OXFORD

Cerebral Cortex, 2024, **34**, bhae368
https://doi.org/10.1093/cercor/bhae368
Advance access publication date 12 September 2024
Original Article

Partisans process policy-based and identity-based messages using dissociable neural systems

Nir Jacoby[1,2,*], Marika Landau-Wells[3], Jacob Pearl[4], Alexandra Paul[4], Emily B. Falk[4,5,6,7], Emile G. Bruneau[4], Kevin N. Ochsner[2]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Moore Hall, 3 Maynard St, Hanover, NH 03755, USA
[2]Department of Psychology, Columbia University, 1190 Amsterdam Ave, New York, NY 10027, USA
[3]Travers Department of Political Science, University of California-Berkeley, 210 Barrows Hall #1950, Berkeley, CA 94720, USA
[4]Annenberg School for Communication, University of Pennsylvania, 3620 Walnut St, Philadelphia, PA 19104, USA
[5]Wharton School, University of Pennsylvania, 3733 Spruce St, Philadelphia, PA 19104, USA
[6]Department of Psychology, University of Pennsylvania, 3720 Walnut St, Philadelphia, PA 19104, USA
[7]Annenberg Public Policy Center, University of Pennsylvania, 202 S 36th St, Philadelphia, PA 19104, USA

*Corresponding authors: Nir Jacoby, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755. Email: nir.jacoby@dartmouth.edu; Kevin N. Ochsner, Department of Psychology, Columbia University, New York, NY 10027. Email: ko2132@columbia.edu

Political partisanship is often conceived as a lens through which people view politics. Behavioral research has distinguished two types of "partisan lenses"—*policy-based* and *identity-based*—that may influence peoples' perception of political events. Little is known, however, about the mechanisms through which partisan discourse appealing to policy beliefs or targeting partisan identities operate within individuals. We addressed this question by collecting neuroimaging data while participants watched videos of speakers expressing partisan views. A "partisan lens effect" was identified as the difference in neural synchrony between each participant's brain response and that of their partisan ingroup vs. outgroup. When processing policy-based messaging, a partisan lens effect was observed in socio-political reasoning and affective responding brain regions. When processing negative identity-based attacks, a partisan lens effect was observed in mentalizing and affective responding brain regions. These data suggest that the processing of political discourse that appeals to different forms of partisanship is supported by related but distinguishable neural—and therefore psychological—mechanisms, which may have implications for how we characterize partisanship and ameliorate its deleterious impacts.

Key words: neural synchrony; neuroimaging; political partisanship; social neuroscience.

Introduction

The past few decades in American politics have seen a notable rise in partisanship among citizens (Abramowitz 2022; Lee et al. 2022), which has been described as one of the most significant threats to democracy (Lupu 2015; Mason 2018; Levitsky and Ziblatt 2019; Finkel et al. 2020). Although individuals do not self-identify as members of one of the two major political parties as much as they have in the past, they exhibit greater party loyalty in their attitudes and behaviors (Sniderman and Stiglitz 2012; Abramowitz and Webster 2016; Finkel et al. 2020).

Yet not all forms of partisanship are inherently negative. "Policy-based partisanship" where alignment with a party's issue positions is the basis of an individual's party attachment (Downs 1957; Shively 1979), can be a constructive element of a vibrant democracy by fostering reasoned discussion (Muirhead and Rosenblum 2020). "Identity-based partisanship" (We note that there are several alternative labeling schemes for distinguishing between partisanship as the product of issue- or ideological agreement and partisanship that is the product of a shared social identity (e.g. instrumental versus expressive; Huddy et al. 2015; issue versus identity; Highton and Kam 2011)), where one's emotional connections to the party as a social identity is the basis of party attachment, is often less constructive (Mason 2018; Iyengar et al. 2019; Finkel et al. 2020).

One major reason for concern about the recent increase in the intensity of partisan attachment is that partisanship can distort how individuals process information (Campbell et al. 1960; Kunda 1990; Bartels 2002; Taber and Lodge 2006; Bisgaard 2015; Theodoridis 2017; Ditto et al. 2019). For example, when exposed to politically relevant media, a partisan will process that content in a way that is more similar to those who share their partisan convictions (partisan ingroup) than those who do not (partisan outgroup) (Coe et al. 2008; Levendusky 2013; Benedictis-Kessner et al. 2019). While this "partisan lens" effect is thought to extend to all manner of politically relevant perceptions, cognitions, emotions, and actions (Campbell et al. 1960; Mason 2018), several fundamental questions remain unanswered, including: How does the "partisan lens" effect work? Does partisanship influence information processing via a single underlying mechanism? Or is partisanship's influence conditional on the nature of the information to which one is exposed (i.e. a policy-based or an identity-based discourse)?

Answers to these questions have implications for how we interpret partisanship's most pernicious societal effects. If partisan processing is conditional on the nature of the information itself, then it would suggest that the information environment should be considered as a target for interventions seeking to attenuate the negative consequences of increasing partisan attachment (Mutz 2015). Answers to these questions could also shed light on

Received: April 1, 2024. **Revised:** August 12, 2024. **Accepted:** August 27, 2024

the nature of partisanship itself and, in particular, how policy-based and identity-based partisanship operate and interact *within* individuals.

Debates about the nature of partisanship are long-standing. On one hand, policy-based partisanship is thought to operate through processes of reasoning and evaluation. That is, a person's partisan affiliation results from assessing the fit between their beliefs, interests, and positions on social and political issues with various party platforms. Policy-based partisanship has an essential role in deliberative processes for individuals and at the societal level (Muirhead and Rosenblum 2020). On the other hand, identity-based partisanship functions as a social identity whose underlying affective mechanisms (Campbell et al. 1960; Tajfel 1979; Green et al. 2004; Greene 2004) can influence identity-based judgments, leading us to misperceive partisan outgroup members and their motivations (Moore-Berg et al. 2020; Lees and Cikara 2021).

Building on this analysis, we used neuroimaging to examine how partisans process political discourse appealing to each form of partisanship. Specifically, we asked whether and how the psychological and neural mechanisms underlying the processing of policy-based discourse and identity-based discourse relate to one another. Would we find evidence for a singular "partisan lens" in processing both types of discourse, or would partisanship shape information processing through multiple neural pathways?

Recent neuroimaging research has shown that information *about* partisan elites, i.e. Democratic and Republican politicians, is processed differently when it does not conform to party-based expectations (Haas et al. 2017; Haas et al. 2021). In addition, information delivered *by* partisan elites is also processed differently, depending on the speaker's partisanship (van Baar et al. 2021). Yet, we still know little about how partisan discourse is processed. Recent work has focused on the role of participants' *ideological* self-identification (i.e. as liberals or conservatives), not their *partisan* self-identification, in the processing of political messages (Leong et al. 2020; van Baar et al. 2021). Although ideological affinity is increasingly aligned with political party self-identification in the United States, the strength of this relationship has varied over time (Abramowitz 2022) and varies also by age, gender, and race (Twenge et al. 2016; Gillion et al. 2020; Jefferson 2020). Our study thus extends previous research in two meaningful ways: first, by measuring participants' partisanship directly rather than substituting ideology for partisanship (Kinder and Kalmoe 2017), and second, by directly manipulating and investigating how the type of partisan discourse affects how it is processed.

Based on prior literature, we identified candidate psychological processes—and associated neural systems—that could underlie the "partisan lenses" that guide people who share partisanship to process politically relevant information in a similar way.

First, we considered processes involved in reasoning, evaluating, and reflecting on one's beliefs, attitudes, interests, and positions on political issues, including deliberating about how they match different party platforms (Muirhead and Rosenblum 2020). Consistent with the involvement of these processes, the handful of neuroimaging studies examining political and related forms of reasoning (Westen et al. 2006; Bruneau and Saxe 2010; Kaplan et al. 2016; Prado et al. 2020) have observed activation of medial frontal, temporal–parietal, and precuneus regions, all of which have been implicated in judgments about mental states (Saxe and Kanwisher 2003; Amodio and Frith 2006; Mar 2011; Zaki and Ochsner 2012). This pattern reflects the fact that reasoning about political and social issues requires the consideration of one's own—as well as other people's—beliefs, attitudes, and feelings about key issues. It is currently unknown whether and how engagement of these systems is related to how people respond to either policy-based or identity-based partisan discourse.

A second set of candidate psychological processes involve those supporting one's attitudes toward, personal valuation of, and self-identification with the party as a social group (Campbell et al. 1960; Tajfel 1979; Greene 2004). Relevant to these processes, imaging studies have identified a set of regions involved in appraising the affective content of stimuli and triggering appropriate responses to them, including the ventral striatum—thought to play key roles in reward learning and reward expectancy (Pessiglione et al. 2006; Ruff and Fehr 2014), the amygdala—thought to be important for detecting goal-relevant stimuli, in general, with a special role in detecting potential threats (Davis and Whalen 2001; Phelps and LeDoux 2005; Pessoa and Adolphs 2010), and the anterior insula—thought to be important for integrating body state information with negative affective states (Craig 2009; Deen et al. 2010; Zaki et al. 2012; Chang et al. 2013). Although exposure to politically relevant information has been shown in prior studies to engage regions associated with affective responding (Westen et al. 2006; Kaplan et al. 2016; Haas et al. 2017; van Baar et al. 2021), only one study measured partisanship (Westen et al. 2006) and its stimuli were unrelated to political discourse. It therefore remains unclear whether and how this engagement is related to responses to either policy-based or identity-based partisan discourse.

With these considerations in mind, we sought to test two hypotheses. The first was that both policy-based and identity-based partisan discourse would activate the same "partisan lens," relying on a *common core* of psychological and neural processes (Van Bavel and Pereira 2018). On this view, both types of partisan discourse should engage *the same set(s) of brain regions*—particularly those implicated in socio-political reasoning/mentalizing and/or affective responding. This account of partisan information processing is consistent with a view that all forms of partisanship ultimately share a unified psychological underpinning. The second, *multiple paths* hypothesis, was that responding to different forms of partisan discourse depends on *different brain regions*. This account is consistent with a view that different forms of partisanship may be rooted in different psychological processes that are activated by different types of discourse in a context-dependent way.

To test these two "partisan lens" hypotheses, we collected whole-brain functional magnetic resonance imaging (fMRI) data while self-identified partisan participants watched video clips that discussed either policy issues (*Policy-Based Discourse* condition), denigrated outgroup partisans (*Identity-Based Discourse* condition), or discussed a nonpolitical topic (*Control* condition). To identify a neural analog of the "partisan lens," we used inter-subject correlation (ISC) analysis (Hasson et al. 2004) to calculate the extent to which participants' brain responses were more similar to their partisan ingroup than to their outgroup. To provide a strong test that each type of partisanship is supported by regions involved in socio-political reasoning, mentalizing, or affective responding, we restricted primary ISC analyses to regions of a priori interest for each type of process. For socio-political reasoning, no studies have directly tested whether it is driven by mentalizing processes per se, as opposed to related but distinct processes (Fedorenko and Kanwisher 2009; Scholz et al. 2009). As such, separate tasks from the literature were selected to localize regions directly involved in socio-political reasoning (using control stimuli from Bruneau and Saxe 2010) and mentalizing (Dodell-Feder et al. 2011). For regions related to affective

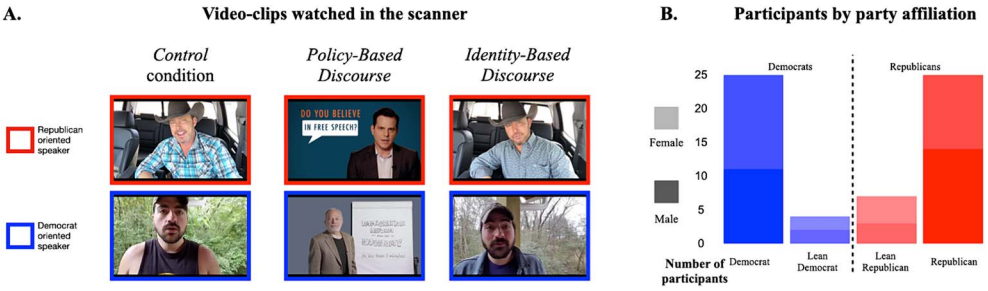

Fig. 1. Experimental procedure and behavioral summary. A) Participants watched six video clips in the scanner. The clips were divided into three conditions—a *Control* condition and two experimental conditions, corresponding to the type of partisan appeal they featured—*Policy-Based Discourse* and *Identity-Based Discourse*. B) Breakdown of participants in the final sample by their political affiliation. "Leaners" are participants who first identified as independent, but when asked, reported which party they felt closer to. For all analyses, leaners were categorized with the party they felt closer to.

responding, key regions (ventral striatum, amygdala, anterior insula) are anatomically circumscribed and were defined structurally. Having identified key regions thusly, we then used ISC to determine whether they were involved in partisan information processing when individuals were exposed to political messaging with policy-based or identity-based partisan appeals and critically, whether results supported the common core or multiple paths hypotheses.

Materials and methods

Participants

Seventy-one participants from across the American political spectrum participated in the experiment. Informed written consent was obtained from all participants. Participants were paid monetarily, in accordance with University of Pennsylvania Institutional Review Board protocol. All participants were scanned between July 2017 and August 2018. Data from two participants were lost altogether; data from another participant were corrupted due to technical errors. Each functional run in which the average framewise displacement (FD, see preprocessing) exceeded 0.3 was discarded as high-motion. Participants who had two or more runs of the main paradigm missing were excluded from all analyses. Such runs were either never completed (3 participants) or excluded for motion (4 participants). The final sample consisted of 61 participants (31 females, 30 males, 0 nonbinary, mean age: 23.6, std: 5.7, range: 18 to 44).

Participants' partisanship was measured in a two-step procedure as is standard in American public opinion research (e.g. the American National Election Studies, ANES). Participants were first asked if they considered themselves a Democrat, Republican, or Independent. Those who identified as Independent were then asked if they felt closer to one of the two major parties. We grouped "leaners" with their respective party in the main analyses. The final sample included 29 Democrats (of whom 4 were "leaners") and 32 Republicans (of whom 7 were "leaners"). The share of "leaners" in our sample (18%) is consistent with their share in nationally representative survey samples from the same period (American National Election Studies [ANES] et al. 2017) (Fig. 1B).

Stimuli and experimental procedure

Stimuli

The stimuli for the main experimental paradigm consisted of 6 short video clips (2:22 to 4:22 min) that varied in their political content, relevance, and orientation. All short clips were taken from the internet and are publicly accessible. This was done to create an ecologically valid experimental environment where participants would see the type of political material that they might encounter when browsing the web. There were three task conditions (*Policy-Based Discourse*, *Identity-Based Discourse*, and *Control*) with two video clips assigned to each one. For each condition, one clip featured a speaker aligned with a Republican partisan orientation and one featured a speaker aligned with a Democrat partisan orientation. In the *Control* condition, two internet vloggers expressed their opinions in a humorous way on arguably nonpartisan topics. The *Policy-Based Discourse* condition featured "myth-busters"-type clips focused on a single topic (e.g. immigration and the economy; what is "progressive"). The *Identity-Based Discourse* condition featured the same vloggers from the *Control* condition expressing their opinions about their respective partisan outgroup in an inflammatory and derogatory manner. A full list of the stimuli, conditions, and topics is shown in Table S1. The videos and their assignment for conditions were all validated prior to the experiment by ratings collected on Amazon's Mechanical Turk platform. Videos were chosen to maximize between group differences in ratings for each item (and minimize such differences for the control condition). The full details of validation procedure can be found in Supplementary Information.

Experimental procedure

Prior to scanning, participants filled out an online questionnaire to assess their political affiliation, attitudes, and activism. The questions regarding political affiliation were used to determine participants' partisan group assignment as described above.

During scanning, participants first underwent an anatomical scan, followed by a calibration of the sound system. Participants passively watched (no explicit task) all the stimuli in a fixed *condition* order (Fig. 1A). *Control* clips were always first, followed by *Policy-Based Discourse* clips and then *Identity-Based Discourse* clips. This order was decided to ensure that participants did not have preconceptions of the partisan orientation of speakers in the *Control* condition (i.e. to avoid contamination from the later partisan arguments made by the same speakers in the *Identity-Based Discourse* condition). The order of clips *within* conditions was altered between participants whereby participants watched either the Democratic- or Republican-oriented speaker first in all of the conditions.

After the main experiment clips, participants watched two more video clips—the first was an edited compilation of excerpts from a presidential debate between Donald Trump and Hilary Clinton. The second was "Partly Cloudy," a short-animated movie (Pixar Studios). The two last video clips were not analyzed for this paper. After watching all the clips, participants completed two functional localizer tasks (see below) used to functionally identify

regions of interest (ROIs) for analysis. Then, if time permitted, a resting-state scan (not analyzed) concluded the scan session.

An additional online questionnaire was filled out by participants after the neuroimaging session, which included additional measurements of political affiliation, along with responses to the (political) clips they watched in the scanner. For each of the four clips in the main experimental conditions, participants were asked about their familiarity with the speaker and specific video. In addition, participants were asked to rate how reasonable they found the argument in the video to be on a sliding scale (0 [very unreasonable] to 100 [very reasonable]). They were also asked how much they felt the following emotions while watching the video: *Irritated; Annoyed; Angry; Satisfied; Validated*. Each emotion was rated on a 5-point scale from (1) *not at all* to (5) *a lot*.

Functional localizers

Socio-political reasoning localizer

Participants completed a task first introduced in Bruneau and Saxe (2010). The task we used was the control task from the experiment that was orthogonal to the specific conflict studied in the paper. Using these stimuli, the task asked participants to read short statements and to judge how reasonable they were. Statements were grouped into two conditions—*Socio-political* and *Generic* (The condition names in the original paper were *Emotional (control)* and *Nonemotional (control)*)—and presented in two functional runs featuring 5 trials per condition (for a total of 10 trials per run). Each trial began with 10 s to read each statement immediately followed by 4 s to respond on a 1 (very unreasonable) to 4 (very reasonable) scale. Trials were presented in a fixed block design with 12 s of fixation time between blocks (and at the beginning and end of the run) for a total of 272 s per run. The task was used to independently localize, at the participant level, brain regions using the *Socio-political* > *Generic* contrast.

Mentalizing localizer

Participants completed one functional run of a standard false-belief localizer from Rebecca Saxe's lab (Dodell-Feder et al. 2011). In this task, participants read short vignettes about outdated representations held either in a character's mind (*Belief* condition) or on an inanimate object (*Photo* condition) and answered a true/false question about them. Participants completed 5 trials of each type; each trial consisted of 10 s to read the story immediately followed by 4 s to answer the question. Those were presented in a fixed block design with 12 s of fixation time between blocks (and at the beginning and end of the run) for a total of 272 s. The task was used to independently localize, at the participant level, brain regions using the *Belief* > *Photo* contrast.

Behavioral manipulation check

Participants' responses to the political clips from the postscan questionnaire were analyzed as a manipulation check. First, we scaled the reasonableness measurement (originally 0 to 100) to a 5-point scale to match the emotion measurements. We then averaged the three negative emotion ratings (irritated, annoyed, angry) into a single negative emotion score. Similarly, we averaged the two positive emotions (satisfied, validated). The balanced design ensured that in each condition, for each participant, one clip represented a partisan ingroup and a partisan outgroup. For each of the measurements (reasonableness, negative emotions, positive emotions), we defined a multilevel model and tested the effect of speaker's partisan alignment (ingroup–outgroup) on the measurement. All statistical analyses were done in R (R Core Team 2013) using lme4 and lmerTest packages (Bates et al. 2015; Kuznetsova et al. 2017).

fMRI acquisition and analysis

Neuroimaging data were acquired using a Siemens Prisma 3T scanner with a 64-channel head/neck array at the University of Pennsylvania. Participants were fitted with an MR-compatible headset for audio. All visuals of the experiment were projected to a screen behind the scanner and viewed via a mirror mounted to the head coil. For each participant, a high-resolution anatomical volume was acquired using a T1-weighted MPRAGE sequence in 176 sagittal slices of 1 mm isotropic voxels with a 256 mm field of view (FoV). All functional runs were acquired using a T2* weighted sequence with multiband factor of 3 simultaneous multi slices and partial Fourier factor of 0.875; repetition time of 1 s; echo time of 30 ms; flip angle of 60°; voxel size of 3.0303 mm on the acquisition plane and 3 mm slice thickness; matrix size of 66, resulting in FoV of 200 mm; each volume consisted of 51 slices providing full brain coverage of blood oxygen level dependent (BOLD). No field maps were collected.

All DICOM images were first converted to 4d NIFTI file formats and arranged to follow the Brain Imaging Data Structure (BIDS) (Gorgolewski et al. 2016). The data were then preprocessed using *fMRIPrep 20.1.0rc3* (Esteban et al. 2018a; Esteban et al. 2018b; Esteban et al. 2020). The preprocessing included anatomical and functional data preprocessing (See *fMRIPrep* boilerplate in Supplementary Information for full description). Anatomical data were segmented, parceled, and normalized to MNI template (MNI152NLin2009cAsym). Functional data were motion-corrected, registered to anatomical, and resliced to both the native space and the normalized MNI space.

ROIs definition

Socio-political reasoning and mentalizing regions

All functional localizer runs were processed in the standard space (MNI152NLin2009cAsym). The (*fMRIPrep*) preprocessed data were smoothed using a 5 mm (FWHM) smoothing Kernel using SPM12 prior to modeling. Each of the localizer tasks was analyzed in a single (first-level) generalized linear model (GLM) using SPM12 (Penny et al. 2011). All localizers' GLMs included two condition regressors for the localizer conditions (i.e. *Socio-political, Generic* in the socio-political reasoning localizer; *Belief, Photo* in the mentalizing localizer). Trials were modeled using a boxcar function convolved with a canonical hemodynamic response function. In addition to conditions of interest, nuisance regressors were included to account for run means along with the following regressors as calculated by *fMRIPrep*: 6 motion parameters, framewise displacement (FD) and delta regressors for time points where FD was larger than 0.8 mm. All models included a cosine-based high-pass filter (cut-off frequency of 1/128 Hz).

The results of the first-level model were then used to define functional ROIs (fROIs) at the individual participant level. The procedure used for picking fROIs followed the outline described in Blank and Fedorenko (2020) and was identical for both localizer tasks. The statistical maps from the critical contrast of the localizer task (i.e. *Socio-political* > *Generic* in the socio-political reasoning localizer; *Belief* > *Photo* in the mentalizing localizer) were used as the input for the process. The maps were first thresholded at $t > 0$ (effect in the correct direction) and then masked by a "search space" to constrain the anatomical region (see Fig. 2B for the search-spaces). The masked maps were then masked again with the participant's gray matter mask (generated by *fMRIPrep*). To ensure that the same number of voxels was picked for all

participants, our algorithm looked for a fixed number of voxels, equivalent to 5% of the search-space size. The script then sorted all the remaining voxels by their *t*-value and picked the 5% with the highest values (most consistently active voxels in the region). In cases where not enough voxels remained in a masked map, no voxels were picked and the procedure is considered failed for the participant (in the ROI). Otherwise, the picked voxels constitute the fROI or further analysis.

For both localizers, the search spaces were taken from a publicly available (https://saxelab.mit.edu/use-our-theory-mind-group-maps) large group analysis of 462 participants who performed the false-belief task (Dufour et al. 2013). The search spaces included the following regions: bilateral temporoparietal junction (L/RTPJ), dorsal ($z > 20$), middle ($20 > z > 0$), and ventral ($z < 0$) portions of the medial prefrontal cortex (D/M/VMPFC), precuneus (PC), and right superior temporal sulcus (RSTS).

The socio-political reasoning localizer (Bruneau and Saxe 2010) was used to define participant-level *socio-political reasoning regions* in the 7 search spaces. The false-belief localizer (Dodell-Feder et al. 2011) was used to define participant-level *mentalizing regions* (fROIs) in the 7 search spaces.

Affective-responding regions

The amygdala and ventral striatum were anatomically defined from participants' T1 image. Voxels were identified by *FreeSurfer* segmentation and parcellation during the preprocessing stage. For the ventral striatum, we used the *accumbens-area* parcels (*FreeSurfer* parcels 26,58). As a result of this process, these regions were personally tailored to each participant. The anterior insula regions, which are not anatomically defined as other affective-responding regions, were taken from a connectivity-based clustering analysis (Deen et al. 2010). We used the clusters representing ventral anterior insula from that paper—the portion most linked to affective processing. The anterior insula clusters were the only regions where we used group regions (i.e. the same voxels were used as the ROI for all participants). Each of the regions above was defined independently in both hemispheres, resulting in 6 regions.

We note that even though the stimuli in the socio-political reasoning localizer vary on the level of affective evocativeness between conditions, it was not used to functionally localize affective-responding regions. In both the reference (Bruneau and Saxe 2010) and in our study, group-level analyses on the task's main contrast do not result in activation of affective regions (see Supplementary Information).

Timecourse processing

All short clip runs were modeled using SPM12. Each functional run (corresponding to a video clip) was modeled separately using a GLM to account for known nuisance regressors. GLMs included a cosine-based high-pass filter (cut-off frequency of 1/128 Hz) and run mean, along with the following regressors as calculated by *fMRIPrep*: 6 motion parameters, FD, 6 first components of aCompCor, and delta regressors for nonsteady-state volumes at the beginning of the run and for every time point where FD was larger than 0.8 mm. The residuals of the model were saved as the new time series for further analysis. Each run was modeled two times for different analysis pipelines: (i) Unsmoothed preprocessed time series in native space were modeled to be used with anatomically defined ROIs (see Affective-responding Regions above). (ii) Unsmoothed preprocessed time series in standard space were modeled to be used with group and functionally defined ROIs.

After modeling all the clips, residualized timecourses were extracted from all voxels in each ROI. Regardless of the type of ROI, similar processing was used on the extracted timecourses. The timecourse from each voxel was first filtered using an ideal band-pass filter between 1/125 and 1/10 Hz (implemented with Matlab's *idealfilter* function), eliminating high frequencies that are unlikely to be from neural sources (Cordes et al. 2001) along with the already filtered low drift frequencies. The filtered timeseries were then averaged between all voxels in the ROI. We then kept only the volumes that reflected the brain response to stimuli (from onset of stimuli + 6 s to offset + 6 s). This was done only after applying the filter because the filtering process creates transient effects at the beginning and end of the timecourses, and our procedure ensured that such effects were not included in our processed timecourse. Resulting timecourses were then z-scored so that averaging between participants wouldn't be biased by participants with higher values. Since each experimental condition had two video clips, we concatenated the timecourse of each pair of timecourses to create a single timecourse per condition (in each ROI for each participant), which we then used for all further analyses. In all the following analyses, those timecourses were used as the main input, and the same analytic procedures were applied to all timecourses.

The partisan lens effect

One way in which partisanship is thought to affect the processing of political information is by creating a "lens" where people who share partisanship process information in a similar way. We defined the partisan lens effect as having response to stimuli that is more similar to partisan ingroup than to that of the partisan outgroup. We operationalized similarity in neural terms using ISC analysis (Hasson et al. 2004). This analysis provides a measure of time-locked synchronization of brain responses for participants watching the same dynamic stimuli. For each participant (in each ROI and condition), we first held out the participant's timecourse. We then calculated the average timecourse of all other members of their political ingroup and defined *ingroup-ISC* as the (Fisher-transformed) timecourse correlation between the participant's timecourse and the average ingroup timecourse. Similarly, *outgroup-ISC* was calculated as the (Fisher-transformed) timecourse correlation with the average timecourse of the political outgroups. For this analysis, we used the binary partisan designation, where "leaners" (participants who first identified as Independent and, when asked again, marked the party they felt closer to) were processed with their preferred party. The partisan lens effect *ingroup-ISC* − *outgroup-ISC* was calculated for each participant and tested for statistical significance across participants. To do so, we used a nonparametric permutation test. In each permutation, we shuffled the original party label of all participants and repeated the partisan lens effect calculation. The process was repeated 10,000 times to create a null distribution. The statistical significance was taken as the proportion of the permutation results larger than the observed effect—akin to a single-tailed test (Fig. 2A). Since our different hypotheses were concerned with three groups of ROIs, we used Bonferroni correction to account for the number of regions in each of the groups (see Disjunction Testing in (Rubin 2021 Jul 6). This resulted in a corrected threshold of $P < 0.0071$ for both socio-political reasoning and mentalizing regions (7 ROIs each) and $P < 0.0083$ for the affective responding regions (6 ROIs).

Our experimental hypotheses went beyond the existence of a significant partisan lens effect within condition (described above). Specifically, our common core hypothesis predicted that we should observe regions where the partisan lens effect is significant for how partisans responded to both discourse

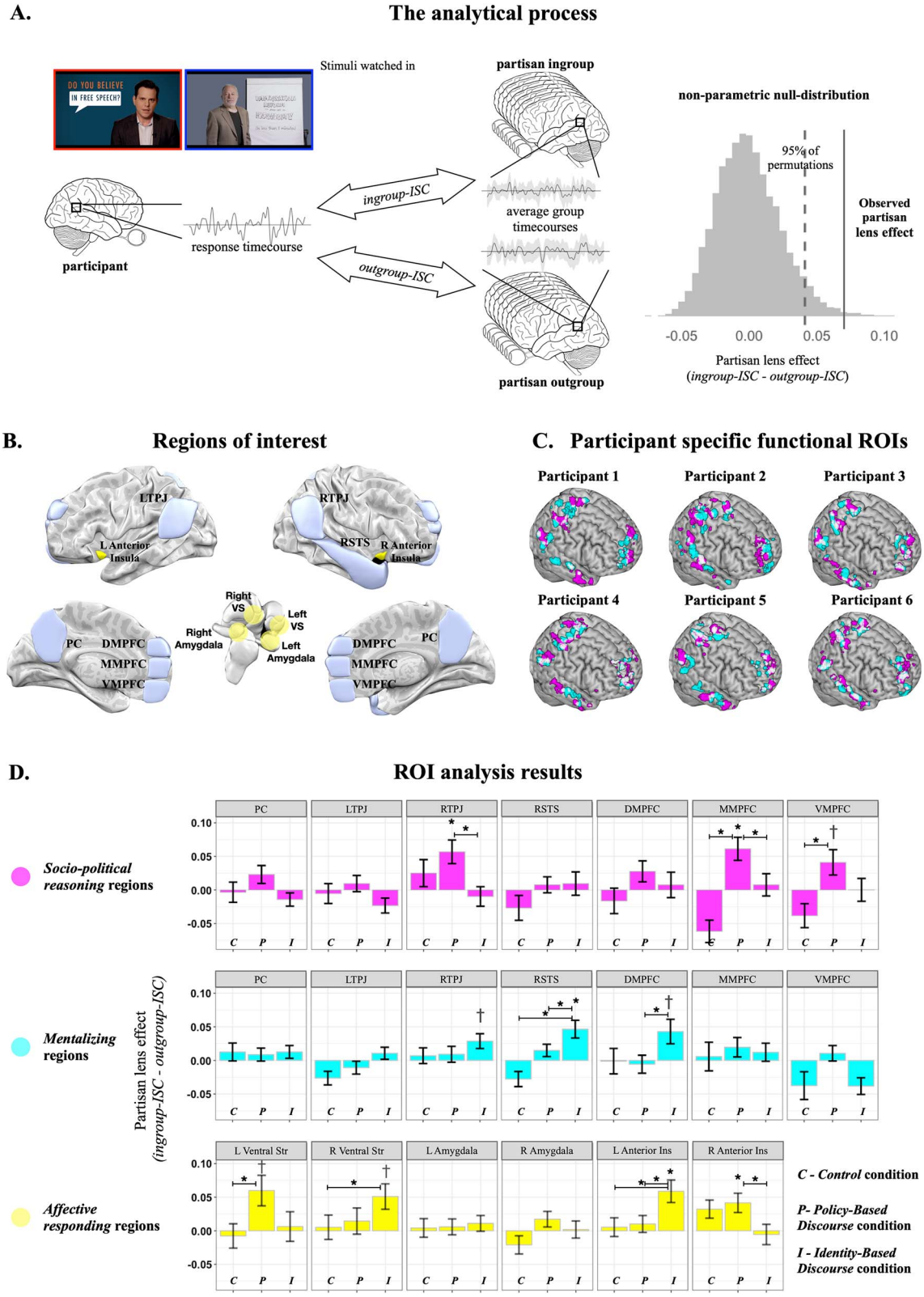

Fig. 2. The partisan lens effect in brain regions. A) The analytic process: BOLD timecourses acquired while participants watched video clips in the scanner were extracted for each participant. Each participant's timecourse was then correlated with the average timecourse of all other participants who shared the same party affiliation (ingroup-ISC) and with the average timecourse of all participants affiliated with the other party (outgroup-ISC). The "partisan lens effect" was defined as the difference between correlations (ingroup-ISC − outgroup-ISC) and statistically tested using a permutation test where party affiliations of participants were shuffled. B) The a priori ROIs used for the study. Magenta—search spaces within which we localized functional regions specific for mentalizing and socio-political reasoning. Yellow—affective responding regions. C) Example of participant-specific functional regions for socio-political (violet) and mentalizing (cyan) processes as identified by the two functional localizer tasks. D) The partisan lens effect in all ROIs and the three task conditions. Abbreviations: PC, precuneus; L/RTP, left/right temporoparietal junctions; RSTS, right superior temporal sulcus; D/M/VMPFC, dorsal/middle/ventral medial prefrontal cortex; Str, striatum; Ins, insula. Statistical significance: *—$P < 0.05$ corrected for number of regions in group (7 for mentalizing/socio-political reasoning, and 6 for affective responding). t—$P < 0.05$ uncorrected.

conditions and that those effects would be significantly larger than the lens effect in the control condition. Our multiple paths hypothesis predicted that we would observe regions where the partisan lens effect is significant in response to only one discourse condition and that this lens effect would be significantly larger than in both the other discourse condition and the control condition. Therefore, after identifying regions where the partisan lens effect was statistically significant (in any condition), we tested whether there was a significant difference in the partisan effect size between conditions. To do so, we used paired *t*-tests between conditions, e.g. $(ingroup\text{-}ISC\text{–}outgroup\text{-}ISC)_{Policy\text{-}Based\ Discourse} - (ingroup\text{-}ISC\text{–}outgroup\text{-}ISC)_{Identity\text{-}Based\ Discourse}$. For each region, we tested the difference in effect size between discourse conditions and between the relevant discourse and *Control* condition.

Results

Behavioral manipulation check

As a manipulation check, ratings of the politically charged clips were collected in the postscan questionnaire. As expected, participants judged speakers whose partisanship aligned with their own to be more reasonable [$t(93.9) = 7.97$, $P < 0.001$]. In addition, participants reported experiencing stronger positive emotions in response to clips featuring speakers of their partisan ingroup [$t(96.7) = 8.42$, $P < 0.001$] and stronger negative emotions in response to clips featuring speakers of their partisan outgroup [$t(92) = 7$, $P < 0.001$]. Taken together, these data confirm that the information in both partisanship conditions evoked partisan-tinted responses.

ROIs definition

To test for the involvement of similar/different brain systems in the different types of partisanship, we defined three independently identifiable groups of brain regions, motivated by theoretical considerations (Fig. 2B). Socio-political reasoning regions were identified at the participants level (Fig. 2C) using the socio-political reasoning localizer (Bruneau and Saxe 2010). The task was completed by 55 participants of the 61 in our final sample. Within those fROIs of MMPFC, VMPFC, PC, LTPJ, and RTPJ were successfully identified in all participants; fROIs of RSTS were identified in 54 participants and fROIs of DMPFC in 52. Mentalizing regions were identified using false-belief localizer (Dodell-Feder et al. 2011). The task was successfully completed by 55 participants. Within those, fROIs of DMPFC, PC, RSTS, LTPJ, and RTPJ were identified in all participants; fROIs of MMPFC were identified in 54 participants; and fROIs of VMPFC were identified in 53. Affective responding regions were defined either from participants' anatomy (bilateral amygdala and ventral striatum) or from an independent group analysis (see Materials and Methods). Therefore, all affective responding regions were successfully defined in all participants. In all of the following ROI analyses, only participants where relevant fROIs could be identified were included in their respective analyses.

The partisan lens effect

Using the procedure outlined above to define the partisan lens effect, we first tested whether neural responses to *Control* clips differed between partisans. As expected, none of the regions showed any significant differences consistent with the partisan lens effect in response to *Control* clips (all regions $P > 0.05$ uncorrected). These results were also consistent with the lack of behavioral difference between partisans in response to these clips during pretesting (see Supplementary Materials). We then applied the same tests to clips in our two discourse conditions to directly test our hypotheses.

No common core for processing partisan discourse

Our Common Core hypothesis suggested that a single set of psychological processes may underlie the partisan lens effect, regardless of the discourse condition. On this view, at least one region should be observed where the partisan lens effect is shown for both the *Identity-Based Discourse* and the *Policy-Based Discourse* conditions, and these effects should be distinguishable from the null effects found for the *Control* condition. In fact, none of the regions we tested showed this pattern of results even at an uncorrected threshold (Fig. 2D).

Different neural pathways for processing partisan discourse

Our multiple paths hypothesis suggested that different types of partisan appeals would engage different psychological and neural processes in a condition-dependent manner. On this view, regions for socio-political reasoning/mentalizing or affective responding should show the partisan lens effect for one of the discourse conditions but not the other, and these effects should be distinguishable from the *Control* condition.

We first tested for regions that demonstrated a partisan lens effect during exposure to the *Policy-Based Discourse* condition only. Consistent with the idea that policy-based partisanship relies in part on reasoning, evaluating, and reflecting on one's own beliefs and attitudes, we observed the partisan lens effect in three of the socio-political reasoning ROIs, namely, RTPJ ($P = 0.003$ uncorrected), MMPFC ($P = 0.003$ uncorrected), and VMPFC ($P = 0.04$ uncorrected), with both RTPJ and MMPFC showing effects at the corrected threshold of $P < 0.0071$. In MMPFC, the partisan lens effect was distinguishable from the lens effect observed in both the *Control* and *Identity-Based Discourse* conditions, while the effect in RTPJ was distinguishable only from *Control* (see Fig. 2C and Table 1 for full between-conditions statistics). For regions associated with affective responding, we observed a partisan lens effect in the right anterior insula ($P = 0.007$ uncorrected) and left ventral striatum ($P = 0.012$ uncorrected), with only the anterior insula effect surviving at the corrected threshold of 0.0083. However, the effect in the insula was distinguishable only from the *Identity-Based Discourse* condition and the effect in the striatum was distinguishable only from the *Control* condition. None of the mentalizing-sensitive ROIs showed significant evidence of a partisan lens effect in the *Policy-Based Discourse* condition.

We then tested for regions that demonstrated the partisan lens effect during the *Identity-Based Discourse* condition only. Here, consistent with the idea that identity-based partisanship is based in part on affective responding (and indeed, our stimuli featured partisan denigrations of the outgroup), we found evidence of the partisan lens effect in the left anterior insula ($P = 0.002$ uncorrected) and right ventral striatum ($P = 0.031$ uncorrected) with the anterior insula surviving at the corrected threshold of 0.0083. In the insula, the partisan lens effect during the *Identity-Based Discourse* condition was distinguishable from the lens effect observed during both the *Control* and *Policy-Based Discourse* conditions. In addition, a partisan lens effect was observed in three mentalizing ROIs—RSTS ($P = 0.005$ uncorrected), RTPJ ($P = 0.024$ uncorrected), and DMPFC ($P = 0.024$ uncorrected), with the RSTS effect surviving at the corrected threshold and significantly greater than the lens effects observed in both the *Control* and *Policy-Based Discourse* conditions. None of the regions sensitive to socio-political reasoning

Table 1. The partisan lens effect, between-condition comparisons.

Regions showing the partisan lens effect during the *Policy-Based Discourse* condition			
Region group	**Region**	***Policy-Based* > *Identity-Based***	***Policy-Based* > *Control***
Affective responding	Left VS	$t(60) = 1.706$, $P = 0.093$	**$t(60) = 2.195$, $P = 0.032$**
Affective responding	Right Ant Ins	**$t(60) = 2.077$, $P = 0.042$**	$t(60) = 0.454$, $P = 0.652$
Socio-political reasoning	RTPJ	**$t(54) = 3.464$, $P = 0.001$**	$t(54) = 1.321$, $P = 0.192$
Socio-political reasoning	MMPFC	**$t(54) = 2.263$, $P = 0.028$**	**$t(54) = 5.58$, $P < 0.001$**
Socio-political reasoning	VMPFC	$t(54) = 1.654$, $P = 0.104$	**$t(54) = 2.792$, $P = 0.007$**
Regions showing the partisan lens effect during the *Identity-Based Discourse* condition			
Region group	**Region**	***Identity-Based* > *Policy-Based***	***Identity-Based* > *Control***
Affective responding	Right VS	$t(60) = 1.614$, $P = 0.112$	**$t(60) = 2.195$, $P = 0.032$**
Affective responding	Left Ant Ins	**$t(60) = 2.153$, $P = 0.042$**	**$t(60) = 2.195$, $P = 0.032$**
Mentalizing	RTPJ	$t(54) = 1.228$, $P = 0.225$	$t(54) = 1.654$, $P = 0.104$
Mentalizing	RSTS	**$t(54) = 2.443$, $P = 0.018$**	**$t(54) = 4.048$, $P < 0.001$**
Mentalizing	DMPFC	**$t(54) = 2.213$, $P = 0.031$**	$t(54) = 1.639$, $P = 0.107$

Statistically significant effects are marked in bold.

showed a partisan lens effect during the *Identity-Based Discourse* condition.

Discussion

American politics have become increasingly partisan (Mason 2018; Finkel et al. 2020), characterized by a tendency to view the world through a "partisan lens" (Mason 2018). The literature currently distinguishes between two main forms of partisanship: policy-based and identity-based. Each centers on different aspects of the partisan political experience, namely, policy preferences or social group identification. In the current study, we used fMRI to test whether processing political discourse that appeals to each form of partisanship is supported by a common core set of brain regions or instead depends on different processes and multiple neural pathways. Our results support the latter view: exposure to examples of policy-based and identity-based partisan discourse created a partisan lens through dissociable neural processes.

Using a theory-driven, ROI-based approach, four key findings emerged. Critically, no single ROI showed the partisan lens effect under both discourse conditions, challenging the idea that partisanship is represented in the brain in a single, unified way (Van Bavel and Pereira 2018). Instead, for the example media used in our *Policy-Based Discourse* condition, partisan lens effects were observed in regions of medial PFC and RPTJ that had been independently defined using a separate socio-political reasoning task (Bruneau and Saxe 2010). For the *Identity-Based Discourse* condition, we observed the partisan lens effect in a right posterior STS region associated with mentalizing (Saxe and Kanwisher 2003; Mar 2011). Finally, though not meeting our strongest criteria for the common core hypothesis, the partisan lens effect was observed under both partisan discourse conditions in the anterior insula (albeit in different hemispheres for each condition), a region associated with affective responding and awareness of bodily states (Zaki et al. 2012).

These findings shed light on the complex nature of contemporary political partisanship. Political scientists continue to debate the relationship between identity-based and policy-based partisanship (Sniderman and Stiglitz 2012; Orr and Huber 2020; Dias and Lelkes 2022), in large part to better understand the dynamics of partisan animosity (Bougher 2017). Although it is tempting to explain partisanship as operating through a simple, core set of mental processes, this is inconsistent with our findings. Instead, using a limited set of stimuli representing only a subsample of the available partisan discourse, we find that different kinds of partisan discourse generated partisan lenses across multiple brain regions associated with a variety of functions. We also found that analogous, but separate, regions involved in affective responding showed the partisan lens effect in both partisan discourse conditions. Our use of neuroimaging to characterize how the brain responds to political discourse with different forms of partisan appeals thus extends work on the implicit measurement of party identification (Theodoridis 2017) to better specify its neural correlates. Taken together, these findings call into question the idea that partisanship is ultimately more rooted in *either* policy preferences *or* affective attachment. Rather, partisanship's influence on the processing of political messages depends on "lenses" with multiple types of components, some of which are involved in reasoning and mentalizing and some of which are involved in affective processing. Indeed, it is also possible that with an even greater range of stimuli (e.g. campaign advertising, speeches by politicians), even more dissociable processing components—i.e. lenses—would be revealed.

The finding that regions for affective responding demonstrated the partisan lens effects under both experimental conditions suggests that even though no singular region met our common core partisan information processing criteria, affective processes seem to be involved in some way across information types. In fact, the strongest partisan lens effects were observed in the anterior insula, which is associated with ongoing tracking of bodily states and affective experience (Craig 2009; Zaki et al. 2012). Although laterality effects in emotion are not always observed in functional imaging and can be subject to multiple interpretations (Canli et al. 1998; Lindquist et al. 2016), we note that the effects were lateralized such that the left anterior insula showed a partisan lens effect during exposure to the *Identity-Based Discourse* condition and the right during exposure to the *Policy-Based Discourse* condition. A weaker partisan lens effect was observed in portions of the ventral striatum (again differing in lateralization between conditions), which can be interpreted as the shared reward experience from political speech. When zooming out, our results in the anterior insula can be seen as consistent with a weaker version of the common core hypothesis. However, taken within the overall pattern of results we observed, the evidence suggests that even if there are some shared or similar processes, overall, the partisan lens effect has many distinguishable components and our results

are not consistent with a singular or principal core process, as has been suggested (Van Bavel and Pereira 2018). Altogether, our findings add to the growing body of theoretical and empirical evidence (Redlawsk 2006; Westen et al. 2006; Kaplan et al. 2016) positing that affective responses are central to biased processing of political information.

Beyond regions for affective responding, we found a double dissociation between regions that showed the partisan lens effect in response to the *Policy-Based Discourse* and the *Identity-Based Discourse* conditions: The effect during the *Policy-Based Discourse* condition was observed in the MMPFC and weakly in the RTPJ, both identified by the socio-political reasoning localizer, whereas the effect during the *Identity-Based Discourse* condition was observed in the RSTS and weakly in the DMPFC, both identified by the mentalizing localizers. This set of observations is interesting in multiple ways.

First, the different regions involved in the partisan lens effect under different conditions strongly support the multiple paths hypothesis, complementing the nuanced results for regions involved in affective responding. Second, the different loci of activation further our understanding of the distribution of labor between prefrontal and parietal regions. Prefrontal regions such as MMPFC and DMPFC are thought to represent social and contextual information at an abstract and schematic level (Tompary and Davachi 2017; Baldassano et al. 2018), while temporal regions such as RSTS are thought to be loci for multimodal integration and the perception of social interaction (Deen et al. 2017; Isik et al. 2017). Third, it is notable that the regions identified by the socio-political reasoning localizer were those that showed the partisan lens effect in the *Policy-Based Discourse* condition. This suggests that the same presumed processes of reasoning and engagement with the content of political speech targeted by the localizer were indeed those underlying the lens effect when the political discourse is about policy. Conversely, regions identified by the mentalizing localizer showed the effect in the *Identity-Based Discourse* condition, suggesting that this form of partisanship involves thinking about a speaker's attitude, beliefs, and goals when they directly denigrate outgroup social identities rather than engage in policy debate. In addition, it is particularly interesting that although both functional localizer tasks identified ROIs within the same broad areas of the brain, there was very little overlap between the functional regions identified for each task within individual participants (see Supplementary Information). This finding is consistent with prior research showing that multiple different, but related, types of mental processes have been associated with group-level activation of the "mentalizing network" (Mar 2011; Schurz et al. 2020), even though the specific regions may be subtly but meaningfully different when studied within an individual (Scholz et al. 2009; DiNicola et al. 2020). Future work should further explore the processes evoked by the different localizers, their downstream consequences, and the distinctiveness/related processes that they evoke.

Taken together, our findings directly relate to and extend recent research in the field. For example, the results of studies by Leong et al. (2020) and van Baar et al. (2021) dovetail with the present findings. Both of these studies used ISC to identify regions involved in biased processing of political information, reporting effects in regions such as DMPFC (Leong et al. 2020), TPJ, and PC (van Baar et al. 2021), along with regions for affective responding such as the anterior insula (van Baar et al. 2021).

The present study extends these findings in at least three meaningful ways. First, our study directly manipulated and compared responding to different forms of partisan discourse. In prior studies, stimuli have either focused on specific issues or on the interactions of politicians during debates. Though these stimuli varied in the way that those topics were addressed (e.g. neutrally or in a provocative way), they all included a significant policy-based partisanship component. Our stimuli, by contrast, allowed us to differentiate policy-based from identity-based appeals, thereby providing unique insights into the ways that they operate and their interactions within individuals. In this, our approach is similar to behavioral experimental work in political science that has sought to systematically pull apart the effects of policy preferences and partisanship (Sniderman and Stiglitz 2012; Lelkes 2021).

Second, whereas prior studies used ideological scales of conservatism/liberalism writ large (van Baar et al. 2021) or for specific policy issues (Leong et al. 2020), we relied on partisan affiliation (i.e. identifying as Republican/Democrat). As discussed in the Introduction, although ideological affinity has become increasingly aligned with partisan identification (Twenge et al. 2016), it is critical that when studying the manifestations and types of partisanship (in experimental conditions) as a social psychological process, we rely on a direct measurement of social group identification rather than substituting it for ideology as a correlated proxy (Kinder and Kalmoe 2017).

Third, our study used an a priori ROI approach with greater power than whole-brain analyses to detect activation in regions associated with processes of a priori interest. In doing so, we were able to identify partisan lens effects in similar but distinct regions that were identified by related but distinct functional localizer tasks (Saxe et al. 2006; Fedorenko and Kanwisher 2009). This turned out to be critical as data-driven, rather than theory-driven, whole-brain analyses might not have been sensitive to detecting separate neural paths correlated with the processing of identity-based and policy-based partisan discourse.

More broadly, our study suggests that it is misleading to assume that partisanship is *either* a policy-based *or* an identity-based phenomenon. Rather, political messages targeting each type of attachment can be processed by dissociable partisan lenses within the same individual. This perspective complements and underscores the importance of research into partisan media and its effects (Stroud 2011; Levendusky 2013; Mutz 2015; Benedictis-Kessner et al. 2019). Specifically, our *Identity-Based Discourse* condition, which featured an outgroup partisan denigrating the ingroup, used stimuli similar to investigations of outrage speech in traditional and social media (Sobieraj and Berry 2011; Berry and Sobieraj 2013). Although this type of discourse generated a partisan lens, consistent with the literature's expectations, we also found that policy-relevant discourse generated a distinguishably different response *within the same individuals*.

Indeed, the results of this study could help inform interventions to bridge the partisan divide. Our results provide insight into why exposure to outgroup messaging might increase polarization by providing a set of neural mechanisms that synchronize information processing within partisan groups, consistent with the patterns observed by Bail et al. (2018). Our findings also suggest why interventions designed to mitigate partisanship's negative effects should take the nature of partisan media content into account (Guess et al. 2021; Broockman and Kalla 2022; Nyhan et al. 2023). Our findings of different neural bases for processing policy-based and identity-based partisan messages suggest that interventions could be designed to target the socio-political reasoning regions underlying people's responses to policy-based partisan messaging and/or mentalizing regions underlying people's responses to identity-based partisan messaging. Given

the behaviors associated with identity-based partisanship in particular, interventions might focus on reducing the differences in copartisan synchrony by targeting processes associated with mentalizing—e.g. by encouraging empathy and perspective-taking. At the same time, the finding that regions involved in affective responding were involved in partisan lens effects for both kinds of partisan discourse highlights the importance of thinking through (and testing) the emotional responses that any intervention evokes regardless of the intervention target. Further, since different forms of partisan discourse are not processed using the same set of mental processes, we should not necessarily expect an intervention designed to mitigate impacts of policy-based partisan messaging to also engage when encountering identity-based partisan messaging (or vice versa). Changing minds about policy might not change hearts committed to interparty animosity.

Despite the insights derived from our study, it is important to acknowledge it's limitations. First, when examining complex social phenomena, there is a tradeoff between experimental control and ecological validity (Zaki and Ochsner 2009). Our use of naturalistic stimuli trades the former for the latter. Indeed, our stimuli sampled only a portion of the wide variety of real-world partisan messages and contain only a few of the "active ingredients" for engaging partisans lenses. Along these lines, other factors we did not control for might affect the processing of our stimuli (such as identity of the speaker, specific topics of discussion, and how "engaging" the speaker is). Second, striving to create a balanced design, we opted for two video clips per condition (one per partisan orientation), thus making our study susceptible to item effects. It is not uncommon to choose a limited amount of treatment stimuli when testing the effect of interventions in political communications (Coppock et al. 2020). Nevertheless, future work is needed to replicate and extend these findings to new instances of partisan discourse. At the same time, we believe that the fact that a limited set of stimuli was sufficient to identify multiple neural paths offers strong evidence against the conceptualization of a common neural core for partisanship and, by extension, the notion that partisanship's influence stems from either a set of policy-based preferences or a social identity (i.e. our results show that both operate within each individual).

In conclusion, in this study, we sought to characterize the neural correlates underlying the "partisan lens" effect. In independently defined ROIs, we used ISC to measure partisans' responses to policy-based and identity-based discourse. We found evidence of dissociable partisan lenses in both conditions. Significantly, partisan processing during our policy-based partisan discourse condition was observed in regions engaged in socio-political reasoning, as well as affective responding. The partisan lens effect was also observed during our identity-based partisan discourse condition in regions related to mentalizing and affective responding. Taken together, this suggests that responding to different types of partisan discourse is supported by a set of related but distinguishable psychological processes.

Acknowledgments

The authors would like to thank Matt O'Donnell for support on this work. We would like to acknowledge that though he did not live to see this work come to fruition, the spirit of Emile Bruneau was an inspiration and a guiding light for us.

Author contributions

Nir Jacoby (Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Marika Landau-Wells (Formal analysis, Methodology, Writing—original draft, Writing—review & editing), Jacob Pearl (Data curation, Investigation, Methodology), Alexandra Paul (Data curation, Investigation), Emily B. Falk (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—review & editing), and Emile G. Bruneau (Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision), Kevin Ochsner (Methodology, Supervision, Writing—review & editing).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

We are grateful for support from Beyond Conflict, The Germanacos Foundation, and the Annenberg School for Communication.

Conflict of interest statement: None declared.

Data and materials availability

Anonymized imaging data extracted from regions of interest, along with all the code required for reproducing the results of the paper, will be publicly available in https://osf.io/4bzkt/. The stimuli used for the main experimental paradigm are publicly available and links were included in the supporting information. The stimuli used for the localizer tasks are vignettes from published research that is all well referenced and cited in the manuscript.

References

Abramowitz AI. The polarized American electorate: the rise of partisan-ideological consistency and its consequences. *Polit Sci Q*. 2022:137(4):645–674. https://doi.org/10.1002/polq.13388.

Abramowitz AI, Webster S. The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Elect Stud*. 2016:41:12–22. https://doi.org/10.1016/j.electstud.2015.11.001.

American National Election Studies (ANES), University Of Michigan and Stanford University. ANES 2016 time series study. *Inter-university Consortium for Political and Social Research [distributor]*. 2017. https://doi.org/10.3886/ICPSR36824.V2.

Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006:7(4):268–277. https://doi.org/10.1038/nrn1884.

Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, Lee J, Mann M, Merhout F, Volfovsky A. Exposure to opposing views on social media can increase political polarization. *Proc Natl Acad Sci*. 2018:115(37):9216–9221. https://doi.org/10.1073/pnas.1804840115.

Baldassano C, Hasson U, Norman KA. Representation of real-world event schemas during narrative perception. *J Neurosci*. 2018:38(45):9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018.

Bartels LM. Beyond the running tally: partisan bias in political perceptions. *Polit Behav*. 2002:24(2):117–150. https://doi.org/10.1023/A:1021226224601.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Soft*. 2015:67(1):1–48. https://doi.org/10.18637/jss.v067.i01.

Benedictis-Kessner JD, Baum MA, Berinsky AJ, Yamamoto T. Persuading the enemy: estimating the persuasive effects of partisan media with the preference-incorporating choice and assignment design. *Am Polit Sci Rev*. 2019:113(4):902–916. https://doi.org/10.1017/S0003055419000418.

Berry JM, Sobieraj S. *The outrage industry: political opinion media and the new incivility*. Oxfortd (UK): Oxford University Press; 2013.

Bisgaard M. Bias will find a way: economic perceptions, attributions of blame, and partisan-motivated reasoning during crisis. *J Polit*. 2015:77(3):849–860. https://doi.org/10.1086/681591.

Blank IA, Fedorenko E. No evidence for differences among language regions in their temporal receptive windows. *NeuroImage*. 2020:219:116925. https://doi.org/10.1016/j.neuroimage.2020.116925.

Bougher LD. The correlates of discord: identity, issue alignment, and political hostility in polarized America. *Polit Behav*. 2017:39(3):731–762. https://doi.org/10.1007/s11109-016-9377-1.

Broockman D, Kalla J. The impacts of selective partisan media exposure: a field experiment with fox news viewers. *OSF Preprints*. 2022:1–68. https://osf.io/preprints/osf/jrw26.

Bruneau EG, Saxe R. Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *NeuroImage*. 2010:52(4):1704–1711. https://doi.org/10.1016/j.neuroimage.2010.05.057.

Campbell A, Converse PE, Miller WE, Stokes DE. *The American voter*. New York: John Wiley & Sons, Ltd; 1960.

Canli T, Desmond JE, Zhao Z, Glover G, Gabrieli JDE. Hemispheric asymmetry for emotional stimuli detected with fMRI. *Neuroreport*. 1998:9(14):3233–3239. https://doi.org/10.1097/00001756-199810050-00019.

Chang LJ, Yarkoni T, Khaw MW, Sanfey AG. Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb Cortex*. 2013:23(3):739–749. https://doi.org/10.1093/cercor/bhs065.

Coe K, Tewksbury D, Bond BJ, Drogos KL, Porter RW, Yahn A, Zhang Y. Hostile news: partisan use and perceptions of cable news programming. *J Commun*. 2008:58(2):201–219. https://doi.org/10.1111/j.1460-2466.2008.00381.x.

Coppock A, Hill SJ, Vavreck L. The small effects of political advertising are small regardless of context, message, sender, or receiver: evidence from 59 real-time randomized experiments. Science. *Advances*. 2020:6(36):eabc4046. https://doi.org/10.1126/sciadv.abc4046.

Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am J Neuroradiol*. 2001:22(7):1326–1333.

Craig AD. How do you feel–now? The anterior insula and human awareness. *Nat Rev Neurosci*. 2009:10(1):59–70. https://doi.org/10.1038/nrn2555.

Davis M, Whalen PJ. The amygdala: vigilance and emotion. *Mol Psychiatry*. 2001:6(1):13–34. https://doi.org/10.1038/sj.mp.4000812.

Deen B, Pitskel NB, Pelphrey KA. Three Systems of Insular Functional Connectivity Identified with cluster analysis. *Cereb Cortex*. 2010:21(7):1498–1506. https://doi.org/10.1093/cercor/bhq186.

Deen B, Richardson H, Dilks DD, Takahashi A, Keil B, Wald LL, Kanwisher N, Saxe R. Organization of high-level visual cortex in human infants. *Nat Commun*. 2017:8(1):13995. https://doi.org/10.1038/ncomms13995.

Dias N, Lelkes Y. The nature of affective polarization: disentangling policy disagreement from partisan identity. *Am J Polit Sci*. 2022:66(3):775–790. https://doi.org/10.1111/ajps.12628.

DiNicola LM, Braga RM, Buckner RL. Parallel distributed networks dissociate episodic and social functions within the individual. *J Neurophysiol*. 2020:123(3):1144–1179. https://doi.org/10.1152/jn.00529.2019.

Ditto PH, Liu BS, Clark CJ, Wojcik SP, Chen EE, Grady RH, Celniker JB, Zinger JF. At least bias is bipartisan: a meta-analytic comparison of partisan bias in liberals and conservatives. *Perspect Psychol Sci*. 2019:14(2):273–291. https://doi.org/10.1177/1745691617746796.

Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. fMRI item analysis in a theory of mind task. *NeuroImage* 2011:55(2):705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040.

Downs A. *An economic theory of democracy*. 1st ed. New York: Harper and Row; 1957.

Dufour N, Redcay E, Young L, Mavros PL, Moran JM, Triantafyllou C, Gabrieli JDE, Saxe R. Similar brain activation during false belief tasks in a large sample of adults with and without autism. Gilbert S, editor. *PLoS One*. 2013:8(9):e75468. https://doi.org/10.1371/journal.pone.0075468.

Esteban O, Blair R, Markiewicz CJ, Berleant SL, Moodie C, Ma F, Isik AI, Erramuzpe A, Kent JD, Goncalves M, et al. FMRIPrep software (20.1.0rc3). 2018a. Zenodo. https://doi.org/10.5281/zenodo.852659.

Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods*. 2018b:16(1):111–116. https://doi.org/10.1038/s41592-018-0235-4.

Esteban O, Ciric R, Finc K, Blair RW, Markiewicz CJ, Moodie CA, Kent JD, Goncalves M, DuPre E, Gomez DEP, et al. Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nat Protoc*. 2020:15(7):2186–2202. https://doi.org/10.1038/s41596-020-0327-3.

Fedorenko E, Kanwisher N. Neuroimaging of language: why Hasn't a clearer picture emerged? *Lang Linguist Compass*. 2009:3(4):839–865. https://doi.org/10.1111/j.1749-818X.2009.00143.x.

Finkel EJ, Bail CA, Cikara M, Ditto PH, Iyengar S, Klar S, Mason L, McGrath MC, Nyhan B, Rand DG, et al. Political sectarianism in America. *Science*. 2020:370(6516):533–536. https://doi.org/10.1126/science.abe1715.

Gillion DQ, Ladd JM, Meredith M. Party polarization, ideological sorting and the emergence of the US partisan gender gap. *Br J Polit Sci*. 2020:50(4):1217–1243. https://doi.org/10.1017/S0007123418000285.

Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. 2016:3(1):160044–160049. https://doi.org/10.1038/sdata.2016.44.

Green DP, Palmquist B, Schickler E. *Partisan hearts and minds: political parties and the social identities of voters*. New Haven (CT): Yale University Press; 2004.

Greene S. Social identity theory and party identification*. *Soc Sci Q*. 2004:85(1):136–153. https://doi.org/10.1111/j.0038-4941.2004.08501010.x.

Guess AM, Barberá P, Munzert S, Yang J. The consequences of online partisan media. *Proc Natl Acad Sci*. 2021:118(14):e2013464118. https://doi.org/10.1073/pnas.2013464118.

Haas IJ, Baker MN, Gonzalez FJ. Who can deviate from the party line? Political ideology moderates evaluation of incongruent policy positions in insula and anterior cingulate cortex. *Soc Just Res*. 2017:30(4):355–380. https://doi.org/10.1007/s11211-017-0295-0.

Haas IJ, Baker MN, Gonzalez FJ. Political uncertainty moderates neural evaluation of incongruent policy positions. *Philos Trans*

R Soc B Biol Sci. 2021:376(1822):20200138. https://doi.org/10.1098/rstb.2020.0138.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. *Science*. 2004:303(5664):1634–1640. https://doi.org/10.1126/science.1089506.

Highton B, Kam CD. The long-term dynamics of partisanship and issue orientations. *J Polit*. 2011:73(1):202–215. https://doi.org/10.1017/S0022381610000964.

Huddy L, Mason L, Aarøe L. Expressive partisanship: campaign involvement, political emotion, and partisan identity. *Am Polit Sci Rev*. 2015:109(1):1–17. https://doi.org/10.1017/S0003055414000604.

Isik L, Koldewyn K, Beeler D, Kanwisher N. Perceiving social interactions in the posterior superior temporal sulcus. *Proc Natl Acad Sci*. 2017:114(43):E9145–E9152. https://doi.org/10.1073/pnas.1714471114.

Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ. The origins and consequences of affective polarization in the United States. *Annu Rev Polit Sci*. 2019:22(1):129–146. https://doi.org/10.1146/annurev-polisci-051117-073034.

Jefferson H. *The curious case of black conservatives: construct validity and the 7-point liberal-conservative scale*, working paper (July 6, 2020). SSRN: https://doi.org/10.2139/ssrn.3602209. https://papers.ssrn.com/abstract=3602209.

Kaplan JT, Gimbel SI, Harris S. Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Sci Rep*. 2016:6(1):39589. https://doi.org/10.1038/srep39589.

Kinder DR, Kalmoe NP. *Neither liberal nor conservative: ideological innocence in the American public*. Chicago(IL): University of Chicago Press; 2017. https://doi.org/10.7208/chicago/9780226452593.001.0001.

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

Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw*. 2017:82(13):1–26. https://doi.org/10.18637/jss.v082.i13.

Lee AH-Y, Lelkes Y, Hawkins CB, Theodoridis AG. Negative partisanship is not more prevalent than positive partisanship. *Nat Hum Behav*. 2022:6(7):951–963. https://doi.org/10.1038/s41562-022-01348-0.

Lees J, Cikara M. Understanding and combating misperceived polarization. *Philos Trans R Soc B Biol Sci*. 2021:376(1822):20200143. https://doi.org/10.1098/rstb.2020.0143.

Lelkes Y. Policy over party: comparing the effects of candidate ideology and party on affective polarization. *Polit Sci Res Med*. 2021:9(1):189–196. https://doi.org/10.1017/psrm.2019.18.

Leong YC, Chen J, Willer R, Zaki J. Conservative and liberal attitudes drive polarized neural responses to political content. *Proc Natl Acad Sci*. 2020:117(44):27731–27739. https://doi.org/10.1073/pnas.2008530117.

Levendusky M. *How partisan media polarize America*. Chicago (IL): University of Chicago Press; 2013 https://doi.org/10.7208/chicago/9780226069159.001.0001

Levitsky S, Ziblatt D. *How democracies die*. New York (NY): Crown; 2019.

Lindquist KA, Satpute AB, Wager TD, Weber J, Barrett LF. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb Cortex*. 2016:26(5):1910–1922. https://doi.org/10.1093/cercor/bhv001.

Lupu N. Party polarization and mass partisanship: a comparative perspective. *Polit Behav*. 2015:37(2):331–356. https://doi.org/10.1007/s11109-014-9279-z.

Mar RA. The neural bases of social cognition and story comprehension. *Annu Rev Psychol*. 2011:62(1):103–134. https://doi.org/10.1146/annurev-psych-120709-145406.

Mason L. *Uncivil agreement: how politics became our identity*. Chicago (IL): University of Chicago Press; 2018 https://doi.org/10.7208/chicago/9780226524689.001.0001.

Moore-Berg SL, Hameiri B, Bruneau EG. The prime psychological suspects of toxic political polarization. *Curr Opin Behav Sci*. 2020:34:199–204. https://doi.org/10.1016/j.cobeha.2020.05.001.

Muirhead R, Rosenblum NL. The political theory of parties and partisanship: catching up. *Annu Rev Polit Sci*. 2020:23(1):95–110. https://doi.org/10.1146/annurev-polisci-041916-020727.

Mutz DC. *In-your-face politics: the consequences of uncivil media. In: In-your-face politics*. Princeton (NJ): Princeton University Press; 2015. https://doi.org/10.23943/princeton/9780691165110.003.0002.

Nyhan B, Settle J, Thorson E, Wojcieszak M, Barberá P, Chen AY, Allcott H, Brown T, Crespo-Tenorio A, Dimmery D, et al. Like-minded sources on Facebook are prevalent but not polarizing. *Nature*. 2023:620(7972):137–144. https://doi.org/10.1038/s41586-023-06297-w.

Orr LV, Huber GA. The policy basis of measured partisan animosity in the United States. *Am J Polit Sci*. 2020:64(3):569–586. https://doi.org/10.1111/ajps.12498.

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical parametric mapping: the analysis of functional brain images*. London (UK): Elsevier; 2011.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*. 2006:442(7106):1042–1045. https://doi.org/10.1038/nature05051.

Pessoa L, Adolphs R. Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. *Nat Rev Neurosci*. 2010:11(11):773–782. https://doi.org/10.1038/nrn2920.

Phelps EA, LeDoux JE. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*. 2005:48(2):175–187. https://doi.org/10.1016/j.neuron.2005.09.025.

Prado J, Léone J, Epinat-Duclos J, Trouche E, Mercier H. The neural bases of argumentative reasoning. *Brain Lang*. 2020:208:104827. https://doi.org/10.1016/j.bandl.2020.104827.

R Core Team. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing, 2013. Available from: http://www.R-project.org/.

Redlawsk D. Feeling politics: New research into emotion and politics. In Redlawsk D, editor. *Feeling politics: emotion in political information processing*. New York (NY): Springer, 2006, p. 1–10. https://doi.org/10.1057/9781403983114.

Rubin M. When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. *Synthese*. 2021:199(3–4):10969–11000. https://doi.org/10.1007/s11229-021-03276-4.

Ruff CC, Fehr E. The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci*. 2014:15(8):549–562. https://doi.org/10.1038/nrn3776.

Saxe R, Kanwisher N. People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind". *NeuroImage*. 2003:19(4):1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1.

Saxe R, Brett M, Kanwisher N. Divide and conquer: a defense of functional localizers. *NeuroImage*. 2006:30(4):1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062.

Scholz J, Triantafyllou C, Whitfield-Gabrieli S, Brown EN, Saxe R. Distinct regions of right temporo-parietal junction are selec-

tive for theory of mind and exogenous attention. Lauwereyns J, editor. *PLoS One*. 2009:4(3):e4869. https://doi.org/10.1371/journal.pone.0004869.

Schurz M, Radua J, Tholen MG, Maliske L, Margulies DS, Mars RB, Sallet J, Kanske P. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol Bull*. 2020:147(3):293–327. https://doi.org/10.1037/bul0000303.

Shively WP. The development of party identification among adults: exploration of a functional model. *Am Polit Sci Rev*. 1979:73(4):1039–1054. https://doi.org/10.2307/1953988.

Sniderman PM, Stiglitz EH. The reputational premium: A theory of party identification and policy reasoning. Princeton (NJ): Princeton University Press; 2012. https://doi.org/10.1515/9781400842551.12.

Sobieraj S, Berry JM. From incivility to outrage: political discourse in blogs, talk radio, and cable news. *Polit Commun*. 2011:28(1):19–41. https://doi.org/10.1080/10584609.2010.542360.

Stroud NJ. *Niche news: the politics of news choice*. USA: Oxford University Press; 2011. https://doi.org/10.1093/acprof:oso/9780199755509.001.0001.

Taber CS, Lodge M. Motivated Skepticism in the evaluation of political beliefs. *Am J Polit Sci*. 2006:50(3):755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x.

Tajfel H. An integrative theory of intergroup conflict In: Austin WG, Worchel S, editors. *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks/Cole; 1979. pp. 33–37.

Theodoridis AG. Me, myself, and (I), (D), or (R)? Partisanship and political cognition through the lens of implicit identity. *J Polit*. 2017:79(4):1253–1267. https://doi.org/10.1086/692738.

Tompary A, Davachi L. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*. 2017:96(1):228–241.e5. https://doi.org/10.1016/j.neuron.2017.09.005.

Twenge JM, Honeycutt N, Prislin R, Sherman RA. More polarized but more independent: political party identification and ideological self-categorization among U.S. adults, college students, and late adolescents, 1970-2015. *Personal Soc Psychol Bull*. 2016:42(10):1364–1383. https://doi.org/10.1177/0146167216660058.

van Baar JM, Halpern DJ, Feldman HO. Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proc Natl Acad Sci*. 2021:118(20):e2022491118. https://doi.org/10.1073/pnas.2022491118.

Van Bavel JJ, Pereira A. The partisan brain: an identity-based model of political belief. *Trends Cogn Sci (Regul Ed)*. 2018:22(3):213–224. https://doi.org/10.1016/j.tics.2018.01.004.

Westen D, Blagov PS, Harenski K, Kilts C, Hamann S. Neural bases of motivated reasoning: an FMRI study of emotional constraints on partisan political judgment in the 2004 U.S. presidential election. *J Cogn Neurosci*. 2006:18(11):1947–1958. https://doi.org/10.1162/jocn.2006.18.11.1947.

Zaki J, Ochsner K. The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci*. 2009:1167(1):16–30. https://doi.org/10.1111/j.1749-6632.2009.04601.x.

Zaki J, Ochsner K. The neuroscience of empathy: progress, pitfalls and promise. *Nat Neurosci*. 2012:15(5):675–680. https://doi.org/10.1038/nn.3085.

Zaki J, Davis JI, Ochsner KN. Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage*. 2012:62(1):493–499. https://doi.org/10.1016/j.neuroimage.2012.05.012.