Policy-based and Affective Partisanship Depend on Dissociable Neural Systems

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

1. Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755.
2. Department of Psychology, Columbia University, New York, NY 19104.
3. Travers Department of Political Science, University of California-Berkeley, Berkeley, CA 94720.
4. Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104.

*Corresponding author:

Nir Jacoby

nir.jacoby@dartmouth.edu

6207 Moore Hall, Dartmouth College,

Hanover, NH 03755

Abstract

Political partisanship is often conceived as a lens through which we view politics. Behavioral research has distinguished two types of “partisan lenses” - *Policy-based* and *Affective* – that may influence our perception of political events. Little is known, however, about the mechanisms through which partisanship operates 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. A policy-based partisanship lens effect was observed in socio-political reasoning and affective responding brain regions. An affective partisanship lens effect was observed in mentalizing and affective responding brain regions. These data suggest that policy-based and affective partisanship are supported by related but distinguishable neural and therefore psychological mechanisms, which may have implications for how we characterize partisanship and ameliorate its deleterious impacts.

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 (Finkel et al., 2020; Levitsky & Ziblatt, 2019; Lupu, 2015; Mason, 2018). 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 (Abramowitz & Webster, 2016; Finkel et al., 2020; Sniderman & Stiglitz, 2012). 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, can be a constructive element of a vibrant democracy by fostering reasoned discussion (Muirhead & Rosenblum, 2020). **Affective partisanship**[1], where one's emotional connections to the party as a social identity is the basis of party attachment, is often less constructive (Finkel et al., 2020; Iyengar et al., 2019; Mason, 2018). Both of these types of partisanship are likely relevant at the societal level when considering partisanship's pernicious effects (e.g. Dias & Lelkes, 2022; Huddy et al., 2015; Orr & Huber, 2020). But a foundational question remains unanswered: how do these two types of partisanship operate and interact *within* individuals? Are they both manifestations of the same underlying mechanisms? Or do they have fundamentally different roots? Addressing this question may be essential for better understanding how different types of political messaging in media might activate partisan information processing, for disrupting partisanship's negative consequences, and for tailoring interventions to particular forms of partisanship.

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 (Downs, 1957; Shively, 1979) . Policy-based partisanship has an essential role in deliberative processes for individuals and at the societal level (Muirhead & Rosenblum, 2020). On the other hand, affective partisanship functions as a type of social identity whose underlying affective mechanisms (Campbell et al., 1960; Greene, 2004; Tajfel & Turner, 1979) can influence identity based judgments, leading us to misperceive partisan outgroup members and their motivations (Lees & Cikara, 2021; Moore-Berg et al., 2020).

Building on this analysis, we used neuroimaging to inform our understanding of how each of these types of partisanship operates within individuals. Specifically, we asked whether and how the psychological and neural mechanisms underlying policy-based and affective partisanship relate to one another. Our method was guided by research documenting a behavioral manifestation of both kinds of partisanship: when facing politically relevant information, a partisan will process the information in a way that is more similar to those who share their partisan convictions (partisan ingroup) than those who do not (partisan outgroup). This "partisan lens" effect is thought to color all manner of politically relevant perceptions, cognitions, emotions and actions (Mason, 2018). Moreover, the strength of such motivated information processing has been shown to be correlated with implicit measurement of party identification (Theodoridis, 2017).

---

[1] 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 & Kam, 2011). Here, we use labels that reflect the nature of the type of partisanship to which the speakers appeal (i.e., policy agreement, positive ingroup/negative outgroup emotions) to "activate" partisan information processing.

Recent neuroimaging research has provided further evidence about the mechanisms by which partisanship can color information processing. 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, 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 average citizens (c.f. elites) process partisan information because recent work has focused on the role of participants' *ideological* self-identification (i.e., as liberals or conservatives), not their *partisan* self-identification (e.g. Haas et al., 2017, 2021; Leong et al., 2020; van Baar et al., 2021). Although ideological affinity is increasingly aligned with political party self-identification in the U.S., the strength of this relationship has varied over time (Abramowitz, 2022), and varies also by age, gender and race (Gillion et al., 2020; Jefferson, 2020; Twenge et al., 2016). Our study thus extends previous research in two meaningful ways: First, by measuring participants' partisanship directly rather than substituting ideology for partisanship (Kinder & Kalmoe, 2017); Second, by directly manipulating and investigating how different types of partisanship affect the processing of political information.

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 & Rosenblum, 2020). Consistent with the involvement of these processes, the handful of neuroimaging studies examining political and related forms of reasoning (Bruneau & Saxe, 2010; Kaplan et al., 2016; Prado et al., 2020; Westen et al., 2006) have observed activation of medial frontal, temporal-parietal and precuneus regions, all of which have been implicated in judgments about mental states (Amodio & Frith, 2006; Mar, 2011; Saxe & Kanwisher, 2003; Zaki & 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 either policy-based or affective partisanship.

A second set of candidate psychological processes involve those supporting one's attitudes towards, personal valuation of, and self-identification with the party as a social group (Campbell et al., 1960; Greene, 2004; Tajfel & Turner, 1979). 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 & Fehr, 2014), the amygdala – thought to be important for detecting goal-relevant stimuli, in general, with a special role in detecting potential threats (Davis & Whalen, 2001; Pessoa & Adolphs, 2010; Phelps & LeDoux, 2005), and the anterior insula – thought to be important for integrating body state information with negative affective states (Chang et al., 2013; Craig, 2009; Deen et al., 2010; Zaki et al., 2012). Notably, although "affective partisanship" as a label is suggestive of affective processes, our condition label pertains to the type of partisan appeal, and not the way that partisans respond to it. Indeed, although exposure to politically-relevant information has been shown in prior studies to engage regions associated with affective responding (e.g. Haas et al., 2017; Kaplan et al., 2016;

van Baar et al., 2021; Westen et al., 2006), it remains unclear whether and how this engagement is related to either policy-based or affective partisanship.

With these considerations in mind, we sought to test two hypotheses. The first was that both policy-based and affective "partisan lens" effects may rely on a **common core** of psychological and neural processes (e.g. Van Bavel & Pereira, 2018). On this view, both types of partisanship rely on similar processes and therefore should engage **the same set(s) of brain regions** – particularly those implicated in socio-political reasoning / mentalizing and/or affective responding. The second, **multiple paths** hypothesis, was that each type of partisanship depends on **different brain regions,** with policy-based partisanship preferentially engaging systems for socio-political reasoning and mentalizing with affective partisanship engaging system for affective responding.

To test these two, "partisan lens", hypotheses, we collected whole-brain fMRI data while self-identified partisan participants watched video clips that discussed either policy issues (*Policy-Based Partisanship* condition), denigrated outgroup partisans (*Affective Partisanship* condition), or discussed a non-political 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 & Kanwisher, 2009; Scholz et al., 2009). As such, separate tasks from the literature were selected to localize regions directly involved in socio-political reasoning (Bruneau & Saxe, 2010) and mentalizing (Dodell-Feder et al., 2011). For regions related to affective 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 aligned with policy-based or affective partisanship, and critically, whether results supported the common core or multiple paths hypotheses.

Results

Sixty-one participants from across the American political spectrum completed an fMRI experiment that took place from July 2017 to August 2018. Participants' partisanship was measured using self-report measures in a pre-scan questionnaire. Participants first were 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, as is standard in American public opinion research (e.g., the American National Election Studies, ANES). 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) (Figure 1.B).

Over the course of the scanning session, each participant watched six short video clips (range: 2:22 – 4:22; min:sec), two per condition. Each of the two political conditions (*Policy-Based*

Partisanship and *Affective Partisanship*) presented a Democrat's position on a specific topic in one clip and a Republican's position on a different topic in the other (Figure 1.A; See Methods, for a full list of stimuli, YouTube links and information about pre-testing). All participants watched the clips in the same **condition** order (*Control*, *Policy-Based Partisanship*, *Affective Partisanship*), though the order of the clips (Republican vs. Democrat) within condition was counter-balanced between participants. First, participants watched clips from the *Control* condition in which two internet vloggers expressed their opinions in a humorous way on arguably non-partisan topics. Next, participants watched clips in the *Policy-Based Partisanship* condition. These clips were instructional videos intended to dispel misconceptions on a single topic (i.e., progressivism from one speaker, and immigration's impact on the economy from the other) and were presented in a relatively unemotional style. Last, participants watched clips in the *Affective Partisanship* condition, where the same vloggers from the *Control* condition expressed their opinions about their respective partisan outgroup in an inflammatory and derogatory manner. Participants then watched two more video clips (a short-animated movie and excerpts from a presidential debate), which we do not discuss here. After watching the videos, participants completed two localizer tasks used to functionally identify regions of interest involved in mentalizing (Dodell-Feder et al., 2011) and socio-political reasoning (Bruneau & Saxe, 2010). Following the scanning session, participants completed a questionnaire about their reactions to the clips they had seen in the scanner.

As a manipulation check, ratings of the politically charged clips were collected in the post-scan 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.

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 underlying types of partisanship – policy-based and affective. (**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.

The partisan lens effect

One way in which partisanship might affect the processing of political information is by creating a "lens" where people who share partisanship process information in a similar way. As noted in the introduction, we operationalized this effect in neural terms using intersubject correlation (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. In our analysis, we extracted signal timecourses from regions of interest. For each participant, we then calculated their *ingroup-ISC* as the correlation between their timecourse and the average (across participants) timecourse of all other participants who shared their partisan identification. In like fashion, we calculated their *outgroup-ISC* using the average timecourse of participants of the opposing partisan identification. We then defined the partisan lens effect as the difference between the two similarity measures – that is: *ingroup-ISC – outgroup-ISC*. To test the statistical significance of this measure we used a permutation analysis whereby a null distribution was created by shuffling the party identification for participants and repeating the calculation of the resulting partisan lens effect 10,000 times. We then compared our average (between all participants) observed partisan lens effect to the null distribution to determine statistical significance (Figure 2.A). Our experimental hypotheses went beyond just the significance of the 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 both partisanship conditions and that those effects would be significantly larger than the lens effect in the control condition. Our multiple paths hypothesis predicted we would observe regions where the partisan lens effect is significant in only one partisanship condition, and that this lens effect would be significantly larger than in both the other partisanship condition and the control condition. Therefore, in regions where we found any partisan lens effect, we followed up with paired *t*-tests of the difference in effect size for the partisan lens effect between conditions.

As noted in the introduction, 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 (Figure 2.B). For affective responding processes, we used 6 regions of interest, including the ventral striatum, amygdala (both anatomically defined separately in each hemisphere per participant) and anterior insula (defined in each hemisphere from a group analysis; see Methods for full ROI definitions). For both mentalizing and socio-political reasoning processes, we used localizer tasks to identify 7 functional regions of interest (per group), in areas of the dorsal, middle and ventral portions of the medial prefrontal cortex (D/M/VMPFC), precuneus (PC), bilateral temporoparietal junction (L/RTPJ), and the superior temporal sulcus (STS). Mentalizing regions were identified using an independent mentalizing localizer task (Dodell-Feder et al., 2011). Socio-political reasoning regions were identified using the reasoning task of (Bruneau & Saxe, 2010). For each participant, we defined two functional regions of interest per region as the voxels most sensitive to each of the localizer tasks within an a-priori search-space (e.g., DMPFC, PC, see methods for full description, (Fedorenko et al., 2010)). Since we evaluated our hypotheses at the neural system level but performed each statistical test at the region level, we report results based on Bonferroni-corrected alpha levels to compensate for multiple testing of regions within systems (See Disjunction Testing in Rubin, 2021).

Using the outlined procedure, 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 pre-testing (See methods). We then applied the same tests to clips in our two partisanship conditions to directly test our hypotheses.

No common core for partisanship

Our Common Core hypothesis suggested that a common core set of psychological processes may underlie the partisan lens effect, regardless of the partisanship condition. On this view, at least one region should be observed where the partisan lens effect is shown for both the *Affective Partisanship* and the *Policy-Based Partisanship* 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 (Figure 2.C).

Different neural paths to partisanship

Our Multiple Paths hypothesis suggested that different types of partisanship would rely on 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 partisanship 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 the *Policy-Based Partisanship* 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 regions of interest, including 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 *Affective Partisanship* conditions, while the effect in RTPJ was distinguishable only from *Control* (See figure 2.C, 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 *Affective Partisanship* condition and the effect in the striatum was distinguishable only from the *Control* condition. None of the mentalizing-sensitive regions of interest showed significant evidence of a partisan lens effect in the *Policy-Based Partisanship* condition.

We then tested for regions that demonstrated the partisan lens effect during the *Affective Partisanship* condition only. Here, consistent with the idea that affective partisanship is based in part on affective responding, 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 *Affective Partisanship condition* was distinguishable from the lens effect observed during both the *Control* and *Policy-Based Partisanship* conditions. In addition, a partisan lens effect was observed in three mentalizing regions of interest – 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 Partisanship* conditions. None of the regions sensitive to socio-political reasoning showed a partisan lens effect during the *Affective Partisanship* condition.

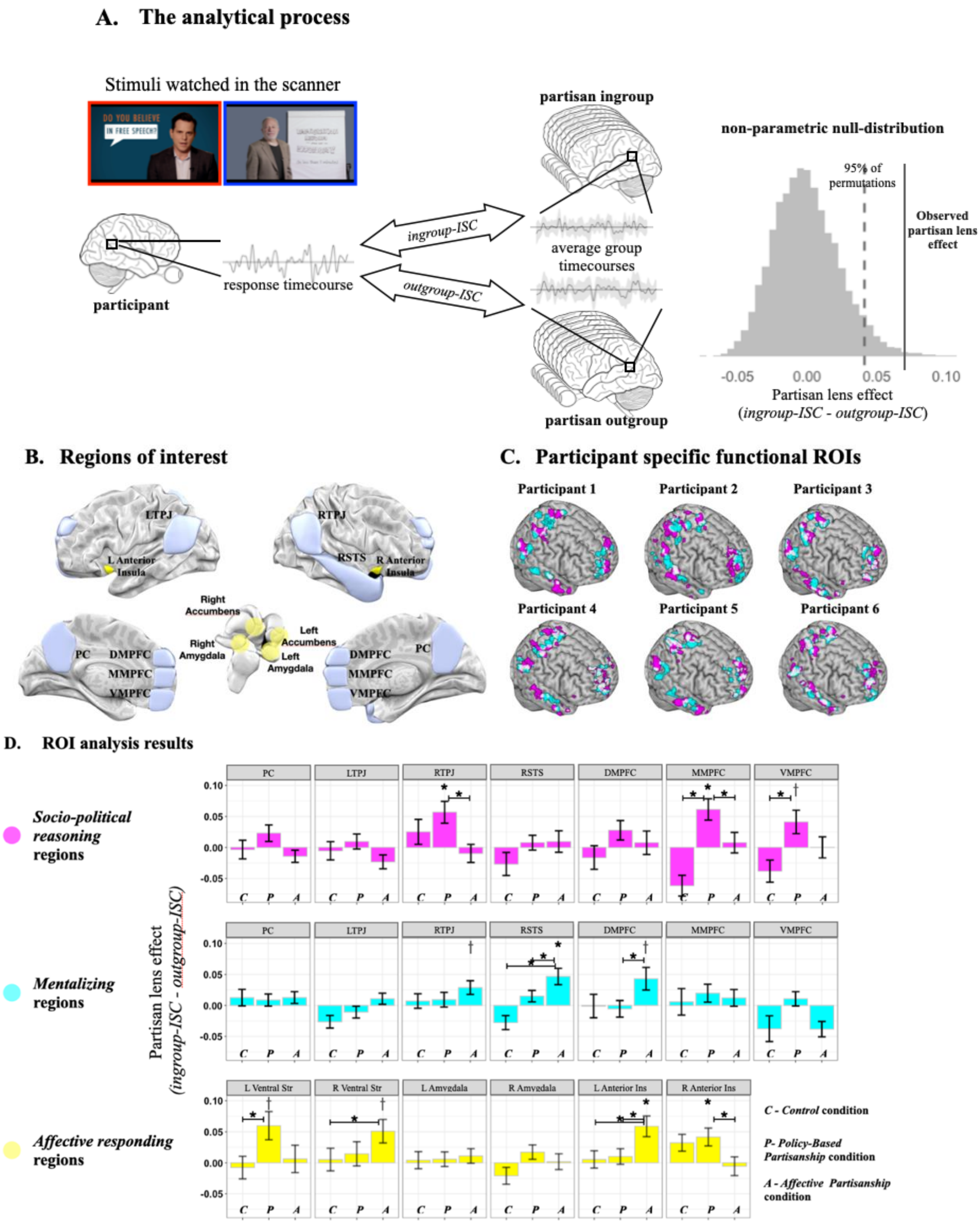
A. The analytical process
Stimuli watched in the scanner
participant
response timecourse
ingroup-ISC
outgroup-ISC
partisan ingroup
average group timecourses
partisan outgroup
non-parametric null-distribution
95% of permutations
Observed partisan lens effect
-0.05
0.00
0.05
0.10
Partisan lens effect
(ingroup-ISC - outgroup-ISC)
B. Regions of interest
LTPJ
L Anterior Insula
RTPJ
RSTS
R Anterior Insula
Right Accumbens
Left Accumbens
Right Amygdala
Left Amygdala
PC
DMPFC
MMPFC
VMPFC
C. Participant specific functional ROIs
Participant 1
Participant 2
Participant 3
Participant 4
Participant 5
Participant 6
D. ROI analysis results
Socio-political reasoning regions
Mentalizing regions
Affective responding regions
Partisan lens effect
(ingroup-ISC - outgroup-ISC)
PC
LTPJ
RTPJ
RSTS
DMPFC
MMPFC
VMPFC
L Ventral Str
R Ventral Str
L Amygdala
R Amygdala
L Anterior Ins
R Anterior Ins
0.10
0.05
0.00
-0.05
C
P
A
C - Control condition
P- Policy-Based Partisanship condition
A - Affective Partisanship condition

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 regions of interest 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 regions of interest 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). † - $p < 0.05$ uncorrected.

Regions showing the partisan lens effect during the *policy-based partisanship* condition			
Region group	**Region**	***policy-based > affective***	***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 *affective partisanship* condition			
Region group	**Region**	***affective > policy-based***	***affective > 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	**$t(54) = 2.213, p = 0.031$**	$t(54) = 1.639, p = 0.107$

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

Discussion

American politics have become increasingly partisan (Finkel et al., 2020; Mason, 2018), characterized by a tendency to view the world through a "partisan lens" (Mason, 2018). The literature distinguishes between two main forms of partisanship: policy-based and affective. 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 those forms of partisanship are supported by similar processes that depend on a common core set of brain regions, or instead depend on different processes and multiple neural pathways. Our results support the latter view: policy-based and affective partisanship create partisan lenses through different neural processes.

Using a theory-driven, region-of-interest-based analysis approach, four key findings emerged. Critically, no single region of interest showed the partisan lens effect under both partisanship conditions, challenging the idea that partisanship is represented in the brain in a single, unified way (Van Bavel & Pereira, 2018). Instead, for the *Policy-Based Partisanship* 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 & Saxe, 2010). For the *Affective Partisanship* condition, we observed the partisan lens effect in a right posterior STS region associated with mentalizing (Mar, 2011; Saxe & Kanwisher, 2003). Finally, though not meeting our strongest criteria for the common core hypothesis, the partisan lens effect was observed under both partisanship 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 affective and policy-based partisanship (Dias & Lelkes, 2022; Orr & Huber, 2020; Sniderman & Stiglitz, 2012), in large part to better understand the dynamics of partisan animosity (Bougher, 2017). Although it is tempting to explain partisanship as the product of a simple, core set of mental processes, this is inconsistent with our findings. Instead, we find that different kinds of partisan content generate 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 partisanship conditions. Our use of neuroimaging to characterize the nature of partisanship 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, partisans interpret both types of content using “lenses” with multiple types of components, some of which are involved in reasoning and mentalizing and some of which are involved in affective processing.

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 the *Affective Partisanship* condition and the right during the *Policy-Based Partisanship* 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 & Pereira, 2018). Altogether, our findings add to the growing body of theory and empirical evidence (Kaplan et al., 2016; Redlawsk, 2006; Westen et al., 2006) 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 the *Policy-Based Partisanship* and the *Affective partisanship* conditions: the effect during the *Policy-Based Partisanship* condition was observed in the MMPFC and weakly in the RTPJ, both identified by the socio-political reasoning localizer, whereas the effect during the *Affective Partisanship* 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 (Baldassano et al., 2018; Tompary & Davachi, 2017), 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 Partisanship* 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 content is about policy ideas. Conversely, regions identified by the mentalizing localizer showed the effect in the *Affective Partisanship* 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 regions of interest 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 (DiNicola et al., 2020; Scholz et al., 2009). 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 presenting 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, however, in at least three meaningful ways. First, 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 & Kalmoe, 2017).

Second, our study directly manipulated and compared different forms of partisan responding. 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 affective partisanship based responding, 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 (Lelkes, 2021; Sniderman & Stiglitz, 2012).

Third, our study used an *a-priori* region of interest 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 (Fedorenko & Kanwisher, 2009; Saxe et al., 2006). This turned out to be critical as data-driven, rather than theory-driven, whole-brain analyses might not have been sensitive to detecting separate affective and policy-based partisanship neural correlates.

More broadly, our study suggests that it is misleading to assume individuals are motivated by one type of partisanship and not the other. Rather, media exposure and political messaging can activate either type of partisan processing within the same individual. This perspective complements and underscores the importance of research into partisan media and its effects (Benedictis-Kessner et al., 2019; Levendusky, 2013; Mutz, 2015; Stroud, 2011), and can inform future research into the influence of partisan media content and behavioral interventions (Broockman & Kalla, 2022; Guess et al., 2021; Nyhan et al., 2023).

Indeed, the results of this study could help inform interventions to bridge the partisan divide. Our findings of different neural bases for policy-based and affective partisanship suggest that interventions could be designed with a specific form of partisanship in mind, targeting the socio-political reasoning regions underlying policy-based partisanship and/or mentalizing regions underlying affective partisanship. For example, policy-specific interventions might be more effective if they alter co-partisan synchrony in socio-political reasoning, though more work is needed to directly test this idea. In like fashion, interventions that target affective partisanship might benefit from a focus on changing co-partisan synchrony in processes associated with mentalizing, such as 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 partisanship 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 partisanship do not rely on a common set of mental processes, we should not necessarily expect an intervention designed to affect policy-based partisanship to influence affective partisanship (or vice versa). As such, changing minds about policy might not change hearts committed to interparty animosity.

In conclusion, in this study, we sought to characterize the neural correlates underlying policy-based and affective partisanship. In independently defined regions of interest, we used inter-subject correlation to measure the partisan lens effect. We found evidence of partisan lenses in the processing of policy-related and affectively-charged partisan content. Significantly, partisan processing during our policy-based partisanship condition was observed in regions identified to be engaged in socio-political reasoning, as well as affective responding. The partisan lens effect was also observed during our affective partisanship condition in regions related to mentalizing and affective responding. Taken together, this suggests that different types of partisanship are supported by a set of related but distinguishable psychological processes.

Materials and Methods

Participants

Seventy-one participants participated in the experiment. 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 was lost altogether; data from another participant was corrupted due to technical errors. Each functional run in which the average Forward 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 non-binary; mean age: 23.6, std: 5.7, range: 18 – 44).

Stimuli and experimental procedure

Stimuli: The stimuli for the main experimental paradigm consisted of 6 short video clips (2:22 – 4:22 minutes) which 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 with two video clips assigned to teach 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, the speakers were YouTube vloggers discussing non-partisan issues. The *Policy-Based Partisanship* condition featured “myth-busters” type clips focused on a single topic. The *Affective Partisanship* condition featured clips by the same vloggers from the *Control* condition, but this time discussing political issues concerning President Obama and the election of President Trump with inflammatory language and spite directed at the “other side”. A full list of the stimuli, conditions and topics is 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. 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. Importantly for the analysis in this paper, participants were 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, as is standard in American public opinion research (e.g., the American National Election Studies, ANES (American National Election Studies (ANES) et al., 2017)).

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. *Control* clips were always first, followed by *Policy-Based Partisanship* clips and then *Affective Partisanship* clips. This order was decided to ensure that participants did not have pre-conceptions 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 *Affective Partisanship* 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) and 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] – 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 two functional runs of a task first introduced in ref.(Bruneau & Saxe, 2010). In the task, participants read short statements and were asked to judge how reasonable those statements were. The statements consisted of two conditions – *Socio-political* and *Generic*[2]. Each run featured 5 trials per condition. Each trial consisted of 10 seconds to read the statement immediately followed by 4 seconds to respond on a 1 (very unreasonable) to 4 (very reasonable) scale. The trials were presented in a fixed block design with 12 seconds of fixation time between blocks (and at the beginning and end of the run) for a total of 272 seconds 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 seconds to read the story immediately followed by 4 seconds to answer the question. Those were presented in a fixed block design with 12 seconds of fixation time between blocks (and at the beginning and end of the run) for a total of 272 seconds. The task was used to independently localize, at the participant level, brain regions using the *Belief* > *Photo* contrast.

Behavioral data analysis

Participants' responses to the political clips from the post-scan questionnaire were analyzed as a manipulation check. First, we scaled the reasonableness measurement (originally 0-100) 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, p. 4; 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 FoV. All functional runs were acquired using a T2* weighted sequence with Multi Band (MB) factor of 3 Simultaneous Multi Slices (SMS) and partial Fourier factor of 0.875;

---

[2] The condition names in the original paper were *Emotional (control)* and *Nonemotional (control)*.

Repetition Time (TR) of 1 sec; Echo Time of 30 ms; Flip Angle of 60º; voxel size of 3.0303 mm on the acquisition plane (AP) and 3 mm slice thickness; Matrix size of 66, resulting in Field of View (FoV) of 200 mm; each volume consisted of 51 slices providing full brain coverage. No fieldmaps 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 was then preprocessed using *fMRIPrep 20.1.0rc3* (Esteban, Blair, et al., 2018; Esteban et al., 2020; Esteban, Markiewicz, et al., 2018). The preprocessing included anatomical and functional data preprocessing (See *fMRIPrep* boilerplate in Supplementary Information for full description). Anatomical data was segmented, parceled and normalized to MNI template (MNI152NLin2009cAsym). Functional data was motion corrected, registered to anatomical and resliced to both the native space and to the normalized MNI space.

Functional localizers analysis

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 (HRF). 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 regions of interest (fROIs) at the individual participant level. The procedure used for picking fROIs followed the outline described in ref.(Blank & 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 Figure 2.B for the search-spaces). The masked maps were then masked again with the participant's grey 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).

Regions of Interest definition

Multiple sets of regions of interest were defined for the main analyses. All the regions were motivated by a theoretical account and represented different hypotheses. However, due to technical constraints, different sets of regions were defined in a different manner. Below are descriptions of the different sets and how they were defined.

Socio-political reasoning regions: The socio-political reasoning localizer (Bruneau & Saxe, 2010) was used to define participant-level fROIs in 7 search-spaces (see *functional localizers analysis* above for full description). The socio-political reasoning localizer 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. Only participants where relevant ROIs could be identified are included in their respective analyses.

Mentalizing regions: The false-belief localizer (Dodell-Feder et al., 2011) was used to define participant level fROIs in 7 search-spaces (see *functional localizers analysis* above for full description). The false-belief localizer was successfully completed by 55 participants. Within those, fROIs of DMPFC, PC, RSTS, LTPJ, 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: Amygdala and ventral striatum were anatomically defined from participants' T1 image. Voxels were identified by *FreeSurfer* segmentation and parcellation during preprocessing stage. For the ventral striatum, we used the *accumbens-area* 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 ref.(Bruneau & 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, Forward displacement (FD), 6 first components of aCompCor, and delta regressors for non-steady state volumes at the beginning of the run and for every time-points 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: (1) Unsmoothed preprocessed time series in native space were modeled to be used with anatomically defined regions of interest (see *Affective-responding regions* above); (2) Unsmoothed preprocessed time series in standard space were modeled to be used with group and functionally defined regions of interest.

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 Hz 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 seconds to offset + 6 seconds). 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 ROIs 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

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. 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 to 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 non-parametric 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. 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). 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).

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., $(\textit{ingroup-ISC} - \textit{outgroup-ISC})_{\textit{Policy-Based Partisanship}} - (\textit{ingroup-ISC} - \textit{outgroup-ISC})_{\textit{Affective Partisanship}}$. For each region, we tested the difference in effect size between partisanship conditions, and between the significant conditions and *Control* condition.

Acknowledgements

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

Competing interests

The authors declare that they have no competing interests.

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 localizers tasks are vignettes from published research that is all well referenced and cited in the manuscript.

References

Abramowitz, A. I. (2022). The Polarized American Electorate: The Rise of Partisan-Ideological Consistency and Its Consequences. *Political Science Quarterly*, *137*(4), 645–674. https://doi.org/10.1002/polq.13388

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12–22. https://doi.org/10.1016/j.electstud.2015.11.001

American National Election Studies (ANES), University Of Michigan, & Stanford University. (2017). *ANES 2016 Time Series Study: Version 2* [dataset]. ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR36824.V2

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), Article 4. https://doi.org/10.1038/nrn1884

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Benedictis-Kessner, J. D., Baum, M. A., Berinsky, A. J., & Yamamoto, T. (2019). Persuading the Enemy: Estimating the Persuasive Effects of Partisan Media with the Preference-Incorporating Choice and Assignment Design. *American Political Science Review*, *113*(4), 902–916. https://doi.org/10.1017/S0003055419000418

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

Bougher, L. D. (2017). The Correlates of Discord: Identity, Issue Alignment, and Political Hostility in Polarized America. *Political Behavior*, *39*(3), 731–762. https://doi.org/10.1007/s11109-016-9377-1

Broockman, D., & Kalla, J. (2022). The impacts of selective partisan media exposure: A field experiment with Fox News viewers. *OSF Preprints*, *1*.

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

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter* (pp. viii, 573). John Wiley & Sons, Ltd.

Canli, T., Desmond, J. E., Zhao, Z., Glover, G., & Gabrieli, J. D. E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *NeuroReport*, *9*(14), 3233.

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex (New York, NY : 1991)*, *23*(3), 739–749. https://doi.org/10.1093/cercor/bhs065

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A., & Meyerand, M. E. (2001). Frequencies Contributing to Functional Connectivity in the Cerebral Cortex in "Resting-state" Data. *American Journal of Neuroradiology*, *22*(7), 1326–1333.

Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*(1), Article 1. https://doi.org/10.1038/sj.mp.4000812

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2010). Three Systems of Insular Functional Connectivity Identified with Cluster Analysis. *Cerebral Cortex*, *21*(7), 1498–1506. https://doi.org/10.1093/cercor/bhq186

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 13995. https://doi.org/10.1038/ncomms13995

Dias, N., & Lelkes, Y. (2022). The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity. *American Journal of Political Science*, *66*(3), 775–790. https://doi.org/10.1111/ajps.12628

DiNicola, L. M., Braga, R. M., & Buckner, R. L. (2020). Parallel distributed networks dissociate episodic and social functions within the individual. *Journal of Neurophysiology*, *123*(3), 1144–1179. https://doi.org/10.1152/jn.00529.2019

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

Downs, A. (1957). *An Economic Theory of Democracy* (1st edition). Harper and Row.

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M., Triantafyllou, C., Gabrieli, J. D. E., & Saxe, R. (2013). Similar Brain Activation during False Belief Tasks in a Large Sample of Adults with and without Autism. *PLoS ONE*, *8*(9), e75468. https://doi.org/10.1371/journal.pone.0075468

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Salo, T., Valabregue, R., Amlien, I. K., Liem, F., … Gorgolewski, K. J. (2018). *FMRIPrep software* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.852659

Esteban, O., Ciric, R., Finc, K., Blair, R. W., Markiewicz, C. J., Moodie, C. A., Kent, J. D., Goncalves, M., DuPre, E., Gomez, D. E. P., Ye, Z., Salo, T., Valabregue, R., Amlien, I. K., Liem, F., Jacoby, N., Stojić, H., Cieslak, M., Urchs, S., … Gorgolewski, K. J. (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature Protocols*, *526*, 1–20. https://doi.org/10.1038/s41596-020-0327-3

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 1–14. https://doi.org/10.1038/s41592-018-0235-4

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitefield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Fedorenko, E., & Kanwisher, N. (2009). Neuroimaging of Language: Why Hasn't a Clearer Picture Emerged? *Language and Linguistics Compass*, *3*(4), 839–865. https://doi.org/10.1111/j.1749-818X.2009.00143.x

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., &

Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715
Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. Yale University Press.
Gillion, D. Q., Ladd, J. M., & Meredith, M. (2020). Party Polarization, Ideological Sorting and the Emergence of the US Partisan Gender Gap. *British Journal of Political Science*, *50*(4), 1217–1243. https://doi.org/10.1017/S0007123418000285
Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 24–29. https://doi.org/10.1038/sdata.2016.44
Greene, S. (2004). Social Identity Theory and Party Identification*. *Social Science Quarterly*, *85*(1), 136–153. https://doi.org/10.1111/j.0038-4941.2004.08501010.x
Guess, A. M., Barberá, P., Munzert, S., & Yang, J. (2021). The consequences of online partisan media. *Proceedings of the National Academy of Sciences*, *118*(14), e2013464118. https://doi.org/10.1073/pnas.2013464118
Haas, I. J., Baker, M. N., & Gonzalez, F. J. (2017). Who Can Deviate from the Party Line? Political Ideology Moderates Evaluation of Incongruent Policy Positions in Insula and Anterior Cingulate Cortex. *Social Justice Research*, *30*(4), 355–380. https://doi.org/10.1007/s11211-017-0295-0
Haas, I. J., Baker, M. N., & Gonzalez, F. J. (2021). Political uncertainty moderates neural evaluation of incongruent policy positions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1822), 20200138. https://doi.org/10.1098/rstb.2020.0138
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity during Natural Vision. *Science*, *303*(5664), 1634–1640. https://doi.org/10.2307/3836448?ref=no-x-route:a5964089433f5df9fd194f2c1a5c7317
Highton, B., & Kam, C. D. (2011). The Long-Term Dynamics of Partisanship and Issue Orientations. *The Journal of Politics*, *73*(1), 202–215. https://doi.org/10.1017/S0022381610000964
Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, *109*(1), 1–17. https://doi.org/10.1017/S0003055414000604
Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145–E9152. https://doi.org/10.1073/pnas.1714471114
Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034
Jefferson, H. (2020). *The Curious Case of Black Conservatives: Construct Validity and the 7-point Liberal-Conservative Scale* (SSRN Scholarly Paper 3602209). https://doi.org/10.2139/ssrn.3602209

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports*, *6*, 39589. https://doi.org/10.1038/srep39589

Kinder, D. R., & Kalmoe, N. P. (2017). *Neither Liberal Nor Conservative: Ideological Innocence in the American Public*. University of Chicago Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, A. H.-Y., Lelkes, Y., Hawkins, C. B., & Theodoridis, A. G. (2022). Negative partisanship is not more prevalent than positive partisanship. *Nature Human Behaviour*, *6*(7), Article 7. https://doi.org/10.1038/s41562-022-01348-0

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1822), 20200143. https://doi.org/10.1098/rstb.2020.0143

Lelkes, Y. (2021). Policy over party: Comparing the effects of candidate ideology and party on affective polarization. *Political Science Research and Methods*, *9*(1), 189–196. https://doi.org/10.1017/psrm.2019.18

Leong, Y. C., Chen, J., Willer, R., & Zaki, J. (2020). Conservative and liberal attitudes drive polarized neural responses to political content. *Proceedings of the National Academy of Sciences*, *117*(44), 27731–27739. https://doi.org/10.1073/pnas.2008530117

Levendusky, M. (2013). *How Partisan Media Polarize America*. University of Chicago Press.

Levitsky, S., & Ziblatt, D. (2019). *How Democracies Die*. Crown.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral Cortex*, *26*(5), 1910–1922. https://doi.org/10.1093/cercor/bhv001

Lupu, N. (2015). Party Polarization and Mass Partisanship: A Comparative Perspective. *Political Behavior*, *37*(2), 331–356. https://doi.org/10.1007/s11109-014-9279-z

Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual Review of Psychology*, *62*(1), 103–134. https://doi.org/10.1146/annurev-psych-120709-145406

Mason, L. (2018). *Uncivil agreement: How politics became our identity*.

Moore-Berg, S. L., Hameiri, B., & Bruneau, E. G. (2020). The prime psychological suspects of toxic political polarization. *Current Opinion in Behavioral Sciences*, *34*, 199–204. https://doi.org/10.1016/j.cobeha.2020.05.001

Muirhead, R., & Rosenblum, N. L. (2020). The Political Theory of Parties and Partisanship: Catching Up. *Annual Review of Political Science*, *23*(1), 95–110. https://doi.org/10.1146/annurev-polisci-041916-020727

Mutz, D. C. (2015). In-Your-Face Politics: The Consequences of Uncivil Media. In *In-Your-Face Politics*. Princeton University Press. https://doi.org/10.1515/9781400865871

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., … Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, *620*(7972), Article 7972. https://doi.org/10.1038/s41586-023-06297-w

Orr, L. V., & Huber, G. A. (2020). The Policy Basis of Measured Partisan Animosity in the United States. *American Journal of Political Science*, *64*(3), 569–586. https://doi.org/10.1111/ajps.12498

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.

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

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a "low road" to "many roads" of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), Article 11. https://doi.org/10.1038/nrn2920

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *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. (2020). The neural bases of argumentative reasoning. *Brain and Language*, *208*, 104827. https://doi.org/10.1016/j.bandl.2020.104827

R Core Team. (2013). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Redlawsk, D. (2006). *Feeling Politics: Emotion in Political Information Processing*. Springer.

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

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews. Neuroscience*, *15*(8), 1–14. https://doi.org/10.1038/nrn3776

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

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*, *4*(3), e4869. https://doi.org/10.1371/journal.pone.0004869

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2020). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293. https://doi.org/10.1037/bul0000303

Shively, W. P. (1979). The Development of Party Identification among Adults: Exploration of a Functional Model. *The American Political Science Review*, *73*(4), 1039–1054. https://doi.org/10.2307/1953988

Sniderman, P. M., & Stiglitz, E. H. (2012). The Reputational Premium: A Theory of Party Identification and Policy Reasoning. In *The Reputational Premium*. Princeton University Press. https://doi.org/10.1515/9781400842551

Stroud, N. J. (2011). *Niche News: The Politics of News Choice*. Oxford University Press, USA.

Tajfel, H., & Turner, J. (1979). *An integrative theory of intergroup conflict* (W. G. Austin & S. worchel, Eds.; pp. 33–37).

Theodoridis, A. G. (2017). Me, Myself, and (I), (D), or (R)? Partisanship and Political Cognition through the Lens of Implicit Identity. *The Journal of Politics*, *79*(4), 1253–1267. https://doi.org/10.1086/692738

Tompary, A., & Davachi, L. (2017). Consolidation Promotes the Emergence of Representational Overlap in the Hippocampus and Medial Prefrontal Cortex. *Neuron*, *96*(1), 228-241.e5. https://doi.org/10.1016/j.neuron.2017.09.005

Twenge, J. M., Honeycutt, N., Prislin, R., & Sherman, R. A. (2016). More Polarized but More Independent: Political Party Identification and Ideological Self-Categorization Among U.S. Adults, College Students, and Late Adolescents, 1970-2015. *Personality and Social Psychology Bulletin*, *42*(10), 1364–1383. https://doi.org/10.1177/0146167216660058

van Baar, J. M., Halpern, D. J., & FeldmanHall, O. (2021). Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proceedings of the National Academy of Sciences*, *118*(20). https://doi.org/10.1073/pnas.2022491118

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

Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An FMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election. *Journal of Cognitive Neuroscience*, *18*(11), 1947–1958. https://doi.org/10.1162/jocn.2006.18.11.1947

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

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