model and the intercept was constrained to zero. The advantage of this parameterization is that the regression coefficients can be interpreted as the average effect sizes for all categories. When comparing multiple categories, the p-values are adjusted using a Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) to decreases the false discovery rate (i.e., the expected proportion of false discoveries among the rejected hypotheses).[3] Furthermore, the Likelihood Ratio Test (LRT) statistics reported in Tables C.1, C.2, and C.3 below refer to the difference between the deviance of the null model and the deviance of the model including the moderator. Hence, this is a formal test of the null hypothesis that the regression coefficient/-s is/are zero (and should not be confused with the LRTs testing the homogeneity of the effect sizes reported in Table B.2 above). Last, although the fitted model differentiates between the variance explained at both the manuscript level ($R^2_{(3)}$) and effect size level ($R^2_{(2)}$) by adding the moderator (Cheung, 2015a; see also Cheung, 2014), the R^2 reported in the Tables below refers to the total variance explained across both levels ($= R^2_{(2)} + R^2_{(3)}$) for the sake of parsimony.

[3] I do not have *a priori* predictions about the differences between categories for all categorical moderators but, instead, I am interested in finding any between-category differences in effect sizes. However, when taking every possible pair of categories and performing a t-test to compare effect sizes on each pair, the family-wise error rates get inflated. There exist various ways to adjust your level of significance such that the family-wise error rate can be controlled, and I use a False Discovery Rate (FDR) approach in this paper by using a Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). For more information on the need and different ways to adjust p-values when making multiple comparisons, see Field and colleagues (2012: 428–432) For more information on the Benjamini-Hochberg procedure used in this paper, see Benjamini and Hochberg (1995).

Table C.1: Moderator Analyses for Outgroup Hostility Hypothesis.

Moderator	k	b	95% LBCI	R^2	LRT Statistic
Panel A: Characteristics of the Independent Variable					
Ideology	645			0.030	$\chi^2(2) = 8.534$, $p = .014$
Islamist$_{a,c}$	471	.122	[.087; .157]		
Non-Islamist$_{b,c}$	59	.056	[–.012; .124]		
No ideology$_{a}$	115	.175	[.124; .227]		
Measurement	645			0.134	$\chi^2(4) = 36.398$, $p < .0001$
Objective exposure$_{a}$	252	.044	[–.005; .092]		
Subjective exposure$_{a,b}$	65	.104	[.055; .155]		
Cognition (threat)$_{c}$	153	.209	[.165; .254]		
Emotion (anger, fear)$_{c,d}$	72	.188	[.138; .239]		
Residual category$_{b,d}$	103	.146	[.097; .195]		
Panel B: Characteristics of the Dependent Variable					
Target outgroup	645			0.045	$\chi^2(2) = 11.437$, $p = .003$
Religious outgroup$_{a}$	333	.143	[.107; .179]		
Immigrant/refugee$_{a}$	208	.129	[.088; .171]		
Other outgroup$_{b}$	104	.073	[.027; .119]		
Guilt-by-association	645			0.056	$\chi^2(2) = 16.462$, $p < .001$
Strong association$_{a}$	299	.155	[.120; .190]		
Moderate association$_{b}$	157	.104	[.063; .146]		
No association$_{b}$	189	.087	[.047: .128]		
Panel C: Characteristics of the Sample and Study					
Research design	645			0.097	$\chi^2(2) = 10.458$, $p = .005$
Experiment$_{a}$	142	.091	[.034; .147]		
Correlational$_{b}$	283	.169	[.128; .209]		
Other$_{a}$	220	.089	[.040; .139]		
Time between IV-DV	645			0.058	$\chi^2(1) = 8.441$, $p = .004$
Yes: Delay$_{a}$	194	.062	[.009; .116]		
No: Direct$_{b}$	451	.144	[.110; .178]		
Sampling protocol	645			0.113	$\chi^2(2) = 29.419$, $p < .0001$
General population$_{a}$	205	.152	[.100; .204]		
Student sample$_{b}$	275	.057	[.011; .102]		
Convenience sample$_{a}$	165	.191	[.143; .240]		
Country	645			0.040	$\chi^2(2) = 7.081$, $p = .029$
US$_{a}$	161	.165	[.114; .217]		
Israel$_{a,b}$	160	.162	[.086; .238]		
Other$_{b}$	324	.084	[.039; .129]		
Mean age	548	.004	[.003; .006]	0.086	
Percentage women	594	.001	[–.001; .003]	0.020	
Data collection year	549	–.002	[–.008; .003]	0.003	
Publication year	645	–.001	[–.008; .005]	0.000	$\chi^2(1) = 0.181$, $p = .671$

Note: k = number of effect sizes in the category. b = unstandardized regression coefficient. The regression coefficients for the categorical variables can be interpreted as the mean effect size (i.e., Fisher's Z correlation coefficient) for each category. Effect sizes belonging to one categorical variable that do not share subscripts$_{a,b,c,d}$ differ at $p < .05$ after Benjamini-Hochberg correction for multiple comparisons. LBCI = Likelihood-Based Confidence Interval. LRT = Likelihood Ratio Test comparing the deviance of the model with the moderator against that of the null model without moderators. R^2 refers to the proportion of the explained total variance across the levels ($= R_2^2 + R_3^2$).

Table C.2: Moderator Analyses for Conservative Shift Hypothesis.

Moderator	k	b	95% LBCI	R^2	LRT Statistic
Panel A: Characteristics of the Independent Variable					
Ideology	728			0.072	$\chi^2(2) = 7.659$, $p = .022$
Islamist$_a$	399	.131	[.103; .159]		
Non-Islamist$_b$	136	.076	[.028; .125]		
No ideology$_a$	193	.157	[.123; .190]		
Measurement	728			0.1710	$\chi^2(4) = 27.673$, $p < .0001$
Objective exposure$_a$	152	.072	[.033; .112]		
Subjective exposure$_a$	95	.107	[.063; .151]		
Cognition (threat)$_b$	248	.160	[.130; .189]		
Emotion (anger, fear)$_b$	128	.180	[.144; .217]		
Residual category$_a$	102	.100	[.061; .141]		
Panel B: Characteristics of the Dependent Variable					
Outcome measure	728			0.090	$\chi^2(5) = 37.327$, $p < .0001$
RWA$_a$	75	.171	[.133; .210]		
SDO$_b$	46	.105	[.059; .151]		
Ideology$_b$	121	.112	[.077; .146]		
Military actions$_{a,b}$	290	.137	[.108; .168]		
Civil liberties$_a$	145	.165	[.131; .201]		
Residual category$_c$	51	.004	[–.056; .063]		
Panel C: Characteristics of the Sample and Study					
Research design	728			0.091	$\chi^2(2) = 12.949$, $p < .001$
Experiment$_a$	200	.101	[.064; .138]		
Correlational$_b$	417	.159	[.131; .186]		
Other$_a$	111	.089	[.049; .130]		
Time between IV-DV	728			0.040	$\chi^2(1) = 3.590$, $p = .058$
Yes: Delay$_a$	78	.085	[.031; .139]		
No: Direct$_a$	650	.138	[.113; .163]		
Sampling protocol	728			0.027	$\chi^2(2) = 8.468$, $p = .014$
General population$_a$	264	.105	[.070; .139]		
Student sample$_{a,b}$	226	.125	[.087; .162]		
Convenience sample$_b$	238	.179	[.139; .221]		
Country	728			0.023	$\chi^2(2) = 2.647$, $p = .266$
US$_a$	278	.147	[.112; .181]		
Israel$_a$	202	.141	[.091; .190]		
Other$_a$	248	.110	[.073; .146]		
Mean age	603	.001	[–.000; .003]	0.015	
Percentage women	660	.000	[–.001; .001]	0.002	
Data collection year	606	–.002	[–.006; .001]	0.026	
Publication year	728	–.004	[–.008; .001]	0.021	$\chi^2(1) = 2.608$, $p = .106$

Note: k = number of effect sizes in the category. b = unstandardized regression coefficient. The regression coefficients for the categorical variables can be interpreted as the mean effect size (i.e., Fisher's Z correlation coefficient) for each category. Effect sizes belonging to one categorical variable that do not share subscripts$_{a,b,c}$ differ at $p < .05$ after Benjamini-Hochberg correction for multiple comparisons. LBCI = Likelihood-Based Confidence Interval. LRT = Likelihood Ratio Test comparing the deviance of the model with the moderator against that of the null model without moderators. R^2 refers to the proportion of the explained total variance across the levels ($= R_2^2 + R_3^2$).

Table C.3: Moderator Analyses for Rally-'Round-the-Flag Hypothesis.

Moderator	k	b	95% LBCI	R^2	LRT Statistic
Panel A: Characteristics of the Independent Variable					
Ideology	360			0.072	$\chi^2(2) = 2.693$, $p = .260$
Islamist$_a$	194	.095	[.056; .135]		
Non-Islamist$_a$	80	.054	[–.002; .111]		
No ideology$_a$	86	.096	[.041; .152]		
Measurement	360			0.017	$\chi^2(4) = 2.434$, $p = .657$
Objective exposure$_a$	122	.100	[.039; .163]		
Subjective exposure$_a$	24	.119	[.045; .194]		
Cognition (threat)$_a$	119	.093	[.044; .142]		
Emotion (anger, fear)$_a$	57	.072	[.020; .125]		
Residual category$_a$	38	.069	[–.009; .147]		
Reference to 9/11	360	.167	[.088; .243]	0.317	$\chi^2(2) = 15.842$, $p < .0001$
Panel B: Characteristics of the Dependent Variable					
Outcome measure	360			0.108	$\chi^2(2) = 3.134$, $p = .209$
Politicians$_a$	58	.124	[.062; .187]		
Political trust$_a$	187	.059	[.011; .110]		
Patriotism$_a$	115	.096	[.054; .139]		
Reference to Republican	360	.060	[.009; .111]	0.071	$\chi^2(1) = 5.461$, $p = .019$
Reference to incumbent	360	.070	[.006; .134]	0.110	$\chi^2(1) = 4.646$, $p = .031$
Reference to pres. Bush	360	.194	[.102; .285]	0.197	$\chi^2(1) = 16.688$, $p < .0001$
Panel C: Characteristics of the Sample and Study					
Research design	360			0.006	$\chi^2(2) = 0.531$, $p = .767$
Experiment$_a$	81	.099	[.029; .172]		
Correlational$_a$	149	.103	[.042; .164]		
Other$_a$	130	.080	[.034; .127]		
Time between IV-DV	360			0.010	$\chi^2(1) = 1.425$, $p = .233$
Yes: Delay$_a$	123	.125	[.058; .195]		
No: Direct$_a$	237	.079	[.040; .120]		
Sampling protocol	360			0.021	$\chi^2(2) = 1.968$, $p = .374$
General population$_a$	186	.072	[.017; .127]		
Student sample$_a$	90	.089	[.029; .148]		
Convenience sample$_a$	84	.132	[.063; .207]		
Country	360			0.233	$\chi^2(2) = 13.634$, $p = .001$
US$_a$	102	.157	[.109; .207]		
Israel$_b$	53	.004	[–.080; .090]		
Other$_b$	205	.053	[.008; .100]		
Mean age	296	.003	[.000; .005]	0.058	
Percentage women	291	.001	[–.001; .003]	0.006	
Data collection year	334	–.009	[–.013; –.004]	0.240	
Publication year	360	–.010	[–.016; –.005]	0.227	$\chi^2(1) = 12.405$, $p < .001$

Note: k = number of effect sizes in the category. b = unstandardized regression coefficient. The regression coefficients for the categorical variables can be interpreted as the mean effect size (i.e., Fisher's Z correlation coefficient) for each category. Effect sizes belonging to one categorical variable that do not share subscripts$_{a,b,c}$ differ at $p < .05$ after Benjamini-Hochberg correction for multiple comparisons. LBCI = Likelihood-Based Confidence Interval. LRT = Likelihood Ratio Test comparing the deviance of the model with the moderator against that of the null model without moderators. R^2 refers to the proportion of the explained total variance across the levels ($= R_2^2 + R_3^2$).

C.2 Sensitivity Analysis

In what follows, I conduct a series of sensitivity analysis to assess whether the results reported in the main paper are sensitive to accounting for additional clustering in the data (C.2.1), to excluding outliers (C.2.2), to a basic assessment of study quality (C.2.3), and to excluding correlations derived from regression coefficients (C.2.4). When accounting for additional clustering, I switched from using the `metaSEM` package to the `metafor` package as the `metaSEM` package does not fit 4-Level models. Table C.4 reproduces the results reported in Table 2 in the main paper using both the `meta3` function from the `metaSEM` and the `rma.mv` function from the `metafor` package, and is used as a benchmark to probe robustness.

Table C.4: Replication of Table 2 Using Both the `meta3` and `rma.mv` Function.

			`meta3`		`rma.mv`		
Outcome Type	k	j	Z_r	LBCI	Z_r	SE	p-value
Outgroup hostility	645	126	.126	[.094; .159]	.126	.017	<.0001
Conservative shift	728	144	.132	[.108; .156]	.132	.012	<.0001
Rally effects	360	72	.090	[.055; .127]	.090	.018	<.0001

Note: k = number of effect sizes. j = number of manuscripts. Z_r = Overall Fisher's Z correlation coefficients. LBCI = Likelihood-Based Confidence Interval. The Z_r estimates are considered significant when the LBCIs do not include zero.

In the next sections, we replicate Table C.4 while dropping particular effect sizes or modifying the model specification. In short, all results reported below produce coefficient estimates that yield substantively similar conclusions to the ones reported in the main paper.

C.2.1 Impact of Additional Clustering

In the main paper, the Level-3 cluster denotes published or unpublished manuscripts. Yet, some manuscripts collect several unique studies (e.g., multiple experiments grouped in one publication or the same/a similar survey fielded in different countries)—which adds another potential source of dependency. However, changing the cluster from manuscripts to studies does not lead to a substantial change in results nor does fitting a 4-Level model accounting for the clustering of studies within manuscripts (Table C.5).

Table C.5: Impact of Additional Clustering.

		Changing Level-3		Adding a Level-4		
Outcome Type	k	Z_r	LBCI	Z_r	SE	p-value
Outgroup hostility	645	.124	[.094; .154]	.127	.016	<.0001
Conservative shift	728	.129	[.108; .150]	.132	.012	<.0001
Rally effects	360	.086	[.055; .118]	.091	.018	<.0001

Note: The right columns ("Adding a Level-4") are fitted with the `rma.mv` function from the `metafor` package and not with the `meta3` function from the `metaSEM` package as all other models. Hence, these coefficients need to be compared to the right columns of Table C.4 to assess robustness.

C.2.2 Impact of Outliers in Effect Size Magnitude and Precision

In Table C.6, I explore the impact of outlying or influential cases. I first look for outliers both in terms of effect size magnitude and precision. Regarding the former, I consider an effect size as an outlier when its confidence interval does not overlap with the confidence interval of the pooled effect (within outcome type) (Harrer et al., 2019). Regarding the latter, I flag variances that exceed the 90% percentile distribution of the variances (again, within outcome type) (Blair et al., 2021). Subsequently, I assess to what extent these outliers influence the overall effect sizes. To do so, I reproduce Table C.4 with the outliers excluded (Panel A) and, second, with the outliers converted to values at the boundaries (i.e., censoring/winsorizing; Panel B). I censor effect sizes to the confidence interval of the pooled effect and variances to the 90% percentile distribution of the variances for each outcome type. At large, the substantive interpretation of the results remains unchanged, although the precision of the estimates improves and the pooled effect for outgroup hostility slightly decreases when excluding outlier effect sizes.

Table C.6: Impact of Studies with Outlier Effect Sizes and Variances.

	Outlier Effect Sizes			Outlier Variances		
Outcome Type	k	Z_r	LBCI	k	Z_r	LBCI
Panel A: Outlier Excluded						
Outgroup hostility	402	.106	[.095; .118]	578	.129	[.094; .164]
Conservative shift	434	.124	[.114; .135]	646	.134	[.108; .161]
Rally effects	213	.087	[.071; .105]	323	.092	[.055; .130]
Panel B: Outliers Winsorized						
Outgroup hostility	645	.114	[.106; .122]	645	.126	[.094; .159]
Conservative shift	728	.124	[.118; .130]	728	.132	[.108; .156]
Rally effects	360	.085	[.074; .098]	360	.090	[.055; .127]

Note: k = number of effect sizes. Z_r = Overall Fisher's Z correlation coefficients. LBCI = Likelihood-Based Confidence Interval. The Z_r estimates are considered significant when the LBCIs do not include zero.

C.2.3 Impact of Risk of Bias

Without criticizing any of the primary studies included in this meta-analysis, risk of bias in the included studies was assessed using various indicators based on the ROBINS-I tool for non-randomized studies (Sterne et al., 2016), the RoB 2 tool for randomized studies (Sterne et al., 2019), and the Study DIAD (Valentine and Cooper, 2008). The following items were used to construct a 7-point sum scale evaluating potential bias (M_{outgroup} = 3.560, SD_{outgroup} = 1.399; $M_{\text{conservatism}}$ = 3.831, $SD_{\text{conservatism}}$ = 1.099; M_{rally} = 3.664, SD_{rally} = 1.125):

- Was the study peer-reviewed?
- Was the study preregistered?
- Was a student sample used (reverse-coded)?
- Did the study clearly state the hypothesis/-es to be tested in this paper?
- Does the study include a discussion of attrition or response rates?
- Was the measure of the independent variable appropriate?

  - Experiments: Was the allocation to the treatment random?
  - Natural experiments: Were treatment and control groups balanced?
  - Observational data: Was the IV measured using a reliable multi-item scale ($\alpha > .70$)?
- Was the dependent variable measured using a reliable multi-item scale ($\alpha > .70$)?

Subsequently, a sum scale is used as an indicator of quality in a meta-regression. Here, a higher value indicates less risk of bias. In addition, Table C.7 reports the results of using the effective sample size (following Paluck, Porat, Clark, and Green 2021) and impact factor of the journal (if the manuscript is a published article) as crude proxies for study quality. In general, Table C.7 does not provide much evidence to worry about the possible effect of lower-quality studies on the results. Still, the results suggest that a lower risk of bias leads to a higher effect size for the sub-sample on outgroup hostility. An item-by-item analysis on the scale indicators reveals that this positive effect is completely driven by the positive effect of using a non-student sample.[4] As Table C.1 above revealed, student samples resulted in a Z_r correlation of .058, other convenience samples in a Z_r of .191, and general population samples in a Z_r of .154.

Table C.7: Impact of Risk of Bias and Study Quality.

	Outgroup Hostility		Conservative Shift		Rally Effects	
Quality Indicator	β(se)	p-value	β(se)	p-value	β(se)	p-value
Risk of Bias	**.040(.014)**	**.004**	–.001(.010)	.930	–.024(.013)	.073
Sample Size	–.001(.007)	.907	–.009(.008)	.247	–.008(.010)	.396
Impact Factor	–.002(.014)	.901	–.012(.013)	.343	.007(.014)	.609

C.2.4 Impact of Multivariate Regression Coefficients

In the pre-processing phase, univariate data were used as much as possible to calculate the common effect size. Studies only reporting multivariate regression coefficients are somewhat problematic for meta-analyses because such coefficients represent partial correlation coefficients controlling for the influence of one or more (different) covariates. Therefore, I contacted authors of studies reporting multivariate regression coefficients to solicit missing zero-order correlations but, if no such data could be extracted, I used imputed standardized (β) regression coefficients using the Peterson and Brown transformation. Peterson and Brown (2005) show how using beta coefficients to impute missing correlations generally produces relatively accurate and precise population effect-size estimates for meta-analyses using correlation coefficients as the effect-size metric. Potential benefits from applying this procedure include smaller sampling errors because of increased numbers of effect sizes and smaller non-sampling errors because of the inclusion of a broader array of research designs.

However, more recently, Aloe (2015) illustrated how the Peterson and Brown (2005) index does not always approximate the bivariate correlation. Therefore, I explore whether the empirical results are robust to excluding data points derived from such regression models. Table C.8 indicates that the results remain quasi-identical to the ones reported in the main paper.

[4] In addition, pre-registered studies on terrorism and outgroup hostility generally result in a lower effect size ($b = –0.138, p = .051$). However, this moderation effect just falls short of the *a priori* set significance level of .05 and is driven by only a handful studies.

Table C.8: Excluding of Multivariate Regression Coefficients.

Outcome Type	k	j	Z_r	LBCI
Out-group hostility	591	115	.127	[.092; .162]
Conservative shift	674	134	.132	[.107; .157]
Rally effects	343	68	.089	[.052; .128]

Note: k = number of effect sizes. j = number of manuscript. Z_r = Overall Fisher's Z correlation coefficients. LBCI = Likelihood-Based Confidence Interval. The Z_r estimates are considered significant when the LBCIs do not include zero.

C.3 Publication Bias Analysis

One commonly acknowledged problem concerning meta-analyses is publication bias. That is, significant results are more likely to get published than null results (Begg, 1994). This creates a "file drawer" problem (Rosenthal, 1979), where studies with weak or null results go unnoticed or unpublished (or are never sent for peer review). As a result, meta-analyses—to a large extent relying on published studies—may overestimate the true mean effect size. I took several steps to mitigate publication bias, applied graphical and statistical techniques to detect publication bias, and assessed the extent to which it skews the estimates reported in the main paper. First, during the data collection phase, special attention was paid to include as many working papers and other unpublished materials (such as raw datasets) as possible. Second, I fitted several meta-regressions to examine the moderating impact of publication status, sample size, and precision on the results and constructed a funnel plot to visually detect publication bias. All these techniques rely in one way or another on the sample size as smaller studies, generating larger standard errors on average, require larger effect sizes to reach significance (see below for more information). Hence, significant relationships between effect sizes and precision measures suggest publication bias. Third, I used two techniques to correct for publication bias: the Trim & Fill method and a PET/PEESE analysis (again, see below for more information).

C.3.1 Detecting Publication Bias

Table C.9 shows the results of a series of diagnostic tests, all of which suggest that there is not much evidence of publication bias. For example, for none of the outcome types, significant differences were found between published and unpublished studies. Furthermore, neither the effective sample size nor the inverse standard error or variance moderated the correlations between terrorism and any of the sociopolitical attitude clusters, except for the inverse standard error for the conservative shift sample which is negatively associated with the overall effect size. Last, the funnel plots (Figure C.1) also suggest minimal publication bias given that effects are distributed approximately symmetrical around the mean effect size. However, this is not true for the funnel plot for studies on rally effects. Here, the funnel-plot is asymmetrical as effect sizes on the left side (and esp. in the left lower corner) are missing.[5] This suggests that (smaller) studies with negative results are less likely to be published.

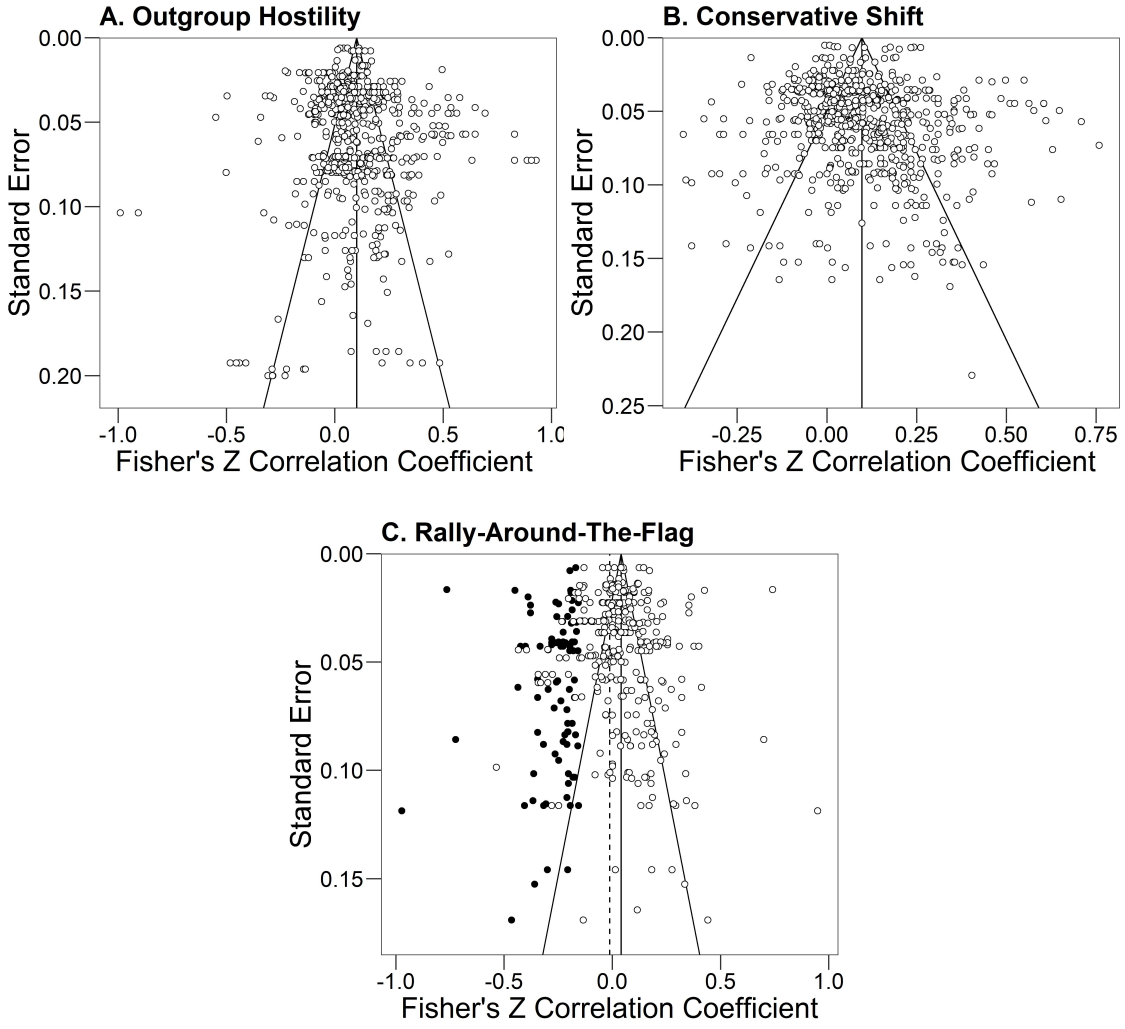
[5] The filled circles in Figure C.1 Panel C represent the missing studies needed to make the funnel plot symmetrical; see also next section.

Table C.9: Diagnostic Tests for Publication Bias

	Outgroup Hostility		Conservative Shift		Rally Effects	
Indicator	β(se)	p-value	β(se)	p-value	β(se)	p-value
Published	.030(.022)	.169	.015(.013)	.262	.023(.018)	.208
Sample Size	–.001(.007)	.907	–.009(.008)	.247	–.008(.010)	.396
Inverse standard error	–.005(.008)	.556	–.017(.008)	.040	–.015(.011)	.192
Inverse Variance	–.001(.007)	.900	–.009(.008)	.249	–.008(.010)	.396

Note: Three-level meta-regression models are used to estimate the standardized effect of the publication status (1=published, 0=unpublished), effective sample size used, inverse standard error, and inverse variance on the average effect size. Results suggest minimal publication bias, as we find no evidence for all but one meta-regression to reject the null hypothesis of no relationship.

Figure C.1: Funnel Plots for **(A)** Outgroup Hostility, **(B)** Conservative Shift, and **(C)** Rally-Around-The-Flag Hypothesis.



Note: Funnel plots are scatterplots displaying the standard errors on a reversed axis against the observed (open circles) and trim-and-fill imputed (filled circles) effect sizes using a two-level random-effects model (see here for more information). Summary effect displayed in solid black lines and summary effect including imputed studies in dotted black lines. An approximately symmetrical distribution of the observed effect sizes (open circles) around the mean effect size indicates minimal publication bias.

C.3.2 Correction for Publication Bias

Although there is not much evidence of publication bias, I also implemented two methods that correct for publication bias. First, Figure C.1 displays a summary effect based on the studies included in the meta-analysis complemented with studies necessary to counter publication bias (imputed by the Trim & Fill method; for more information, see Cooper 2017; Duval and Tweedie, 2000). While imputing studies does not affect the result for the outgroup hostility and conservative shift hypothesis, there is some indication that this would substantially decrease the rally-around-the-flag effect. This suggests a relevant publication bias for rally effect studies, which adds to the overarching conclusion that rally effects might be idiosyncratic. However, it is important to note that the results in Figure C.1, including the adjusted effect size, are based on a conventional two-level meta-analytic model. As explained above, such a model violates the important assumption of independent effect sizes. Therefore, I complement the Trimm & Fill results with a PET-PEESE analysis that allows to fit three-level models.

PET-PEESE models (Stanley and Doucouliagos, 2014) give an estimate of the effect size when the standard error (PET) or the variance (PEESE) is zero (i.e., a study with infinite sample size). In other words, in a PET-PEESE analysis, two meta-regressions are conducted in which the standard error and sampling variance act as moderators, respectively, and the intercepts, β_0, represent the estimates of interest (i.e., the estimated effect when the SE or variance is zero). When the intercepts are of a similar magnitude and significance of the overall effect sizes, the results prove to be robust. Table C.10 suggests that this is the case for all three hypotheses.

Table C.10: PET-PEESE Analysis to Correct for Publication Bias.

	Outgroup Hostility		Conservative Shift		Rally Effects	
	β(se)	p-value	β(se)	p-value	β(se)	p-value
Panel A: PET						
Intercept	.126(.017)	$<$ .0001	.133(.012)	$<$ .0001	.086(.018)	$<$ .0001
Standard Error	–.003(.014)	.859	.021(.009)	.019	.025(.015)	.096
Panel B: PEESE						
Intercept	.125(.017)	$<$ .0001	.133(.012)	$<$ .0001	.088(.018)	$<$ .0001
Variance	–.015(.017)	.367	.016(.009)	.087	.020(.015)	.169

Note: Three-level meta-regression models use to estimate the effect size when the standard error (se) or variance is zero (i.e., for a study with a hypothetical infinite sample size). The parameters of interest in this table are the intercepts. The results suggest that a hypothetical study with an infinite sample size would yield a similar effect size as the overall effect sizes reported in the paper.

D PRISMA Checklist

Section/topic	#	Checklist item	Place*
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	i
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	iii
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1-2, 7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2, 7, SI (table) B.1
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	ii, SI A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Table B.1
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	8, SI B.1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	SI B.1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8-9, SI B.1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9, Rep. Mat.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	C.2.3

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9-10
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.	10-12, SI B.2
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	SI C.2, C.3
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	12-13, SI C.2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Fig.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	13-15, App
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	SI C.2.3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	App
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Table 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table C.7
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	SI C.2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	16-17, 23
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	23-26
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	23-26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	ii

* Currently, the numbers refer to pages in the unpublished manuscript (Word version uploaded for publication). This will be updated upon publication; App = The Shiny App complementing this study; Fig. = Figure; Rep. Mat. = Replication Materials; SI = Supporting Information.

References

Aloe, A. M. (2015). Inaccuracy of regression results in replacing bivariate correlations. *Research Synthesis Methods*, *6*(1), 21–27.

Ayer, L., Venkatesh, B., Stewart, R., Mandel, D., Stein, B., & Schoenbaum, M. (2017). Psychological Aspects of the Israeli–Palestinian Conflict: A Systematic Review. *Trauma, Violence, & Abuse*, *18*(3), 322–338.

Begg, C. B. (1994). Publication Bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. Russel Sage Foundation.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Blair, G., Christensen, D., & Rudkin, A. (2021). Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments. *American Political Science Review*, 1–8.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley.

Burke, B. L., Kosloff, S., & Landau, M. J. (2013). Death goes to the polls: A meta-analysis of mortality salience effects on political attitudes. *Political Psychology*, *34*(2), 183–200.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229.

Cheung, M. W.-L. (2015a). *Meta-analysis: A Structural Equation Modeling Approach*. Wiley.

Cheung, M. W.-L. (2015b). metaSEM: an R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*(1521), 1–7.

Cooper, H. M. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). SAGE Publications.

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, *75*(4), 872–886.

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463.

Dwan, K., Altman, D. G., Cresswell, L., Blundell, M., Gamble, C. L., & Williamson, P. R. (2011). Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database of Systematic Reviews*.

Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2018). A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behavior Research Methods*.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics Using r*. SAGE Publications Ltd.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). Doing Meta-Analysis in R: A Hands-on Guide.

Jost, J. T., Stern, C., Rule, N. O., & Sterling, J. (2017). The Politics of Fear: Is There an Ideological Asymmetry in Existential Motivation? *Social Cognition*, *35*(4), 324–353.

Lambert, A. J., Eadeh, F. R., & Hanson, E. J. (2019). Anger and its consequences for judgment and behavior: Recent developments in social and political psychology. *Advances in Experimental Social Psychology*, *59*, 103–173.

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(0021), 1–9.

On the Use of Beta Coefficients in Meta-Analysis. (2005). *Journal of Applied Psychology*, *90*(1), 175–181.

Onraet, E., Van Hiel, A., Dhont, K., & Pattyn, S. (2013). Internal and External Threat in Relationship With Right-Wing Attitudes. *Journal of Personality*, *81*(3), 233–248.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology*, *72*(1), 533–560.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, *90*(5), 751–783.

Quintana, D. S. (2015). From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology*, *6*.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Schmid, A. P. (2011). *The Routledge handbook of terrorism research*. Routledge.

Sniderman, P. M., Petersen, M. B., Slothuus, R., Stubager, R., & Petrov, P. (2019). Reactions to Terror Attacks: A Heuristic Model. *Political Psychology*, *40*(S1), 245–258.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78.

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*, 1–7.

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*, *355*, l4898.

Valentine, J. C., & Cooper, H. (2008). A Systematic and Transparent Approach for Assessing the Methodological Quality of Intervention Effectiveness Research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods*, *13*(2), 130–149.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594.

Vasilopoulos, P., Marcus, G. E., Valentino, N. A., & Foucault, M. (2019). Fear, Anger, and Voting for the Far Right: Evidence From the November 13, 2015 Paris Terror Attacks. *Political Psychology*, *40*(4), 679–704.

Warren, M. (2018). First analysis of 'pre-registered' studies shows sharp rise in null findings. *Nature*.