Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 11, 3116–3134
https://doi.org/10.1037/xge0001436

The Role of Political Devotion in Sharing Partisan Misinformation and Resistance to Fact-Checking

Clara Pretus[1, 2, 3], Camila Servin-Barthet[1, 2], Elizabeth A. Harris[4, 5], William J. Brady[6], Oscar Vilarroya[1, 2], and Jay J. Van Bavel[4, 5]

[1] Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona
[2] Programa en Neurociències, Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, Spain
[3] Center of Conflict Studies and Field Research, ARTIS International, St Michaels, Maryland, United States
[4] Department of Psychology, New York University
[5] Center for Neural Science, New York University
[6] Management and Organizations Department, Kellogg School of Management, Northwestern University

Online misinformation is disproportionality created and spread by people with extreme political attitudes, especially among the far-right. There is a debate in the literature about why people spread misinformation and what should be done about it. According to the purely cognitive account, people largely spread misinformation because they are lazy, not biased. According to a motivational account, people are also motivated to believe and spread misinformation for ideological and partisan reasons. To better understand the psychological and neurocognitive processes that underlie misinformation sharing among the far-right, we conducted a cross-cultural experiment with conservatives and far-right partisans in the Unites States and Spain ($N = 1{,}609$) and a neuroimaging study with far-right partisans in Spain ($N = 36$). Far-right partisans in Spain and U.S. Republicans who highly identify with Trump were more likely to share misinformation than center-right voters and other Republicans, especially when the misinformation was related to sacred values (e.g., immigration). Sacred values predicted misinformation sharing above and beyond familiarity, attitude strength, and salience of the issue. Moreover, far-right partisans were unresponsive to fact-checking and accuracy nudges. At a neural level, this group showed increased activity in brain regions implicated in mentalizing and norm compliance in response to posts with sacred values. These results suggest that the two components of political devotion—identity fusion and sacred values—play a key role in misinformation sharing, highlighting the identity-affirming dimension of misinformation sharing. We discuss the need for motivational and identity-based interventions to help curb misinformation for high-risk partisan groups.

This article was published Online First June 22, 2023.

Clara Pretus https://orcid.org/0000-0003-2172-1184

Elizabeth A. Harris is now at Annenberg Public Policy Center, University of Pennsylvania

We want to thank all participants who took part in this study, especially those who participated in the fMRI study. We thank Marta Cazorla for support in developing materials, Luis Marcos for advice on the neuroimaging analysis, and Georgina Benet for assistance in the neuroimaging sessions.

This project has received funding from the European Union under the European Innovation Council research and innovation program. Ref. No. 101070930. The authors declare that they have no competing interests.

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). The data and analysis code employed in the analyses of the three presented studies are available at this link (https://osf.io/twr6b/). Research materials can be found in Table S3 in the online supplemental materials. Due to privacy concerns, the neuroimaging data can be obtained from the authors upon request with an approved IRB (we will send a password-protected link to scientists with ethics review board approval). Data were analyzed using R, Version 4.2.1 (R Core Team, 2022) and the package ggplot2, Version 3.3.6 (Wickham, 2016). The preregistration of the experiment design, hypotheses, and analysis plan for Experiment 1 can be found following this link. Experiments 2 and 3 were not preregistered.

The preregistration of the study design, hypotheses, and analysis plan for Experiment 1 can be found following this link. The data and analysis code employed in the analyses of the three presented studies are available at this link (https://osf.io/twr6b/). A preprint of this manuscript has been made publicly available and can be found following this link (https://psyarxiv.com/7k9gx/). The data and ideas of this manuscript have been presented at the European Society of Cognitive and Affective Neuroscience (ESCAN) Conference (Vienna, Austria) on July 21, 2022.

Clara Pretus served as lead for conceptualization, formal analysis, investigation, methodology, and writing–original draft. Camila Servin-Barthet served in a supporting role for investigation. Elizabeth A. Harris served in a supporting role for methodology. William J. Brady served in a supporting role for conceptualization and methodology. Oscar Vilarroya served as lead for funding acquisition and served in a supporting role for supervision. Jay J. Van Bavel served as lead for supervision and contributed equally to conceptualization. Camila Servin-Barthet, Elizabeth A. Harris, William J. Brady, Oscar Vilarroya and Jay J. Van Bavel contributed equally to writing–review and editing.

Correspondence concerning this article should be addressed to Clara Pretus, Programa en Neurociències, Institut Hospital del Mar d'Investigacions Mèdiques, Carrer Doctor Aiguader, 88, 08003 Barcelona, Spain, or Jay J. Van Bavel, Department of Psychology and Center for Neural Science, 6 Washington Pl, New York, NY 10003, United States. Email: cpretus@imim.es or jay.vanbavel@nyu.edu

Public Significance Statement

Given the spread of misinformation among far-right partisans, it is critical to understand the social and neural processes underlying misinformation sharing in these groups. Across two online experiments and a neuroimaging study with conservatives and far-right partisans in the Unites States and Spain, we found that far-right partisans were more likely to share misinformation relevant to conservative sacred values and were resistant to fact-checks and accuracy nudges. At a brain level, we observed a strong response in brain regions implicated in norm compliance and mentalizing to misinformation that included sacred values (vs. nonsacred values) among far-right partisans. Our work provides new theoretical and practical insights into misinformation sharing and resistance to fact-checking among far-right partisans.

Keywords: misinformation, sacred values, identity fusion, fact-checking, social media

Supplemental materials: https://doi.org/10.1037/xge0001436.supp

Over half of adults use social media as a source of news (Shearer, 2021), and malicious agents are using this opportunity to spread misinformation to larger audiences faster than ever before (Allcott & Gentzkow, 2017; Nyilasy, 2019). Attempts to quantify this phenomenon suggest that only 0.1% of social media users, mostly among the far-right, are responsible for sharing 80% of fake news, while 70% of Twitter users are exposed to fake news (Grinberg et al., 2019). Critically, online misinformation can impact real-world outcomes such as fueling political polarization (Au et al., 2021; Lee, 2016; Spohr, 2017), threatening democracy (Haslam et al., 2023; Piazza, 2022), and reducing vaccination intentions (Loomba et al., 2021). Thus, it is critical to understand the psychological processes associated with sharing online misinformation as well as potential interventions that may help counteract its spread. In the present work, we use a behavioral and neural approach to examine the role of political devotion in the spread of misinformation and immunity to fact-checks and accuracy nudges.

One of the most prominent theoretical frameworks used to understand why people share misinformation is that they simply fail to engage in analytical reasoning, they are "lazy, not biased" (Pennycook & Rand, 2019). Several studies have found that a lower propensity to engage in analytical thinking is associated with poor truth discernment in news headlines (Pennycook et al., 2020). According to this cognitive framework, interventions that "shift users' attention toward the concept of accuracy" decrease likelihood of sharing by 10% compared to control, regardless of the content of the headlines (Pennycook & Rand, 2022). Although the cognitive account and accuracy nudge has been highly influential in the misinformation literature, there is a growing debate about the theoretical and practical utility of this approach (Batailler et al., 2022; Borukhson et al., 2022; Gawronski, 2021; Pennycook & Rand, 2021). For instance, several labs have failed to replicate the relationship between cognitive reflection and misinformation sharing (Osmundsen et al., 2021) and the effect of accuracy nudges (Roozenbeek et al., 2021). Moreover, a recent meta-analysis found that accuracy nudges produce small effects, especially among far-right-wing participants ($d = 0.11$; Rathje et al., 2022). This is especially problematic since the same group of right-wing Americans is seven times more likely to share misinformation than moderates (Guess et al., 2019). This suggests that a purely cognitive approach to understanding or combatting misinformation may be relatively impotent.

In this respect, partisanship or identification with a political party has been identified as one of the main factors driving belief and willingness to share misinformation (Osmundsen et al., 2021; Pereira et al., 2023; Sternisko et al., 2023). According to the Identity-based Model of Political Belief (Van Bavel & Pereira, 2018), individuals are incentivized to believe (mis)information that affirms their partisan identities; that is, beliefs that serve belonging, epistemic, status, and moral goals associated with their social identity. Thus, for accuracy nudges to work, the benefits of being accurate need to outweigh the partisan motives (see Rathje et al., 2023), or accuracy needs to be linked to a particular social identity (e.g., as with scientists; Reinero et al., 2020). The model also makes key predictions about the neural processes underlying partisan motivations. Particularly, these value computations may be mediated by the orbitofrontal cortex (Van Bavel & Pereira, 2018), a richly interconnected brain region thought to integrate overall value during decision-making and generate evaluations (Cunningham & Zelazo, 2007; Rangel et al., 2008). As such, this model provides a multilevel framework for understanding how partisanship shapes political beliefs.

The influence of identity-motivated cognition may be especially relevant for people with extreme identities (Van Bavel & Pereira, 2018). Research on extremism often refers to extreme partisans as "devoted" actors (Atran & Ginges, 2015) who are characterized by two identity motives: identity fusion and willingness to sacrifice for sacred values. Identity fusion is a visceral feeling of oneness with a group that results from the merging between their social and personal identities (Swann et al., 2009). Unlike group identification (Hogg & Reid, 2006), which involves a process of depersonalization and adjustment to group norms (Hogg et al., 1993), identity fusion preserves a strong sense of self. It is precisely this "heightened sense of self" that enables devoted actors to initiate extraordinary actions such as risking their lives on behalf of the group and collective values (Swann et al., 2009). If devoted actors are willing to risk their lives to fight for in-group values, they may also be willing to share misinformation to promote their cause—possibly as a form of information warfare.

Neuroimaging studies have revealed that devoted actors deactivate neural networks associated with deliberation during decisions that involve high (vs. low) willingness to fight and die for in-group values (Pretus et al., 2019). Moreover, devoted actors are more likely to engage in these behaviors when the values at stake are perceived to be sacred (Atran & Ginges, 2015). Sacred values are strongly held

beliefs resistant to economic tradeoffs (Baron & Spranca, 1997; Tetlock, 2003) that often lie at the heart of intractable conflict (Atran, 2016), such as conflict over the "holy land" in the middle east. At a brain level, sacred values have been associated with increased neural activity in norm compliance networks, especially the left interior frontal cortex (Pretus et al., 2018). Thus, rule-bound thinking seems to prevail over deliberation during high-stake decisions involving sacred in-group values in devoted actors.

We propose that most partisan misinformation is shared by a small number of super spreaders (in line with Grinberg et al., 2019) who are devoted actors. As such, we hypothesized that the two components of political devotion—identity fusion and sacred values—would be critical identity motives that help drive the spread of partisan misinformation. In terms of the Identity-based Model of Political Belief (Van Bavel & Pereira, 2018), identity fusion can be understood as a form of extreme partisanship, where sacred values are highly moralized beliefs that serve partisan goals. If identity motives play any role in misinformation sharing, then two consequences should follow. First, devoted actors (e.g., fused individuals) should weigh partisan goals more heavily than accuracy goals and willingly spread misinformation when it aligns with highly relevant partisan goals (e.g., promoting sacred values). Second, attempts to enhance accuracy concerns (e.g., using fact-checks or accuracy nudges against misinformation) should be less effective among devoted actors when their identity motives are most salient (e.g., when misinformation is related to sacred values). As such, sharing partisan misinformation should be higher whenever identity motives overshadow accuracy concerns.

Current Research

In the present work, we explore how political devotion impacts misinformation sharing using a behavioral and neural approach. First, we assess how the two fundamental identity motives of political devotion—sacred values and identity fusion—impact the likelihood of sharing misinformation across two countries (the Unites States and Spain). As conservatives and far-right partisans disproportionally share far more misinformation (Garrett & Bond, 2021; Grinberg et al., 2019; Guess et al., 2019), we focused on these populations. Moreover, we assessed the efficacy of different popular interventions, including *fact-checks* used by social media companies (Porter & Wood, 2021; Walter et al., 2020), *accuracy nudges* supported by widely cited papers (Pennycook et al., 2020), and *media literacy nudges* (Jones-Jang et al., 2021).

Finally, to identify which brain networks are involved in processing misinformation related to sacred (vs. nonsacred) values, as well as to evaluate the brain response to the tested intervention, we conducted a neuroimaging study with far-right partisans in Spain ($N = 36$). Testing the functional neuroarchitecture is central to evaluating the Identity-based Model of Political Belief (Van Bavel & Pereira, 2018), and no published studies have formally tested the neural predictions from the model. Moreover, if an intervention is ineffective, the recorded brain response could provide clues as to why it is ineffective (e.g., the neurofunctional analysis could implicate different neurocognitive processes). As such, this approach would help clarify several theoretically important questions.

In line with the idea that identity motives should outweigh accuracy concerns whenever partisan goals are invoked, we predicted that misinformation relevant to sacred values (vs. nonsacred values) would be associated with greater likelihood of sharing. In addition, because sacred values are particularly resistant to tradeoffs and social influence (Sheikh et al., 2013), we expected a reduced effect of fact-checking for misinformation relevant to sacred (vs. nonsacred values). In terms of brain activity, we expected the orbitofrontal cortex to be involved in sharing messages relevant to sacred values in line with the higher subjective value for extreme partisans involved in these decisions (Van Bavel & Pereira, 2018). Finally, we expected interventions to be associated with increased activity in brain regions that support cognitive control such the dorsolateral prefrontal cortex (Weissman et al., 2008), or prediction error such as the anterior cingulate, which have been associated with subsequent behavioral adjustment (Cohen & Ranganath, 2007).

Experiment 1: Misinformation Sharing Among Conservatives in Spain

In our first experiment, we investigated whether the two components of political devotion—sacred values and identity fusion—predict the likelihood of sharing misinformation and resistance to fact-checks among far-right and center-right voters in Spain. For that, we launched a survey in Spain ($N = 812$) asking conservatives and far-right partisans to rate the likelihood of sharing a series of social media posts that contained misinformation. The misinformation looked like it was posted by different political leaders and public figures affiliated with the political party participants voted for in the last election. We also examined participants' responses to *fact-checks* used by social media companies, *accuracy nudges*, and *media literacy nudges*. Our goal was to understand if sacred values and identity fusion affect people's likelihood of sharing misinformation and their response to popular interventions against misinformation.

Method

This research was approved by the Ethics Committee on Human and Animal Experimentation at the Universitat Autònoma de Barcelona according to the Declaration of Helsinki guidelines (Ref. 5385 and 5388).

Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow Journal article reporting standards (Kazak, 2018). The data and analysis code employed in the analyses of the three presented studies are available at this link (https://osf.io/twr6b/). Research materials can be found in Table S3 in the online supplemental materials. Due to privacy concerns, the neuroimaging data can be obtained from the authors upon request with an institutional review board approval (we will send a password-protected link to scientists with ethics review board approval). Data were analyzed using R, Version 4.2.1 (R Core Team, 2022) and the package ggplot2, Version 3.3.6 (Wickham, 2016). The preregistration of the experiment design, hypotheses, and analysis plan for Experiment 1 can be found following this link. Experiments 2 and 3 were not preregistered.

A power analysis conducted with the R package "simr" (Green et al., 2016) revealed that, based on 1,000 simulations, recruiting 800 participants would enable the detection of small effects (e.g.,

a slope of 0.1) with an $\alpha = .05$ for the interaction term between (a) within-subjects value type (sacred vs. nonsacred) and fact-checking (fact-checked vs. control) with a statistical power of 99.6% (95% CI [98.98, 99.89]) with four observations per participant, and (b) within-subjects fact-checking (baseline, control, and fact-checked trials) and fact-checking group (Twitter vs. Accuracy vs. Literacy fact-check) with a statistical power of 79.1% (95% CI [76.45, 81.58]) with three observations per participant.

Participants

Inclusion criteria included being 18 years old or older and having the intention to vote for either a far-right party ("Vox") or a center-right party ("Partido Popular") in the next presidential election. Four hundred and eight Spanish far-right voters and 404 center-right voters were recruited by means of an online panel. The demographic data of the participants in all experiments is presented in Table 1. Informed written consent was obtained from all participants at the start of the study. Participants were debriefed at the end of the study.

Procedure

Participants completed a survey where they asked to rate the likelihood of sharing a series of social media posts on issues related to sacred and nonsacred values. Prior work finds that self-reported sharing intentions are highly correlated with real online news sharing (Mosleh et al., 2020). In our experiment, participants first completed a baseline block which did not include any fact-checks, followed by an experimental block that contained social media posts with and without fact-checks. Social media posts were randomized across blocks. Participants were split in three groups and each group was exposed to a different fact-check in the experimental block: the Twitter fact-check ($n = 270$, "This claim about…is disputed."), an accuracy-based fact-check ($n = 277$, "To the best of your knowledge, is the above statement accurate?" based on Pennycook et al., 2020), and a media literacy-based fact-check ($n = 265$, "What techniques are used in this Tweet to attract your attention?" based on materials by the Center of Media Literacy; www.medialit.org).

Social Media Posts

Half of the posts, designed to look like Tweets, included conservative sacred values (immigration, nationalism, and women and family values) and the other half nonsacred values (roads and infrastructure, foreign affairs, and waste management and materials). To confirm that the sacred values we tackled were judged as sacred by our experimental samples, the value sacredness of each issue was measured in the survey, as described in the "Measures" section. As expected, the three proposed sacred values were rated as sacred by a large segment of the sample (see Table S1 in the online supplemental materials). The social media posts were designed by researchers to contain false information and, whenever possible, they were inspired by real Tweets. For instance:

Real Tweet (Santiago Abascal, *Vox* leader, December 1, 2020): "On Thursday, Friday and Saturday I will visit … to learn first-hand about the concerns of Canarians in the face of the avalanche of *illegal immigrants who have assaulted our coasts* and who put the Canarian *way of life at risk*."

Designed Tweets: "This year alone, more than 100,000 *illegal immigrants have assaulted our coasts* because of the handouts this government gives them," and "3 out of 4 illegal immigrants who enter our country end up in criminal gangs, *endangering our way of life*."

Posts in both the sacred and nonsacred value conditions contained a critique of the current liberal government, which allowed us to control for out-group animosity effects (Rathje et al., 2021). As previous studies suggest an effect of moral–emotional language in online content sharing (Brady et al., 2017), all items were formulated twice, once using moral–emotional language and once using neutral language. We conducted pilot study ($N = 45$) to ensure that people perceived the two language formulations as different in terms of moral and emotional language. Results of this pilot study can be found in

Table 1

Demographic Information of the Study Samples in Spain and the United States

	Spanish sample ($N = 812$)		U.S. sample ($N = 797$)		fMRI sample ($N = 36$)
	Center-right voters	Far-right voters	Republicans not fused with Trump	Republicans fused with Trump	Spanish far-right voters
Demographic information	February 2021		July 2021		April–June 2021
N	404	408	675	122	36
% Women[a]	60	48	46	47	33
% Men	40	52	54	53	67
% Other	0	0	0.1	0	0
Age (years)	45.6 (15.6)	44.6 (13.8)	41.7 (13.7)	43.7 (13.6)	23.1 (4.8)
% Higher education	56	48	65	56	19
% Low household income (<1,200 €/month)	28[b]	32[c]	n/a	n/a	24[d]
Political orientation (1 = *left* to 10 = *right*)	7.15 (1.58)	7.55 (1.84)	7.46 (1.76)	8.5 (1.85)	8.27 (1.52)
Means of recruitment	YouGov Spain		Prolific		WhatsApp Groups

Note. fMRI = functional magnetic resonance imaging; SES = socioeconomic status.
[a] Gender was assessed with the question "Which gender do you identify with?" and the responses included "Male," "Female." and "Other." [b] SES data not available for 63 participants. [c] SES data not available for 61 participants. [d] SES data not available for six participants. No data on ethnicity was obtained.

Figure S1 in the online supplemental materials. The final set of posts included 32 items. Social media posts with sacred (vs. nonsacred) values were matched for number of likes, number of retweets, character length, and leaders tweeting the post (see details in the Supplemental Methods in the online supplemental materials).

Measures

Value Sacredness. We assessed value sacredness at the start of the survey by means of economic tradeoff scenarios, following the operationalization of sacred values used in studies on devoted actors (for instance, see Hamid et al., 2019). We asked participants if they would be "willing to give up (value) if that involved a great benefit to Spanish families such as a better economy, more jobs, better schools and hospitals, and in general, a better quality of life for all Spanish families including yours." For instance, participants were asked whether they would allow the free entry of immigrants in their country if that involved a great benefit for Spanish families. Possible responses included "Yes," "No," and "Maybe." A value was considered sacred if participants refused to give up that value in exchange for material benefits.

Identity Fusion. We assessed identity fusion using the pictorial identity fusion measure (Swann et al., 2009), which includes five pairs of circles with different degrees of overlap representing the relationship between the participant (small circle) and the group (big circle). Respondents were asked to convey which pair of circles best represents their relationship with the group. Identity fusion was assessed in relation to the political party participants supported ("Vox" for far-right voters and "Partido Popular" for center-right voters).

Likelihood of Sharing. We assessed the likelihood of sharing each of the social media posts using a probability scale "If you were to see the above post on social media, how likely would you be to share it?" (Pennycook et al., 2020). Responses could range from 1 "Extremely unlikely" to 6 "Extremely likely."

Cognitive Reflection Test. We assessed analytical (vs. intuitive) thinking styles with the cognitive reflection test (Frederick, 2005). This questionnaire includes three mathematical problems such as "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" Correct answers were added into an Analytical thinking score with values ranging between 0 and 3 ($M = 0.67$, $SD = 0.96$).

We also obtained measures of scientific curiosity, intellectual humility, and media literacy (see Supplemental Methods in the online supplemental materials).

Statistical Analysis

To evaluate the effect of sacred values, political affiliation, identity fusion, and fact-checking, we conducted a series of mixed effects models with random intercepts for participants using REML (afex package in R; Singmann et al., 2015). To assess differences in the effect of sacred values on likelihood of sharing between far-right and center-right voters, we added an interaction term for group (far-right vs. center-right voters) and type of value (sacred vs. nonsacred value). Of note, the identity fusion pictorial measure (Swann et al., 2009) was analyzed as a dichotomous variable following previous literature, with the maximal score coded as "fused" and the rest as "nonfused."

To test the effect of fact-checks in trials with and without fact-checks, we created a mixed effects model with interaction terms for fact-check group (Twitter fact-check, accuracy fact-check, and literacy fact-check) and presence of fact-check (baseline trials, fact-checked trials, and control trials). All employed tests were two-sided.

Results

Effect of Sacred Values

In line with our preregistered hypothesis, the presence of sacred values was associated with higher likelihood of sharing misinformation, $M_{diff} = 0.54$, 95% CI [0.48, 0.60], $t(811) = 18.80$, $p < .001$, $d = 0.41$, 95% CI [0.31, 0.51] (see Figure 1A). Moreover, the impact of sacred values increased as a function of the number of sacred values held by Spanish conservatives (out of the three proposed ones; see Figure 1B), $B = 0.18$, 95% CI [0.13, 0.24], $t(810) = 6.60$, $p < .001$. Thus, participants were more likely to share misinformation about sacred issues (e.g., immigration) compared to nonsacred issues (e.g., infrastructure), especially if they personally held these values as sacred, and even if both types of issues were formulated as a critique to the government or the status quo.

Effects of Political Affiliation and Identity Fusion

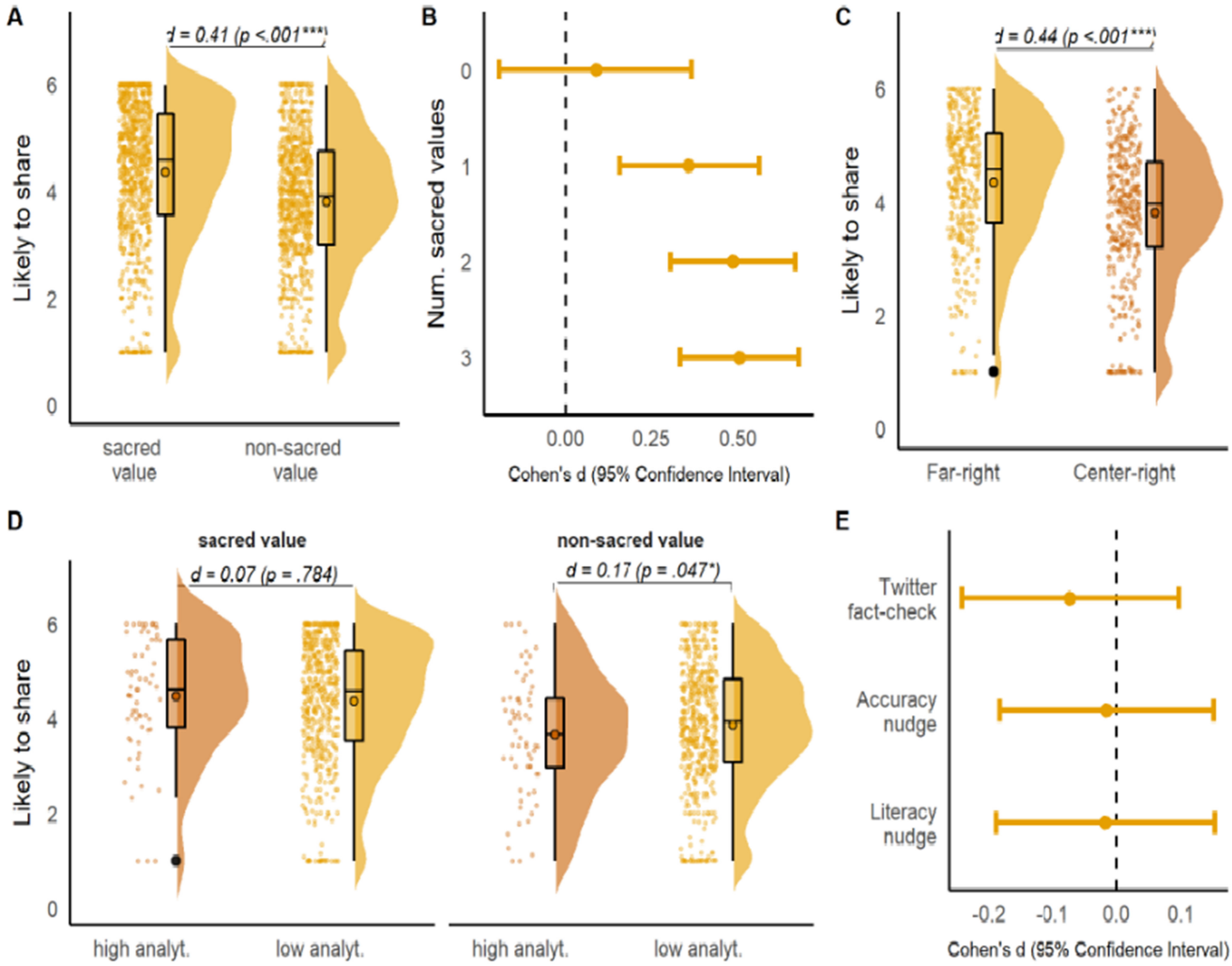
Identity fusion was more frequent among the far-right in Spain (15%, 60 out of 408) compared to the Spanish center-right (6%, 35 out of 404). Analyses of political affiliation and identity fusion effects were preregistered as exploratory. Spanish far-right voters reported more sacred values ($M = 2.09$, $SD = 0.95$) compared to Spanish center-right voters ($M = 1.62$, $SD = 1.03$), $M_{diff} = 0.46$, 95% CI [0.33, 0.60], $t(811) = 6.65$, $p < .001$, $d = 0.47$, 95% CI [0.33, 0.61], and were also more likely to share misinformation that included sacred values, compared to center-right voters (see Table 2; Figure 1C). Political affiliation with the Spanish far-right and identity fusion across groups was associated with a greater likelihood of sharing misinformation—political affiliation with the Spanish far-right vs. center-right: $M_{diff} = 0.54$, 95% CI [0.37, 0.71], $t(811) = 6.28$, $p < .001$, $d = 0.41$, 95% CI [0.31, 0.51]; fused with the Spanish far-right vs. nonfused: $M_{diff} = 0.92$, 95% CI [0.61, 1.24], $t(406) = 5.76$, $p < .001$, $d = 0.81$, 95% CI [0.52, 1.09]; fused with the Spanish center-right vs. nonfused: $M_{diff} = 0.74$, [0.31, 1.17], $t(402) = 3.41$, $p < .001$, $d = 0.60$, 95% CI [0.25, 0.95].

Limited Efficacy of Fact-Checking

Contrary to our preregistered hypothesis and prior research (Porter & Wood, 2021; Walter et al., 2020), fact-checks and the accuracy nudge did not have any overall effect within the experimental block, $M_{diff} = -0.05$, 95% CI [−0.11, 0.02], $t(1,622) = -1.73$, $p = .20$, $d = -0.04$, 95% CI [−0.13, 0.06], and the different types of fact-checks and accuracy nudges did not affect likelihood of sharing misinformation during fact-checked trials compared to control trials (literacy nudge: $p = .88$, $d = -0.02$, 95% CI [−0.19, 0.15]; accuracy nudge: $p = .89$, $d = -0.02$, 95% CI [−0.18, 0.15]; Twitter fact-check: $p = .10$, $d = -0.07$, 95% CI [−0.24, 0.10]; see Figure 1E). Across blocks, respondents actually *increased* their

Figure 1

Experiment 1 Results (Spanish Sample)



Note. (A) Participants reported higher likelihood of sharing social media posts with misinformation relevant to sacred versus nonsacred values; (B) likelihood of sharing misinformation relevant to sacred (vs. nonsacred) values increased as a function of the number of sacred values held by participants; (C) far-right voters reported higher likelihood of sharing misinformation than center-right voters; (D) analytical thinking style was associated with reduced sharing of misinformation about nonsacred values but was unrelated to sharing misinformation about sacred values; and (E) fact-checks were largely ineffective, with no effect of type of fact-check for fact-checked compared to control trials. Boxplot bounds represent the interquartile range, and the circle within the box represents the means. See the online article for the color version of the figure.

likelihood of sharing during fact-checked trials in the experimental block compared to control trials in the previous baseline block, $M_{diff} = 0.16$, 95% CI [0.10, 0.22], $t(1,622) = 5.87$, $p < .001$, suggesting either a temporal effect or a backfire effect of fact-checking (see Nyhan & Reifler, 2010). Therefore, we included a control group in Experiment 2 to account for potential temporal effects. Overall, three popular interventions aimed to reduce misinformation sharing were notably ineffective (ds = 0.02–0.07) among conservatives and far-right voters in Spain.

Moderation Effects

The small effect sizes of fact-checks (average $d = 0.04$) made it difficult to detect moderator effects. For instance, contrary to our preregistered hypothesis, type of value (sacred vs. nonsacred) did not moderate the effect of fact-checks (sacred values: $p = .20$, $d = -0.04$, 95% CI [−0.14, 0.05]; nonsacred values: $p = .76$, $d = -0.02$, 95% CI [−0.11, 0.08]).

Other Variables of Interest

Analytical thinking style, as measured by correct responses in the cognitive reflection test, has been previously found to predict reduced sharing of fake news (Pennycook & Rand, 2019). While analytical thinking style was weakly associated with reduced sharing of misinformation about nonsacred values, $M_{diff} = -0.29$, 95% CI [−0.57, −0.003], $t(982) = -1.98$, $p = .047$, $d = -0.16$, 95% CI [−0.43, 0.09], it did not affect sharing of misinformation about

Table 2
Moderation Effect of Political Affiliation and Identity Fusion on Willingness to Share Misinformation Relevant to Sacred and Nonsacred Values in the Spanish Sample

Effect	*M*	*SD*	M_{diff}	95% CI	*t*	*p*	Cohen's *d*	95% CI
(a) Political affiliation (Spain)								
Interaction: Political Affiliation × Sacred Values			0.24	[0.14, 0.35]	4.14	<.001		
Far-right								
SV	4.67	1.51						
Non-SV	4.02	1.51	0.66	[0.58, 0.74]	16.38	<.001	0.52	[0.38, 0.66]
Center-right								
SV	4.02	1.68						
Non-SV	3.60	1.52	0.42	[0.34, 0.50]	10.46	<.001	0.32	[0.19, 0.46]
(b) Fusion with far-right (Spain)								
Interaction: Identity Fusion × Sacred Values			0.08	[−0.15, 0.32]	0.68	*n.s.* (.50)		
Fused								
SV	5.43	1.08						
Non-SV	4.84	1.30	0.59	[0.37, 0.81]	5.31	<.001	0.62	[0.25, 0.99]
Nonfused								
SV	4.54	1.54						
Non-SV	3.87	1.50	0.67	[0.58, 0.76]	14.55	<.001	0.53	[0.38, 0.68]
(c) Fusion with center-right (Spain)								
Interaction: Identity Fusion × Sacred Values			0.03	[−0.23, 0.29]	0.23	*n.s.* (.82)		
Fused								
SV	4.68	1.60						
Non-SV	4.29	1.54	0.39	[0.14, 0.65]	3.05	.002	0.34	[−0.14, 0.82]
Nonfused								
SV	3.96	1.67						
Non-SV	3.53	1.50	0.42	[0.35, 0.50]	10.71	<.001	0.32	[0.18, 0.47]

Note. CI = confidence interval; SV = sacred values.

sacred values, $M_{diff} = -0.04$, 95% CI [−0.32, 0.24], $t(982) = -0.28$, $p = .78$, $d = 0.08$, 95% CI [−0.18, 0.34] (see Figure 1D). Since analytical thinking was not a very effective buffer against spreading misinformation, the cognitive reflection test was not administered in Experiment 2.

In line with our preregistered hypothesis, presence of moral–emotional language increased sharing in the Spanish sample (see Supplemental Results in the online supplemental materials). With regards to other variables of interest, both media literacy and humility, but not scientific curiosity, were associated with increased sharing of posts with sacred values compared to nonsacred values (see Supplemental Results and Supplementary Discussion in the online supplemental materials).

Discussion

In our first experiment, we evaluated whether sacred values and identity fusion impact people's likelihood of sharing misinformation and resistance to fact-checking among far-right and center-right voters in Spain. As expected, we found that far-right voters reported a higher likelihood of sharing misinformation than center-right voters, especially when that misinformation was relevant to sacred values (e.g., immigration) compared to nonsacred values (e.g., infrastructure), and even if both types of misinformation were formulated as a critique to the government. Identity fusion with a political party (far-right and center-right parties) also predicted higher likelihood of sharing misinformation, regardless of whether the content was relevant to sacred or nonsacred issues. While analytical thinking style was weakly associated with decreased likelihood of sharing misinformation relevant to nonsacred values, it did not influence disposition to share misinformation related to sacred values. Contrary to our predictions, popular interventions such as fact-checks and accuracy nudges did not reduce likelihood of sharing misinformation among either far-right or center-right voters in Spain. This suggests that these interventions might not be well suited for these high-risk populations (see also Rathje et al., 2022).

Our findings suggest that devoted actors and appeals to sacred values in political online messages increases people's likelihood of sharing misinformation, especially among far-right voters. Of note, analytical thinking does not seem to reduce people's likelihood of engaging with misinformation when it appeals to sacred values. Moreover, people who feel a visceral connection with a political party (identity fusion) seem to have a higher disposition to share online misinformation posted by their political leaders regardless of whether the content is relevant to sacred values or not. As fact-checks and accuracy nudges did not change participants' likelihood of sharing misinformation, we were not able to evaluate whether sacred values and identity fusion moderated people's resistance to fact-checks among conservatives and far-right voters in Spain.

One of the most striking findings is the failing of several population interventions to stem the spread of misinformation. While none of the employed interventions yielded significant effects, the Twitter fact-check exhibited the largest (nonsignificant) effect with a very small effect size of d of 0.07 compared to $d = 0.02$ for the accuracy nudge and the media literacy nudge. Likewise, a recent meta-analysis found that accuracy nudges have very small effects on right-wing partisans ($d = 0.11$; Rathje et al, 2022). Therefore, we focused on the Twitter fact-check in the rest of the experiments since it seemed the most promising approach for stemming the spread of misinformation in this population.

Experiment 2: Misinformation Sharing Among Republicans in the United States

In our second experiment, we sought to replicate the effects of sacred values and identity fusion on the likelihood of sharing misinformation among Republicans in the United States. We conducted an experiment in the United States ($N = 797$) asking Republican party voters to rate the likelihood of sharing a series of social media posts on issues related to sacred and nonsacred values designed to appear as though they were posted by different Republican party leaders and conservative public figures. To evaluate the role of sacred values and identity fusion in Republicans' response to interventions against misinformation we focused on the intervention that yielded the largest (nonsignificant) effect size in Experiment 1, the Twitter fact-check. To control for temporal effects across the baseline and the experimental block, we included a control group that was not exposed to the intervention. Moreover, because misinformation interventions could be less effective for plausible compared to implausible misinformation, we evaluated the effect of perceived accuracy of each social media post on participants' response to the intervention. Finally, because sacred values are associated with greater attitude strength, familiarity, and salience compared to nonsacred values, we also controlled for these confounds in this experiment.

Method

A power analysis conducted with the R package "simr" (Green et al., 2016) revealed that, based on 1,000 simulations, recruiting 800 participants would enable the detection of small effects (e.g., a slope of 0.1) with an $\alpha = 0.05$ for the interaction term between (a) within-subject value type (sacred vs. nonsacred) and fact-checking (fact-checked vs. control) with a statistical power of 99.6% (95% CI [98.98, 99.89]) with four observations per participant, and between (b) within-subject block (baseline vs. experimental block) and fact-checking group (fact-checked group vs. control group) with a statistical power of 82.9% ([80.42, 85.18]) with two observations per participant.

Participants

The demographic data of the study participants is presented in Table 1. We recruited a sample of 797 participants via an online panel, who had reported voting for Republican Donald J. Trump in the two previous U.S. presidential elections (2016 and 2020). Informed written consent was obtained from all participants at the start of the study. Participants were debriefed at the end the study.

Procedure

Participants were asked to rate the likelihood of sharing a series of social media posts on issues related to sacred and nonsacred values composed by different Republican political leaders. Similar to Experiment 1, participants first completed a baseline block which did not include any fact-checks, followed by an experimental block that contained social media posts with and without fact-checks. Social media posts were randomized across blocks. In light of the null effects ($d = 0.02$) of the accuracy and the media literary nudge in Experiment 1, we only used the Twitter fact-check in this study since the effect size was slightly larger ($d = 0.07$). Contrary to Experiment 1, half of the sample in Experiment 2 was exposed to the Twitter fact-check and the other half did not see any fact-checks (control group).

To control for the effect of potential confounds associated with sacred values and the effect of the intervention, we also obtained perceived accuracy, attitude strength, familiarity, and salience scores for each item in a pilot study on a sample of 80 U.S. Republicans who were prescreened using the same criteria as the Experiment 2 participants (see the Supplemental Methods in the online supplemental materials).

Social Media Posts

The 32 social media posts employed in Experiment 2 were designed following the same procedure used in Experiment 1. Half of the posts included conservative sacred values (immigration, nationalism, and women and family values) and the other half nonsacred values (matched for number of likes, number of retweets, character length, and leaders tweeting the post, see details in the Supplementary Methods in the online supplemental materials). Value sacredness was measured in the survey (see "Measures" section and Table S1 in the online supplemental materials). Similarly to Experiment 1, most posts across conditions contained critiques to the current liberal government (see complete list of items employed in the U.S. study in Table S2 in the online supplemental materials) and all items were formulated twice, once using moral–emotional language and once using neutral language, a distinction that was confirmed by means of a pilot study ($N = 45$ each; see Figure S1 in the online supplemental materials).

Measures

Similarly to Experiment 1, we obtained *Value sacredness* measures and *Identity fusion* with the Republican party. In addition, U.S. Republicans were also asked to rate their identity fusion with Trump specifically (following Kunst et al., 2019). Unlike fusion with the Republican party, fusion with Trump has been found to predict willingness to engage in political violence (Kunst et al., 2019). *Likelihood of sharing* each social media post was assessed with the same 6-point Likert scale used in Experiment 1. Media literacy measures were also obtained (see Supplemental Methods in the online supplemental materials).

Statistical Analysis

We followed the same analysis strategy used in Experiment 1. To assess differences in the effect of sacred values on likelihood of sharing between Republicans fused and nonfused with Trump, we added an interaction term for group (fused vs. nonfused with Trump) and type of value (sacred vs. nonsacred value). To test the effect of fact-checks, we created a mixed effects model with interaction terms for the between-subjects fact-check condition (fact-checked group vs. control group) and the within-subject fact-check condition (baseline block vs. experimental block). All statistical tests were two-sided.

Results

Effect of Sacred Values

Replicating Experiment 1, the presence of sacred values was associated with a higher likelihood of sharing misinformation among

U.S. Republicans, $M_{diff} = 0.56$, 95% CI [0.50, 0.63], $t(796) = 16.03$, $p < .001$, $d = 0.38$, 95% CI [0.28, 0.48] (see Figure 2A). Similarly to Experiment 1, the impact of sacred values also increased as a function of the number of sacred values held by Republicans, $B = 0.19$, 95% CI [0.13, 0.25], $t(795) = 6.41$, $p < .001$ (see Figure 2B).

To assess and control the effect of perceived accuracy, attitude strength, familiarity, and salience on participants' likelihood of sharing social media posts relevant to sacred (vs. nonsacred) values in Experiment 2 (see Figure 2F), we reran the analyses accounting for the effect of each of these variables (exploratory analysis). When assessed separately, all variables influenced the likelihood of sharing social media posts—perceived accuracy: $B = 1.13$, 95% CI [0.94, 1.32], $t(950) = 11.67$, $p < .001$; attitude strength: $B = 0.78$, 95% CI [0.67, 0.89], $t(842) = 13.88$, $p < .001$; familiarity: $B = 0.79$, 95% CI [0.68, 0.90], $t(857) = 14.14$, $p < .001$; and salience: $B = 0.84$, 95% CI [0.72, 0.95], $t(867) = 13.83$, $p < .001$. However, when we added all of these variables into the same model together with value sacredness, value *sacredness* remained the only significant predictor of likelihood of sharing social media posts—value sacredness: $B = 0.62$, 95% CI [0.44, 0.80], $t(1,091) = 6.78$, $p < .001$; accuracy: $B = 0.26$, 95% CI [−0.15, 0.66], $t(967) = 1.24$, $p = .216$; attitude strength: $B = 0.02$, 95% CI [−0.45, 0.48], $t(1,004) = 0.07$, $p = .943$; familiarity: $B = 0.06$, 95% CI [−0.36, 0.48], $t(938) = 0.29$, $p = .772$; and salience: $B = -0.32$, 95% CI [−0.77, 0.12], $t(1,140) = -1.42$, $p = .157$. Since both value sacredness and the variable scores for each social media post were obtained from sample averages, differences in individual-level versus sample-level measurements did not appear to influence these results. Thus, sacred (vs. nonsacred) values were associated with increased likelihood of sharing social media posts above and beyond differences in attitude strength, familiarity, and salience (which were all nonsignificant once accounting for sacredness).

Effects of Identity Fusion

Identity fusion with Trump (15%, 122 out of 797) was more frequent among U.S. Republicans than identity fusion with the Republican party (6%, 51 out of 797, were fused with the Republican party but not Trump, and 14% were fused with both). Republicans fused with Trump reported more sacred values ($M = 2.09$, $SD = 1.19$) compared to Republicans not fused with Trump ($M = 1.18$, $SD = 1.14$), $M_{diff} = 0.24$, 95% CI [0.02, 0.46], $t(796) = 2.11$, $p = .035$, $d = 0.21$, 95% CI [0.01, 0.40]. Republicans fused with Trump were also more likely to share misinformation that included sacred values compared to Republicans not fused with Trump (see Table 3; Figure 2C). Identity fusion across groups was associated with a greater likelihood of sharing misinformation—fused with the Republican party vs. nonfused, excluding fused with Trump: $M_{diff} = 1.25$, 95% CI [0.89, 1.61], $t(673) = 6.77$, $p < .001$, $d = 0.99$, 95% CI [0.69, 1.28]; fused with Trump vs. nonfused: $M_{diff} = 0.89$, 95% CI [0.63, 1.15], $t(795) = 6.72$, $p < .001$, $d = 0.66$, 95% CI [0.47, 0.86].

Limited Efficacy of Fact-Checking

Within the experimental block, participants in the fact-checked group exhibited very small, but significant, decreases in their likelihood of sharing misinformation in fact-checked trials compared to control trials, $M_{diff} = 0.16$, 95% CI [0.07, 0.25], $t(377) = 3.47$, $p = .006$, $d = 0.11$, 95% CI [−0.04, 0.25]. Experiment 2 also revealed the presence of a temporal effect across blocks: While people in the control group (no intervention) were more likely to share social media posts in the second experimental block compared to the baseline block, $M_{diff} = 0.14$, 95% CI [0.06, 0.23], $t(1,173) = 3.33$, $p < .001$, $d = 0.10$, 95% CI [−0.04, 0.24], the experimental group did not increase their likelihood of sharing in the experimental block compared to baseline, $M_{diff} = 0.01$, 95% CI [−0.06, 0.09], $t(1,173) = 0.37$, $p = .71$, $d = 0.01$, 95% CI [−0.11, 0.13]. Thus, fact-checks led to a very small ($d = 0.10$) reduction in sharing in the experimental block.

Moderation Effects

Similarly to Experiment 1, type of value (sacred vs. nonsacred) did not moderate the effect of fact-checks in the U.S. sample (sacred values: $p = .03$, $d = -0.08$, 95% CI [−0.22, 0.06]; nonsacred values: $p = .03$, $d = -0.09$, 95% CI [−0.23, 0.05]). However, an exploratory analysis revealed that identity fusion with Trump (but not identity fusion with the Republican party) moderated the effect of fact-checks in the fact-checked group. Republicans fused with Trump did not reduce misinformation sharing during fact-checked compared to control trials, $M_{diff} = -0.02$, 95% CI [−0.24, 0.19], $t(376) = -0.21$, $p = .84$, $d = -0.01$, 95% CI [−0.35, 0.32], while the rest of Republicans did, $M_{diff} = -0.19$, 95% CI [−0.29, 0.09], $t(376)$[1] $= -3.74$, $p < .001$, $d = -0.13$, 95% CI [−0.29, 0.02] (see Figure 2D). Hence, while Twitter fact-checks were somewhat effective in reducing misinformation sharing among nonfused Republicans, they were ineffective among highly identified Trump supporters.

Moderation Effect of Perceived Accuracy

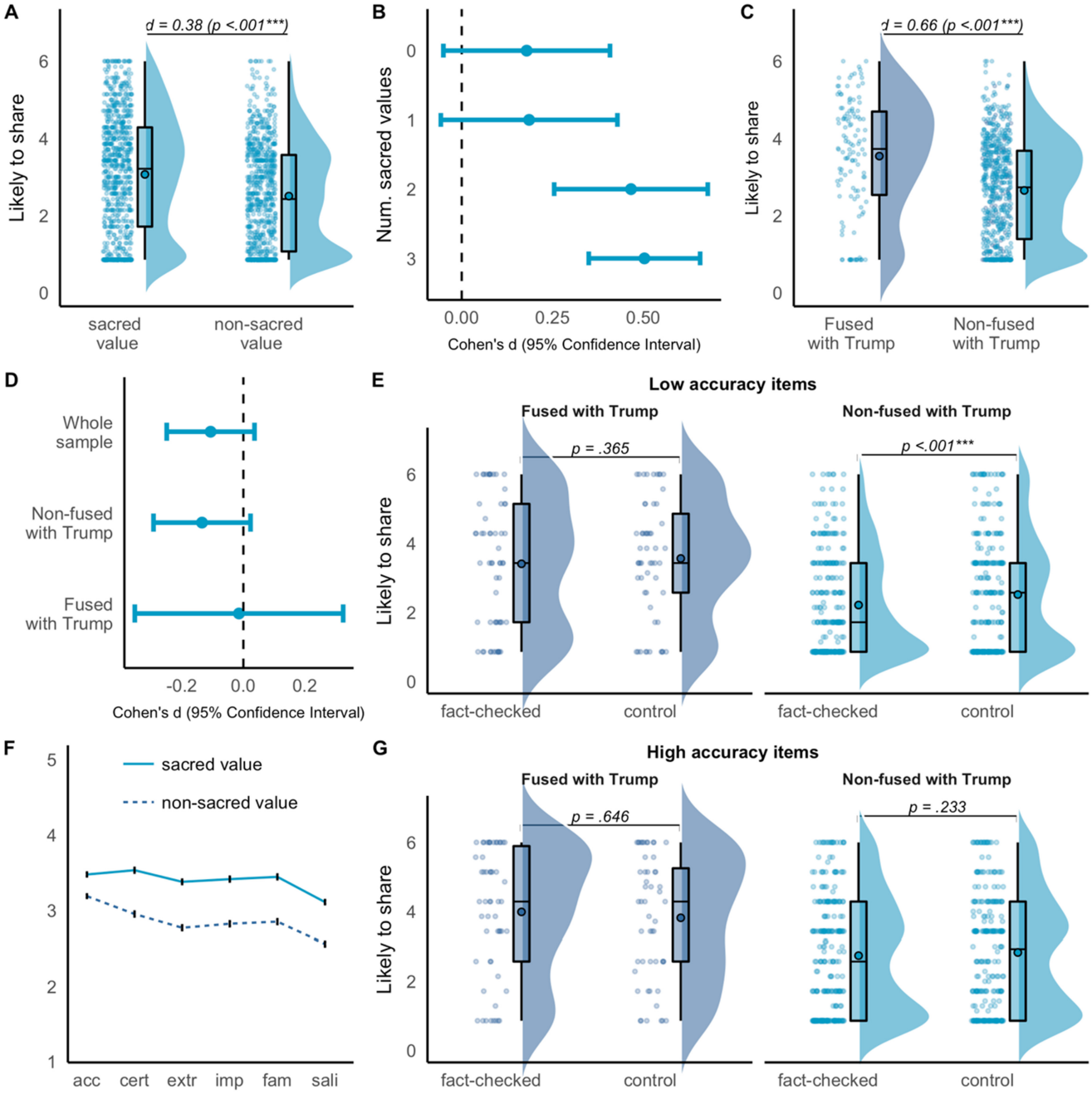
To explore possible psychological processes underlying the reduced effect of fact-checks in Republicans fused with Trump, we assessed how these participants responded to fact-checks in social media posts with different levels of perceived accuracy (exploratory analysis). To maximize differences between social media posts associated with high and low perceived accuracy, we selected only the top 10 and the bottom 10 social media posts (out of 32 displayed posts) which were associated, respectively, with the highest ($M = 3.96$, $SD = 0.19$) and lowest ($M = 2.71$, $SD = 0.29$) perceived accuracy ratings obtained in a pilot study with a different sample of Republicans ($N = 80$). We then ran a model comparing the effects of fact-checking on the likelihood of sharing high (vs. low) perceived accuracy social media posts in participants fused with Trump (vs. nonfused). We did not find any evidence of a three-way interaction ($p = .546$) but given the relatively low statistical power, we decided to analyze simple effects. Specifically, we found that fact-checking was (weakly) effective only for low perceived accuracy posts among Republicans who were not fused with Trump, $M_{diff} = -0.30$, 95% CI [−0.47, −0.12], $t(905)$[2] $=$

[1] Only participants in the fact-checked group ($n = 378$) were included in this analysis.

[2] Only participants in the fact-checked group ($n = 378$) were included in this analysis, with four observations per participant (Fact-Checked Trials versus Control Trials × High versus Low Accuracy Items).

Figure 2

Experiment 2 Results (U.S. Sample)



Note. (A) Participants reported higher likelihood of sharing social media posts with misinformation relevant to sacred versus nonsacred values; (B) likelihood of sharing misinformation relevant to sacred (vs. nonsacred) values increased as a function of the number of sacred values held by participants; (C) Republicans fused with Trump reported higher likelihood of sharing misinformation than other Republican voters; (D) the Twitter fact-check had a small effect in the experimental block (vs. baseline) in fact-checked participants compared to control participants, though the effect was nonsignificant in Republicans fused with Trump; (E) Republicans fused with Trump were unresponsive to fact-checks even when the social media posts were among the least plausible ones (as rated by other Republicans); (F) social media posts relevant to sacred values were associated with higher attitude strength ("cert" = certainty, "extr" = extremity, and "imp" = importance), familiarity ("fam"), and salience ("sal") scores (but similar "acc" = perceived accuracy scores) than social media posts relevant to nonsacred values as rated by other Republicans in a pilot study; and (G) fact-checks were ineffective for both Republicans fused and nonfused with Trump for the top 10 most plausible social media posts as rated by other Republicans in a pilot study. Boxplot bounds represent the interquartile range and the circle within the box represents the means. See the online article for the color version of the figure.

Table 3
Moderation Effect of Identity Fusion on Willingness to Share Misinformation Relevant to Sacred and Nonsacred Values in the U.S. Sample

Effect	*M*	*SD*	M_{diff}	95% CI	*t*	*p*	Cohen's *d*	95% CI
(a) Fusion with Trump (United States)								
Interaction: Identity Fusion × Sacred Values			0.32	[0.14, 0.52]	3.43	<.001		
Fused								
SV	3.97	1.99						
Non-SV	3.15	1.86	0.84	[0.67, 1.02]	9.47	<.001	0.52	[0.27, 0.78]
Nonfused								
SV	2.89	1.80						
Non-SV	2.39	1.56	0.51	[0.44, 0.59]	13.51	<.001	0.37	[0.26, 0.47]
(b) Fusion with the Republican party, excluding those fused with Trump (United States)								
Interaction: Identity Fusion × Sacred Values			0.15	[0.06, 0.34]	1.08	*n.s.* (.28)		
Fused								
SV	4.00	1.69						
Non-SV	3.63	1.65	0.37	[0.11, 0.64]	2.78	.006	0.30	[−0.10, 0.69]
Nonfused								
SV	2.81	1.79						
Non-SV	2.29	1.51	0.52	[0.45, 0.60]	13.69	<.001	0.38	[0.27, 0.50]

Note. CI = confidence interval; SV = sacred values.

−3.35, $p < .001$, $d = -0.19$, 95% CI [−0.36, −0.02] (see Figure 2E). Conversely, fact-checks were ineffective for low perceived accuracy posts among Republicans fused with Trump, $M_{diff} = -0.20$, 95% CI [−0.58, 0.18], $t(909) = -1.02$, $p = .307$, $d = -0.09$, 95% CI [−0.46, 0.29], and for high perceived accuracy posts across groups—fused with Trump: $M_{diff} = 0.17$, 95% CI [−0.23, 0.57], $t(918) = 0.83$, $p = .405$, $d = 0.09$, 95% CI [−0.29, 0.48]; nonfused with Trump: $M_{diff} = -0.12$, 95% CI [−0.30, 0.06], $t(909) = -1.29$, $p = .196$, $d = -0.05$, 95% CI [−0.23, 0.12] (see Figure 2G). Thus, our exploratory analysis tentatively suggests that fact-checks are modestly effective for highly implausible posts, though not for Republicans fused with Trump.

Other Variables of Interest

In the U.S. sample, moral–emotional language only increased sharing among Republicans fused with Trump (see the Supplemental Results in the online supplemental materials). Media literacy did not moderate sharing of posts with sacred values compared to nonsacred values (see the Supplemental Results in the online supplemental materials).

Discussion

In Experiment 2, we aimed to replicate the effects of sacred values and identity fusion in a sample of Republicans in the United States. Similar to far-right voters in Spain, Republicans fused with Trump in the United States exhibited a higher likelihood of sharing misinformation than the rest of Republicans, especially when that misinformation was relevant to sacred values (e.g., immigration) compared to nonsacred values (e.g., infrastructure). Of note, the effect of sacred values on Republicans intention to share misinformation was independent of attitude strength, familiarity, and salience effects. Identity fusion with the Republican party was also associated with a higher likelihood of sharing misinformation, regardless of whether values were sacred or not. The Twitter fact-check had a small but significant effect in reducing participants' likelihood of sharing misinformation among Republicans. However, there was no effect for Republicans fused with Trump. Moreover, we found preliminary evidence that Republicans fused with Trump were resistant to interventions against misinformation even when that misinformation was perceived as implausible by other Republicans.

These results suggest that the two components of political devotion—sacred values and identity fusion—are relevant predictors of sharing partisan misinformation in the United States and Spain. Appealing to sacred values in online messages seems particularly effective in increasing misinformation sharing among Republicans fused with Trump in the Unites States (vs. other Republicans) and among far-right partisans in Spain (vs. center-right voters). In addition, Republicans fused with Trump and far-right partisans in Spain were particularly resistant to fact-checking, even for social media posts judged to be implausible by fellow party members. This calls into question the efficacy of fact-checking strategies among these high-risk groups and raises important questions about the underlying cognitive processes underlying these decisions. In the following experiment, we used neuroimaging to better understand why political devotion was driving the spread of misinformation.

Experiment 3: Neural Correlates of Misinformation Sharing in Far-Right Partisans

In our third experiment, we aimed to identify the neural activity and functional connectivity underlying the spread of misinformation relevant to sacred values. Moreover, we were interested in investigating whether these neural correlates were influenced by exposing people to the Twitter fact-check. We conducted a neuroimaging study with far-right partisans in Spain ($N = 36$). We did not include a control group because we were interested in the effect of sacred values and fact-checking on misinformation sharing in participants who we knew would respond strongly to the presented stimuli rather than identifying neural activity patterns specific to these participants in comparison to others (and because between subject comparisons are highly underpowered in neuroimaging research, making it difficult to make strong inferences; see Yarkoni, 2009). Similarly to Experiments 1 and 2, participants were asked to rate the likelihood of sharing social media posts composed by far-right party leaders that included misinformation relevant to sacred and nonsacred values, with and without Twitter fact-checks. This time, we obtained

images of their brain activity using functional magnetic resonance imaging (fMRI) while they were completing the task.

We aimed to evaluate if online messages relevant to sacred (vs. nonsacred) values modulated brain activity and functional connectivity patterns, especially in regions involved in social decision-making. For that, we analyzed whole brain and region-of-interest (ROI) activity differences while participants were being exposed to online messages with misinformation relevant to sacred and nonsacred values. The ROI analyses included brain regions previously associated with sacred compared to nonsacred values, such as the left inferior frontal gyrus (Pretus et al., 2018), and brain regions associated with integrating subjective value during decision-making, such as the orbitofrontal cortex (Cunningham & Zelazo, 2007; Rangel et al., 2008). Finally, we conducted a functional connectivity analysis to evaluate how the sacred (vs. nonsacred) value condition modulated functional connectivity between brain regions. We examined task-dependent changes in functional connectivity between neural networks associated with social cognition (the default mode network), executive control (the frontoparietal network), and two attentional networks (salience and dorsal attention networks).

Method

A power analysis based on 1,000 simulations and an $\alpha = 0.05$ showed that with 40 participants and 80 observations per participant (40 trials per run in two runs), we would be able to detect small effects (e.g., a slope of 0.1) for the interaction between within-subjects value type (sacred vs. nonsacred) and within-subjects fact-checking (control vs. fact-checked trials) with a statistical power of 99.80% (95% CI [99.28, 99.98]).

Participants

We recruited 36 far-right partisans using social media. Initially, we contacted Twitter followers of far-right accounts based in Barcelona via direct message and recruited them to participate in the study. Ultimately, most of our participants were recruited through other participants that shared our study ad in far-right youth WhatsApp groups that are used to organize political actions locally and usually include a few hundred members. Participants were selected based on their response to the question "Which political party best represents your values and beliefs at this moment?" Only participants who responded with a far-right party ("Vox") and were 18 years old or older were recruited for the study. Candidates with neurological and psychiatric disorders, taking psychiatric medication, claustrophobic, or with metal parts in their body incompatible with a magnetic resonance scanner were excluded. Selected participants were at the far-right end of the political spectrum (8.27 out of 10 points on a liberal to conservative scale, see demographics in Table 1), similar to U.S. Republicans fused with Trump who participated in Experiment 2 (8.50 out of 10). Informed written consent was obtained from all participants at the start of the study. Participants were debriefed at the end of the study and were given the chance to discuss and ask questions about the study.

Procedure

Participants completed an event-related fMRI paradigm where they had to convey how likely they would be to share a series of social media posts using a 6-point Likert scale adapted for fMRI. The fMRI paradigm was designed using MATLAB 2020a (The MathWorks, Inc., Natick, Massachusetts, United States) with Psychtoolbox extensions (Kleiner et al., 2007). Participants completed two 6-min runs without any fact-checks (baseline block) followed by another two 6-min runs with Twitter fact-checks (experimental block). The order of these two blocks was reversed in half of the sample ($N = 18$). Stimuli were presented randomly with jittered intertrial intervals that varied randomly from 0.5 up to 3.70 s. Each run included 40 7-s trials (20 trials for the sacred value condition and 20 trials for the nonsacred value condition) and in each trial, participants were given 4.5 s to read a social media post, and 2.5 s to respond how likely they would be to share it using a 6-point Likert scale adapted for fMRI using left/right buttons (see Figure S2). The orientation of the scale was reversed in half of the trials and the cursor starting point was randomized. Half of the presented social media posts included sacred values and the other half nonsacred values. The total duration of the paradigm was 24 min. Participants had the chance to practice the task thoroughly on a laptop before going into the scanner. No posts with fact-checks or sacred values were shown during the training phase.

Materials

The 32 social media posts employed in Experiment 3 were simplified versions of the items used in Experiment 1. As participants were given 4.5 s to read each statement, the items were shortened from an average character length of 134 ($SD = 2.97$) to an average character length of 93.4 ($SD = 2.83$) while preserving the substantive content.

Measures

Similarly to Experiment 1, we obtained *Value sacredness* measures and assessed *Identity fusion* with the political party participants supported ("Vox"). Likelihood of sharing was assessed with the same 6-point Likert scale used in Experiments 1 and 2.

Neuroimaging Data. Magnetic resonance images were obtained in a Philips 3 T scanner. T1-weighted images were acquired using a FSPGR sequence (Repetition Time [TR]: 9.9 ms, Time to Echo [TE]: 4.6 ms, Fractional anisotropy [FA]: 8, matrix size: 240 × 240, 180 slices, and slice thickness: 1.00 mm) and functional volumes were obtained using a EPI-T2* sequence (TR: 1,750 ms, TE: 35 ms, FA: 70, matrix size: 76 × 76, 46 slices, and slice thickness: 3.1 mm).

Statistical Analysis

Image processing was conducted using SPM12 (Wellcome Trust, University College London, UK) on MATLAB. Functional images were realigned and coregistered to the structural images, which were then segmented into white matter, gray matter, and cerebrospinal fluid. The forward deformation fields generated during the segmentation of the structural images were used to normalize the functional images. Finally, images were smoothed using an 8-mm FWHM kernel.

The first-level general linear model (GLM) included two regressors for stimulus presentation (for sacred vs. nonsacred value trials) with a duration of 4.5 s each and two regressors for the response period (for sacred vs. nonsacred value trials) with a duration of

2.5 s each. The response regressors were parametrically modulated by the number of keypresses in each trial. Six nonconvolved head movement regressors were also included. In addition, first-level GLMs included scrubbing regressors that accounted for outlier volumes as required by each participant. Outliers included volumes with a framewise displacement larger than 0.9 mm or global BOLD signal changes above 0.5 *SD*.

Group-level effects of sacred versus nonsacred values during stimulus presentation time-locked to the start of the onset stimulus were evaluated by means of a *t*-test on individual-level contrasts. Group differences related to the order of the runs (half of the sample started with fact-checks and the other half without fact-checks) were assessed using a two-sample *t*-test on individual-level contrasts. The fact-check by value interaction contrast was computed by assigning counterbalanced weights to fact-checked (fc) versus control (ctr) posts with sacred values versus nonsacred values ([1 × ctr sacred values − 1 × ctr nonsacred values − 1 × fc sacred values 1 × fc nonsacred values]).

Post hoc analyses included parameter estimate extractions of the clusters of activity obtained in the group-level contrasts using Marsbar for visualization purposes (see Figure 3B). We also conducted ROI analyses to evaluate neural activity differences in the sacred versus nonsacred value condition specifically in the left inferior frontal gyrus, previously associated with sacred versus nonsacred values (Pretus et al., 2018) and the orbitofrontal cortex, a region thought to integrate subjective value during decision-making (Cunningham & Zelazo, 2007; Rangel et al., 2008). The ROI masks were anatomical and based on the Neuromorphometrics atlas for SPM12. The inferior frontal gyrus mask included the (bilateral) orbital, opercular, and triangular part of the inferior frontal gryus; the orbitofrontal cortex mask included the (bilateral) lateral, anterior, and medial orbital gyrus based on neuromorphometrics labels. Reported results were corrected for multiple comparisons using family-wise error (FWE) correction at a cluster level, with a peak-level contrast of $p < .001$.

Functional Connectivity Analysis

A functional connectivity analysis was conducted on SPM12 using the connectivity toolbox CONN (Whitfield-Gabrieli & Nieto-Castanon, 2012). We ran a type of functional connectivity analysis known as generalized psychophysiological interaction analysis (gPPI) that allows assessing how a given experimental condition modulates functional connectivity between regions of interest (ROIs). gPPI allows assessing task-dependent functional connectivity changes between pairs of ROIs by estimating the effect of the interaction between task effects and a ROI's time series on another ROI's time series using a multiple regression model. Here, we assessed changes in functional connectivity associated with the sacred value condition compared to the nonsacred value condition. We included all 19 ROIs defined as part of our target networks (default mode, frontoparietal, salience, and dorsal attention networks) by CONN (see the Montreal Neurological Institute coordinates for each ROI in Table S3 in the online supplemental materials). Group-level differences in functional connectivity between the sacred and nonsacred value condition were assessed using a multivariate pattern analysis omnibus test that measures the strength of all connections from each ROI. Results were corrected with an uncorrected region-to-region connection threshold of $p < .01$ and a multiple comparisons correction of $p < .05$ false discovery rate (FDR) at an ROI level.

Results

Effect of Sacred Values

Replicating Experiments 1 and 2, appealing to sacred values was associated with higher likelihood of sharing misinformation in the neuroimaging sample, $M_{diff} = 1.13$, 95% CI [0.81, 1.45], $t(35) = 7.19$, $p < .001$, $d = 1.07$, [0.57, 1.58]. Though the number of sacred values held by participants positively influenced the likelihood of sharing misinformation, this effect was not statistically significant in the smaller fMRI sample, $B = 0.14$, 95% CI [−0.14, 0.41], $t(34) = 0.97$, $p = .34$.

Brain Response to Sacred Values

The neuroimaging study revealed very strong brain activity in response to social media posts containing sacred values compared to nonsacred values among far-right partisans. Due to generalized activation across the brain at a standard threshold ($p < .001$ FWE cluster level) in this contrast, we had to use a more stringent statistical threshold ($p < .05$ FWE peak level) to differentiate several prominent clusters of activity in the left middle temporal gyrus, left dorsomedial prefrontal cortex, left precuneus, left middle and bilateral inferior frontal gyrus, and the left inferior parietal (thresholded at $T = 5.54$, $p < .05$ FWE; see Figure 3A and Table S4a in the online supplemental materials). A location analysis with *Neurosynth* (Yarkoni et al., 2011) indicated that the peak activations in each cluster were broadly associated with terms related to language and social inferences (see the Supplemental Results in the online supplemental materials), and the obtained pattern of activation considerably overlapped with a *Neurosynth* mask associated with *theory of mind* (see Figure 3D).

ROI Analyses

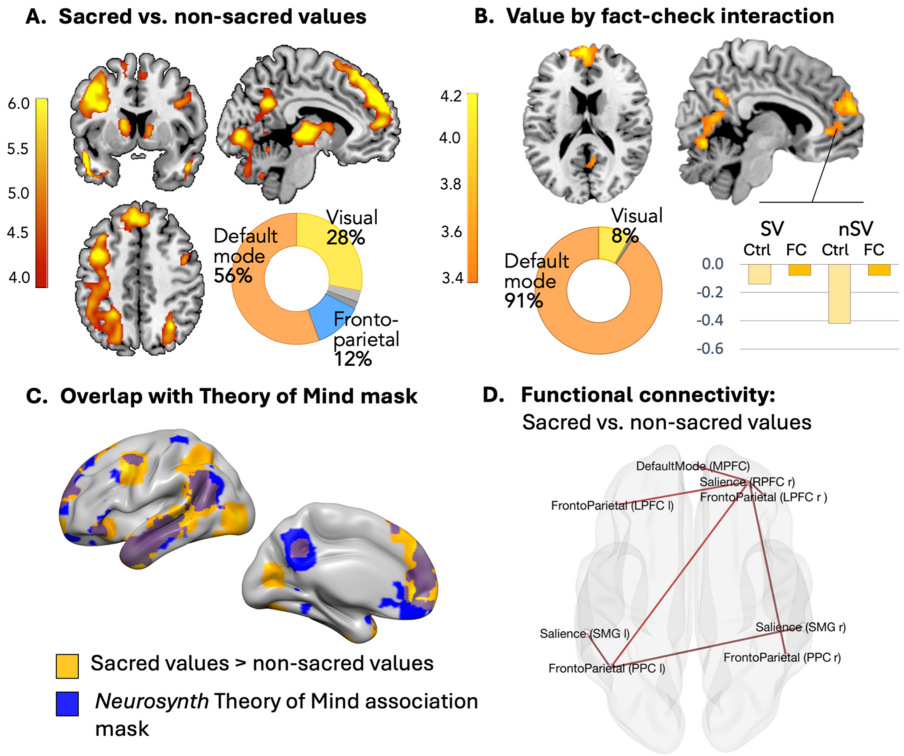
The involvement of the left inferior frontal gyrus, previously associated with sacred compared to nonsacred values (Pretus et al., 2018) was detected at a whole brain level (see above) and confirmed with a post hoc ROI analysis with anatomical masks for the left inferior frontal gyrus. The ROI analysis revealed higher neural activity in the left inferior frontal gyrus (pars orbitalis) in response to social media posts relevant to sacred versus nonsacred values ($k = 813$, $T = 6.37$, $p = .006$ FWE) with a peak activation ($x = -42$, $y = 26$, $z = -8$) similar to the one reported in Pretus et al. ($x = -46$, $y = 26$, $z = -2$). The anatomical mask for the orbitofrontal cortex did not yield any significant results in the sacred versus nonsacred value contrast ($p > .62$).

Functional Connectivity

The general psychophysiological interaction analysis revealed functional connectivity changes in the sacred versus nonsacred value condition among regions of interest (ROIs) in the evaluated neural networks (the default mode, frontoparietal, salience, and dorsal attention networks). Specifically, the right rostral prefrontal cortex (salience network) exhibited increased functional connectivity with frontoparietal nodes, including the bilateral prefrontal and

Figure 3

Neuroimaging Results



Note. (A) Social media posts with sacred values yielded much greater brain activation compared to posts with nonsacred values, especially in default mode regions, as revealed by the degree of overlap with the seven functional networks described in Yeo et al. (2011); (B) a similar brain activation pattern was found for fact-checked trials compared to control trials, an increase that was more prominent for nonsacred value trials, which had notably lower activation in control trials than sacred value trials; (C) comparison with a *Neurosynth* functional mask suggests that the activity detected in the sacred versus nonsacred value contrast largely overlaps with functional activations associated with the term *theory of mind*; and (D) a generalized psychophysiological interaction analysis revealed increased functional connectivity between salience and both frontoparietal and default mode network nodes in response to social media posts relevant to sacred compared to nonsacred values. SV = sacred values; nSV = nonsacred values; Ctr = control trials; FC = fact-checked trials. Brain regions in Figure D: LPFC = lateral prefrontal cortex; MPFC = medial prefrontal cortex; RPFC = rostral prefrontal cortex; SMG = supramarginal gyrus; PPC = posterior parietal cortex; l = left; r = right. See the online article for the color version of the figure.

posterior parietal cortex ($T > 3.22$, $p = .012$ FDR), and one default mode network node, the medial prefrontal cortex ($T = 3.12$, $p = .013$ FDR; see Figure 3D and Table S5 in the online supplemental materials). Moreover, the posterior parietal cortex (frontoparietal network) also exhibited increased functional connectivity with salience nodes, including the bilateral supramarginal gyrus and the right rostral prefrontal cortex ($T > 3.23$, $p = .016$ FDR). Thus, functional connectivity increased between frontoparietal and salience network nodes, as well as from salience to default mode network nodes in the sacred value compared to the nonsacred value condition.

Effects of Identity Fusion

Only five participants in the neuroimaging study were completely fused with the far-right party, a similar proportion to that found among far-right voters in the Spanish survey (14%). The reduced number of fused participants in the fMRI study precluded any comparisons between fused and nonfused individuals.

Fact-Checking in the Scanner

In the fMRI study, the effect size of the Twitter fact-check was nearly identical to the very small effect size found in Experiments 1 and 2 ($d = 0.07$) and also failed to reach statistical significance, $M_{diff} = -0.06$, 95% CI [−0.15, 0.04], $t(35) = -1.15$, $p = .26$, $d = -0.06$, 95% CI [−0.53, 0.41]. At a neural level, the interaction between sacred values and fact-checking revealed a network comprising the bilateral dorsomedial prefrontal cortex, the right posterior cingulate, the left parahippocampal gyrus, and the bilateral cerebellum cortex (thresholded at $T = 3.35$, $k = 376$, $p < .001$ FWEc). At baseline, these areas were more active during the sacred value

condition, but they became similarly active for sacred and nonsacred values when they were fact-checked (see Figure 3B; Table S4b in the online supplemental materials). A meta-analysis of locations using *Neurosynth* (Yarkoni et al., 2011) revealed associations between the obtained cluster peaks and terms related to *theory of mind* and *autobiographical recall* (see the Supplemental Results in the online supplemental materials). Thus, rather than eliciting activation in brain regions associated with cognitive control and behavioral adjustment, fact-checks resulted in increased activity in brain areas associated with *theory of mind*.

Neural activity in similar areas was higher in participants who started getting fact-checked in the second half of the fMRI session compared to those who were fact-checked from the beginning (see the Supplemental Results and Table S4c in the online supplemental materials; Figure 3C), suggesting a decreased response to fact-checking in participants who were habituated to fact-checks. This might also have important implications for the sustained efficacy of fact-checking.

Discussion

In a neuroimaging experiment, we aimed to characterize the neural activity and functional connectivity correlates of processing misinformation relevant to sacred compared to nonsacred values, as well as to identify changes in this neural activation pattern in response to fact-checking. We found a very strong neural response to misinformation relevant to sacred values compared to nonsacred values, especially in the dorsomedial prefrontal cortex, the bilateral inferior frontal cortex, and the precuneus. This pattern of activation overlapped with *theory of mind* neurofunctional correlates using *Neurosynth*. The sacred value condition also elicited higher functional connectivity between the salience, frontoparietal, and default mode network nodes compared to the nonsacred value condition.

The fact that messages relevant to sacred (vs. nonsacred) values elicited neural activation in lateral (e.g., orbital part of the inferior frontal gyrus) rather than medial aspects of the orbitofrontal cortex is aligned with previous studies that directly compare sacred and nonsacred values (Pretus et al., 2018). Left lateral aspects of the orbitofrontal cortex have been described as an important neural network for social norm compliance, which helps people respond appropriately to the threat of social punishment (Nestor et al., 2013; O'Doherty et al., 2001; Spitzer et al., 2007). Meanwhile, right medial aspects are thought to encode reward value during social decision-making (Nakamura et al., 2007; O'Doherty et al., 2001). Of note, we analyzed neural activity changes during stimulus presentation (4.5 s), rather than the response period (2.5 s). Therefore, our analyses may not have been suitable to capture neural activity changes in brain regions associated with later stages of decision-making, such as the orbitofrontal cortex (Cunningham & Zelazo, 2007; Rangel et al., 2008).

While we cannot perfectly infer the mental state of our participants from neural activity patterns, our results suggest that norm compliance and social punishment may play a role in how people evaluate in-group online messages that appeal to sacred values. Moreover, the dorsomedial prefrontal cortex has been involved in theory of mind functions, especially in belief rather than emotion attribution (Abu-Akel & Shamay-Tsoory, 2011; Corradi-Dell'Acqua et al., 2014). This function may be particularly helpful for someone who is attempting to gauge and generate norm-appropriate behavior. Since conforming to sacred (vs. nonsacred) values is critical to affirm one's group membership, group members may share messages relevant to sacred values in an attempt to conform with the group rather than as a function of the affective or personal value of these messages.

This idea is reinforced by the functional connectivity findings, which reveal greater functional connectivity between frontoparietal network nodes (lateral prefrontal and posterior prefrontal) and salience network nodes (supramarginal gyrus and rostral prefrontal), and from salience nodes (rostral prefrontal) to default mode network nodes (medial prefrontal cortex). This pattern of results suggests the recruitment of networks involved in cognitive control (frontoparietal) and social cognition (default mode) to appropriately respond to online messages relevant to sacred (vs. nonsacred) values. If norm compliance drives sharing in this high-stakes situation (i.e., affirming group membership), then group members should infer what the appropriate behavior is and comply with it to the best of their ability.

Finally, we identified an interaction effect between sacred versus nonsacred values and fact-checked versus control trials. Particularly, the dorsomedial prefrontal gyrus and the posterior cingulate were more active for sacred (vs. nonsacred) values in control trials, while these regions were similarly active for both sacred and nonsacred values in fact-checked trials. These findings suggest that fact-checks in the nonsacred value condition elicit a neural response similar to the sacred value condition. Specifically, brain regions previously involved in mentalizing (e.g., dorsomedial prefrontal cortex) were recruited in the face of messages relevant to nonsacred values if these included fact-checks. Aside from the inefficacy of fact-checks in far-right partisans, they appear to elicit cognitive processes (e.g., mentalizing) that may have not been triggered had the fact-check not been there. These neural patterns do not correspond to those predicted in response to fact-checking (i.e., cognitive control and behavioral adjustment; see, for instance, Cohen & Ranganath, 2007; Weissman et al., 2008). Thus, we found no evidence of a link between fact-checking and enhanced analytical thinking processes at a brain level in this sample.

General Discussion

We investigated the role of political devotion—in terms of sacred values and identity fusion—in the spread of political misinformation among far-right partisans in Spain and fused Republicans in the United States. Across three experiments, we found that appealing to sacred values in political messages increases the likelihood of sharing misinformation on social media, even after controlling for attitude strength, familiarity, and salience. The effect of sacred values was particularly strong among people who have a fused sense of identity with the far-right in Spain and with Donald Trump in the United States. We also found that a variety of fact-checks and accuracy nudges employed by social media companies and supported by widely cited papers (e.g., Pennycook et al., 2020) were ineffective at reducing the spread of misinformation. In our neuroimaging study, we found that messages relevant to sacred (vs. nonsacred) values elicited activation in norm compliance (i.e., orbital part of the left inferior frontal gyrus) and mentalizing networks (e.g., dorsomedial prefrontal cortex), as well as higher functional connectivity between attentional, executive, and social cognition networks. Fact-checks elicited neural activity in mentalizing networks (i.e., dorsomedial prefrontal cortex) instead of deliberation networks (Cohen &

Ranganath, 2007; Weissman et al., 2008). This suggests that social cognitive process might be more important that analytic thinking among groups at higher risk of spreading misinformation.

Our findings are consistent with the idea that the spread of misinformation is driven by politically devoted partisans who are resistant to fact-checks. Our results emphasize the role of partisan identity in misinformation sharing (Van Bavel et al., 2021), which stems from the idea that extreme partisans are often motivated to believe and share information that affirms their social identity, with little regard for accuracy (Kahan, 2013, 2017; Van Bavel & Pereira, 2018). Accordingly, while analytical thinking was weakly associated with reduced sharing of misinformation about identity-irrelevant topics (i.e., nonsacred values such as infrastructure), it made no difference for identity-relevant topics (i.e., sacred values such as immigration). These results suggest that extreme partisans have a strong drive to share identity-relevant information regardless of whether they are intuitive or analytical. Several other studies find that partisan motives override accuracy concerns when it comes to sharing news online (Osmundsen et al., 2021; Pereira et al., 2023).

The role of political devotion in partisan misinformation sharing suggests that motivated cognition may be at play during these decisions. Motivated cognition—and especially motivated reasoning—is a cognitive process through which people make decisions based on desirability rather than an unbiased analysis of the evidence (Kunda, 1990). As such, fused partisans may use motivated cognition to a greater extent than nonfused individuals, as they are more likely to share partisan misinformation and ignore available evidence (e.g., fact-checks). Following this logic, how much people's choices are biased towards outcomes that are desirable for the group may be a function of how much they identify with the group. Fused individuals may thus use motivated cognition in conscious or unconscious ways to affirm their group membership, especially when the values at stake are critical for the group (i.e., sacred).

To better understand the neurocognitive processes underlying the spread of misinformation we conducted a neuroimaging study among far-right voters in Spain. We found a very strong neural response to messages involving conservative sacred values (vs. nonsacred values) in far-right partisans. In particular, we found activity in norm compliance networks (i.e., orbital part of the left inferior frontal gyrus), which was very similar to previous studies that directly compare sacred and nonsacred values (Pretus et al., 2018). In the same contrast, we also found strong activation in brain regions typically associated with social cognition and theory of mind, such as the dorsomedial prefrontal cortex, the middle temporal gyrus, and the precuneus (Abu-Akel & Shamay-Tsoory, 2011; Corradi-Dell'Acqua et al., 2014). The medial prefrontal cortex was also more functionally connected to salience network nodes during the sacred (vs. nonsacred) value condition. These neural findings suggest that individuals' disposition to share online messages relevant to sacred values may be driven by norm compliance. A similar neural activation pattern emerged for sacred and nonsacred values that were being fact-checked. These results suggest that fact-checks may mobilize mentalizing networks (for instance, for assessing intentions) among extreme partisans rather than brain regions that support behavioral adjustment or deliberation.

The limited effectiveness of fact-checks among our right-wing samples adds nuance to previous literature on the effectiveness of fact-checks (Porter & Wood, 2021; Walter et al., 2020) and is in line with studies showing that those at the far-right of the political spectrum are more prone to believe and share misinformation (Garrett & Bond, 2021; Grinberg et al., 2019; Guess et al., 2019). However, because conservatives and liberals perform similarly well on the cognitive reflection test (Kahan, 2013), lack of analytical reasoning cannot explain these differences (indeed we found that individual differences on the cognitive reflection were weakly related or unrelated to sharing misinformation). As identity fusion predicts misinformation sharing and susceptibility to fact-checks, the capabilities to elicit identity fusion among extreme partisans could contribute to political asymmetries in the spread of misinformation. For devoted partisans, the spread of misinformation about sacred values might be seen as a form of justifiable information warfare.

One possible explanation for the limited effect of fact-checks across the three experiments is that the stimuli we designed were too plausible. Previous studies have found that the disposition for analytical thinking is unlikely to help reduce belief in misinformation that is plausibly true (Pennycook & Rand, 2019). Thus, it could be that fact-checks and nudges are helpful specifically when the presented misinformation is clearly implausible. We tested this idea in an exploratory analysis comparing the effect of fact-checks on high and low perceived accuracy items in the U.S. sample. As asking people to make accuracy judgements could influence subsequent likelihood of sharing (Pennycook et al., 2020), we obtained accuracy judgements for each social media post from a separate, but ideologically similar sample. Our results are consistent with the notion that fact-checks are most useful in reducing sharing of misinformation that is perceived to be less accurate on average. However, this effect was absent in Republicans fused with Trump, who still did not decrease sharing of fact-checked posts containing claims deemed to be implausible by other Republicans. Thus, there appears to be an important role for social identity in the efficacy of these interventions (see also Rathje et al., 2022). Future research should test this hypothesis with larger samples of individuals fused with Trump to confirm these preliminary results.

Given the potential shortcomings of traditional misinformation interventions among devoted actors, we believe there is an urgent need to develop and test identity-based interventions. The fact that sharing misinformation relevant to sacred (vs. nonsacred) values relies on neural activity in norm compliance and mentalizing networks emphasizes the role of the in-group in modulating these decisions. Since extreme partisans likely value identity-based motives over accuracy concerns, they should be more responsive to normative information—especially if it reflects the beliefs and values of their in-group. For instance, interventions based on in-group norms have been found to decrease willingness to fight and die in defense of sacred values in devoted actors (Hamid et al., 2019). When it comes to misinformation, interventions and platform features that appeal to identity and group-based norms might be more effective than identity-neutral alternatives for extreme partisans (Pretus et al., 2022).

Constraints on Generality

Because previously reported political asymmetries in information sharing (e.g., Guess et al., 2019), we decided to focus on partisans of right-wing parties since they represented a high-risk group for the spread of misinformation. Therefore, our results generalize to

these populations, and further research is needed to elucidate whether the identified processes apply to moderates and liberals in Spain, the United States, and elsewhere. Our data indicate that extreme political options elicit higher levels of identity fusion among partisans, suggesting that far-left partisans could possibly show similar patterns of misinformation sharing relative to sacred values. Similarly, we did not include a control group in the neuroimaging study, so it is unclear if our results are specific to far-right partisans. In fact, we would predict similar neural results in other groups exposed to misinformation relevant to their own sacred values. For instance, a recent nationally representative sample of Americans found that people on the extreme left score higher in dogmatism (albeit not as far as those on the far-right) and they may therefore be resistant to information that challenges their beliefs (see Harris & Van Bavel, 2021).

To understand the spread of misinformation, we focused on false statements in this research. We did not include true statements because we were primarily interested in testing the impact of using sacred moral values as a vehicle to spread misinformation, rather than trying to identify how people differentiate false from true information. Ultimately, misinformation in its most compelling form is only distinguishable from true information in that it does not correspond to reality. This design feature has two implications: first, participants in this series of studies were exposed to an unrealistic volume of misinformation, considering that misinformation in the real world coexists with true information (Guess et al., 2019), and second, our findings do not preclude that appealing to sacred values is associated with higher likelihood of sharing any type of information, regardless of whether it is true or false. Finally, our study was a controlled experiment with artificially designed social media posts. Future work will benefit from field studies looking at the relationship between identity fusion, sacred values, and fact-checks in real social media ecosystems.

Conclusion

Overall, our findings suggest that believing and sharing fake news is, at least in part, a socio-cognitive process serving partisan identity-affirming goals (Van Bavel & Pereira, 2018). Critically, individuals' willingness to spread misinformation relevant to sacred values appears resistant to analytical thinking and nudges aimed to improve accuracy, especially among extreme partisans. This is particularly important since extreme partisans—especially from far-right groups—appear to spread a disproportionate amount of misinformation. As such, there is an urgent need to design and test interventions that are more effective among extreme partisans. Strategies aimed to reduce misinformation should thus address the devotional aspect of misinformation sharing.

References

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, *49*(11), 2971–2984. https://doi.org/10.1016/j.neuropsychologia.2011.07.012

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Atran, S. (2016). The devoted actor: Unconditional commitment and intractable conflict across cultures. *Current Anthropology*, *57*(S13), S192–S203. https://doi.org/10.1086/685495

Atran, S., & Ginges, J. (2015). Devoted actors and the moral foundations of intractable intergroup conflict. In J. Decety & T. Wheatley (Eds.), *The moral brain: A multidisciplinary perspective* (pp. 69–85). Boston Review.

Au, H., Ho, K. W., & Chiu, K. W. (2021). The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implications. *Information Systems Frontiers*, *24*(4), 1331–1354. https://doi.org/10.1007/S10796-021-10133-9

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*(1), 1–16. https://doi.org/10.1006/obhd.1997.2690

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A signal detection approach to understanding the identification of fake news. *Perspectives on Psychological Science*, *17*(1), 78–98. https://doi.org/10.1177/1745691620986135

Borukhson, D., Lorenz-Spreen, P., & Ragni, M. (2022). When does an individual accept misinformation? An extended investigation through cognitive modeling. *Computational Brain and Behavior*, *5*(2), 244–260.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, *27*(2), 371–378. https://doi.org/10.1523/JNEUROSCI.4421-06.2007

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2014). Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *9*(8), 1175–1184. https://doi.org/10.1093/scan/nst097

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. https://doi.org/10.1016/j.tics.2006.12.005

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, *7*(23), Article eabf1234. https://doi.org/10.1126/sciadv.abf1234

Gawronski, B. (2021). Partisan bias in the identification of fake news. *Trends in Cognitive Sciences*, *25*(9), 723–724. https://doi.org/10.1016/j.tics.2021.05.001

Green, P., MacLeod, C. J., & Nakagawa, S. (2016). SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/mee3.2016.7.issue-4

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), Article eaau4586. https://doi.org/10.1126/sciadv.aau4586

Hamid, N., Pretus, C., Atran, S., Crockett, M. J., Ginges, J., Sheikh, H., Tobeña, A., Carmona, S., Gómez, A., Davis, R., & Vilarroya, O. (2019). Neuroimaging "will to fight" for sacred values: An empirical case study with supporters of an Al Qaeda associate. *Royal Society Open Science*, *6*(6), Article 181585. https://doi.org/10.1098/rsos.181585

Harris, E. A., & Van Bavel, J. J. (2021). Preregistered replication of "feeling superior is a bipartisan issue: Extremity (not direction) of political views predicts perceived belief superiority". *Psychological Science*, *32*(3), 451–458. https://doi.org/10.1177/0956797620968792

Haslam, S. A., Reicher, S. D., Selvanathan, H. P., Gaffney, A. M., Steffens, N. K., Packer, D., Van Bavel, J. J., Ntontis, E., Neville, F., Vestergren, S., Jurstakova, K., & Platow, M. J. (2023). Examining the role of Donald Trump and his supporters in the 2021 assault on the US Capitol: A dual-

agency model of identity leadership and engaged followership. *The Leadership Quarterly*, *34*(2), Article 101622. https://doi.org/10.1016/j.leaqua.2022.101622

Hogg, M. A., CooperShaw, L., & Holzworth, D. W. (1993). Group prototypically and depersonalized attraction in small interactive groups. *Personality and Social Psychology Bulletin*, *19*(4), 452–465. https://doi.org/10.1177/0146167293194010

Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, *16*(1), 7–30. https://doi.org/10.1111/j.1468-2885.2006.00003.x

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371–388. https://doi.org/10.1177/0002764219869406

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision making*, *8*(4), 407–424. https://doi.org/10.2139/SSRN.2182588

Kahan, D. M. (2017, May 24). *Misconceptions, misinformation, and the logic of identity-protective cognition*. Cultural Cognition Project Working Paper Series No. 164, Yale Law School, Public Law Research Paper No. 605, Yale Law & Economics Research Paper No. 575. https://doi.org/10.2139/ssrn.2973067

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kunst, J. R., Dovidio, J. F., & Thomsen, L. (2019). Fusion with political leaders predicts willingness to persecute immigrants and political opponents. *Nature Human Behaviour*, *3*(11), 1180–1189. https://doi.org/10.1038/s41562-019-0708-1

Lee, F. L. F. (2016). Impact of social media on opinion polarization in varying times. *Communication and the Public*, *1*(1), 56–71. https://doi.org/10.1177/2057047315617763

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Mosleh, M., Pennycook, G., Rand, D. G., & Jankowski, J. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLoS ONE*, *15*(2), Article e0228882. https://doi.org/10.1371/journal.pone.0228882

Nakamura, M., Nestor, P. G., Levitt, J. J., Cohen, A. S., Kawashima, T., Shenton, M. E., & McCarley, R. W. (2007). Orbitofrontal volume deficit in schizophrenia and thought disorder. *Brain*, *131*(1), 180–95. https://doi.org/10.1093/brain/awm265

Nestor, P. G., Nakamura, M., Niznikiewicz, M., Thompson, E., Levitt, J. J., Choate, V., Shenton, M. E., & McCarley, R. W. (2013). In search of the functional neuroanatomy of sociality: MRI subdivisions of orbital frontal cortex and social cognition. *Social Cognitive and Affective Neuroscience*, *8*(4), 460–467. https://doi.org/10.1093/scan/nss018

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Nyilasy, G. (2019). Fake news: When the dark side of persuasion takes over. *International Journal of Advertising*, *38*(2), 336–342. https://doi.org/10.1080/02650487.2019.1586210

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, *4*(1), 95–102. https://doi.org/10.1038/82959

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, *115*(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2021). Lack of partisan bias in the identification of fake (versus real) news. *Trends in Cognitive Sciences*, *25*(9), 725–726. https://doi.org/10.1016/j.tics.2021.06.003

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-30073-5

Pereira, A., Harris, E., & Van Bavel, J. J. (2023). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes and Intergroup Relations*, *26*(1), 24–47. https://doi.org/10.1177/13684302211030004

Piazza, J. A. (2022, August 3). *Stop the steal!: Allegations of election cheating and support for political violence among US conservatives*. SSRN. https://doi.org/10.2139/ssrn.4179900

Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, *118*(37), Article e2104235118. https://doi.org/10.1073/pnas.2104235118

Pretus, C., Hamid, N., Sheikh, H., Ginges, J., Tobeña, A., Davis, R., Vilarroya, O., & Atran, S. (2018). Neural and behavioral correlates of sacred values and vulnerability to violent extremism. *Frontiers in Psychology*, *9*, Article 2462. https://doi.org/10.3389/fpsyg.2018.02462

Pretus, C., Hamid, N., Sheikh, H., Gómez, Á, Ginges, J., Tobeña, A., Davis, R., Vilarroya, O., & Atran, S. (2019). Ventromedial and dorsolateral prefrontal interactions underlie will to fight and die for a cause. *Social Cognitive and Affective Neuroscience*, *14*(6), 569–577. https://doi.org/10.1093/scan/nsz034

Pretus, C., Javeed, A., Hughes, D. R., Hackenburg, K., Tsakiris, M., Vilarroya, O., & Van Bavel, J. J. (2022, July 19). *The misleading count: An identity-based intervention to counter partisan misinformation sharing*. https://doi.org/10.31234/osf.io/7j26y

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Rathje, S., Roozenbeek, J., Steenbuch Traberg, C., Van Bavel, J. J., & van der Linden, S. (2022). Letter to the editors of psychological science: Meta-analysis reveals that accuracy nudges have little to no effect for U.S. conservatives: Regarding Pennycook et al. (2020). *Psychological Science*. Advance online publication. https://doi.org/10.25384/SAGE.12594110.V2

Rathje, S., Roozenbeek, J., Van Bavel, J., van der Linden, S. (2023) Accuracy and social motivations shape judgements of (mis)Information. *Nature Human Behaviour*. Advance online publication. https://doi.org/10.1038/s41562-023-01540-w

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), Article e2024292118. https://doi.org/10.1073/pnas.2024292118

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reinero, D. A., Wills, J. A., Brady, W. J., Mende-Siedlecki, P., Crawford, J. T., & Van Bavel, J. J. (2020). Is the political slant of psychology research related to scientific replicability? *Perspectives on Psychological Science*, *15*(6), 1310–1328. https://doi.org/10.1177/1745691620924463

Roozenbeek, J., Freeman, A. L. J., & van der Linden, S. (2021). How accurate are accuracy-nudge interventions? A preregistered direct replication of Pennycook et al. (2020). *Psychological Science*, *32*(7), 1169–1178. https://doi.org/10.1177/09567976211024535

Shearer, E. (2021). *More than eight-in-ten Americans get news from digital devices*. Pew Research Center.

Sheikh, H., Ginges, J., & Atran, S. (2013). Sacred values in the Israeli–Palestinian conflict: Resistance to social influence, temporal discounting, and exit strategies. *Annals of the New York Academy of Sciences*, *1299*(1), 11–24. https://doi.org/10.1111/nyas.12275

Singmann, H., Bolker, B., & Westfall, J. (2015). *Afex: Analysis of factorial experiments* (R Package Version 0.15-2). http://CRAN.R-project.org/package=afex

Spitzer, M., Fischbacher, U., Hermberger, B., Groen, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*(1), 185–96. https://doi.org/10.1016/j.neuron.2007.09.011

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, *34*(3), 150–160. https://doi.org/10.1177/0266382117722446

Sternisko, A., Cichocka, A., Cislak, A., & Bavel, J. J. V. (2023). National narcissism predicts the belief in and the dissemination of conspiracy theories during the COVID-19 pandemic: Evidence from 56 countries. *Personality and Social Psychology Bulletin*, *49*(1), 48–65. https://doi.org/10.1177/01461672211054947

Swann, W. B., Gómez, A., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, *96*(5), 995–1011. https://doi.org/10.1037/a0013668

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis) Information age: A model of news belief and sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.12077

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

Weissman, D. H., Perkins, A. S., & Woldorff, M. G. (2008). Cognitive control in social situations: A role for the dorsolateral prefrontal cortex. *NeuroImage*, *40*(2), 955–962. https://doi.org/10.1016/j.neuroimage.2007.12.021

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Wickham, H. (2016). *ggplot2: Elegant rraphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al.(2009). *Perspectives on Psychological Science*, *4*(3), 294–298. https://doi.org/10.1111/j.1745-6924.2009.01127.x

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Received October 6, 2022
Revision received March 31, 2023
Accepted April 17, 2023 ■