

Vision and Division: Partisan Identity and Visual Processing of Political Imagery

Elena Sirotkina^{*} Olga Gasparyan[†]

Abstract

Contemporary politics is increasingly fought through visuals, yet we know little about how partisans take in what they see. Partisan evaluative divergence may come from visual processing itself or emerge only in interpretation; distinguishing these pathways is essential for understanding how visuals contribute to polarization. We present the first multi-topic eye-tracking study linking gaze to political evaluations in a diverse U.S. sample, measuring visual engagement across four polarizing issues and more than 50,000 participant-image observations. Partisans allocate attention to the same objects within the same images, with no selective attention to attitude-consistent elements. Divergence instead appears in processing depth. Viewers whose position on the depicted issue aligns with their party's position process the whole image more deeply, give more extreme evaluations, and this deeper processing accounts for a meaningful share of the partisan evaluative gap. Political images can therefore sustain polarization under shared exposure by drawing party-congruent viewers into more intensive processing and more extreme judgment.

^{*}Assistant Professor/Faculty Fellow, Center for Data Science, New York University. First and corresponding author: elena.sirotkina@nyu.edu

[†]Assistant Professor, Florida State University.

Introduction

Political images circulate as a primary format of political content online (Lu and Peng 2024; Casas and Williams 2019; Grabe and Bucy 2009). They reach viewers faster than text (Todorov et al. 2005; Thorpe et al. 1996; Potter et al. 2014), persist longer in memory (Graber 1990; Gerber et al. 2011), and sometimes can shift attitudes more effectively than equivalent verbal frames (Powell et al. 2015; Druckman 2003; Ballew and Todorov 2007). This combination of speed, memorability, and persuasive force means that political images may operate on viewers through processing stages that occur before deliberate evaluation begins, stages that behavioral (Balcetis and Dunning 2006) and neural (Leong et al. 2019) research has documented but that political science has not yet connected to the evaluative divergence it routinely observes.

As growing evidence suggests, millions of people scroll through the same political images and arrive at opposite conclusions about what those images show (Webb Williams et al. 2025; Gasparyan and Sirotkina 2025), which means that an image of a migrant family at a border crossing registers as evidence of humanitarian crisis for one viewer and as a security threat for another. Such evaluative divergence itself is among the most replicated findings in political science (Lodge and Taber 2013; Taber and Lodge 2006; Iyengar and Westwood 2015; Mason 2018). What remains almost entirely understudied is whether this divergence originates in how partisans see political imagery or in how they interpret what they have seen.

This distinction matters because if partisans attend to different elements of a political image, the roots of polarization reach into visual perception itself (Balcetis and Dunning 2006; Leong et al. 2019; Van Bavel and Pereira 2018), and what people believe they have seen is already shaped by who they are before any evaluation begins (Xiao et al. 2016; Caruso et al. 2009). If partisans attend to the same elements but process them with different intensity and evaluate them differently, then political identity operates at processing depth and interpretation rather than at perception (Leong et al. 2020; Lodge and Taber 2013). Distinguishing empirically between these possibilities requires tracking what partisans look

at, for how long, and what they conclude, all within the same behavioral stream.

We answer this question with the first large-scale webcam-based eye-tracking study across three waves on a demographically and geographically diverse U.S. sample, covering four topics on which Democrats and Republicans strongly disagree and over 50,000 participant-image observations. Participants used their everyday devices, mobile phones, laptops, and tablets, with fixations identified through the I-VT (Velocity-Threshold Identification) algorithm. This setup approximates how people encounter political images online, and we assess data quality through wave-specific validation and robustness checks to ensure that the results are not driven by device type, calibration error, or low-quality gaze data.

We measure visual engagement at two levels of analysis. At the image level, eye-tracking metrics capture how long and how deeply each participant processes each photograph (see Jenke and Sullivan (2025)). At the object level, we develop a computational pipeline that isolates each participant's fixation hotspots from their heatmap, matches them to persons and objects detected in the image by deep-learning models, and computes the share of each participant's total attention directed at each detected element. This allows us to test whether Democrats and Republicans selectively attend to different objects within the same political scene.

We find that partisans viewing the same political image look at the same elements within it, and their gaze distribution across the scene is identical. The divergence appears instead in image-level processing depth, and this differential processing depth turns out to mediate the relationship between partisan identity and evaluative judgment. Because partisans already attend to the same parts of the scene, the polarizing force of political images enters downstream, during the processing that follows initial visual exposure. We show that partisanship and issue attitudes jointly predict this processing depth, and the deepest engagement occurs when the two align. The deeper engagement, in turn, transmits the effect of partisan identity onto evaluation, which means that processing depth is the pathway through which identical images may produce opposing political judgments.

Taken together, these findings open a new dimension in the study of partisan political cognition. Research on motivated reasoning has established that partisans evaluate identical information differently (Taber and Lodge 2006; Lodge and Taber 2013; Bisgaard and Slothuus 2018), and this paper extends that tradition by tracing the evaluative divergence backward through the processing stream to the attentional stage that precedes it.

Political neuroscience has localized partisan divergence across several neural systems including narrative interpretation and mentalizing regions when partisans watch identical political content (Leong et al. 2020), conflict monitoring during domain-general cognitive tasks (Amodio et al. 2007), and reward and affective valuation circuits during processing of political concepts (Bruin et al. 2023). These studies identify where partisan processing diverges neurally but not through which behavioral steps it reaches evaluative output; the present study measures attention allocation and evaluative judgment in the same participants viewing the same images, which provides a behavioral counterpart to those neural findings.

Moreover, the identity-based model of political belief proposes that partisanship can modulate cognition at multiple dissociable stages from attention and perception through memory and evaluation (Van Bavel and Pereira 2018); this paper provides a direct behavioral test of that proposal by simultaneously capturing two of those stages and estimating the mediating pathway between them.

Together, these contributions offer the first simultaneous measurement of within-scene attention allocation, image-level processing depth, and evaluative judgment in a large diverse sample. By locating where partisan divergence enters the visual processing stream, our study clarifies how political images contribute to polarization and provides a foundation for thinking more precisely about when shared visual exposure can narrow disagreement, when it can deepen it, and how visually driven polarization might be fueled but also curbed.

Partisan Vision, Figuratively and Literally

Decades of research in political science have shown that partisan identity shapes how citizens evaluate political information, biasing what evidence they accept and how they interpret it, and what judgments they ultimately report (Green et al. 2002; Mason 2018; Taber and Lodge 2006; Taber et al. 2009; Kunda 1987; Kunda 1990). That Democrats and Republicans can reach different conclusions from the same political content is now one of the field's most robust findings (Ditto et al. 2019). Yet most existing evidence observes this process at its endpoint, after citizens have already transformed exposure into verbal evaluation. Survey measures therefore tell us that partisan divergence exists, but reveal much less about how it unfolds across the processing stages that precede judgment (Druckman and Lupia 2000; Bruin et al. 2023).

The growing evidence suggests that partisan divergence indeed appears across multiple processing stages, from perceptual encoding and attention allocation through source credibility judgments and deliberate evaluation (Derreumaux et al. 2022; Bullock et al. 2015), which means that citizens may literally *process and see* the political world differently with the upstream stages that produce those divergent evaluations remaining largely unobserved (Balcetis and Dunning 2006; Caruso et al. 2009; Katabi et al. 2023).

Political neuroscience provides further evidence that ideological and partisan differences can reach basic cognitive and physiological systems (Jost et al. 2014; Knutson et al. 2006). People who support socially protective policies show heightened physiological reactivity to threatening stimuli (Oxley et al. 2008), liberals and conservatives produce divergent neural responses to identical political content (Leong et al. 2020), and neural activation during risk-taking decisions tracks ideological orientation (Schreiber et al. 2013). These findings suggest that partisan vision may be literal as well as figurative. Political identity may shape how citizens attend to and process what they see, before they articulate an evaluation.

This possibility is especially important for political images. Images are processed quickly, often holistically, and through attentional mechanisms that operate before deliberate reason-

ing. Divergence between partisans can enter at two distinct stages. At the perceptual stage, attentional allocation together with top-down modulation from prior beliefs and motivational states can produce different perceptual content from the same exposure, so shared visual exposure does not guarantee shared perceptual input. If perceptual input is shared across partisans, attitude divergence enters downstream, through partisan interpretive frames operating during meaning construction or through evaluative judgments grounded in political identity.

A further question is whether partisan identity alone drives visual processing or whether issue-specific attitudes condition its effect. Partisan identity generates group-based motivation, including in-group loyalty, out-group hostility, and automatic affective responses to political objects (Green et al. 2004; Iyengar and Westwood 2015; Lodge and Taber 2005; Lodge and Taber 2013). But party identification does not fully determine citizens' positions on every issue. Individuals vary in how closely their views align with their party, and this alignment may shape how strongly political motivation is activated by issue-relevant content. Evidence from adjacent domains supports this possibility showing, for instance, that pro-environmental attitudes predict longer fixation on climate images (Sollberger et al. 2017) and xenophobic attitudes predict selective neural activation in response to refugee faces (Kesner et al. 2020).

The theoretical question is therefore twofold. First, does partisan divergence in response to political imagery appear in visual attention itself, or does it emerge later, after citizens have seen the same visual content? Second, do partisan identity and issue-specific attitudes contribute separately to visual engagement, or does their alignment intensify or curb the depth with which citizens process political images? This framework links the figurative claim that partisans see politics differently to the literal question of whether they actually look at political images differently.

Where to Look for Partisan Vision? Two Levels of Partisan Visual Engagement

The literature on motivated cognition and visual perception identifies two levels at which partisan identity could shape how people process political images. At the level of the image as a whole, partisans may differ in how long and how intensely they engage with political content. At the level of objects within the image, partisans may allocate attention to different regions of the same scene. Each level corresponds to a distinct mechanism with distinct empirical predictions.

Motivated Seeing versus Motivated Reasoning

Motivated reasoning explains why partisans exposed to identical political information often reach divergent judgments. For visual politics, this account leaves a prior step unresolved. Before citizens can interpret an image, they must visually sample it. If Democrats and Republicans process the same visual input and diverge only afterward, partisan disagreement enters through downstream interpretation. If political motivation changes how long they look or how deeply they encode the image, divergence enters earlier, during visual processing itself.

Research on motivated perception makes this earlier pathway plausible. Motivation can alter what people perceive from ambiguous visual input before deliberate evaluation begins (Balcetis and Dunning 2006). Liberals and conservatives watching identical political videos produce divergent neural responses at moments carrying risk-related and moral-emotional content, and these differences predict subsequent attitude shifts (Leong et al. 2020). Related work shows that attitudes can guide attention and sensory encoding. For instance, climate concern predicts detection of climate-related words (Whitman et al. 2018), identification with police predicts fixation patterns that amplify disagreement in legal judgment (Granot et al. 2014), and motivation can enhance sensory representations independently of response bias

(Leong et al. 2019; Kim et al. 2026).

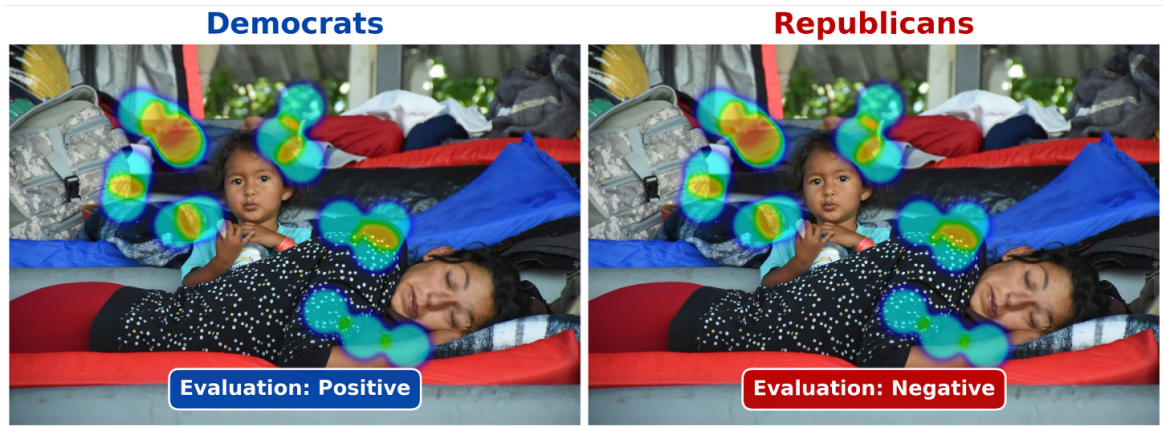
For political images, this motivation is tied to the issue depicted in the image. Partisan identity gives political content group meaning, but images of immigration or gun control also activate viewers' substantive positions on those issues. A Democratic viewer who holds a Democratic-aligned position on immigration, or a Republican viewer who holds a Republican-aligned position on guns, encounters the image through both group identity and issue commitment. Evidence that policy-based and identity-based partisan agreement recruit dissociable neural systems during naturalistic political viewing (Jacoby et al. 2024), and that neural responses to policy incongruence vary with ideological commitment (Haas et al. 2017), suggests that this alignment can intensify visual engagement with issue-relevant content.

These accounts generate the first image-level contrast. Under **H1** (Motivated Reasoning), partisans should visually engage with the same political images in similar ways, with divergence appearing mainly in subsequent evaluations. Under **H2** (Motivated Seeing), partisan identity should predict image-level visual engagement, especially when viewers' issue positions are congruent with their partisan identity, and this visual engagement should account for part of the relationship between partisanship and evaluation.

Figure 1 illustrates two schematic examples of possible viewing patterns. In the motivated reasoning pattern, Democrats and Republicans visually engage with the same political image in similar ways, but their evaluations diverge after viewing. Here, the visual processing channel is shared, and partisan divergence appears at the stage of interpretation and attitude expression. In the motivated seeing pattern, evaluative divergence is accompanied by different image-level engagement where one group fixates longer or more intensely on the same image. Here, partisan motivation enters earlier, during visual processing itself, and differential engagement becomes part of the pathway through which political images produce divergent evaluations.

Figure 1: Motivated Reasoning versus Motivated Seeing

(a) Motivated Reasoning: Similar Viewing, Different Evaluations



(b) Motivated Seeing: Different Viewing, Different Evaluations



Note: The panels illustrate hypothetical viewing patterns. Each row shows a potential distribution of fixation density for Democrats and Republicans viewing the same political image. Warmer colors indicate longer fixation duration. Under motivated reasoning, Democrats and Republicans visually engage with the image similarly, while their evaluations diverge after viewing. Under motivated seeing, evaluative divergence is accompanied by different image-level engagement, consistent with partisan motivation entering during visual processing itself.

Visual Confirmation Bias versus Holistic Processing

H1 and H2 test the motivated perception versus motivated reasoning distinction at the image level, using overall engagement metrics. The same distinction can be tested at a within-image

level by asking whether partisans allocate attention differently across objects within a single scene. H3 and H4 discussed below are the object-level counterparts of H1 and H2. Together, the four hypotheses provide two complementary tests of where partisan motivation enters the visual processing stream.

Within-image visual attention is governed by two competing systems that operate in parallel. Bottom-up processing extracts global scene structure from low-level spectral and spatial features within the first 200 milliseconds of exposure, generating a holistic gist representation that guides subsequent fixations independently of prior beliefs (Oliva and Torralba 2001; Bar 2004). Top-down processing, by contrast, imposes goal-directed templates that bias attention toward schema-consistent objects in a scene, an effect that is involuntary and robust across domains (Rajsic et al. 2015; Rajsic et al. 2018). These two systems are not mutually exclusive, and much of ordinary scene viewing reflects their interaction (Bar 2004; Sheridan and Reingold 2017). The question for partisan visual processing is which system dominates when the scene carries political meaning.

Motivation has been shown to extend its influence down into preconscious visual processing, shaping which objects in a scene reach conscious awareness at all (Balcetis and Dunning 2006; Dunning and Balcetis 2013). When social identity is activated, individuals from different groups selectively attend to distinct regions of identical visual evidence, producing divergent judgments from nominally shared exposure (Qu-Lee and Balcetis 2024).

Confirmation bias in visual search manifests as preferential fixation on template-matching objects, such that attitude-consistent information can be inspected first, longer, and more frequently than disconfirming information, an asymmetry detectable through eye blink rate and fixation duration (Rajsic et al. 2015; Artemenko et al. 2025).

Issue-specific attitudes amplify evaluations of attitude-congruent photographs and attenuate responses to incongruent ones (Krause and Bucy 2018). However, while attitude consistency predicts which political posts people choose to engage with, it does not predict how long they look at them (Sülflow et al. 2019).

Figure 2: Holistic Processing: Equal Allocation Across Areas of Interest

Democrats



(a) Area of Interest: Child

(b) Area of Interest: Crowd

Republicans



(a) Area of Interest: Child

(b) Area of Interest: Crowd

Note: The figure illustrates holistic processing, in which attention is distributed similarly across areas of interest within the same image.

Returning to the immigration example introduced above, immigration imagery in American political communication circulates through two dominant visual frames that activate opposing attitudinal predispositions (Sides et al. 2018; Valentino et al. 2013; Gasparyan and Sirotkina 2025). The first presents migrants as vulnerable individuals in distress (Parrott et al. 2019; Esses et al. 2013), a portrayal likely to resonate with viewers who already hold

pro-immigration positions. The second frames immigration as collective threat (Brader et al. 2008; Esses et al. 2013), typified by large crowds carrying foreign flags (Valentino et al. 2013; Sides et al. 2018), a portrayal likely to resonate with viewers who perceive immigration as a challenge to national sovereignty. Democrats' attitudes on immigration are therefore should be anchored in humanitarian concern and Republicans' attitudes in threat perception (Sides et al. 2018; Tesler 2016; Gasparyan and Sirotkina 2025).

Figure 3: Visual Confirmation Bias: Differential Allocation Across Areas of Interest

Democrats

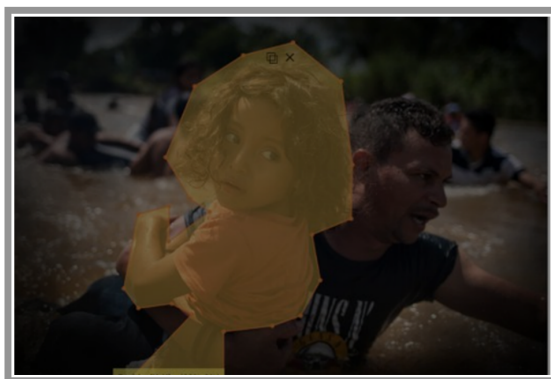


(a) Area of Interest: Child



(b) Area of Interest: Crowd

Republicans



(a) Area of Interest: Child



(b) Area of Interest: Crowd

Note: The figure illustrates visual confirmation bias, in which attention is disproportionately allocated to attitude-consistent elements within the same image.

If visual confirmation bias dominates, each group should allocate disproportionate attention to the object that is consistent with their prior attitudes, Democrats to the child and Republicans to the crowd, and this differential encoding should mediate the partisan attitude gap. Issue importance amplifies the link between gaze and evaluative judgment (Yang et al. 2026; Jacoby et al. 2024), what it implies is that the motivated seeing pathway is activated most strongly when partisan identity and issue position are aligned, producing a conditional rather than uniform effect of partisanship on visual engagement.

We therefore advance two hypotheses at the within-image level: Under **H3** (Holistic Processing), Democrats and Republicans allocate fixation time equally across areas of interest within immigration images, with no partisan difference in attention to the child versus crowd regions (see Figure 2). Under **H4** (Visual Confirmation Bias), Democrats fixate longer on attitude-consistent vulnerable figures while Republicans fixate longer on attitude-consistent crowd regions, and these attentional asymmetries mediate partisan differences in image evaluation (see Figure 3).

Empirical Strategy and Data

Study Waves

To test these hypotheses, we conducted a three-wave webcam-based eye-tracking study (see e.g., Jenke and Sullivan (2025)) on multiple crowdsourcing platforms. The study covers four political issues on which Democrats and Republicans hold sharply different attitudes. Participants completed the task on their own devices, including laptops, tablets, and smartphones, which allowed us to observe visual engagement in a setting closer to how people ordinarily encounter political content online (Simmelmann and Weigelt 2017).¹ Table 1 summarizes the design of each wave.

¹Webcam-based eye-tracking improves ecological validity but reduces some measurement precision; we address this limitation through the quality-filtering procedures described in Appendix A.1, Overlap resolution

Table 1: Data Overview

Wave	Period	N	Obs	Images	Platform	Devices	Primary purpose
1	Jun. 2023	1,082	17,254	16	Lucid (Cint)	Desktop, tablet, smart-phone	Baseline partisan visual engagement with immigration imagery across three device types
2	Mar. 2024	924	28,425	31	Prolific	Desktop, tablet	Replication on a higher-quality panel, with the image set expanded to 31 stimuli
3	Aug. 2025	200	4,756	48	Prolific	Desktop, tablet	Multi-topic extension across four polarizing issues (immigration, gun control, January 6, LGBT rights), with each participant viewing 12 images on each of two assigned topics

Note: N refers to unique participants in the analytic sample for each wave. Obs refers to participant $\times$ image observations. Images counts unique stimuli shown in each wave (Wave 3 covers 12 images on each of four topics). The full analytic sample contains 50,435 observations from 2,206 unique participants. All waves recruited equal quotas of self-identified Democrats and Republicans.

Waves 1 and 2 focused on immigration. Wave 1 was fielded on Lucid Theorem (Cint) in June 2023 and recruited participants across desktop, tablet, and smartphone devices. Wave 2 replicated the core design on Prolific in March 2024 and expanded the image set from 16 to 31 stimuli. Both waves presented images depicting diverse visual representations of immigration sourced from the X (formerly Twitter) feeds and Getty Image archives of major US news outlets, so that stimuli reflected authentic media environments without explicit partisan cues such as party logos or color schemes.² Wave 3, fielded in August 2025, extended the paradigm to a multi-topic design covering four polarizing issues, immigration, gun control, the January 6 Capitol attack, and LGBT rights, each of which generates substantial partisan disagreement (Mason 2018; Sides et al. 2018) and has a well-developed visual media footprint. Each Wave 3 participant viewed 12 images on each of two assigned topics.

All three waves recruited equal quotas of self-identified Democrats and Republicans. The combined sample was geographically and demographically diverse, representing all 50 U.S. states, ranging in age from 18 to over 85. Participant-level summary statistics appear in and ghost filtering section.

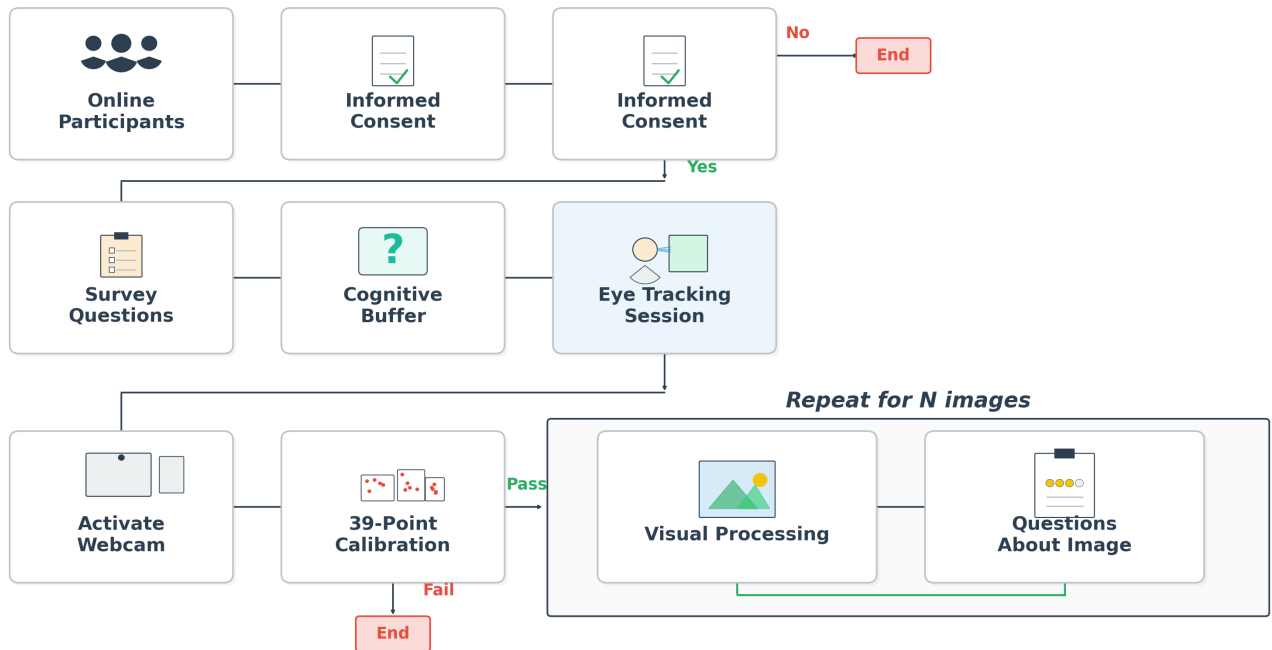
²In waves 1 and 2, participants were informed that the images depicted immigration.

appendix Tables A.1 and A.2, and the comparison to U.S. Census benchmarks in Table A.3.

Study Pipeline

Figure 4 illustrates the study pipeline. Participants were recruited through the respective crowdsourcing platform, screened for U.S. residency and age over 18, and provided two separate informed consents.³

Figure 4: Study Pipeline from Participant Recruitment to Image Viewing and Attitude Measurement



Note: The figure illustrates the study pipeline from participant recruitment through consent, Qualtrics survey, eye-tracking calibration, image viewing task, and attitude measurement.

Participants completed a Qualtrics survey on social demographics, political orientation, and attitudes toward political issues including immigration, abortion, and same-sex marriage

³The first consent covered the study and participation requirements. The second covered GDPR compliance for the eye-tracking component, informing participants that no video, images, or audio were recorded or stored, and that only predicted gaze coordinates were retained as text data.

(see Appendix B), followed by a non-political open-ended prompt that served as a cognitive buffer between the political survey and the eye-tracking session.⁴ They were then redirected to the eye-tracking platform for a one-minute, 39-point calibration task. Images were displayed in a rotated order and one at a time for exactly five seconds,⁵ followed by a one-second black screen. After each image, participants rated their attitudes toward the subjects portrayed on a 7-point scale from extremely negative to extremely positive.

Eye-Tracking Phase

Eye-tracking data were collected remotely using participants' built-in webcams. The platform processes video frames through a two-stage machine learning pipeline running entirely within the participant's web browser. A deep neural network first detects facial landmarks and isolates pupil and eye corner positions; a regression model then maps these features to predicted gaze coordinates, calibrated individually for each participant, and achieves high spatial accuracy on different device displays (Papoutsaki et al. 2016). Raw gaze streams are denoised with a moving median filter; the platform then segments them into fixations using a velocity-threshold algorithm.⁶

⁴Participants answered how they would spend lottery winnings. This reduced the risk that political attitudes activated during the survey would carry over into the image viewing task and artificially inflate partisan differences in visual engagement.

⁵An average typical exposure duration (Holmqvist et al. 2011; Maughan et al. 2007).

⁶The velocity-threshold identification (I-VT) algorithm classifies gaze samples as fixations or saccades based on eye movement speed. Consecutive samples below a predefined velocity threshold are grouped into a fixation, faster movements are treated as saccades. Only gaze point coordinates are stored as text data, with no video or audio recorded, in compliance with privacy regulations (Sammelmann and Weigelt 2017; Papoutsaki et al. 2016). Image-level measures (AFD, MFD, TFD) are computed from these platform-produced fixations, with points defined as periods of gaze stability lasting at least 60-80 milliseconds (Holmqvist et al. 2011) and durations exceeding 800 ms excluded as momentary disengagement (Negi and Mitra 2020), following standards in the eye-tracking literature (Jenke and Sullivan 2025). The raw stream is retained separately and re-segmented with a dispersion-threshold algorithm in the object-level pipeline (Stage C below).

Table 2: Image-level eye-tracking outcome measures

Average fixation duration (AFD)	Median fixation duration (MFD)	Total fixation duration (TFD)
Mean duration of individual fixations (ms). Capped at 800 ms.	Median duration of individual fixations (ms). Capped at 800 ms.	Sum of all fixation durations on an image (ms).
Cognitive interpretation Longer fixations signal deeper semantic encoding and greater cognitive load (Just and Carpenter 1980; Rayner 1998). The mean is sensitive to outlier fixations in the right tail of the distribution (Manor and Gordon 2003).	Captures the same per-fixation processing depth as AFD but is robust to the right-skewed distribution characteristic of fixation duration data (Pannasch et al. 2008). Tracks the central component of the distribution rather than the tail (see a similar argument in (Staub et al. 2010)).	Overall time budget allocated to an image. Conflates depth with return-visit frequency: $TFD = AFD \times N_{\text{fix}}$ (Holmqvist et al. 2011; Fitts et al. 1950). Can remain stable when depth and frequency change in opposite directions (Liu et al. 2022).
Prior use in political science Primary DV in political eye-tracking research (Jenke and Sullivan 2025). Closely related dwell-time measures used in partisan attention studies (Dodd et al. 2012; Schmuck et al. 2020).	Recommended over the mean for skewed fixation data (Manor and Gordon 2003; Pannasch et al. 2008). Not previously used in political eye-tracking research.	Common DV in political communication research (Schmuck et al. 2020; Vraga et al. 2016). TFD on political ads predicted selective exposure ($\eta_p^2 = .19$; (Marquart et al. 2016)). TFD on ballot text predicted aggregate voting outcomes (Coronel et al. 2021).

Note. All measures computed per participant $\times$ image. AFD and MFD index per-fixation processing depth. TFD indexes cumulative engagement.

Image-Level Attention Measurement

We measure visual engagement at the image level with three eye-tracking outcomes. Fixation duration has been the primary dependent variable in political eye-tracking research (Jenke and Sullivan 2025; Dodd et al. 2012; Schmuck et al. 2020), grounded in the eye-mind assumption that the duration of a fixation reflects the depth of cognitive processing at the fixated location (Just and Carpenter 1980; Rayner 1998). Fixation duration distributions are characteristically right-skewed (Manor and Gordon 2003), and mean and median can diverge substantially when occasional long fixations inflate the right tail (Negi and Mitra 2020; Galley et al. 2015).

We therefore report both average fixation duration (AFD) and median fixation duration (MFD), with MFD providing the more robust and distributional-assumption-free estimate of per-fixation processing depth (Manor and Gordon 2003). Total fixation duration (TFD) captures the cumulative time budget allocated to each image. Because TFD is the product of fixation count and average duration, it conflates per-fixation processing depth with the frequency of return visits (Holmqvist et al. 2011; Fitts et al. 1950).⁷

Object-Level Attention Measurement

Testing whether partisans selectively attend to attitude-consistent elements within a single image (H4) or process all elements equally and diverge only in evaluation (H3) requires decomposing each participant’s gaze record into object-level dwell measures. We developed a four-stage computational pipeline that proceeds automatically from raw images and denoised gaze samples to participant-level dwell shares.

Appendix A.1 details the full pipeline and the theoretical justification for each object

⁷Fixation count and fixation duration are empirically uncorrelated (Sharafi et al. 2015), and higher cognitive load produces longer but fewer fixations while higher perceptual load produces shorter but more frequent fixations (Liu et al. 2022), which means that TFD can remain stable under conditions that produce opposite changes in AFD. Reporting all three measures allows us to separate whether partisanship shapes the depth of processing per fixation (AFD, MFD), the cumulative engagement with the stimulus (TFD), or both. Table 2 summarizes each measure.

category and frame aggregation, drawn from political-communication research on the visual elements most polarizing for each topic. Here we describe the logic of the pipeline in Algorithm 1 with reference to Figure 5, which illustrates all four stages on a single immigration photograph.

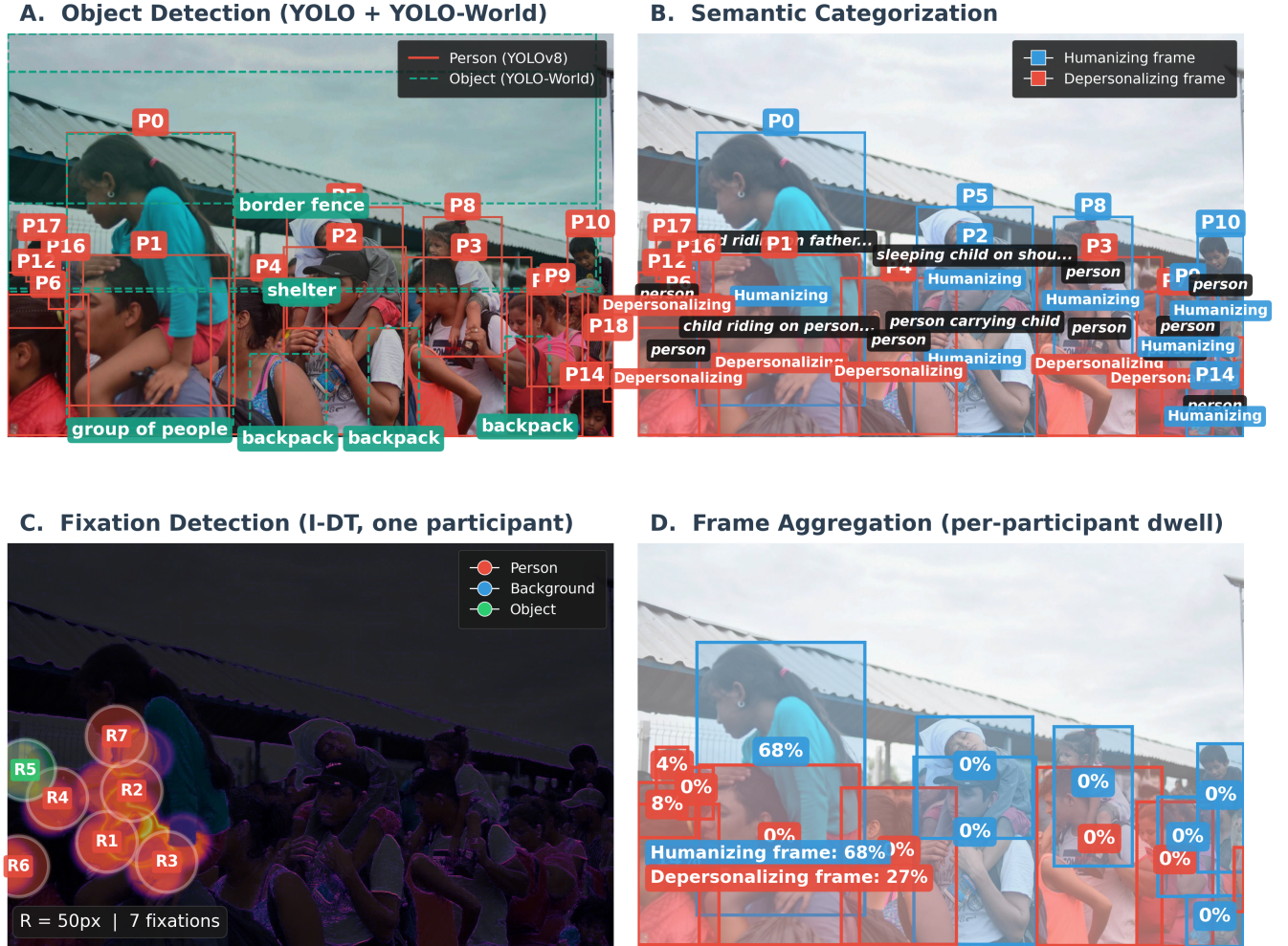
Algorithm 1 Computing Object-Level Dwell Shares

```

1: Stages A and B (image-level)
2: for each image  $j = 1, \dots, J$  do
3:   Detect persons and topic-relevant objects  $\rightarrow$  bounding boxes  $\{B_{j1}, \dots, B_{jK_j}\}$ 
4:   Assign each detection a category  $c_{jk}$  from a closed list (child, crowd, weapon, partisan flag, ...)
5: end for
6: Stage C (gaze-level)
7: for each participant  $i = 1, \dots, N$  and each image  $j$  they viewed do
8:   Extract fixations  $r = 1, \dots, R_{ij}$  from the gaze stream via the dispersion-threshold algorithm
9: end for
10: Stage D (linking and frame aggregation)
11: for each  $(i, j)$  pair do
12:   Assign each fixation  $r$  to the smallest  $B_{jk}$  containing its center (else uncategorized)
13:    $dwell_{ijk} = \sum_{r \in B_{jk}} duration_{ijr}$ 
14: end for
15: for each topic  $t$  with frames  $F_t = \{f_{t,1}, f_{t,2}\}$  (Appendix Table A.26) do
16:   for each  $(i, j)$  on topic  $t$  and each frame  $f \in F_t$  do
17:      $frame\_dwell_{ijf} = \sum_{k: c_{jk} \in f} dwell_{ijk}$ 
18:      $dwell\_share_{ijf} = frame\_dwell_{ijf} / \sum_{f' \in F_t} frame\_dwell_{ijf'}$ 
19:   end for
20: end for

```

Figure 5: Object-Level Attention Measurement Pipeline



Note. Panels A–D show four views of the same hypothetical immigration photograph at successive stages of the pipeline. Panel A (Stage A) marks every detection, using solid red boxes for persons returned by YOLOv8 and dashed green boxes for non-person topic objects returned by YOLO-World; anything outside every box is treated as background. Panel B (Stage B) recolors the same detections by their Gemini-assigned category, blue for the humanizing frame and red for the depersonalizing frame. Panel C (Stage C) overlays one participant’s fixations on the image, with circle area proportional to fixation duration. Panel D (Stage D) labels each detection with that participant’s dwell share, and the 68% and 27% totals at the bottom sum these shares within the humanizing and depersonalizing frames defined in Appendix Table A.26.

Stage A: Detecting persons and objects. Before we can ask where participants looked, we need to know what each image contains, both the people in the scene and the topic-specific objects that carry political meaning (border fences in immigration, partisan flags in January 6,

firearms in gun imagery, pride symbols in LGBT). YOLOv8 (Jocher et al. 2023), a computer-vision algorithm, scans an image and identifies visible persons and draws a box around each detection. YOLO-World (Cheng et al. 2024), a related algorithm that can search for objects named in ordinary text prompts, then locates topic objects such as firearms, flags, border walls, tactical gear, etc. Gemini⁸ (Gemini Team, Google 2024) flags people and objects YOLOs may have missed and checks whether each YOLO-World crop actually matches its proposed label.⁹ Panel A of Figure 5 shows the result.

Stage B: Semantic labeling. Once each object has a descriptive label, its political meaning still depends on the surrounding scene. A red baseball cap reads as MAGA in January 6 imagery and as an ordinary hat at a gun range; tactical gear codes as police if a badge is visible on the same person and as militant otherwise. We therefore hand every detection to Gemini, which sees the full image with the Stage A bounding boxes and labels overlaid and assigns each detection to a category from a fixed closed list of theory-motivated categories.¹⁰ Panel B of Figure 5 shows the assignments.

Stage C: Extracting fixations for object-level analysis. Stage C moves from the image side of the data to the participant side. Image-level outcomes (AFD, MFD, TFD) use the platform’s I-VT fixations from the Eye-Tracking Phase above. For object-level attribution we re-segment the same raw gaze stream with a dispersion-threshold algorithm¹¹ (Salvucci and Goldberg 2000), calibrated to the lower spatial accuracy of webcam tracking.¹² The output is

⁸Gemini is a multimodal AI model that can inspect images and respond to written instructions about what is depicted in them.

⁹The rectangular region around a detected person or object is called a bounding box. YOLOv8 is used here for person detection. YOLO-World is for objects and, e.g., a crowd. Gemini serves as a review layer by checking missed detections and mislabeled object crops against the image itself. Full pipeline and thresholds appear in Appendix A.1.

¹⁰The closed list, topic-conditional rules, and full definitions appear in Appendix Table A.25.

¹¹The algorithm groups nearby gaze points into a fixation when the eyes remain spatially stable for a short period of time. Rapid movements between stable periods are treated as saccades.

¹²Fixations alternate with saccades, the rapid jumps during which little information is acquired (Just and Carpenter 1980; Rayner 1998). The dispersion-threshold algorithm groups consecutive samples that stay clustered within a small region of the screen; closed windows of maximum 100–2000 ms are kept as fixations. Webcam offsets are on the order of four degrees of visual angle (Kaduk et al. 2024; Yang and Krajbich 2021),

a discrete list of fixations, the short intervals during which the eyes hold still on one location and visual information is taken in. Fixation duration indexes the cognitive processing at the point of gaze. Panel C of Figure 5 shows the fixation list for one participant and image.

Stage D: Attention assignment and frame aggregation. With objects categorized in Stage B and fixations identified in Stage C, Stage D ties the two sides together. Each fixation is assigned to the smallest bounding box whose extent contains its center, so a fixation on a child held by an adult counts toward the child and a fixation on a face inside a crowd region counts toward that individual person. Per-object dwell sums the durations of the fixations assigned to that object.¹³ The Stage B categories are then aggregated into topical visual frames explained in Appendix Table A.26, and each frame's dwell share is its dwell over the participant's total categorized dwell on the trial.¹⁴ Panel D of Figure 5 shows the per-detection dwell shares for one participant on one image.

The dependent variable for H3 and H4 is each participant's proportional dwell on each frame, compared across Democrats and Republicans.

Results

1. Partisan Identity and Issue Attitudes Jointly Shape Visual Processing

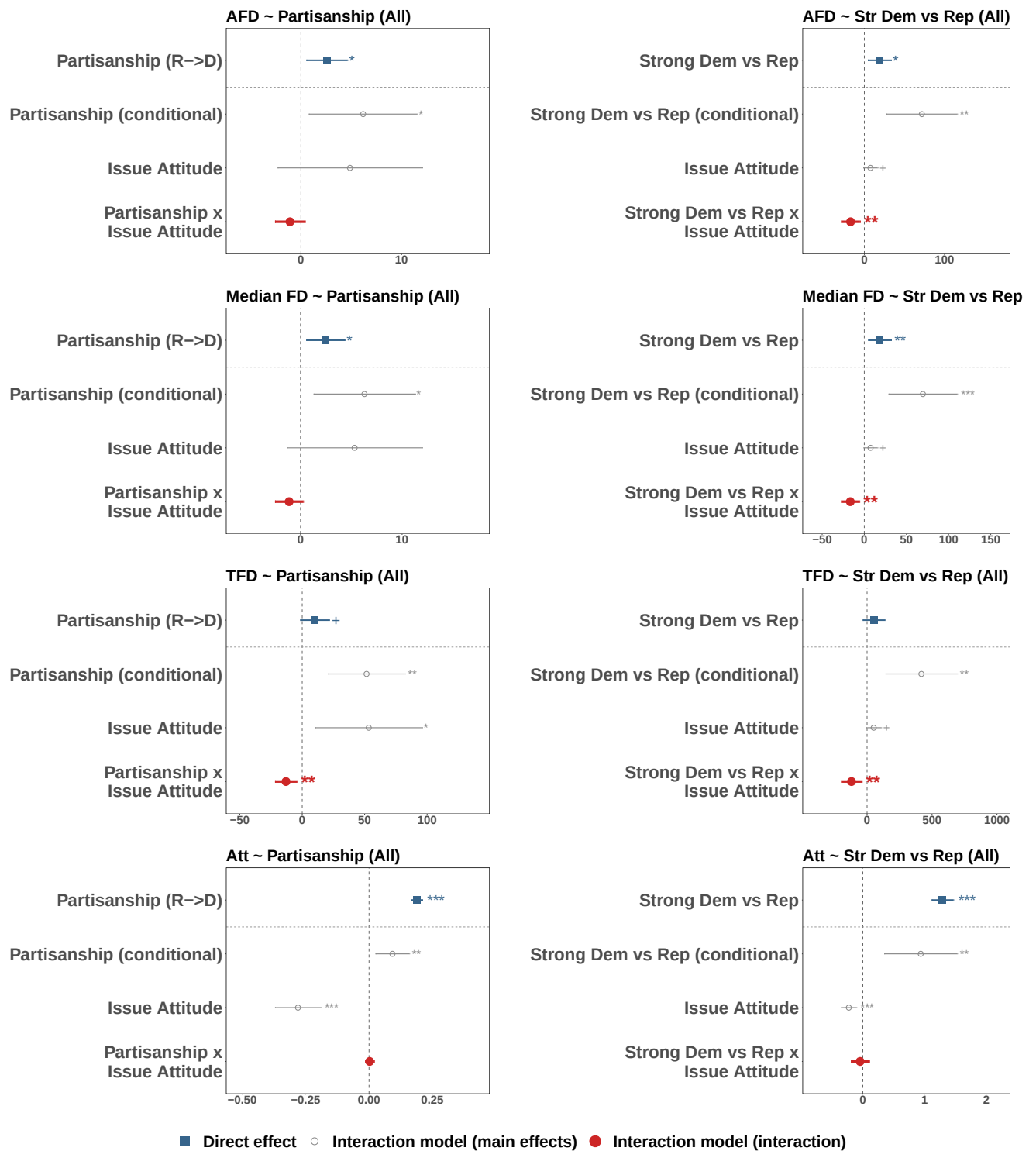
We test H1 (Motivated Reasoning) against H2 (Motivated Seeing) at the image level using three eye-tracking measures. Average fixation duration (AFD) and median fixation duration (MFD) index per-fixation processing depth. Total fixation duration (TFD) captures cumulative viewing time. All models include controls for age, gender, education, income, and race,

which precludes fine scan-path analysis. Full thresholds in Appendix A.1.

¹³Topic framings were defined by linking theory to image content. We first identified theoretical accounts that specify where Democrats and Republicans should diverge in their reading of political images (see Table A.26 in Appendix). We then used Gemini to describe what was depicted in each image. Those descriptions were mapped back to the theoretical accounts, producing categories that cover most of the identified objects, people, and visible cues.

¹⁴The smallest-box rule follows standard area-of-interest guidance (Holmqvist et al. 2011; Hessels et al. 2016); the proportional dwell share controls for individual differences in total viewing time and calibration quality.

Figure 6: Coefficient Estimates for Partisanship Effects on Visual Engagement and Attitudes



Note: Blue markers show direct-model partisanship coefficients with individual controls; gray markers show conditional main effects from the interaction model; red markers show partisanship $\times$ issue-attitude interactions. All models include image fixed effects and participant-clustered standard errors; 95% confidence intervals are shown.

with device type and survey wave indicators, image fixed effects, and standard errors clustered at the participant level. Partisanship is measured on a seven-point scale (1 = Strong Republican, 7 = Strong Democrat). The strong partisan comparison uses a binary variable restricted to Strong Republicans and Strong Democrats only (0 = Strong Republican, 1 = Strong Democrat), with all other respondents excluded from this subset analysis. Issue position is the participant's self-reported attitude on the depicted topic.

In the immigration waves (Waves 1 and 2, $N_{\text{obs}} = 45,679$, $N_{\text{participants}} = 2,006$), each unit toward Democrat is associated with approximately 2.8 ms of additional MFD ($b = 2.789$, $P = 0.007$) and 2.9 ms of AFD ($b = 2.919$, $P = 0.008$). Among strong partisans, the gap reaches 22 ms ($P = 0.003$). These effects persist after controlling for immigration attitudes, indicating that the visual engagement gap is not reducible to pre-existing policy positions. At the image level, we therefore reject H1 as we find evidence in favor of motivated seeing. Full estimates for all three eye-tracking outcomes appear in appendix Tables [A.4–A.11](#).

The interaction between partisanship and issue position is significant for MFD ($P = 0.032$) and AFD ($P = 0.035$), and strengthens among strong partisans ($P < 0.001$). Participants whose issue position aligns with their partisan group fixate longer on imagery depicting that issue. TFD shows no direct partisan effect but a significant interaction for strong partisans ($P = 0.001$), because the increase in per-fixation depth among Democrats is offset by a slight decrease in fixation count.

Pooling across all four topics ($N_{\text{obs}} = 50,435$, $N_{\text{participants}} = 2,206$), the direct partisan effect on MFD survives ($b = 2.468$, $P = 0.011$; strong partisans: $b = 18.206$, $P = 0.009$) but is attenuated. This attenuation is structural meaning that the sign of the per-topic partisan coefficient is positive for immigration (Democrats fixate longer) and negative for gun control, January 6, and LGBT (Republicans fixate longer), and opposite slopes partially cancel in a pooled model with a single coefficient. The interaction captures what the direct effect does not (strong partisans: $b = -16.604$, $P = 0.003$), which confirms that partisan visual engagement is activated in a direction determined by the alignment between the viewer's group identity

and position on the depicted issue (Figure 6; per-topic estimates in appendix Table A.12).

The evidence supports a refined H2. Partisanship is associated with per-fixation processing depth, but the direction of the effect depends on the viewer's position on the specific issue depicted in the image. The image-level partisan effect is robust to alternative fixation-duration thresholds, device subsets, calibration-quality cutoffs, alternative political predictors, partisan subgroups, and per-wave estimation (appendix Tables A.19–A.24).

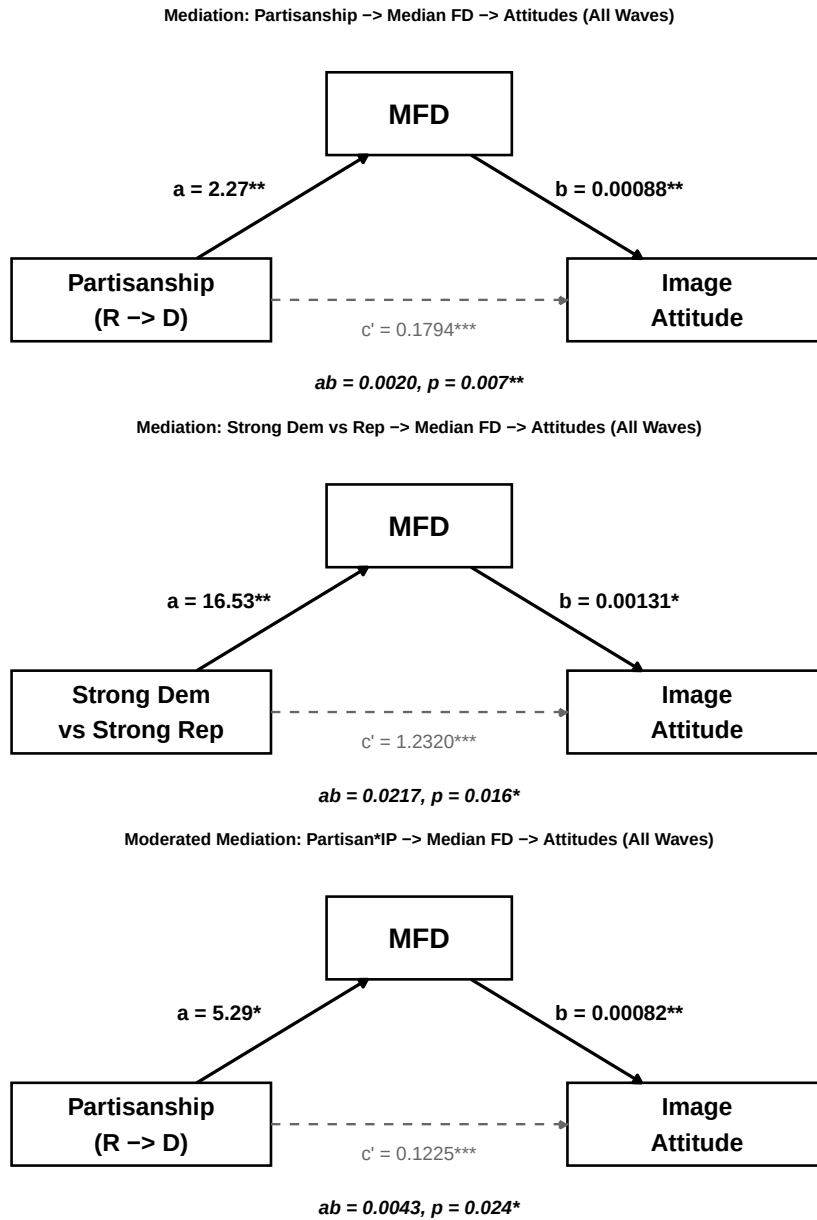
2. Visual Processing Mediates the Association Between Partisanship and Attitudes

H2 makes two predictions. The first, on image-level engagement, was tested in Section 1. The second is that visual engagement transmits part of the partisan-attitude association, and that this transmission strengthens when the viewer's issue position aligns with the party's position on the depicted topic. We therefore test the second prediction with a mediation analysis followed by a moderated mediation that captures the conditional pathway H2 specifies. Both per-fixation depth measures capture per-fixation processing depth, the construct on which H2 rests. The fixation duration distribution is right-skewed, which exposes the mean to outlier fixations, and we therefore use median fixation duration (MFD) as the primary mediator and report mean fixation duration (AFD) as robustness in the appendix.¹⁵

The simple mediation specification provides the baseline test. Across all waves ($N = 2,152$ participants), path $a = 2.27$ ($P = 0.009$), path $b = 0.00088$ ($P = 0.002$), and the indirect association is $ab = 0.0020$ (bootstrap $P = 0.007$), with 1.1% of the partisanship-attitude association passing through MFD (Figure 7, top panel).

¹⁵Models aggregate to the participant level because partisanship is a between-person attribute, which matches the level at which the predictor varies and reduces within-person noise across 16–48 images per participant. Standard errors are heteroskedasticity-robust (HC1); bootstrap confidence intervals resample participants (2,000 iterations). All models include the same demographic controls and wave indicators as the regression specifications.

Figure 7: Mediation Path Diagrams for Partisan Visual Processing



Note: Top panel shows the simple mediation with partisanship (7-point scale) as the independent variable across all waves. Middle panel uses the strong Democrat vs. strong Republican comparison, also across all waves. Bottom panel shows the moderated mediation across all waves, where the association between partisanship and MFD depends on the participant's issue attitude. Coefficients from person-level OLS with HC1 robust standard errors. Significance from 2,000-iteration participant bootstrap. All models include demographic controls and wave indicators. + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Restricting the sample to strong partisans ($N = 809$, Figure 7, middle panel) yields a larger indirect association of $ab = 0.022$ (bootstrap $P = 0.016$, 1.7% mediated), consistent with stronger identity-driven engagement at the extremes of the partisan scale.

The moderated specification provides the direct test of H2's conditional pathway. Allowing the partisanship-MFD association to depend on issue position across all waves ($N = 2,152$ participants) produces path $a = 5.29$ ($P = 0.023$) and an indirect association $ab = 0.0043$ (bootstrap $P = 0.024$), with the mediated share rising to 3.4% (Figure 7, bottom panel). AFD as mediator produces a comparable estimate ($ab = 0.0044$, bootstrap $P = 0.030$; Appendix Tables A.13 and A.14). Restricting the moderated specification to the immigration waves produces a larger indirect association ($ab = 0.0052$, bootstrap $P = 0.020$, 4.3% mediated; Appendix Table A.14).

So we find that the mediation effect is just 1–2% (4.3% in the moderated specification) - is it large? Fixation duration is not a survey measure. It captures the depth of cognitive processing at the level of the oculomotor system (Just and Carpenter 1980), a layer of information processing that operates automatically, without conscious control, and that is inaccessible to self-report instruments. And research across multiple domains demonstrates that visual attention differences of about this scale predict consequential behavioral outcomes. For instance, across 21 experiments, exogenously manipulated gaze shifted choice probability by three to four percentage points per decision (Bhatnagar and Orquin 2022). Fixation durations recorded from 120 laboratory participants predicted voting outcomes across 64 ballot measures representing 137.6 million actual votes (Coronel et al. 2021). Snap impressions formed from 100-millisecond face exposures predicted 60% of real gubernatorial election winners (Ballew and Todorov 2007). The present study captures a single viewing session, but citizens encounter political images daily (Farkas and Bene 2021; Casas and Williams 2019), which means that a processing pathway activating automatically on each encounter has ample opportunity to accumulate (Coppock et al. 2020; Gerber et al. 2011).

3. Partisans Allocate Gaze Identically Across Objects but Produce Divergent Evaluations

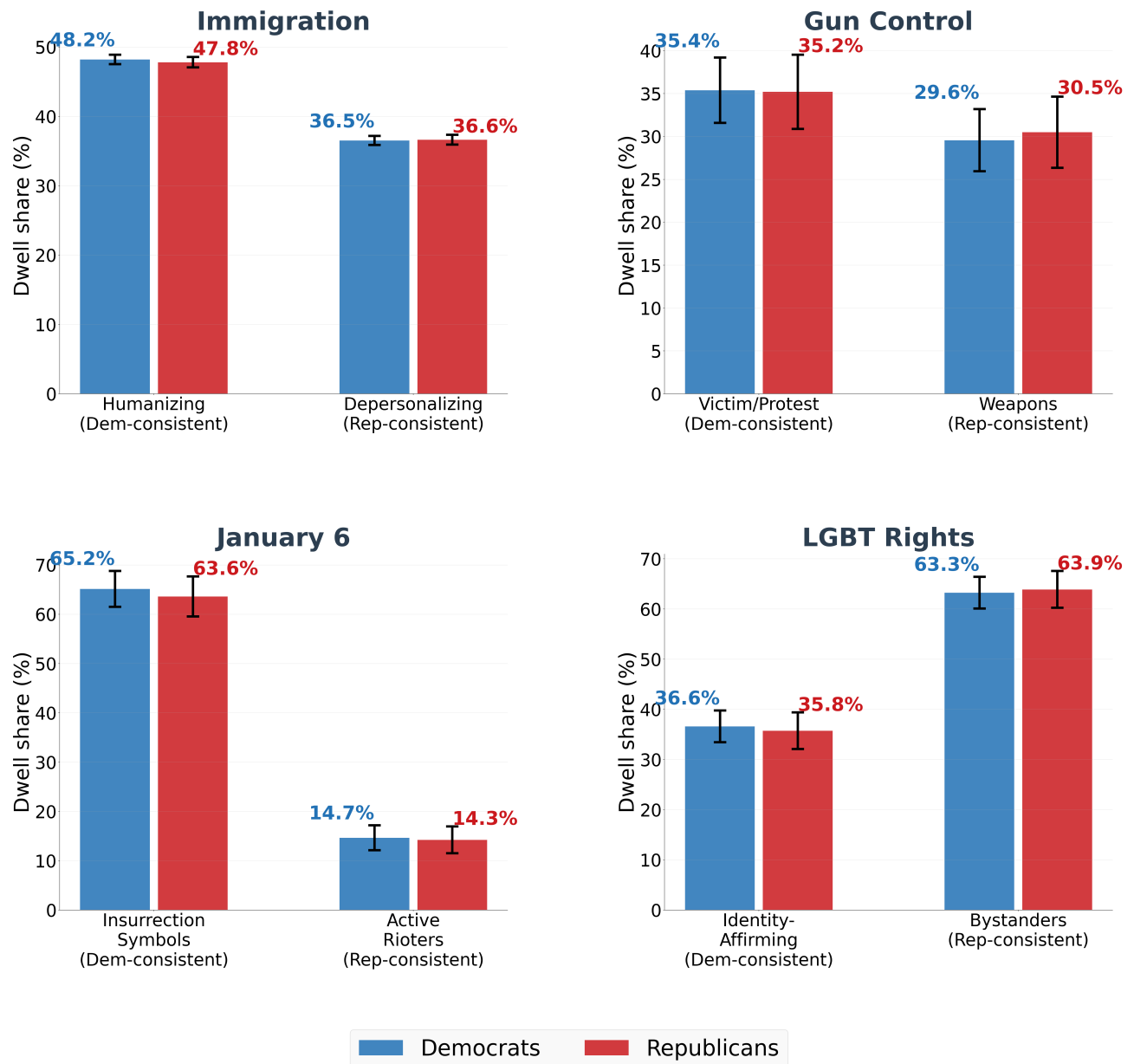
The image-level analyses treat each image as a unitary stimulus. Testing whether partisans selectively direct their gaze toward attitude-consistent objects within a single image (H4, visual confirmation bias) or allocate attention equally across objects (H3, holistic processing) requires decomposing gaze streams into object-level dwell times. Using the semantic categories and visual frames described in the Empirical Strategy section, we assigned 143,092 participant-object dwell records (per-category detection and dwell counts in appendix Table A.28) from 42,830 participant-image observations to theoretically motivated frames (e.g., humanizing vs depersonalizing for immigration, victim vs weapons for gun control, etc.) (see the detailed explanation in Appendix Table A.26).¹⁶

Figure 8, with the underlying proportions in appendix Table A.15, reports the proportion of within-image object dwell time allocated to each frame, separately for Democrats and Republicans. The evidence uniformly supports H3 and provides no support for H4. In immigration images ($N_{\text{obs}} = 39,673$), Democrats allocated 48.2% of their object dwell time to humanizing objects and Republicans 47.8% ($\Delta = +0.4$ pp, $P = .44$). Both parties allocated approximately 37% to depersonalizing objects ($P = .84$). In gun control images ($N_{\text{obs}} = 955$), dwell shares on victim objects were 35.4% for Democrats and 35.2% for Republicans, and shares on weapons objects were 29.6% and 30.5%. In January 6 images ($N_{\text{obs}} = 1,037$), shares on insurrection symbols were 65.2% and 63.6%, and shares on active-rioter cues were 14.7% and 14.3%. In LGBT images ($N_{\text{obs}} = 1,165$), shares on identity-affirming objects were 36.6% and 35.8%, and shares on bystanders were 63.3% and 63.9%. No difference reached statistical significance in any of the frame-level comparisons (all $P > .44$). Democrats and Republicans viewing the same political image distribute their gaze across constituent objects

¹⁶Of the 50,435 image-level observations in the analytic sample, 3,049 trials failed I-DT fixation extraction and 4,556 further participant-image pairs were dropped because no fixations landed inside any detected bounding box, leaving 42,830 pairs with at least one categorized dwell. Per-frame coverage of these pairs appears in appendix Table A.30.

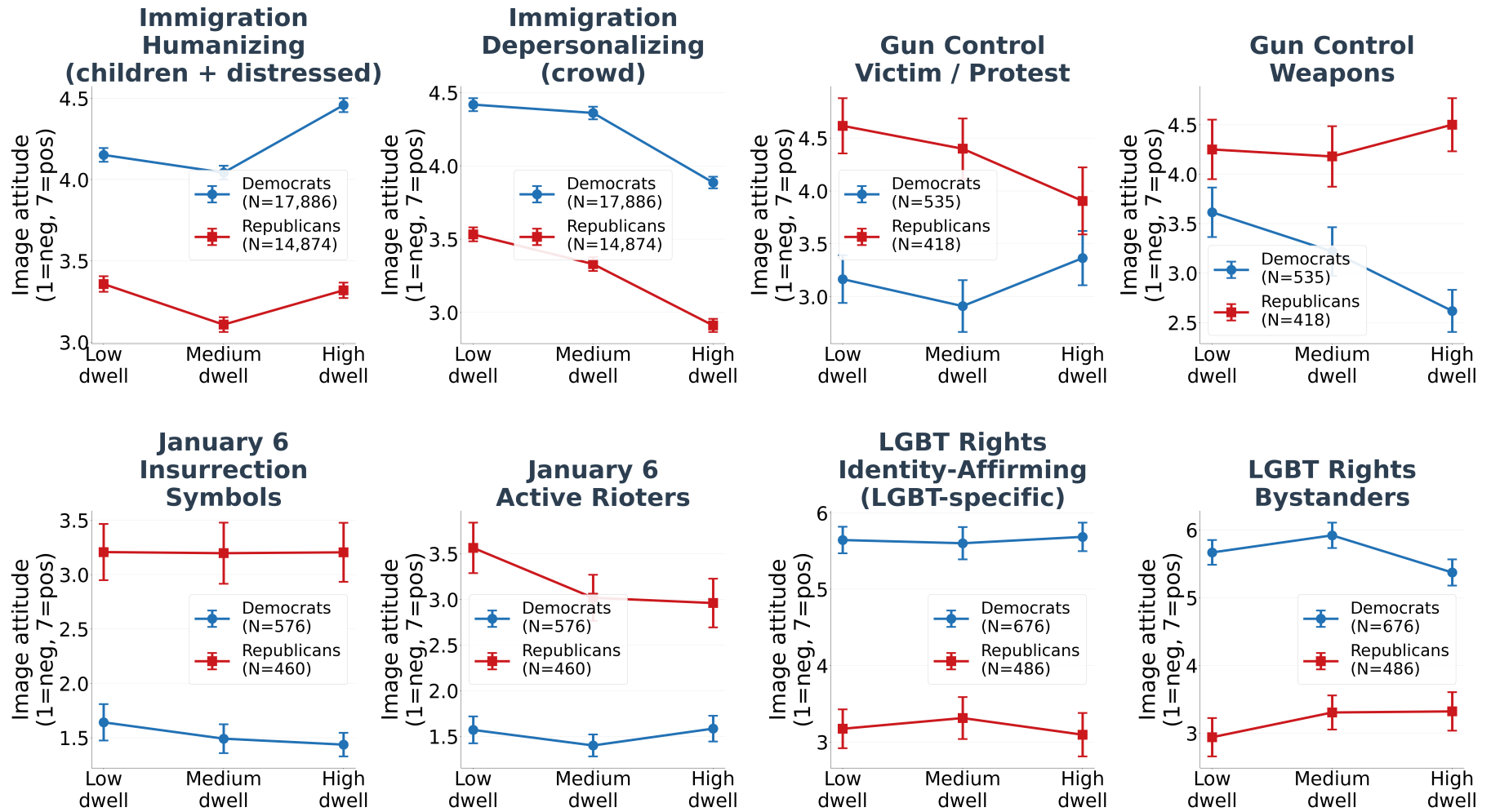
in indistinguishable proportions.

Figure 8: H3 vs H4: Do Partisans Selectively Attend to Attitude-Consistent Objects?



Note: Bars show the proportion of total within-image object dwell time allocated to each theoretically motivated visual frame, separately for Democrats (blue) and Republicans (red). Error bars are 95% confidence intervals. H4 (visual confirmation bias) predicts differential allocation by party. H3 (holistic processing) predicts similar allocation. All eight comparisons are statistically indistinguishable ($P > .44$).

Figure 9: Differential Processing by Visual Frames and Party



Note: Each panel shows mean image attitudes (1 = extremely negative, 7 = extremely positive) by tercile of dwell time on the specified object frame, separately for Democrats (blue circles) and Republicans (red squares). Error bars are 95% confidence intervals. Where lines diverge across terciles, the partisan attitude gap shifts with the level of object-level visual processing. Where they remain parallel, the gap holds constant.

Figure 9, with the underlying cell means in appendix Table A.18, plots mean image attitudes by tercile of dwell time on each theoretically motivated visual frame, separately by party. The partisan attitude gap remains stable across terciles for January 6, both for insurrection symbols and for active-rioter cues. For immigration humanizing, for gun control (both victim and weapons frames), and for LGBT contrast cues, the gap shifts with the tercile of frame dwell, indicating that visual frame engagement modulates evaluative response. The regression tests below quantify these interactions (appendix Tables A.16 and A.17).

When we separate dwell on humanizing objects from dwell on depersonalizing objects in the immigration stimuli and interact each with partisanship ($\text{att}_{ij} = \beta_1 \text{party}_{7i} \times \text{dwell}_{z,ij}^{\text{type}} + \beta_2 \text{party}_{7i} + \beta_3 \text{dwell}_{z,ij}^{\text{type}} + \gamma \text{att_topic}_i + \alpha_j + \varepsilon_{ij}$, standard errors clustered on participant), the humanizing interaction is significant in the expected direction ($b = +0.023$, $P < .001$). Democrats are more pro-immigration when they spend more time on humanizing objects. The depersonalizing interaction is in the expected negative direction but does not reach conventional significance ($b = -0.007$, $P = .13$). The contrast score (humanizing minus depersonalizing dwell) produces a significant interaction with partisanship ($b = +0.011$, $P < .001$).¹⁷

The object-level null carries an implication that the image-level results alone do not. If shared visual sampling were enough to align judgment, the absence of partisan difference in where the eye goes would predict convergence in what partisans conclude, yet the same dwell on the same frame yields opposing evaluations on three of the four topics. Equalizing what citizens look at within a political image therefore does not equalize what they take from it, which locates the polarizing force of visual content not in the distribution of attention but in the meaning assigned to a commonly sampled scene.

¹⁷The same model applied to the other three topics yields a significant contrast interaction for gun control ($b = +0.073$, $P < .001$) and for LGBT rights ($b = +0.018$, $P = .025$), and a non-significant interaction for January 6 ($b = -0.011$, $P = .371$). For gun control both individual frame interactions are significant ($P < .001$ for victim and for weapons). For LGBT, the contrast significance comes from the joint pattern, with individual frame interactions falling short of conventional thresholds ($P = .35$ for identity-affirming, $P = .054$ for bystanders). Even with about 1,000 observations per topic, the same visual content carries different evaluative weight for the two parties on three of the four topics studied.

Discussion and Conclusion

Democrats and Republicans often reach different conclusions from the same political information (Taber and Lodge 2006; Lodge and Taber 2013; Ditto et al. 2019). For political images, the key question is where this divergence begins. It could arise because partisans look at different parts of the same scene, because they process the same scene with different intensity, or because they interpret the image differently only after viewing it. The evidence here points to the middle pathway. Across four polarizing issues, partisans allocate gaze to the same people, objects, and visual frames, yet differ in how deeply they process the image as a whole.

This finding clarifies the mechanism of how visuals may contribute to polarization. The strongest visual-confirmation-bias account predicts that partisans should direct attention toward attitude-congruent elements of a shared scene (Rajic et al. 2015; Qu-Lee and Balcetis 2024; Balcetis and Dunning 2006). The data do not support that prediction at the level of objects and frames measured here. Democrats and Republicans visually sample the same content. The same dwell on the same frames produces different evaluations for the two parties in three of four topics (immigration, gun control, LGBT rights), with January 6 the only exception. However, this shared sampling does not mean that visual processing is politically neutral. Viewers whose position on the depicted issue aligns with their party's position fixate longer on the image, and this pattern reverses across topics as party-issue alignment reverses. Partisan visual engagement therefore depends on the match between identity and issue position, rather than on a stable tendency of one party to look more.

The mediation analysis shows why this processing difference matters for political judgment. Per-fixation duration accounts for a small but measurable share of the partisan evaluative gap, with a stronger pathway when issue alignment is modeled directly. The main study is observational, so this result should be read as evidence of a behavioral pathway rather than as proof of causality. Still, the pathway is important because it links partisan identity, processing depth, and evaluation within the same visual encounter. A follow-up gaze-contingent experiment that directly manipulates fixation duration further supports this

interpretation by testing the causal direction experimentally (Authors, N.d.).

This mechanism also helps explain why shared exposure may fail to reduce disagreement. Recent field experiments changed what users saw in their feeds and found limited effects on political attitudes (Guess et al. 2023; Nyhan et al. 2023; Bail et al. 2018). The present results suggest that exposure is only part of the story. Citizens may see the same image and attend to the same objects, while party-congruent viewers process that shared content more deeply and form more extreme evaluations. In visual media environments, polarization can therefore persist even when exposure is equalized, because the same content is not processed with the same intensity.

The findings also connect behavioral evidence to political neuroscience. Neuroscience studies show that partisans watching identical political videos can have similar responses in sensory regions and divergent responses in systems involved in interpretation and valuation (Leong et al. 2020; Bruin et al. 2023; Amodio et al. 2007). The behavioral pattern observed here follows the same logic. Spatial gaze is shared, while fixation duration diverges. Because fixation duration is a standard behavioral index of processing depth at the point of gaze (Just and Carpenter 1980; Rayner 1998), this study shows how a related process can be measured at scale, in diverse samples, on participants' own devices, and linked directly to evaluative outcomes.

Several questions remain, however, open. The mediation evidence is consistent with the proposed pathway, but the main study cannot rule out unmeasured factors such as emotional arousal or issue interest, which means that future work should measure affective responses alongside gaze. The study also captures single five-second viewings, although political images are encountered repeatedly in news and social media environments. Longitudinal designs could test whether small processing differences accumulate, fade, or depend on issue salience. Finally, the object-level null applies to the spatial grain available in webcam eye-tracking. Higher-resolution laboratory studies could test whether partisans differ in attention to finer features such as facial expressions, posture, text, or color.

Taken together, these findings imply that visual politics cannot be understood by cataloging image content or measuring exposure alone. Political images matter because they create a shared perceptual situation in which citizens attend to the same visible material while, as we show, investing in it unequally. Their force therefore lies in the gap between common visual input and unequal processing. Once an image activates an issue position tied to partisan identity, the viewer does more cognitive work on the same scene, and that added work helps turn the image into political evidence. This makes visual polarization a problem of uptake as much as content. The relevant question is therefore how political identity changes the depth with which shared visual material is processed. The study shifts visual politics from a theory of exposure and depiction toward a theory of politically patterned engagement with common visual worlds.

AI Use Disclosure

This disclosure documents the use of large language models in the study and follows the GUIDE-LLM reporting checklist for large language models in behavioural and social science (Feuerriegel et al. 2026).

Scope of LLM use

(A.1) Purpose. A multimodal large language model, Google Gemini, was used inside the object-level attention pipeline for five circumscribed tasks, namely recovering person detections that the object detectors missed, verifying and relabelling each open-vocabulary object crop against the image, generating per-image prompt augmentations for the detector, producing short descriptive captions for detected persons, and assigning every detection to one of 23 theory-motivated semantic categories in Stage B. The initial localisation of persons and objects was performed by two non-LLM computer-vision detectors, YOLOv8m and YOLO-World. The model produced none of the analysed outcome variables, and the image-level and object-level eye-tracking measures were computed entirely by the gaze-processing

platform and the dispersion-threshold algorithm.

(A.2) Degree of automation. The pipeline operated with a human in the loop, and every bounding box and every assigned category label was manually reviewed and corrected by the research team before any data entered the statistical analysis, with the Stage B categorisation run only after the manual review of all Stage A outputs was complete.

Model or system details

(B.1) Model identity. The Stage B categorisation used Google Gemini 2.5 Flash, and the Stage A detection-assistance and captioning calls used a Google Gemini model of the same family, while the two computer-vision detectors were YOLOv8m and the YOLO-World open-vocabulary detector. The exact pinned model-version identifiers are specified in the replication code, and the corresponding dates of access are reported in the replication documentation.

(B.2) Access and context mode. Gemini was accessed programmatically through its application programming interface, and every request was issued as an independent per-image or per-crop call that carried no conversational state from one call to the next, with the Stage B categorisation issued as a single batched call per image.

(B.3) Configuration. The computer-vision thresholds are reported in full in the appendix, including the YOLOv8m confidence values of 0.25 and 0.20, the YOLO-World confidence value of 0.10, the intersection-over-union suppression values of 0.40 and 0.50, and the minimum normalised bounding-box areas of 0.005 and 0.003. The Gemini generation settings, covering temperature, maximum output tokens, any fixed seed, and the number of runs, are recorded in the released configuration files.

(B.4) Customisation. The models were used off the shelf without fine-tuning or other weight modification and without retrieval augmentation, and they were steered only through the documented task prompts and the closed list of 23 categories defined in the appendix.

(B.5) Persistent memory. The calls used no persistent memory, and no state was carried

from one image or session to the next, so each categorisation and caption was produced from a single self-contained request.

Prompts

(C.1) Prompts. The YOLO-World topic prompt lists are reproduced verbatim in the appendix, and the Gemini prompt logic for the missing-person pass, the per-crop verification, the per-image prompt augmentation, the person captioning, and the Stage B categorisation is described in the appendix, with the full verbatim prompt strings included in the replication materials.

(C.2) System-wide instructions. No separate system-level prompt governed the model, and all instructions, including the closed category list and the topic-conditional decision rules, were supplied inside the per-call task prompts documented in the appendix.

Data inputs and privacy

(D.1) Personal and sensitive data. The only inputs sent to the model were the stimulus images, which were drawn from public news outlets and commercial stock archives, and no participant identifiers, gaze coordinates, survey responses, or webcam video were ever transmitted to the model, consistent with the GDPR consent under which no video or images of participants were recorded or stored.

Validation and interpretation

(E.1) Human validation. Every detection box and every assigned category was manually inspected and corrected by the research team before analysis, and Gemini additionally functioned as a verification layer for the open-vocabulary detector by confirming or rejecting each proposed object crop against the full image.

(E.2) Post-processing. Model outputs were post-processed through overlap resolution that removed object boxes overlapping a person box at intersection-over-union of at least 0.50 and then deduplicated the remaining detections greedily at 0.40 by label specificity, through

ghost-bounding-box filtering that removed whole-image background detections, through truncation of free-text labels to 40 characters, through screening of returned labels for refusal patterns, and through a smallest-containing-box rule that attributed each fixation to a single object.

Reproducibility

(F.1) Code availability. The code, prompt files, and configuration settings for all model calls will be released in the replication archive on Dataverse upon publication.

Competing interests

(G.1) Relationships. The authors declare no competing interests, and they received no funding or compute credits and no in-kind model access from any large-language-model or artificial-intelligence provider in connection with this work.

References

- Amodio, David M. et al. (2007). "Neurocognitive Correlates of Liberalism and Conservatism". In: *Nature Neuroscience* 10.10, pp. 1246–1247.
- Artemenko, Elena, Anastasiia Zhitkova, and Maksim Terpilowsky (2025). "In the Blink of an Eye: Behavioural Correlates of the Confirmation Bias Effect". In: *Cognitive Processing* 26, pp. 567–575.
- Bail, Christopher A. et al. (2018). "Exposure to Opposing Views on Social Media Can Increase Political Polarization". In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9216–9221.
- Balcetis, Emily and David Dunning (2006). "See What You Want to See: Motivational Influences on Visual Perception". In: *Journal of Personality and Social Psychology* 91.4, pp. 612–625.
- Ballew, Charles C. and Alexander Todorov (2007). "Predicting Political Elections from Rapid and Unreflective Face Judgments". In: *Proceedings of the National Academy of Sciences* 104.46, pp. 17948–17953.
- Bar, Moshe (2004). "Visual Objects in Context". In: *Nature Reviews Neuroscience* 5.8, pp. 617–629.
- Bhatnagar, Ritwik and Jacob L. Orquin (2022). "A Meta-Analysis on the Effect of Visual Attention on Choice". In: *Journal of Experimental Psychology: General* 151.10, pp. 2372–2394.
- Bisgaard, Martin and Rune Slothuus (2018). "Partisan Elites as Culprits? How Party Cues Shape Partisan Perceptual Gaps". In: *American Journal of Political Science* 62.2, pp. 456–469.
- Bleiker, Roland et al. (2013). "The visual dehumanisation of refugees". In: *Australian Journal of Political Science* 48.4, pp. 398–416.
- Brader, Ted, Nicholas A. Valentino, and Elizabeth Suhay (2008). "What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration Threat". In: *American Journal of Political Science* 52.4, pp. 959–978.

- Brewer, Paul R. (2003). "Values, Political Knowledge, and Public Opinion about Gay Rights: A Framing-Based Account". In: *Public Opinion Quarterly* 67.2, pp. 173–201.
- Bruin, Willa R. de et al. (2023). "Shared Neural Representations and Temporal Segmentation of Political Content Predict Ideological Similarity". In: *Science Advances* 9, eabq5920.
- Bullock, John G. et al. (2015). "Partisan Bias in Factual Beliefs about Politics". In: *Quarterly Journal of Political Science* 10.4, pp. 519–578.
- Caruso, Eugene M, Nicole L Mead, and Emily Balcetis (2009). "Political Partisanship Influences Perception of Biracial Candidates' Skin Tone". In: *Proceedings of the National Academy of Sciences* 106.48, pp. 20168–20173.
- Casas, Andreu and Nora Webb Williams (2019). "Images That Matter: Online Protests and the Mobilizing Role of Pictures". In: *Political Research Quarterly* 72.2, pp. 360–375.
- Chavez, Leo R. (2013). *The Latino threat: Constructing immigrants, citizens, and the nation*. 2nd ed. Stanford, CA: Stanford University Press.
- Cheng, Tianheng et al. (2024). "YOLO-World: Real-Time Open-Vocabulary Object Detection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16901–16911.
- Coppock, Alexander, Seth J. Hill, and Lynn Vavreck (2020). "The Small Effects of Political Advertising Are Small Regardless of Context, Message, Sender, or Receiver: Evidence from 59 Real-Time Randomized Experiments". In: *Science Advances* 6.36, eabc4046.
- Coronel, Jason C. et al. (2021). "Eye Movements Predict Large-Scale Voting Decisions". In: *Psychological Science* 32.6, pp. 836–848.
- Craig, Alison and Bethany Albertson (2024). "Congressional support for democratic norms on January 6th". In: *Political Science Research and Methods*. Published online 16 September 2024, pp. 1–10.
- Derreumaux, Yannick, Robin Bergh, and Brent L. Hughes (2022). "Partisan-Motivated Sampling: Re-Examining Politically Motivated Reasoning across the Information Processing Stream". In: *Journal of Personality and Social Psychology* 123.2, pp. 316–336.

- Ditto, Peter H. et al. (2019). "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives". In: *Perspectives on Psychological Science* 14.2, pp. 273–291.
- Dodd, Michael D. et al. (2012). "The Political Left Rolls with the Good and the Political Right Confronts the Bad: Connecting Physiology and Cognition to Preferences". In: *Philosophical Transactions of the Royal Society B* 367.1589, pp. 640–649.
- Druckman, James N (2003). "The Power of Television Images: The First Kennedy-Nixon Debate Revisited". In: *The Journal of Politics* 65.2, pp. 559–571.
- Druckman, James N. and Arthur Lupia (2000). "Preference Formation". In: *Annual Review of Political Science* 3, pp. 1–24.
- Dunning, David and Emily Balciis (2013). "Wishful Seeing: How Preferences Shape Visual Perception". In: *Current Directions in Psychological Science* 22.1, pp. 33–37.
- Entman, Robert M. (1993). "Framing: Toward Clarification of a Fractured Paradigm". In: *Journal of Communication* 43.4, pp. 51–58.
- Esses, Victoria M., Stelian Medianu, and Andrea S. Lawson (2013). "Uncertainty, Threat, and the Role of the Media in Promoting the Dehumanization of Immigrants and Refugees". In: *Journal of Social Issues* 69.3, pp. 518–536.
- Farkas, Xénia and Márton Bene (2021). "Images, Politicians, and Social Media: Patterns and Effects of Politicians' Image-Based Political Communication Strategies on Social Media". In: *The International Journal of Press/Politics* 26.1, pp. 119–142.
- Feuerriegel, Stefan et al. (2026). "A reporting checklist for large language models in behavioural science". In: *Nature Human Behaviour*.
- Fitts, Paul M., Richard E. Jones, and John L. Milton (1950). "Eye Movements of Aircraft Pilots during Instrument-Landing Approaches". In: *Aeronautical Engineering Review* 9.2, pp. 24–29.
- Flores, Lisa A. (2003). "Constructing rhetorical borders: Peons, illegal aliens, and competing narratives of immigration". In: *Critical Studies in Media Communication* 20.4, pp. 362–387.

- Galley, Niels, Daniel Betz, and Christoph Biniossek (2015). "Fixation Durations — Why Are They So Highly Variable?" In: *Advances in Visual Perception Research*. Nova Science Publishers.
- Gasparyan, Olga and Elena Sirotkina (2025). "Media Choice and Audience Perceptions: Evidence from Visual Framing of Immigration in News Stories". In: *PLoS ONE* 20.9, e0331219.
- Gemini Team, Google (2024). "Gemini: A Family of Highly Capable Multimodal Models". In: *arXiv preprint arXiv:2312.11805*.
- Gerber, Alan S. et al. (2011). "How Large and Long-Lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment". In: *American Political Science Review* 105.1, pp. 135–150.
- Grabe, Maria Elizabeth and Erik Page Bucy (2009). *Image Bite Politics: News and the Visual Framing of Elections*. Oxford University Press.
- Graber, Doris A. (1990). "Seeing Is Remembering: How Visuals Contribute to Learning from Television News". In: *Journal of Communication* 40.3, pp. 134–155.
- Granot, Yael et al. (2014). "Justice Is Not Blind: Visual Attention Is Directed by Preexisting Attitudes in Evidence Evaluation". In: *Psychological Science* 25.2, pp. 455–462.
- Green, Donald P, Bradley Palmquist, and Eric Schickler (2004). *Partisan hearts and minds: Political parties and the social identities of voters*. Yale University Press.
- (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.
- Guess, Andrew M. et al. (2023). "How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?" In: *Science* 381.6656, pp. 398–404.
- Haas, Ingrid J., Melissa N. Baker, and Frank J. Gonzalez (2017). "Who Can Deviate from the Party Line? Political Ideology Moderates Evaluation of Incongruent Policy Positions in Insula and Anterior Cingulate Cortex". In: *Social Justice Research* 30.4, pp. 355–380.

- Haider-Markel, Donald P. and Mark R. Joslyn (2001). "Gun policy, opinion, tragedy, and blame attribution: The conditional influence of issue frames". In: *The Journal of Politics* 63.2, pp. 520–543.
- Harrison, Brian F. and Melissa R. Michelson (2017). *Listen, We Need to Talk: How to Change Attitudes about LGBT Rights*. New York: Oxford University Press.
- Hessels, Roy S. et al. (2016). "The Area-of-Interest Problem in Eyetracking Research: A Noise-Robust Solution for Face and Sparse Stimuli". In: *Behavior Research Methods* 48.4, pp. 1694–1712.
- Holmqvist, Kenneth et al. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- Hooge, Ignace T. C. et al. (2026). "The Fundamentals of Eye Tracking Part 6: Working with Areas of Interest". In: *Behavior Research Methods* 58.3, p. 65.
- Iyengar, Shanto and Sean J. Westwood (2015). "Fear and Loathing across Party Lines: New Evidence on Group Polarization". In: *American Journal of Political Science* 59.3, pp. 690–707.
- Jacoby, Nir et al. (2024). "Partisans Process Policy-Based and Identity-Based Messages Using Dissociable Neural Systems". In: *Cerebral Cortex* 34.9, bhae368.
- Jenke, Libby and Nicolette Sullivan (2025). "Attention and political choice: A Foundation for eye Tracking in political science". In: *Political Analysis* 33.4, pp. 298–314.
- Jocher, Glenn, Ayush Chaurasia, and Jing Qiu (2023). *Ultralytics YOLOv8*. Version 8.0.0.
- Joo, Jungseock and Zachary C. Steinert-Threlkeld (2022). "Image as Data: Automated Content Analysis for Visual Presentations of Political Actors and Events". In: *Computational Communication Research* 4.1, pp. 11–67.
- Jost, John T. et al. (2014). "Political Neuroscience: The Beginning of a Beautiful Friendship". In: *Political Psychology* 35.S1, pp. 3–42.
- Just, Marcel A and Patricia A Carpenter (1980). "A Theory of Reading: From Eye Fixations to Comprehension". In: *Psychological Review* 87.4, p. 329.

- Kaduk, Tobias et al. (2024). "Webcam Eye Tracking Close to Laboratory Standards: Comparing a New Webcam-Based System and the EyeLink 1000". In: *Behavior Research Methods* 56.5, pp. 5002–5022.
- Katabi, Noa et al. (2023). "Deeper Than You Think: Partisanship-Dependent Brain Responses in Early Sensory and Motor Brain Regions". In: *Journal of Neuroscience* 43.6, pp. 1027–1037.
- Kesner, Ladislav et al. (2020). "Fusiform activity distinguishes between subjects with low and high xenophobic attitudes toward refugees". In: *Frontiers in Behavioral Neuroscience* 14, p. 98.
- Kilgo, Danielle K. and Rachel R. Mourão (2021). "Protest Coverage Matters: How Media Framing and Visual Communication Affects Support for Black Civil Rights Protests". In: *Mass Communication and Society* 24.4, pp. 576–596.
- Kim, Haena, Ian C. Ballard, and Yuan Chang Leong (2026). "A Neurocomputational Account of Motivated Seeing". In: *Trends in Cognitive Sciences* 30.1, pp. 26–39.
- Knutson, Kristine M. et al. (2006). "Politics on the Brain: An fMRI Investigation". In: *Social Neuroscience* 1.1, pp. 25–40.
- Krause, Nicole M. and Erik P. Bucy (2018). "Interpreting Image-Based Frames of Contested Issues". In: *International Journal of Communication* 12, pp. 4769–4791.
- Kunda, Ziva (1987). "Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories". In: *Journal of Personality and Social Psychology* 53.4, p. 636.
- (1990). "The Case for Motivated Reasoning". In: *Psychological Bulletin* 108.3, p. 480.
- Lacombe, Matthew J. (2019). "The Political Weaponization of Gun Owners: The National Rifle Association's Cultivation, Dissemination, and Use of a Group Social Identity". In: *The Journal of Politics* 81.4, pp. 1342–1356.
- Qu-Lee, Jennie and Emily Balcetis (2024). "Shielded Perspectives: How Visual Attention Moderates the Link Between Social Identity and Biased Judgments about Police". In: *Social and Personality Psychology Compass* 18.8, e12994.

- Leong, Yuan Chang et al. (2019). "Neurocomputational Mechanisms Underlying Motivated Seeing". In: *Nature Human Behaviour* 3.9, pp. 962–973.
- Leong, Yuan Chang et al. (2020). "Conservative and Liberal Attitudes Drive Polarized Neural Responses to Political Content". In: *Proceedings of the National Academy of Sciences* 117.44, pp. 27731–27739.
- Liu, Jung-Chun et al. (2022). "Assessing Perceptual Load and Cognitive Load by Fixation-Related Information of Eye Movements". In: *Sensors* 22.3, p. 1187.
- Lodge, Milton and Charles S. Taber (2005). "The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis". In: *Political Psychology* 26.3, pp. 455–482.
- (2013). *The Rationalizing Voter*. Cambridge University Press.
- Lu, Yingdan and Yilang Peng (2024). "The mobilizing power of visual media across stages of social-mediated protests". In: *Political Communication* 41.4, pp. 531–558.
- Manor, Brenda R. and Evian Gordon (2003). "Defining the Temporal Threshold for Ocular Fixation in Free-Viewing Visuocognitive Tasks". In: *Journal of Neuroscience Methods* 128.1–2, pp. 85–93.
- Marquart, Franziska, Jörg Matthes, and Elisabeth Rapp (2016). "Selective Exposure in the Context of Political Advertising: A Behavioral Approach Using Eye-Tracking Methodology". In: *International Journal of Communication* 10, p. 20.
- Marwick, Alice et al. (2024). "Child-sacrificing drag queens: Historical antecedents in disinformative narratives supporting the Drag Queen Story Hour moral panic". In: *Women's Studies in Communication* 47.4, pp. 459–479.
- Mason, Lilliana (2018). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.
- Maughan, Lizzie, Sergei Gutnikov, and Rob Stevens (2007). "Like More, Look More. Look More, Like More: The Evidence from Eye-Tracking". In: *Journal of Brand Management* 14.4, pp. 335–342.

- Melzer, Scott (2009). *Gun crusaders: The NRA's culture war*. New York: New York University Press.
- Negi, Shruti and Rahul Mitra (2020). "Fixation Duration and the Learning Process: An Eye Tracking Study with Subtitled Videos". In: *Journal of Eye Movement Research* 13.6.
- Nyhan, Brendan et al. (2023). "Like-Minded Sources on Facebook Are Prevalent but Not Polarizing". In: *Nature* 620, pp. 137–144.
- Oliva, Aude and Antonio Torralba (2001). "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope". In: *International Journal of Computer Vision* 42.3, pp. 145–175.
- Orquin, Jacob L., Nathaniel J. S. Ashby, and Alasdair D. F. Clarke (2016). "Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research". In: *Journal of Behavioral Decision Making* 29.2–3, pp. 103–115.
- Oxley, Douglas R. et al. (2008). "Political Attitudes Vary with Physiological Traits". In: *Science* 321.5896, pp. 1667–1670.
- Pannasch, Sebastian et al. (2008). "Visual Fixation Durations and Saccade Amplitudes: Shifting Relationship in a Variety of Conditions". In: *Journal of Eye Movement Research* 2.2.
- Pape, Robert A. and Keven Ruby (2022). *American face of insurrection: Analysis of individuals charged for storming the US Capitol on January 6, 2021*. Tech. rep. Chicago Project on Security and Threats, University of Chicago.
- Papoutsaki, Alexandra et al. (2016). "WebGazer: Scalable Webcam Eye Tracking Using User Interactions". In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3839–3845.
- Parrott, Scott et al. (2019). "Portrayals of Immigrants and Refugees in U.S. News Media: Visual Framing and Its Effect on Emotions and Attitudes". In: *Journal of Broadcasting & Electronic Media* 63.4, pp. 677–697.
- Potter, Mary C et al. (2014). "Detecting Meaning in RSVP at 13 ms per Picture". In: *Attention, Perception, & Psychophysics* 76, pp. 270–279.

- Powell, Thomas E et al. (2015). "A Clearer Picture: The Contribution of Visuals and Text to Framing Effects". In: *Journal of Communication* 65.6, pp. 997–1017.
- Rajsic, Jason, Daryl E. Wilson, and Jay Pratt (2015). "Confirmation Bias in Visual Search". In: *Journal of Experimental Psychology: Human Perception and Performance* 41.5, pp. 1353–1364.
- (2018). "The Price of Information: Increased Inspection Costs Reduce the Confirmation Bias in Visual Search". In: *Quarterly Journal of Experimental Psychology* 71.4, pp. 832–849.
- Rayner, Keith (1998). "Eye Movements in Reading and Information Processing: 20 Years of Research". In: *Psychological Bulletin* 124.3, pp. 372–422.
- Redmon, Joseph et al. (2016). "You Only Look Once: Unified, Real-Time Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788.
- Salvucci, Dario D and Joseph H Goldberg (2000). "Identifying fixations and saccades in eye-tracking protocols". In: *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78.
- Santa Ana, Otto (2002). *Brown tide rising: Metaphors of Latinos in contemporary American public discourse*. Austin, TX: University of Texas Press.
- Schmuck, Desirée, Jörg Matthes, and Hajo G. Boomgaarden (2020). "Issue Attitudes and Visual Attention to Immigration News Coverage". In: *International Journal of Public Opinion Research* 32.3, pp. 466–487.
- Schreiber, Darren et al. (2013). "Red Brain, Blue Brain: Evaluative Processes Differ in Democrats and Republicans". In: *PLoS ONE* 8.2, e52970.
- Semmelmann, Kilian and Sarah Weigelt (2017). "Online psychophysics: Reaction time effects in cognitive experiments". In: *Behavior Research Methods* 49.4, pp. 1241–1260.
- Sharafi, Zohreh et al. (2015). "Eye-Tracking Metrics in Software Engineering". In: *Proceedings of the 22nd Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, pp. 96–103.
- Sheridan, Heather and Eyal M. Reingold (2017). "The Holistic Processing Account of Visual Expertise in Medical Image Perception: A Review". In: *Frontiers in Psychology* 8, p. 1620.

- Sides, John, Michael Tesler, and Lynn Vavreck (2018). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton University Press.
- Slovic, Paul et al. (2017). "Iconic photographs and the ebb and flow of empathic response to humanitarian disasters". In: *Proceedings of the National Academy of Sciences* 114.4, pp. 640–644.
- Sollberger, Silja, Thomas Bernauer, and Ulrike Ehlert (2017). "Predictors of Visual Attention to Climate Change Images: An Eye-Tracking Study". In: *Journal of Environmental Psychology* 51, pp. 46–56.
- Staub, Adrian et al. (2010). "Distributional Effects of Word Frequency on Eye Fixation Durations". In: *Journal of Experimental Psychology: Human Perception and Performance* 36.5, pp. 1280–1293.
- Sülflow, Michael, Svenja Schäfer, and Stephan Winter (2019). "Selective Attention in the News Feed: An Eye-Tracking Study on the Perception and Selection of Political News Posts on Facebook". In: *New Media & Society* 21.1, pp. 168–190.
- Taber, Charles S., Damon Cann, and Simona Kucsova (2009). "The Motivated Processing of Political Arguments". In: *Political Behavior* 31.2, pp. 137–155.
- Taber, Charles S. and Milton Lodge (2006). "Motivated Skepticism in the Evaluation of Political Beliefs". In: *American Journal of Political Science* 50.3, pp. 755–769.
- Tesler, Michael (2016). *Post-Racial or Most-Racial? Race and Politics in the Obama Era*. University of Chicago Press.
- Thorpe, Simon, Denis Fize, and Catherine Marlot (1996). "Speed of Processing in the Human Visual System". In: *Nature* 381.6582, pp. 520–522.
- Todorov, Alexander et al. (2005). "Inferences of Competence from Faces Predict Election Outcomes". In: *Science* 308.5728, pp. 1623–1626.
- Valentino, Nicholas A., Ted Brader, and Ashley E. Jardina (2013). "Immigration Opposition Among U.S. Whites: General Ethnocentrism or Media Priming of Attitudes About Latinos?" In: *Political Psychology* 34.2, pp. 149–166.

- Van Bavel, Jay J and Andrea Pereira (2018). "The Partisan Brain: An Identity-Based Model of Political Belief". In: *Trends in Cognitive Sciences* 22.3, pp. 213–224.
- Vansteenkiste, Pieter et al. (2015). "Measuring Dwell Time Percentage from Head-Mounted Eye-Tracking Data — Comparison of a Frame-by-Frame and a Fixation-by-Fixation Analysis". In: *Ergonomics* 58.5, pp. 712–721.
- Vraga, Emily, Leticia Bode, and Sonya Troller-Renfree (2016). "Beyond Self-Reports: Using Eye Tracking to Measure Topic and Style Differences in Attention to Social Media Content". In: *Communication Methods and Measures* 10.2-3, pp. 149–164.
- Webb Williams, Nora et al. (2025). "When Conservatives See Red but Liberals Feel Blue: Labeler Characteristics and Variation in Content Annotation". In: *The Journal of Politics*. Forthcoming; SSRN 4540742.
- Whitman, Julie C. et al. (2018). "Political Orientation and Climate Concern Shape Visual Attention to Climate Change". In: *Climatic Change* 147.3–4, pp. 383–394.
- Wozniak, Kevin H. (2017). "Public opinion about gun control post–Sandy Hook". In: *Criminal Justice Policy Review* 28.3, pp. 255–278.
- Xiao, Y. Jenny, Géraldine Coppin, and Jay J. Van Bavel (2016). "Perceiving the World Through Group-Colored Glasses: A Perceptual Model of Intergroup Relations". In: *Psychological Inquiry* 27.4, pp. 255–274.
- Yamane, David (2023). "Gun Culture 2.0: The evolution and contours of defensive gun ownership in America". In: *The Annals of the American Academy of Political and Social Science* 704.1, pp. 20–43.
- Yang, Xiaozhi and Ian Krajbich (2021). "Webcam-Based Online Eye-Tracking for Behavioral Research". In: *Judgment and Decision Making* 16.6, pp. 1485–1505.
- Yang, Xiaozhi et al. (2026). "Issue Importance Amplifies the Effect of Gaze on Voting Decisions". In: *Cognition* 268, p. 106376.

Zulli, Diana, Kevin Coe, and Zachary Isaacs (2023). "News Framing in the Aftermath of the January 6 Attacks on the U.S. Capitol: An Analysis of Labels, Definitional Uncertainty, and Contextualization". In: *American Behavioral Scientist* 67.6, pp. 702–720.

A Appendix

Table A.1: Participant-Level Summary Statistics by Wave

	Wave 1 (<i>N</i> = 1,082)	Wave 2 (<i>N</i> = 924)	Wave 3 (<i>N</i> = 200)	All (<i>N</i> = 2,206)
<i>Demographics</i>				
Age	4.66 (1.64)	4.31 (1.41)	3.85 (1.37)	4.44 (1.54)
Female (%)	58.3	53.0	60.0	56.3
Education	3.74 (1.48)	4.32 (1.37)	4.45 (1.31)	4.05 (1.45)
Household income	5.82 (3.24)	7.40 (3.39)	7.97 (3.46)	6.68 (3.43)
Hispanic (%)	9.0	10.1	14.0	9.9
White (%)	80.2	77.1	74.0	78.3
Black (%)	9.7	10.1	11.0	10.0
Asian (%)	3.1	7.4	6.5	5.2
Other race (%)	7.0	5.5	8.5	6.5
<i>Political Identity</i>				
Partisanship	4.20 (2.23)	4.08 (2.03)	4.40 (2.45)	4.17 (2.17)
Strong Republican (%)	18.5	13.3	16.0	16.0
Not strong Republican (%)	12.8	17.1	24.0	15.7
Lean Republican (%)	6.4	8.1	3.0	6.8
Independent (%)	16.6	16.9	0.0	15.2
Lean Democrat (%)	7.3	16.4	8.0	11.3
Not strong Democrat (%)	15.3	11.0	16.0	13.5
Strong Democrat (%)	23.1	17.2	33.0	21.5
Ideology	3.93 (1.82)	4.24 (1.84)	4.20 (2.10)	4.08 (1.86)
Feeling thermometer, Democrats	50.1 (33.1)	49.7 (28.0)	53.1 (31.9)	50.2 (30.9)
Feeling thermometer, Republicans	44.8 (32.6)	40.8 (30.5)	43.1 (34.8)	42.9 (32.0)

Note. Continued in Table A.2. Cells report mean (SD) for continuous variables and the percentage of non-missing respondents for categorical levels. Age is coded 1 = Under 18, 2 = 18–24, ..., 9 = 85+. Education is coded 1 = Less than high school, ..., 7 = Doctoral degree. Household income is coded 1 = Under \$10,000, ..., 12 = Over \$150,000. Partisanship is the survey's 7-point party identification scale (1 = Strong Republican, ..., 7 = Strong Democrat). Waves 1–2 use a standard 7-point item; Wave 3 uses a 6-point item without an Independent category, so percent at category 4 is zero for Wave 3 by construction. Ideology is coded 1 = Very conservative, ..., 7 = Very liberal; text labels in Waves 1–2 are recoded to the same numeric scale. Race is recoded to four levels; respondents selecting multiple categories are placed in "Other."

Table A.2: Participant-Level Summary Statistics by Wave (continued)

	Wave 1 (<i>N</i> = 1,082)	Wave 2 (<i>N</i> = 924)	Wave 3 (<i>N</i> = 200)	All (<i>N</i> = 2,206)
<i>Issue Positions</i>				
Immigration	2.86 (1.38)	2.76 (1.42)	2.83 (1.53)	2.82 (1.41)
Same-sex marriage	3.50 (1.48)	3.87 (1.44)	3.65 (1.50)	3.67 (1.48)
Abortion	3.77 (1.45)	3.96 (1.42)	3.69 (1.56)	3.84 (1.45)
Affirmative action	3.24 (1.31)	3.11 (1.45)	3.16 (1.54)	3.18 (1.39)
State-sponsored college	4.03 (1.10)	4.16 (1.03)	4.21 (1.02)	4.10 (1.07)
Gun permit	—	—	2.69 (1.47)	—
January 6	—	—	4.24 (1.13)	—
<i>Eye-Tracking Measures</i>				
Average fixation duration (ms)	185.7 (108.0)	381.2 (110.3)	401.8 (116.5)	287.2 (148.3)
Median fixation duration (ms)	155.4 (89.3)	323.0 (105.4)	330.5 (103.8)	241.5 (129.1)
Total fixation duration (ms)	2,710 (1,259)	4,280 (487)	4,292 (517)	3,511 (1,232)
<i>Study Design</i>				
Desktop (%)	29.7	72.4	68.8	51.1
Smartphone (%)	53.0	0.0	0.0	26.0
Tablet (%)	17.4	27.6	31.2	22.9
Quality grade	4.31 (1.28)	4.95 (1.04)	4.62 (1.11)	4.61 (1.21)
Images viewed	15.9 (0.7)	30.8 (1.6)	23.8 (1.4)	22.9 (7.2)
Image attitude	3.57 (1.15)	3.88 (1.11)	3.61 (1.30)	3.70 (1.16)

Note. Continued from Table A.1. All issue positions are scaled 1 = most conservative, 5 = most liberal. Gun permit and January 6 items are available only in Wave 3. Eye-tracking measures are averaged to the participant level across images viewed. Average and median fixation duration are capped at 800 ms; durations above the cap are set to missing prior to averaging. Quality grade is the platform's per-participant calibration rating (1–6, higher is better). Image attitude is the participant's average post-exposure evaluation (1–7 scale).

Table A.3: Sample Composition vs. U.S. Census Benchmarks

	Sample (%)	Census (%)
<i>Gender</i>		
Male	43.7	49.0
Female	56.3	51.0
<i>Age</i>		
18–24	9.5	10.6
25–34	20.7	14.7
35–44	24.0	14.8
45–54	19.5	14.9
55–64	15.5	17.8
65–74	8.6	15.9
75–84	1.9	8.3
85+	0.3	3.1
<i>Hispanic Ethnicity</i>		
Not Hispanic	90.1	86.0
Hispanic	9.9	14.0

Note. Sample percentages are computed from the full analytic sample of 2,206 participants pooled across Waves 1–3. Census benchmarks are from the U.S. Census Bureau American Community Survey, Public Use Microdata Sample (PUMS), 5-year estimates reported in 2023, restricted to U.S. adults aged 18 and over. Nine participants with age category “Under 18” (age coded as 1) are excluded from the age comparison.

Table A.4: Full Estimates Underlying Figure 6. Median Fixation Duration (ms), 7-Point Partisanship

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Partisanship (1–7)	2.789** (1.032)	8.730** (2.842)	2.468* (0.971)	6.292* (2.553)
Issue position		8.342* (3.693)		5.324 (3.401)
Partisanship × Issue pos.		−1.676* (0.780)		−1.125 (0.708)
Female	−5.957 (4.553)	−6.211 (4.559)	−6.682 (4.342)	−6.858 (4.350)
Age category	−2.585 (1.503)	−2.860 (1.528)	−2.857* (1.437)	−3.032* (1.464)
Education	−0.689 (1.739)	−0.694 (1.740)	−0.270 (1.666)	−0.296 (1.668)
Household income	2.195** (0.760)	2.155** (0.759)	2.299** (0.717)	2.273** (0.716)
Black	−24.386** (8.171)	−22.552** (8.124)	−24.545** (7.609)	−23.333** (7.591)
Asian	−24.408* (9.530)	−23.766* (9.569)	−22.874** (8.843)	−22.456* (8.866)
Other race	4.580 (8.234)	4.174 (8.184)	9.123 (7.843)	8.925 (7.811)
Smartphone	−137.134*** (4.971)	−136.582*** (4.976)	−137.971*** (4.866)	−137.645*** (4.868)
Tablet	−37.028*** (6.274)	−37.440*** (6.261)	−38.533*** (5.811)	−38.826*** (5.806)
Wave 2	58.832*** (5.602)	58.711*** (5.611)		
Wave 1			−252.110*** (25.817)	−250.292*** (25.825)
Wave 2			−194.149*** (25.601)	−192.411*** (25.615)
Intercept	235.477*** (11.454)	205.449*** (17.656)	488.806*** (27.049)	468.451*** (30.068)
<i>N</i>	43,211	43,211	47,054	47,054
<i>R</i> ²	0.385	0.386	0.389	0.390
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). In W1+2 columns, Wave 1 is the reference; in All Waves columns, Wave 3 is the reference. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.5: Full Estimates Underlying Figure 6. Median Fixation Duration (ms), Strong Dem vs Strong Rep

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Strong Democrat	21.900** (7.320)	89.995*** (22.334)	18.206** (6.921)	69.519*** (20.776)
Issue position		10.232* (4.386)		7.424 (4.209)
Strong Dem $\times$ Issue pos.		-21.087*** (5.978)		-16.604** (5.533)
Female	-11.620 (7.117)	-13.730 (7.118)	-9.407 (6.788)	-10.993 (6.798)
Age category	-1.502 (2.282)	-0.848 (2.299)	-1.144 (2.176)	-0.640 (2.196)
Education	-0.458 (2.733)	-1.146 (2.714)	-0.036 (2.569)	-0.592 (2.557)
Household income	1.517 (1.276)	1.753 (1.275)	1.800 (1.167)	1.937 (1.166)
Black	-34.738* (13.565)	-25.827* (13.137)	-32.746** (12.107)	-25.398* (11.917)
Asian	-39.574* (16.579)	-37.283* (16.561)	-29.760* (14.977)	-27.050 (15.083)
Other race	15.493 (15.315)	11.559 (14.915)	16.083 (14.614)	12.775 (14.296)
Smartphone	-144.934*** (7.817)	-141.621*** (7.928)	-146.015*** (7.596)	-143.398*** (7.682)
Tablet	-33.767*** (9.879)	-35.991*** (9.790)	-37.611*** (8.921)	-39.265*** (8.862)
Wave 2	49.056*** (8.811)	48.683*** (8.728)		
Wave 1			-242.479*** (36.888)	-237.250*** (36.818)
Wave 2			-194.670*** (36.483)	-189.761*** (36.445)
Intercept	242.813*** (17.719)	197.857*** (24.667)	481.886*** (38.697)	444.859*** (42.522)
N	15,081	15,081	16,963	16,963
R^2	0.400	0.407	0.408	0.411
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Sample restricted to Strong Democrats (partisan = 7) and Strong Republicans (partisan = 1). OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.6: Full Estimates Underlying Figure 6. Average Fixation Duration (ms), 7-Point Partisanship

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Partisanship (1–7)	2.920** (1.105)	9.123** (3.070)	2.578* (1.039)	6.179* (2.752)
Issue position		8.559* (3.990)		4.871 (3.665)
Partisanship × Issue pos.		–1.774* (0.841)		–1.085 (0.760)
Female	–7.759 (4.869)	–8.065 (4.875)	–8.746 (4.640)	–8.952 (4.648)
Age category	–2.936 (1.631)	–3.193 (1.655)	–3.101* (1.558)	–3.236* (1.585)
Education	–0.875 (1.876)	–0.904 (1.876)	–0.403 (1.793)	–0.449 (1.793)
Household income	2.312** (0.817)	2.271** (0.816)	2.380** (0.768)	2.356** (0.768)
Black	–26.034** (9.051)	–24.090** (9.003)	–26.205** (8.439)	–25.019** (8.422)
Asian	–25.152* (10.138)	–24.456* (10.182)	–23.075* (9.338)	–22.664* (9.361)
Other race	4.876 (8.806)	4.459 (8.755)	9.804 (8.325)	9.616 (8.297)
Smartphone	–176.246*** (5.560)	–175.655*** (5.567)	–177.055*** (5.449)	–176.733*** (5.454)
Tablet	–42.914*** (6.823)	–43.313*** (6.817)	–44.644*** (6.304)	–44.900*** (6.307)
Wave 2	63.000*** (6.197)	62.910*** (6.209)		
Wave 1			–283.931*** (18.129)	–282.397*** (18.145)
Wave 2			–221.734*** (17.707)	–220.241*** (17.731)
Intercept	289.820*** (12.366)	259.186*** (19.112)	574.698*** (20.348)	556.408*** (24.534)
<i>N</i>	44,110	44,110	48,249	48,249
<i>R</i> ²	0.495	0.496	0.500	0.501
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). In W1+2 columns, Wave 1 is the reference; in All Waves columns, Wave 3 is the reference. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.7: Full Estimates Underlying Figure 6. Average Fixation Duration (ms), Strong Democrat vs. Strong Republican

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Strong Democrat	22.543** (7.827)	96.602*** (24.184)	18.911* (7.387)	71.799** (22.525)
Issue position		11.199* (4.751)		7.614 (4.574)
Strong Dem $\times$ Issue pos.		-22.895*** (6.515)		-17.174** (6.014)
Female	-11.788 (7.606)	-14.053 (7.580)	-10.174 (7.270)	-11.845 (7.264)
Age category	-1.577 (2.441)	-0.885 (2.452)	-1.166 (2.340)	-0.645 (2.359)
Education	-0.048 (2.910)	-0.792 (2.883)	0.086 (2.731)	-0.493 (2.714)
Household income	1.403 (1.366)	1.658 (1.361)	1.747 (1.251)	1.885 (1.247)
Black	-39.186** (14.688)	-29.507* (14.293)	-37.360** (13.085)	-29.705* (12.944)
Asian	-43.522* (18.418)	-40.958* (18.333)	-33.351* (16.542)	-30.417 (16.607)
Other race	16.292 (15.819)	12.053 (15.412)	17.287 (14.956)	13.855 (14.648)
Smartphone	-186.136*** (8.614)	-182.558*** (8.754)	-187.537*** (8.407)	-184.818*** (8.518)
Tablet	-38.892*** (10.669)	-41.235*** (10.589)	-43.665*** (9.624)	-45.301*** (9.572)
Wave 2	50.306*** (9.587)	49.838*** (9.493)		
Wave 1			-288.244*** (26.258)	-284.215*** (26.411)
Wave 2			-239.151*** (25.546)	-235.498*** (25.748)
Intercept	293.555*** (19.381)	244.464*** (26.884)	579.471*** (29.708)	542.906*** (35.418)
<i>N</i>	15,333	15,333	17,347	17,347
<i>R</i> ²	0.512	0.518	0.518	0.522
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Sample restricted to Strong Democrats and Strong Republicans. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.8: Full Estimates Underlying Figure 6. Total Fixation Duration (ms), 7-Point Partisanship

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Partisanship (1–7)	10.068 (6.493)	66.288*** (18.302)	10.025 (5.915)	51.508** (15.808)
Issue position		73.378** (24.723)		53.260* (21.877)
Partisanship × Issue pos.		–16.740** (5.103)		–12.987** (4.480)
Female	–6.209 (26.122)	–9.959 (26.251)	–13.185 (24.461)	–16.335 (24.572)
Age category	–16.641 (8.608)	–18.135* (8.740)	–16.482* (8.112)	–17.500* (8.256)
Education	–10.699 (10.614)	–11.489 (10.566)	–7.631 (10.053)	–8.524 (10.016)
Household income	11.643* (4.668)	11.318* (4.640)	10.781* (4.348)	10.528* (4.333)
Black	–78.907 (51.639)	–60.407 (52.527)	–95.587* (48.358)	–80.906 (49.269)
Asian	–86.447 (51.034)	–80.269 (50.985)	–77.750 (46.033)	–72.946 (45.923)
Other race	–20.212 (52.748)	–23.816 (52.270)	6.749 (47.437)	4.684 (47.108)
Smartphone	–2160.955*** (46.733)	–2155.051*** (46.486)	–2159.299*** (46.407)	–2155.314*** (46.206)
Tablet	–228.326*** (37.315)	–231.657*** (37.193)	–224.839*** (33.987)	–227.750*** (33.931)
Wave 2	326.278*** (39.010)	326.514*** (38.852)		
Wave 1			–884.084*** (43.595)	–868.226*** (43.366)
Wave 2			–558.036*** (32.064)	–541.975*** (32.118)
Intercept	3945.764*** (70.454)	3687.139*** (115.596)	4824.925*** (63.208)	4629.761*** (104.051)
<i>N</i>	44,772	44,772	49,495	49,495
<i>R</i> ²	0.703	0.705	0.697	0.698
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device).

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.9: Full Estimates Underlying Figure 6. Total Fixation Duration (ms), Strong Democrat vs. Strong Republican

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Strong Democrat	58.647 (48.972)	530.149*** (158.996)	53.746 (44.270)	418.000** (140.755)
Issue position		67.425* (32.760)		50.797 (29.791)
Strong Dem $\times$ Issue pos.		-149.143** (45.583)		-119.868** (40.637)
Female	0.954 (44.687)	-13.819 (44.427)	3.425 (41.227)	-8.445 (41.088)
Age category	-2.354 (14.420)	2.680 (14.714)	2.095 (13.646)	5.884 (13.933)
Education	-2.407 (17.036)	-7.308 (16.627)	-1.415 (15.738)	-5.471 (15.420)
Household income	5.751 (8.026)	7.432 (7.844)	5.465 (7.249)	6.393 (7.129)
Black	-189.578* (92.927)	-124.466 (98.778)	-197.096* (83.154)	-142.498 (88.957)
Asian	-165.117 (93.279)	-147.520 (90.912)	-123.621 (80.475)	-101.602 (79.744)
Other race	-7.717 (105.547)	-36.385 (100.881)	5.552 (95.674)	-18.077 (92.319)
Smartphone	-2296.313*** (69.338)	-2272.155*** (68.914)	-2295.463*** (68.831)	-2275.956*** (68.487)
Tablet	-206.198** (64.539)	-221.574*** (64.184)	-214.205*** (55.979)	-225.016*** (55.594)
Wave 2	222.493*** (57.516)	219.380*** (57.094)		
Wave 1			-855.012*** (64.247)	-835.616*** (63.917)
Wave 2			-632.117*** (55.420)	-615.324*** (54.519)
Intercept	3986.249*** (117.941)	3687.046*** (184.722)	4819.004*** (107.480)	4581.735*** (170.064)
N	15,506	15,506	17,815	17,815
R^2	0.711	0.716	0.703	0.707
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Sample restricted to Strong Democrats and Strong Republicans. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.10: Full Estimates Underlying Figure 6. Image Attitude (1–7), 7-Point Partisanship

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Partisanship (1–7)	0.208*** (0.012)	0.139*** (0.038)	0.187*** (0.012)	0.092** (0.034)
Issue position		–0.256*** (0.051)		–0.280*** (0.046)
Partisanship × Issue pos.		–0.006 (0.010)		0.002 (0.009)
Female	0.055 (0.049)	0.020 (0.046)	0.063 (0.048)	0.033 (0.045)
Age category	–0.063*** (0.017)	–0.027 (0.016)	–0.066*** (0.016)	–0.032* (0.016)
Education	0.055** (0.019)	0.035 (0.018)	0.040* (0.018)	0.022 (0.017)
Household income	–0.013 (0.008)	–0.011 (0.008)	–0.016* (0.008)	–0.014 (0.007)
Black	–0.298** (0.091)	–0.290** (0.089)	–0.298*** (0.084)	–0.281*** (0.084)
Asian	–0.449*** (0.098)	–0.457*** (0.093)	–0.378*** (0.097)	–0.386*** (0.093)
Other race	0.024 (0.106)	0.023 (0.101)	0.015 (0.098)	0.011 (0.093)
Smartphone	0.058 (0.075)	0.075 (0.072)	0.054 (0.074)	0.065 (0.072)
Tablet	–0.030 (0.059)	–0.009 (0.055)	–0.004 (0.057)	0.015 (0.053)
Wave 2	0.182** (0.068)	0.224*** (0.064)		
Wave 1			–0.130 (0.142)	–0.142 (0.140)
Wave 2			0.054 (0.134)	0.081 (0.133)
Intercept	3.314*** (0.127)	4.383*** (0.224)	3.608*** (0.169)	4.779*** (0.241)
<i>N</i>	42,048	42,048	46,763	46,763
<i>R</i> ²	0.207	0.245	0.209	0.243
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device).

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.11: Full Estimates Underlying Figure 6. Image Attitude (1–7), Strong Democrat vs. Strong Republican

	W1+2 Direct	W1+2 Interaction	All Waves Direct	All Waves Interaction
Strong Democrat	1.437*** (0.091)	1.382*** (0.321)	1.291*** (0.089)	0.938** (0.302)
Issue position		−0.191** (0.066)		−0.227*** (0.063)
Strong Dem × Issue pos.		−0.138 (0.081)		−0.044 (0.076)
Female	0.133 (0.089)	0.100 (0.085)	0.139 (0.087)	0.108 (0.085)
Age category	−0.084** (0.030)	−0.051 (0.029)	−0.092** (0.029)	−0.064* (0.029)
Education	0.017 (0.034)	0.002 (0.032)	0.002 (0.033)	−0.009 (0.032)
Household income	−0.015 (0.015)	−0.012 (0.015)	−0.021 (0.015)	−0.019 (0.015)
Black	−0.702*** (0.147)	−0.556*** (0.141)	−0.699*** (0.130)	−0.575*** (0.130)
Asian	−0.556** (0.215)	−0.499* (0.197)	−0.486* (0.208)	−0.434* (0.196)
Other race	0.013 (0.235)	−0.062 (0.237)	−0.035 (0.218)	−0.095 (0.217)
Smartphone	−0.026 (0.129)	0.034 (0.126)	−0.025 (0.128)	0.022 (0.125)
Tablet	−0.062 (0.107)	−0.080 (0.098)	−0.013 (0.104)	−0.016 (0.099)
Wave 2	0.121 (0.113)	0.120 (0.107)		
Wave 1			0.073 (0.216)	0.105 (0.218)
Wave 2			0.204 (0.207)	0.237 (0.208)
Intercept	3.769*** (0.224)	4.449*** (0.341)	3.897*** (0.281)	4.724*** (0.382)
<i>N</i>	14,676	14,676	16,982	16,982
<i>R</i> ²	0.257	0.288	0.251	0.275
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Sample restricted to Strong Democrats and Strong Republicans. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). **p* < .05, ***p* < .01, ****p* < .001.

Table A.12: Per-Topic Estimates Underlying Section 1 Per-Topic Claims. Median Fixation Duration (ms), 7-Point Partisanship $\times$ Issue Position

	Immigration	Gun Control	January 6	LGBT
Partisanship (1–7)	8.698** (2.811)	–15.053 (8.714)	–6.166 (9.069)	–13.956 (10.014)
Issue position	8.361* (3.665)	–21.933 (12.465)	–16.008 (16.429)	–25.528 (15.485)
Partisanship $\times$ Issue pos.	–1.721* (0.773)	3.534 (2.613)	2.502 (6.171)	5.453 (3.311)
Female	–5.882 (4.516)	–36.324 (20.097)	–16.065 (20.884)	–26.176 (23.237)
Age category	–2.986* (1.514)	–3.034 (7.934)	–1.897 (8.698)	–2.691 (6.685)
Education	–0.629 (1.723)	9.098 (7.598)	13.588 (9.088)	–3.206 (7.943)
Household income	2.228** (0.750)	1.138 (2.659)	1.159 (3.019)	6.191* (3.034)
Black	–21.581** (8.012)	–34.709 (26.005)	–42.780 (30.980)	–51.187 (34.547)
Asian	–22.905* (9.459)	–15.157 (24.559)	–2.459 (33.676)	–29.233 (24.516)
Other race	5.121 (8.132)	41.145 (29.069)	48.634 (42.101)	25.659 (37.620)
Smartphone	–136.986*** (4.948)			
Tablet	–38.047*** (6.153)	–31.145 (19.058)	–40.137 (22.818)	–63.071** (22.621)
Wave 1	–249.884*** (25.824)			
Wave 2	–191.546*** (25.641)			
Intercept	455.795*** (30.891)	329.128*** (71.682)	435.830*** (65.511)	445.220*** (63.986)
<i>N</i>	43,856	1,003	1,105	1,090
<i>R</i> ²	0.398	0.374	0.241	0.214
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	No	No	No

Note. OLS with standard errors clustered at the participant level in parentheses. All specifications include image fixed effects (coefficients suppressed). Reference categories are White (race), Desktop (device). Immigration column pools all three waves (immigration observations only) with wave fixed effects (Wave 3 is the reference). Gun Control, January 6, and LGBT columns use Wave 3 data only; smartphone is absent from Wave 3. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.13: Full Mediation Results Underlying Section 2 and Figure 7

Specification	<i>N</i>	<i>a</i>	<i>b</i>	<i>ab</i>	Boot 95% CI	Boot <i>p</i>	% Med
<i>Simple specification, all waves</i>							
Partisan → MFD → Att	2,152	2.268**	0.00088**	0.00199	[0.0004, 0.0043]	0.007**	1.10
Str Dem → MFD → Att	809	16.525**	0.00131*	0.02166	[0.0028, 0.0515]	0.016*	1.73
Partisan → AFD → Att	2,152	2.368*	0.00085**	0.00202	[0.0004, 0.0043]	0.011*	1.11
Str Dem → AFD → Att	809	17.448*	0.00113*	0.01977	[0.0020, 0.0487]	0.021*	1.58
<i>Moderated specification (partisanship × issue position), all waves</i>							
Partisan×IP → MFD → Att	2,152	5.294*	0.00082**	0.00432	[0.0004, 0.0100]	0.024*	3.41
Partisan×IP → AFD → Att	2,152	5.826*	0.00076**	0.00444	[0.0004, 0.0105]	0.030*	3.51
<i>Moderated specification (partisanship × issue position), W1+2</i>							
Partisan×IP → MFD → Att	1,954	7.653**	0.00068*	0.00521	[0.0006, 0.0115]	0.020*	4.27
Partisan×IP → AFD → Att	1,954	8.452**	0.00061*	0.00519	[0.0004, 0.0118]	0.031*	4.26

Note. Person-level mediation (Baron-Kenny). Path *a* is the effect of partisanship on fixation duration; path *b* is the effect of fixation duration on image attitude controlling for partisanship; *ab* is the indirect effect. Bootstrap confidence intervals and *p*-values based on 2,000 nonparametric bootstrap replications. All specifications include demographic controls (gender, age, education, income, race) and fixed effects for device and wave. Heteroskedasticity-robust (HC1) standard errors. “Str Dem” restricts the sample to Strong Democrats (partisan = 7) and Strong Republicans (partisan = 1). “Partisan×IP” is the moderated specification where path *a* includes the partisanship × issue position interaction. % Med is the proportion of the total effect (*c*) transmitted through the mediator. **p* < .05, ***p* < .01, ****p* < .001.

Table A.14: Full Mediation Results Underlying Section 2, All Specifications

Specification	<i>N</i>	<i>a</i>	<i>b</i>	<i>ab</i>	Boot 95% CI	Boot <i>p</i>	% Med
<i>All waves, simple</i>							
Partisan → MFD → Att	2,152	2.268**	0.00088**	0.00199	[0.0004, 0.0043]	0.007**	1.10
Partisan → AFD → Att	2,152	2.368*	0.00085**	0.00202	[0.0004, 0.0043]	0.011*	1.11
Partisan → TFD → Att	2,152	11.549 ⁺	0.00009*	0.00099	[−0.0001, 0.0028]	0.096 ⁺	0.55
Str Dem → MFD → Att	809	16.525**	0.00131*	0.02166	[0.0028, 0.0515]	0.016*	1.73
Str Dem → AFD → Att	809	17.448*	0.00113*	0.01977	[0.0020, 0.0487]	0.021*	1.58
Str Dem → TFD → Att	809	57.370	0.00004	0.00217	[−0.0090, 0.0150]	0.743	0.17
<i>All waves, moderated (partisan × issue position)</i>							
Partisan×IP → MFD → Att	2,152	5.294*	0.00082**	0.00432	[0.0004, 0.0100]	0.024*	3.41
Partisan×IP → AFD → Att	2,152	5.826*	0.00076**	0.00444	[0.0004, 0.0105]	0.030*	3.51
Partisan×IP → TFD → Att	2,152	61.897***	0.00007 ⁺	0.00453	[−0.0002, 0.0112]	0.058 ⁺	3.57
<i>W1+2, simple</i>							
Partisan → MFD → Att	1,954	2.558**	0.00063*	0.00160	[0.0001, 0.0039]	0.041*	0.81
Partisan → AFD → Att	1,954	2.639**	0.00058*	0.00153	[−0.0000, 0.0038]	0.051 ⁺	0.78
Partisan → TFD → Att	1,954	11.745 ⁺	0.00007 ⁺	0.00086	[−0.0002, 0.0026]	0.138	0.43
Str Dem → MFD → Att	712	18.587**	0.00093 ⁺	0.01724	[−0.0033, 0.0477]	0.108	1.27
Str Dem → AFD → Att	712	18.981**	0.00081	0.01542	[−0.0041, 0.0453]	0.139	1.14
Str Dem → TFD → Att	712	57.315	0.00002	0.00112	[−0.0129, 0.0134]	0.910	0.08
<i>W1+2, moderated (partisan/str_dem × issue position)</i>							
Partisan×IP → MFD → Att	1,954	7.653**	0.00068*	0.00521	[0.0006, 0.0115]	0.020*	4.27
Partisan×IP → AFD → Att	1,954	8.452**	0.00061*	0.00519	[0.0004, 0.0118]	0.031*	4.26
Partisan×IP → TFD → Att	1,954	78.454***	0.00007 ⁺	0.00570	[−0.0002, 0.0133]	0.062 ⁺	4.68
Str Dem×IP → MFD → Att	712	73.696***	0.00078	0.05778	[−0.0221, 0.1582]	0.180	5.06
Str Dem×IP → AFD → Att	712	80.028***	0.00068	0.05442	[−0.0283, 0.1596]	0.225	4.77
Str Dem×IP → TFD → Att	712	550.988***	−0.00001	−0.00523	[−0.1002, 0.0814]	0.843	−0.46

Note. Person-level mediation (Baron-Kenny). Path *a* is the effect of the political predictor on fixation duration; path *b* is the effect of fixation duration on image attitude controlling for the political predictor; *ab* is the indirect effect. Bootstrap confidence intervals and *p*-values based on 2,000 nonparametric bootstrap replications (percentile method). All specifications include demographic controls (gender, age, education, income, race) and fixed effects for device and wave. HC1 robust standard errors. “Str Dem” restricts the sample to Strong Democrats (partisan = 7) and Strong Republicans (partisan = 1). “×IP” denotes the moderated specification where path *a* includes the interaction with issue position. % Med is the proportion of the total effect (*c*) passing through the mediator. ⁺*p* < .10, **p* < .05, ***p* < .01, ****p* < .001.

Table A.15: Proportional Dwell Shares by Party, Underlying Figure 8

Topic	Frame	Dem (%)	Rep (%)	Δ (pp)	p	N_{Dem}	N_{Rep}
Immigration	Humanizing	48.2	47.8	+0.4	0.441	17,887	14,875
Immigration	Depersonalizing	36.5	36.6	−0.1	0.837	17,887	14,875
Gun Control	Victim	35.4	35.2	+0.2	0.951	536	419
Gun Control	Weapons	29.6	30.5	−0.9	0.739	536	419
January 6	Insurrection symbols	65.2	63.6	+1.5	0.582	576	461
January 6	Active rioters	14.7	14.3	+0.4	0.827	576	461
LGBT	Identity-affirming	36.6	35.8	+0.9	0.726	678	487
LGBT	Bystanders	63.3	63.9	−0.7	0.789	678	487

Note. Dwell shares are the percentage of total within-image categorized object dwell time allocated to each theoretically motivated frame. Shares within a topic do not sum to 100% because some categorized objects fall outside the two main frames. Democrats are defined as partisanship ≥ 5 on the 7-point scale; Republicans as ≤ 3 . Δ is the Democrat minus Republican difference in percentage points. p -values from two-sample t -tests. All eight comparisons are statistically indistinguishable ($p > .44$), supporting H3 (holistic processing) over H4 (visual confirmation bias).

Table A.16: Object-Level OLS Underlying Section 3, Individual Frame Interactions

	Immigration		Gun Control		January 6		LGBT	
	Human.	Depers.	Victim	Weapons	Insurr.	Rioters	Ident.	Bystd.
Partisanship	0.118*** (0.014)	0.117*** (0.014)	−0.183*** (0.044)	−0.183*** (0.042)	−0.209*** (0.038)	−0.209*** (0.039)	0.274*** (0.053)	0.275*** (0.052)
Dwell (z)	−0.002 (0.033)	0.062* (0.025)	−0.474** (0.153)	0.440*** (0.119)	0.076 (0.116)	0.027 (0.124)	−0.102 (0.106)	−0.013 (0.139)
Issue position	0.283*** (0.022)	0.285*** (0.022)	−0.221*** (0.066)	−0.237*** (0.064)	−0.561*** (0.085)	−0.558*** (0.086)	0.584*** (0.084)	0.579*** (0.082)
Partisan $\times$ Dwell	0.023*** (0.006)	−0.007 (0.005)	0.098*** (0.027)	−0.075*** (0.020)	−0.011 (0.018)	0.009 (0.021)	0.015 (0.016)	−0.042 ⁺ (0.022)
Intercept	2.876*** (0.066)	2.816*** (0.064)	5.312*** (0.254)	5.434*** (0.254)	5.883*** (0.351)	5.908*** (0.355)	1.536*** (0.280)	1.675*** (0.270)
N	38,995	38,995	953	953	1,036	1,036	1,162	1,162
R^2	0.249	0.247	0.255	0.252	0.425	0.426	0.539	0.546
Image FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. Dependent variable is image attitude (1–7). Dwell is z-scored within each frame. All specifications include image fixed effects (coefficients suppressed) and issue position as a control. Partisanship is the 7-point scale (1 = Strong Republican, 7 = Strong Democrat). ⁺ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.17: Object-Level OLS Underlying Section 3, Contrast Score Interactions

	Immigration Human.–Depers.	Gun Control Victim–Weapons	January 6 Insurr.–Rioters	LGBT Ident.–Bystd.
Partisanship	0.117*** (0.014)	−0.182*** (0.042)	−0.208*** (0.039)	0.275*** (0.052)
Contrast dwell (z)	−0.021 (0.016)	−0.399*** (0.095)	0.022 (0.079)	0.008 (0.053)
Issue position	0.285*** (0.022)	−0.230*** (0.064)	−0.559*** (0.085)	0.580*** (0.083)
Partisan $\times$ Contrast	0.011*** (0.003)	0.073*** (0.016)	−0.011 (0.012)	0.018* (0.008)
Intercept	2.826*** (0.064)	5.349*** (0.253)	5.894*** (0.354)	1.680*** (0.276)
N	38,995	953	1,036	1,162
R^2	0.248	0.263	0.425	0.542
Image FE	Yes	Yes	Yes	Yes

Note. OLS with standard errors clustered at the participant level in parentheses. Dependent variable is image attitude (1–7). Contrast dwell is the difference between z-scored dwell on the positive frame and z-scored dwell on the negative frame (e.g., humanizing minus depersonalizing for immigration). All specifications include image fixed effects (coefficients suppressed) and issue position as a control. Significant contrast $\times$ partisanship interactions in three of four topics (immigration $p < .001$, gun control $p < .001$, LGBT $p = .025$), with January 6 the exception ($p = .371$). $^+p < .10$, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table A.18: Mean Image Attitudes by Tercile of Frame Dwell, Underlying Figure 9

Topic, Frame		Democrats			Republicans		
		Low	Mid	High	Low	Mid	High
<i>Immigration</i>							
Humanizing	Mean	4.152	4.043	4.458	3.357	3.108	3.318
	SD	(1.663)	(1.672)	(1.703)	(1.723)	(1.634)	(1.689)
	N	5,894	5,935	6,058	5,027	4,985	4,863
Depersonalizing	Mean	4.418	4.361	3.886	3.532	3.329	2.912
	SD	(1.725)	(1.701)	(1.588)	(1.715)	(1.686)	(1.593)
	N	5,868	5,991	6,028	5,053	4,929	4,893
<i>Gun Control</i>							
Victim	Mean	3.161	2.917	3.365	4.615	4.399	3.905
	SD	(1.527)	(1.668)	(1.757)	(1.596)	(1.715)	(1.894)
	N	175	180	181	144	138	137
Weapons	Mean	3.615	3.223	2.610	4.248	4.180	4.500
	SD	(1.682)	(1.695)	(1.450)	(1.835)	(1.808)	(1.620)
	N	174	184	178	145	134	140
<i>January 6</i>							
Insurrection symbols	Mean	1.642	1.492	1.437	3.224	3.178	3.204
	SD	(1.177)	(0.924)	(0.782)	(1.652)	(1.796)	(1.676)
	N	190	187	199	156	158	147
Active rioters	Mean	1.571	1.400	1.584	3.556	3.024	2.961
	SD	(0.998)	(0.823)	(1.061)	(1.829)	(1.654)	(1.530)
	N	177	180	219	169	165	127
<i>LGBT</i>							
Identity-affirming	Mean	5.644	5.608	5.680	3.178	3.309	3.094
	SD	(1.365)	(1.568)	(1.444)	(1.595)	(1.871)	(1.831)
	N	237	212	229	152	176	159
Bystanders	Mean	5.667	5.919	5.371	2.940	3.305	3.321
	SD	(1.433)	(1.393)	(1.492)	(1.773)	(1.708)	(1.833)
	N	238	211	229	151	177	159

Note. Mean image attitude (1–7 scale) by tercile of z-scored within-image dwell on the specified frame, separately for Democrats (partisanship ≥ 5) and Republicans (≤ 3). Terciles computed by ranking dwell values within each frame (ties broken by row order). Diverging patterns across terciles (e.g., Gun Control victim, Gun Control weapons, Immigration humanizing) indicate that the evaluative consequence of dwell depends on partisanship. Stable gaps across terciles (e.g., January 6 insurrection symbols) indicate that dwell does not moderate the partisan attitude gap for that frame.

Table A.19: Robustness to Fixation Duration Threshold. Dependent Variable is Mean Fixation Duration at Each Threshold, All Waves

	80 ms	90 ms	100 ms	110 ms
<i>Panel A. Direct effect</i>				
Partisanship (1–7)	2.725*	2.683*	2.726*	2.644*
	(1.100)	(1.090)	(1.082)	(1.071)
<i>N</i>	49,445	49,342	49,138	48,889
Strong Democrat	17.786*	17.473*	17.560*	16.898*
	(7.786)	(7.735)	(7.682)	(7.609)
<i>N</i>	17,793	17,749	17,657	17,537
<i>Panel B. Interaction with issue position</i>				
Partisanship	6.033*	6.008*	5.875*	5.766*
	(2.867)	(2.841)	(2.815)	(2.793)
Issue position	4.217	4.274	3.945	3.956
	(3.839)	(3.808)	(3.774)	(3.748)
Partisan $\times$ Issue pos.	−1.041	−1.040	−1.004	−0.988
	(0.795)	(0.788)	(0.782)	(0.775)
<i>N</i>	49,445	49,342	49,138	48,889
Controls	Yes	Yes	Yes	Yes
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Each column uses a different minimum fixation duration threshold (80, 90, 100, 110 ms) for computing average fixation duration. Controls, image FE, wave FE, and clustered SE (participant level) as in Table A.4. The partisan direct effect is stable across thresholds. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.20: Robustness by Device Type. MFD (ms), 7-Point Partisanship

	Desktop	Smartphone	Tablet
<i>Panel A. Direct effect</i>			
Partisanship (1–7)	4.064** (1.432)	0.029 (0.270)	0.226 (2.221)
<i>N</i>	26,791	8,724	11,539
<i>R</i> ²	0.210	0.029	0.225
<i>Panel B. Interaction with issue position</i>			
Partisanship	6.741 ⁺ (3.737)	1.530* (0.703)	8.355 (5.818)
Issue position	3.603 (4.899)	2.359* (0.963)	11.175 (8.055)
Partisan × Issue pos.	−0.814 (1.023)	−0.417* (0.204)	−2.369 (1.615)
<i>N</i>	26,791	8,724	11,539
Controls	Yes	Yes	Yes
Image FE	Yes	Yes	Yes
Wave FE	Yes ^a	No ^b	Yes

Note. Each column restricts the sample to a single device type. Controls and clustered SE as in Table A.4.

^aDesktop subsample spans all three waves. ^bSmartphones appear only in Wave 1, so wave FE is dropped. The desktop subsample drives the main effect; the smartphone subsample, while showing near-zero direct effect, has a significant interaction. ⁺ $p < .10$, $*$ $p < .05$, $**p < .01$, $***p < .001$.

Table A.21: Robustness by Calibration Quality Cutoff. MFD (ms), 7-Point Partisanship, All Waves

	All (≥ 1)	≥ 3	≥ 4	≥ 5
<i>Panel A. Direct effect</i>				
Partisanship (1–7)	2.468* (0.971)	2.383* (0.991)	2.451* (1.043)	1.613 (1.138)
<i>N</i>	47,054	43,259	40,294	33,291
<i>Panel B. Interaction with issue position</i>				
Partisanship	6.292* (2.553)	4.501 ⁺ (2.636)	4.594 (2.793)	1.637 (3.062)
Issue position	5.324 (3.401)	2.845 (3.519)	3.236 (3.718)	0.263 (4.045)
Partisan $\times$ Issue pos.	−1.125 (0.708)	−0.637 (0.723)	−0.576 (0.764)	0.039 (0.847)
<i>N</i>	47,054	43,259	40,294	33,291
Controls	Yes	Yes	Yes	Yes
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Each column restricts the sample to participants at or above the specified quality grade (1–6 scale, platform calibration rating). The direct partisan effect is robust through quality ≥ 4 but attenuates at quality ≥ 5 , consistent with lower-quality sessions contributing genuine variance in fixation duration.

⁺ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.22: Alternative Political Predictors of Median Fixation Duration (ms), All Waves

	Ideology	Ideology $\times$ IP	Therm. Dem	Therm. Rep
Ideology (1–7)	2.397* (1.141)	6.046* (2.929)		
Issue position		5.133 (3.780)		
Ideology $\times$ Issue pos.		–1.081 (0.829)		
Therm. Democrats (0–100)			0.113 (0.069)	
Therm. Republicans (0–100)				–0.203** (0.067)
<i>N</i>	47,891	47,891	46,756	46,566
<i>R</i> ²	0.392	0.392	0.384	0.385
Controls	Yes	Yes	Yes	Yes
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Each column replaces 7-point partisanship with an alternative political predictor. Ideology is coded 1 = Very conservative, 7 = Very liberal. Feeling thermometers range 0–100 (higher = warmer). Controls, image FE, wave FE, and clustered SE as in Table A.4. The ideology effect mirrors the partisanship effect in direction and magnitude. The Republican thermometer significantly predicts MFD (warmer toward Republicans $\rightarrow$ shorter fixations), while the Democrat thermometer is positive but marginally insignificant. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.23: MFD (ms) Within Partisan Subgroups, All Waves

	Issue position $\rightarrow$ MFD		Partisanship $\rightarrow$ MFD	
	Dem only	Rep only	Dem (5–7)	Rep (1–3)
Issue position	–2.408 (2.568)	2.050 (2.739)		
Partisanship			0.473 (4.055)	6.818 (4.470)
<i>N</i>	21,568	18,245	21,568	18,245
<i>R</i> ²	0.392	0.407	0.392	0.407
Controls	Yes	Yes	Yes	Yes
Image FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes

Note. Democrats defined as partisanship ≥ 5 ; Republicans as ≤ 3 . Neither subgroup shows a significant within-group effect of issue position or partisanship gradient on MFD, indicating that the main effect in the pooled sample is driven by the between-group difference. Controls, image FE, wave FE, and clustered SE as in Table A.4. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A.24: MFD (ms) by Wave, 7-Point Partisanship

	Wave 1	Wave 2	Wave 3
<i>Panel A. Direct effect</i>			
Partisanship (1–7)	1.601 ⁺ (0.911)	3.493* (1.648)	–0.935 (2.851)
<i>N</i>	16,446	26,765	3,843
<i>R</i> ²	0.388	0.158	0.376
<i>Panel B. Interaction with issue position</i>			
Partisanship	5.348* (2.244)	10.210* (4.597)	–7.870 ⁺ (4.779)
Issue position	5.249 ⁺ (2.846)	9.445 (5.848)	–14.671* (7.313)
Partisan × Issue pos.	–1.078 ⁺ (0.651)	–1.854 (1.252)	1.453 (1.590)
<i>N</i>	16,446	26,765	3,843
Controls	Yes	Yes	Yes
Image FE	Yes	Yes	Yes
Wave FE	No	No	No

Note. Each column estimates the model within a single wave. Wave 1 (Lucid, Jun 2023, immigration), Wave 2 (Prolific, Mar 2024, immigration), Wave 3 (Prolific, Aug 2025, multi-topic). Controls and clustered SE as in Table A.4; wave FE omitted (within-wave estimation). Wave 3 pools across four topics, and the negative (nonsignificant) partisan coefficient reflects the sign reversal across topics documented in the per-topic analysis. ⁺ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

A.1 Stage A: Object Detection

Person detection. Person bounding boxes are obtained through a three-pass procedure. The first pass uses YOLOv8m (Jocher et al. 2023), which belongs to the YOLO (“You Only Look Once”) family of neural network models (Redmon et al. 2016) that process an entire image in a single forward pass to simultaneously locate and classify objects. On the COCO benchmark, the standard evaluation dataset for object detection containing 80 categories, YOLOv8 achieves a mean average precision (mAP) of 53.9% across all categories in its largest variant, and substantially higher accuracy on the person class specifically, because persons are large, visually distinct, and overrepresented in the training data. The first pass restricts the model to the COCO `person` class and runs with a confidence threshold of 0.25¹⁸; person detections with confidence below 0.30 are additionally flagged as low-confidence for downstream visual review but are not dropped at this stage. The output forms the base set of person detections.

A second pass uses YOLOv8m restricted to the COCO `person` class with a permissive confidence threshold of 0.20 and IoU non-maximum-suppression threshold of 0.50 to recover persons missed by the first pass. Candidate detections from the second pass are merged with the base set only if they do not duplicate an existing person box (pairwise IoU < 0.40).

A third pass uses Gemini (Gemini Team, Google 2024) to identify persons missed by both YOLO passes. Gemini is shown the image together with the already-detected bounding boxes overlaid, and is prompted to flag persons not covered, with explicit attention to figures that are lying or sleeping, children, babies on backs, persons under blankets, and partially occluded foreground figures. Gemini returns normalized bounding-box coordinates with a candidate label that must begin with one of `man`, `woman`, `boy`, `girl`, `child`, `baby`,

¹⁸The confidence threshold sets the minimum model score required for a proposed detection to be retained. We use a low threshold of 0.10-0.30 in this paper because YOLO and YOLO-World are an open-vocabulary detector and may assign lower confidence scores to topic-specific objects than closed-set detectors. The IoU non-maximum-suppression threshold controls how much two predicted boxes can overlap before the lower-scoring box is removed. An IoU threshold of, for instance, 0.50 keeps the highest-confidence box among substantially overlapping detections while retaining distinct nearby objects.

toddler. The returned coordinates are merged with the existing person set under the same IoU-duplication rule (pairwise IoU < 0.40), and detections smaller than 0.5% of the image area are discarded.

Topic-specific object detection. Non-person objects relevant to each political topic are detected by YOLO-World (Cheng et al. 2024), an open-vocabulary detector that pairs an image encoder with a text encoder trained on millions of image-caption pairs. The model matches visual regions against researcher-specified text prompts, eliminating the need for labeled training data for each new object type. For a discussion of open-vocabulary and zero-shot methods in political image analysis, see (Joo and Steinert-Threlkeld 2022).

The text prompts supplied to YOLO-World for each image are formed as the union of a topic-specific base list and a per-image augmentation generated by Gemini, meaning we passed all images through Gemini and asked for key words. The topic-specific (human objects are excluded) base lists are reported verbatim below.

Immigration (migrants):

flag, banner, fence, barrier, wall, truck, bus, car, motorcycle, tent, shelter, backpack, water bottle, baby stroller, sign, bag, suitcase, razor wire, barbed wire, blanket, sleeping bag, mattress, Honduran flag, Guatemalan flag, Salvadoran flag, American flag, border wall, border fence, palisade fence, metal fence, military, chain link fence, concertina wire, stroller, wheelchair, police, baby carrier, barbed wire fence, metal panel, metal gate, military, trees, sky, mountain, grass, ground, dirt road, paved road, river, water, clouds, bushes, hill, crowd of people, police, crowd of people in background, building, house, bridge

Gun control (guns):

rifle, handgun, pistol, gun, firearm, weapon, shotgun, AR-15, ammunition, magazine, bullet, holster, scope, trigger guard,

gun barrel, gun grip, ammunition box, gun safe, gun rack,
gun case, target, shooting target, shooting range, range bench,
earmuffs, ear protection, safety goggles, shooter stance,
gun show table, table, floor, wall, sign, banner, flag,
American flag, trees, sky, grass, ground, building, room,
indoor scene, outdoor scene, crowd of people, group of people,
child holding gun, boy holding gun, girl holding gun,
child with rifle, small hands, child hands, kid hands,
text on shirt, logo on shirt, writing on shirt, print on shirt,
graphic tee, earplug, earplugs, ear protection device,
trigger, muzzle, firing position, aimed weapon

January 6 (jan6):

American flag, Trump flag, MAGA flag, Trump banner, Trump 2020 flag,
Confederate flag, Gadsden flag, Don't Tread on Me flag, patriot flag,
MAGA hat, red MAGA cap, baseball cap, helmet, tactical helmet,
gas mask, Capitol building, Capitol dome, stone pillar, stone steps,
stairs, scaffolding, metal scaffolding, construction scaffold,
platform, police barricade, metal barricade, fence, metal fence,
barrier, tear gas, smoke, pepper spray, baton, stick, pole, flagpole,
riot police, police officer, police shield, riot shield, body armor,
megaphone, microphone, camera, sign, placard, banner,
crowd of people, large crowd, group of people, sky, clouds, building,
text on sign, writing on sign, poster text, banner text, slogan

LGBT rights (lgbt):

rainbow flag, pride flag, LGBT flag, gay pride flag, trans flag,
transgender flag, bisexual flag, nonbinary flag, rainbow banner,
rainbow umbrella, rainbow heart, rainbow elements, rainbow scarf,

rainbow beads, rainbow pin, rainbow flag pole, drag queen costume, feather boa, pride parade float, colorful wig, glitter outfit, face paint, body paint, painted face, pride sign, rainbow sign, LGBTQ banner, equality sign, balloons, rainbow balloons, streamer, confetti, stage, platform, microphone, megaphone, camera, trees, sky, clouds, building, city street, crowd of people, large crowd, group of people, banner, flag, sign, pole, text on sign, slogan, poster text, logo on shirt, graphic tee

The per-image augmentation is generated by a single Gemini call that views the full stimulus image together with the already-detected person bounding boxes and is prompted to list every distinct significant element visible, which include natural features, infrastructure, country flags, vehicles, bags, bottles, signs, background crowds, weapons held by persons, accessories, and any visible text or logos. The prompt explicitly excludes generic clothing labels, body parts, and bare “person” labels. The Gemini-generated list is unioned with the topic base list to form the final set of YOLO-World text classes for that image, ensuring that visually-evident but list-omitted entities (e.g., a Honduras flag in a migrants image, or a yellow ribbon at an LGBT parade) are still detected.

YOLO-World runs with a confidence threshold of 0.10 and IoU non-maximum-suppression threshold of 0.50. The permissive confidence threshold is chosen because the open-vocabulary detector’s calibration varies substantially across prompt types and is filtered downstream by phantom detection and per-crop verification (below). Detections with normalized bounding-box area below 0.003 (approximately 0.3% of the image) are discarded as too small to support reliable attention attribution.

Per-detection verification and labeling. Each YOLO-World detection is forwarded to Gemini for verification. Gemini is shown the cropped bounding-box region (with an 8% padding margin on each side) together with the YOLO-World class name, and is asked to return a Boolean indicator of whether the label correctly identifies the depicted object,

accompanied by a refined free-text label if appropriate. Detections that Gemini flags as incorrect are dropped; detections whose verification call fails (network error, malformed response) are retained only if YOLO-World’s original confidence is at least 0.25. The refined free-text label, truncated to 40 characters and screened for refusal patterns (e.g., “cannot determine”, “insufficient detail”), replaces the YOLO-World class name as the detection’s final caption.

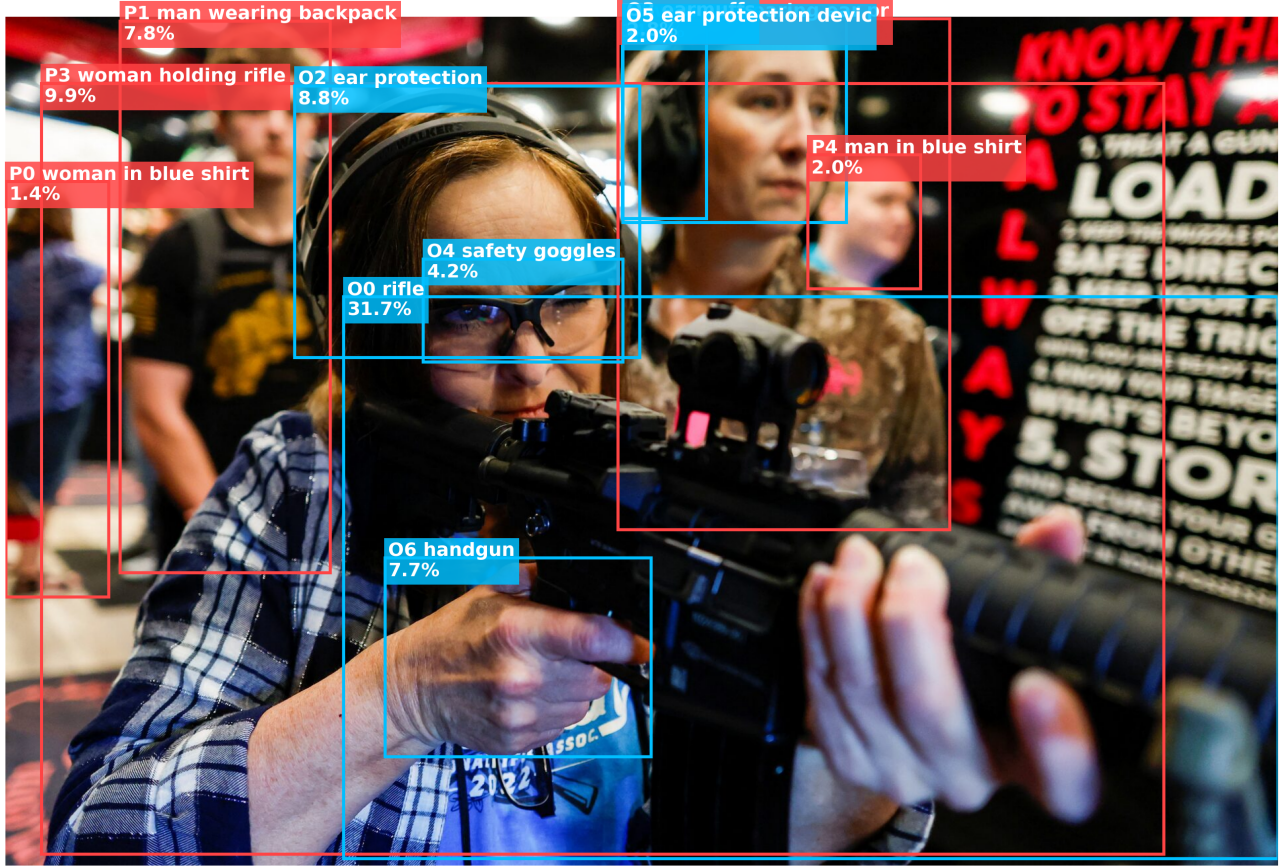
Each person bounding box also receives a Gemini-generated caption. The cropped region is passed to Gemini with a prompt directing it to produce a 3–6 word description starting with one of `man`, `woman`, `boy`, `girl`, `child`, `toddler`, `baby` (selected by Gemini from visible body proportions) and adding a single distinguishing detail such as clothing color, posture, or held object (e.g., “boy holding rifle”, “woman carrying baby”, “man in red shirt”). These person captions are the textual units used as input to the semantic categorization.

Overlap resolution and ghost filtering. Importantly, before moving to the next step, all labels are manually reviewed to verify both the bounding boxes and the assigned labels. The combined set of person and non-person detections is then deduplicated. Object detections (non-person) that overlap substantially with any person detection ($\text{IoU} \geq 0.50$) are removed. These overlapping boxes typically pick up the person’s clothing or held accessory, which the person box already covers. The remaining detections are sorted by label specificity (operationalized as string length excluding generic placeholders such as `person`, `object`, `background`, `thing`), and a greedy IoU-based pass retains the most specific detection in each within-type cluster of pairwise $\text{IoU} > 0.40$.

A second filter removes *ghost-bbox* detections. These detections typically reflect detector artifacts and do not correspond to any discrete entity in the image. A detection is flagged as a ghost-bbox if (i) its label belongs to a topic-specific background category (e.g., `sky`, `ground`, `wall`, `building`, `trees`, `Capitol building in jan6`, `city street in lgbt`) and its normalized bounding-box area exceeds 0.95 (essentially the full image); (ii) its label is not a background or crowd category and its area exceeds 0.50 (more than half the image);

or (iii) at least three of its four normalized coordinates fall within 0.02 of a round value in $\{0, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, 1.0\}$ and the area exceeds 0.30 (a heuristic for whole-image bounding boxes returned at the detector's coordinate-grid boundaries).

Figure A.1: Example Stage A detection output on a guns-topic stimulus.



Note. Red bounding boxes (P0–P4) denote person detections from the three-pass YOLOv8 plus Gemini procedure; cyan bounding boxes (O0–O6) denote non-person object detections from YOLO-World augmented with per-image Gemini prompts and subsequently verified by Gemini per-crop. Each label combines the detection identifier with the Gemini-generated free-text caption (for example, P3 woman holding rifle, O2 ear protection, O4 safety goggles). The percentage beneath each label is the mean blob-attention fraction attributed to that object across the $N = 94$ participants who viewed this image; it previews the Stage D attribution output (Stage D below) and is shown here for completeness, though it is not produced by the detection stage itself. Detection identifiers prefixed P (persons) and O (non-person objects) are assigned in the order detections are appended during the audit pipeline and do not encode any visual or semantic ranking. The example shown contains six retained non-person detections numbered O0, O2, O3, O4, O5, O6; identifier O1 is absent because the corresponding YOLO-World detection was removed during overlap resolution as a redundant object box overlapping a person box (IoU at least 0.50). Two additional detections labelled `room` and `indoor scene` were filtered as ghost-bbox detections (whole-image background bounding boxes with area exceeding 0.95) and likewise do not appear here. The labels for P2 (woman wearing ear protection) and O3 (earmuffs) are visually occluded by overlapping labels for O5 and O2 in the upper-right portion of the image, but both bounding boxes are present in the rendered figure and in the underlying detection data.

Crowd detections (matching any of crowd, crowd of people, group of people, dense crowd, crowd of people in background, plus topic-specific synonyms such as rioters, protesters, mob in jan6 images) are exempted from ghost-bbox filtering on the grounds that crowds genuinely occupy large fractions of these stimulus images. Weapon labels and a small whitelist of small-but-informative items (scope, holster, earplugs, ear protection device, text on shirt, logo on shirt, writing on shirt, print on shirt, graphic tee, text on sign, writing on sign, poster text, banner text, slogan) are also exempted. All flagged ghost-bbox detections are logged for visual inspection before exclusion. Figure A.1 shows an example of the Stage A detection output.

A final pixel-assignment step is applied at the attention-attribution stage, after detection has completed. When two retained bounding boxes spatially overlap, each pixel inside their intersection is attributed to the smaller of the two boxes for purposes of attention scoring (see Stage D below). This implements a specificity heuristic consistent with the standard recommendation that areas of interest in eye-tracking research should not overlap, so that each fixation is attributed to exactly one region (Holmqvist et al. 2011; Hooge et al. 2026; Orquin et al. 2016).

A.2 Stage B — Semantic Labeling

Gemini theory-driven categorization. Stage A produces a free-text caption for each detection (for example, woman holding rifle, ear protection, rainbow flag on parade float). Stage B converts these captions into the 23 theory-motivated semantic categories that feed the framing-level hypotheses in the frame-aggregation step at the end of Stage D. For each stimulus image, Gemini 2.5 Flash receives the full image annotated with Stage A bounding boxes and identifiers, together with the Stage A free-text labels, and in a single batched call returns one category per detection drawn from a fixed closed list of 23 categories (Figure A.2 shows an example on a guns-topic stimulus). The full-image context allows the assignment to depend on cues that lie outside any individual bounding box. A person

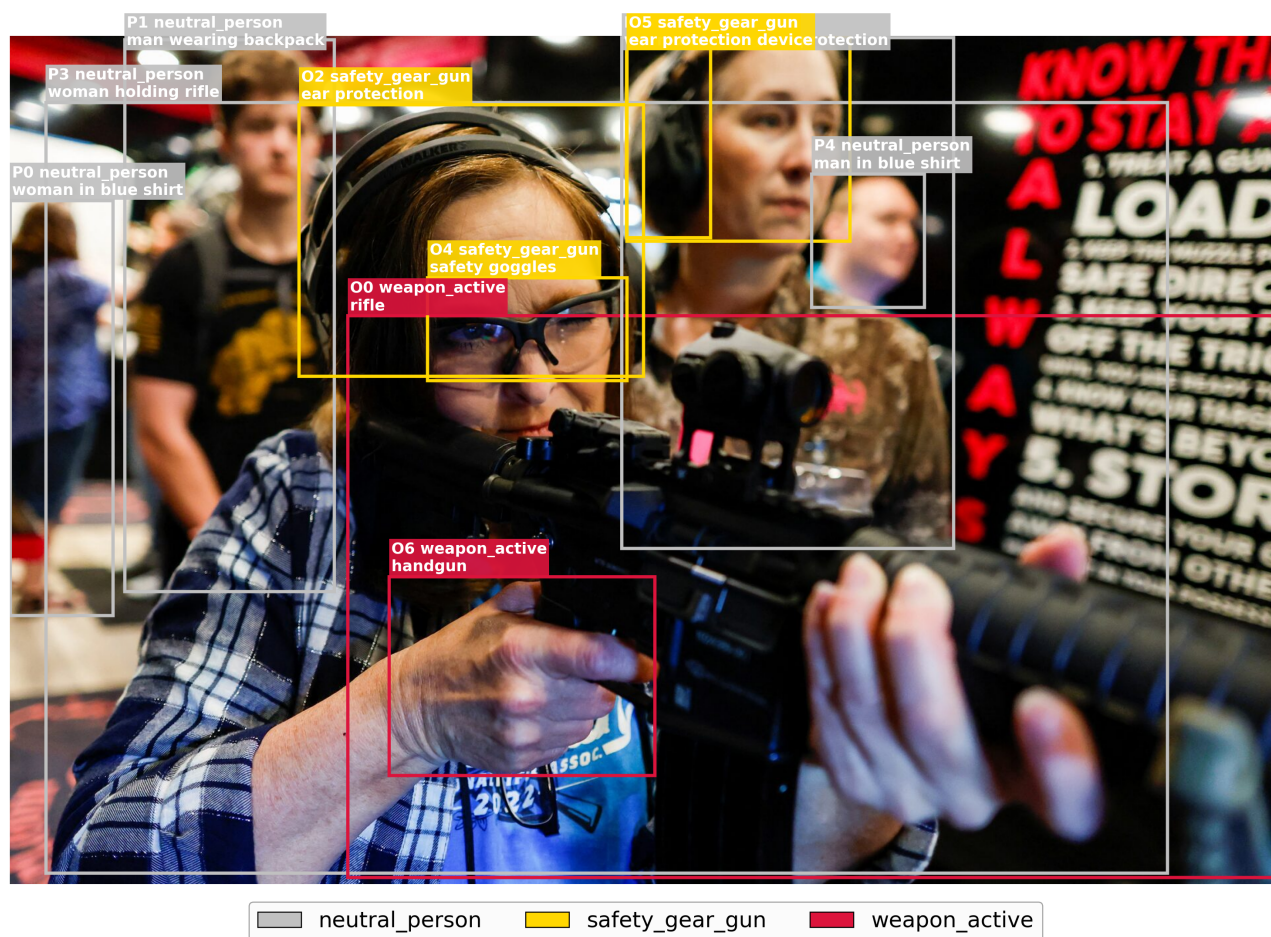
tagged man in tactical gear is assigned `police_uniform` when a police shield or badge is visible on the same person. When no law-enforcement markers are visible the same label maps to `tactical_militant_gear`. A figure described as kneeling woman is assigned `vulnerable_adult` when visible injury or surrender markers are present. When no distress is visible the same label maps to `neutral_person`. A red hat is assigned `political_clothing_partisan` (MAGA proxy) when the image topic is January 6. In immigration, gun control, and LGBT-rights topics a red hat alone maps to `neutral_person`.

Closed category list. The 23 categories partition the conceptual space into eleven theoretically motivated groups, listed with their constituent categories in Table A.25. Vulnerability and humanity cues capture children, distressed adults, and personal belongings carried by people. Mass imagery is collapsed into a single crowd category. Weapons are split by whether the firearm is aimed and being used, displayed at rest, or accompanied by range safety equipment. Flags are split four ways across American, Confederate, Trump or MAGA, and foreign national flags. Gear distinguishes tactical militant attire from law-enforcement uniforms. Partisan civilian clothing is captured in a separate category, with the red-hat-as-MAGA-proxy rule applied only in the January 6 topic. LGBT visibility is split between symbolic markers and performative cues. Breach signals are split between active breaching action and the static aftermath of damage. Setting categories cover Capitol-specific architecture in January 6 and border-control infrastructure in immigration. The `protest_sign` category captures placards and banners with writing. Persons whose label carries no specific theoretical signal default to `neutral_person`, and detections that fit no group default to `unmapped`.

Why a vision-language model for this step. The choice of Gemini at this stage is motivated by three considerations that follow from the design of the visual stimulus set. First, the same Stage A label can carry different theoretical meaning across topics; a category assignment that uses whole-image context is needed to disambiguate. Second, person-attached attributes that are not part of the Stage A label, such as a police shield held by a person tagged man in tactical gear, become visible to Gemini through the bounding-box overlay

even when they were missed at the labeling step. Third, the closed list constrains the model to a finite, theory-anchored vocabulary, which makes the output auditable category by category and reproducible up to the residual stochasticity of the underlying model. We address that residual stochasticity in the limitations section of the main text.

Figure A.2: Example Stage B Gemini categorization on the same guns-topic stimulus as Stage A.



Note. Bounding boxes are coloured by the theory category assigned by Gemini given the full image and the Stage A label. Labels combine the detection identifier with the assigned category and the Stage A free-text caption (for example, O0 weapon_active - rifle, O2 safety_gear_gun - ear protection). The legend lists only the categories present in this image. The full 23-category closed list with definitions appears in Table A.25.

Supplementary detections. Stage A includes a Gemini missing-persons pass and a Gemini per-image augmentation of the YOLO-World prompt list. These supplementary detections

enter Stage B in the same way as YOLO-detected objects and the three-pass person detections. Stage B does not introduce new bounding boxes; it categorizes those produced by Stage A. All Stage A labels and bounding boxes are manually reviewed before Stage B is run.

Table A.25: Stage B closed list of 23 theory-motivated categories used to convert Stage A free-text labels into framing units. Topic-conditional rules are noted where applicable.

Category	Definition
child	Anyone visibly a minor (child, boy, girl, baby, toddler, infant).
vulnerable_adult	Adult showing distress, injury, surrender, or mourning markers (raised hands, hands on head, lying down, crying, wounded, bleeding, “don’t shoot” gesture, carrying child or baby).
personal_belongings	Bags, backpacks, tents, blankets, sleeping bags, strollers, water bottles, baby carriers, hanging laundry.
crowd_mass	Large groups, caravans, gatherings of people, dense crowds, sea of people, mobs, marching protesters.
weapon_active	Firearm aimed, pointed, fired, or held in shooter stance. Bear spray actively used as an assault weapon.
weapon_passive	Firearm displayed, holstered, on a table or rack, at a gun show, on a shooting-range bench.
safety_gear_gun	Ear protection, earmuffs, earplugs, safety or shooting goggles, shooting-range targets.
american_flag	US flag or US-flag-patterned clothing.
confederate_flag	Confederate flag or banner.
partisan_flag_trump	Trump, Trump 2020, MAGA, or Women-for-Trump flag or banner.
foreign_flag	National flag other than US (Honduras, El Salvador, Guatemala, Mexico, Venezuela, Nicaragua, etc.).

Table A.25 continued from previous page

Category	Definition
tactical_militant_gear	Body armor, tactical vest, gas mask, tactical helmet, riot gear, plate carrier, balaclava, camouflage. In January 6 this captures rioter gear; in immigration this also captures border-patrol tactical gear (state-control signal).
police_uniform	Officer with visible police uniform, shield, badge, duty belt, or Capitol Police marking.
political_clothing_partisan	MAGA hat or cap, Trump-branded apparel, Confederate clothing. The red-hat-as-MAGA rule is applied only in the January 6 topic; in immigration, gun control, and LGBT a red hat alone is <code>neutral_person</code> .
lgbt_symbol	Rainbow, pride, trans, bisexual, or nonbinary flag, rainbow scarf or beads or paint, pride sign, LGBTQ banner, equality sign.
drag_performance	Drag queen, feather boa, sequined or glitter outfit, elaborate parade costume, painted face, pride parade float.
breach_act	Active climbing, smashing, breaking, breaching, crossing border or river, jumping fence, scaling wall.
breach_aftermath	Broken glass, glass debris, broken or shattered window, smashed door panel, smoke, tear gas, overturned barricade.
capitol_architecture	Capitol building, dome, stone pillars, stone steps, scaffolding, rotunda, colonnade, historical paintings. Valid only in the January 6 topic.
border_infrastructure	Border wall, border fence, palisade or chain-link fence at border, razor wire, barbed wire, concertina wire, border-control gate.
protest_sign	Placard, banner, poster, or sign with writing. Object-only category.
neutral_person	Person whose label and image evidence carry no specific theoretical signal.
unmapped	Detection that fits no category in the closed list. Background, sky, ground, and generic buildings (other than Capitol architecture) default here.

Stage C — Fixation Cluster Detection

This stage converts each participant’s raw gaze sample stream into a discrete set of fixations. The input is denoised raw gaze data from the eye-tracking platform, which for each (participant, stimulus image) trial contains a JSON-encoded list of triples $[x, y, t]$ produced by the webcam-based eye-tracker. Coordinates x and y are reported in percent of screen width and height; t is milliseconds from trial onset.

Webcam-based eye-tracking carries spatial offsets on the order of 4 degrees of visual angle, against 0.1–0.5 degrees for commercial infrared trackers sampling above 120 Hz¹⁹ (Kaduk et al. 2024; Yang and Krajbich 2021; Semmelmann and Weigelt 2017). Spatial noise at this magnitude rules out laboratory-grade scan-path analysis. It is compatible with the object-level attribution we perform downstream, where the question is which person, object, or background region in the stimulus a participant looked at.

Identification by dispersion-threshold (I-DT). Fixations are extracted from the raw sample stream using the dispersion-threshold algorithm of (Salvucci and Goldberg 2000). The algorithm walks through the time-sorted samples and accumulates a growing window. At each step we compute the dispersion of the window as the sum of its horizontal and vertical extents,

$$\Delta = (x_{\max} - x_{\min}) + (y_{\max} - y_{\min}),$$

where x and y are expressed as fractions of the screen, so $\Delta = 0.05$ means the points span a combined horizontal-plus-vertical extent of 5% of the screen. The window expands while $\Delta \leq 0.05$ (the dispersion threshold) and closes the moment the next sample would push

¹⁹One degree of visual angle corresponds to roughly the width of a thumbnail held at arm’s length, and to roughly 35–40 pixels on a typical participant monitor viewed at 60 cm. A 4-degree offset means a raw gaze estimate may land on the wrong face in a two-person scene, or on background sky adjacent to a person’s head, even when the participant is fixating reliably. A 0.1–0.5-degree offset, characteristic of commercial infrared trackers, keeps the estimate well within the boundaries of an individual facial feature. Sampling rate matters because commercial infrared trackers sample at 120–1000 Hz, which permits direct measurement of saccade trajectories and microfixations, whereas webcam trackers typically sample at 25–30 Hz, which suffices for fixation localization. Our analysis operates at the level of object categories and frames (Stage D below), at which webcam-grade accuracy is sufficient.

it past this threshold. A closed window is recorded as a fixation if it contains at least two samples and its temporal extent (last sample time minus first sample time) lies between 100 ms and 2000 ms; windows outside these duration bounds are discarded as saccades or as tracking losses. Each recorded fixation stores the mean x and y of the samples in its window, the timestamp of the first sample, the fixation duration, and the sample count.²⁰

The output of Stage C, for each (participant, stimulus) trial, is the list of fixations identified by I-DT. These per-trial fixation lists feed directly into the object-assignment step in Stage D below.

Stage D — Attention Assignment

Stage D takes the per-trial fixation lists from Stage C above and the bounding-box detections from Stage A (with Stage B supplementary detections where applicable) and produces the participant-level object dwell measure used as the building block for the frame attention computed in the frame-aggregation step below.

Fixation-to-object assignment. Each fixation produced by Stage C carries a normalized centre coordinate (x, y) in screen fractions. For each stimulus image, the bounding boxes for that image are sorted by normalized area in ascending order, so the smallest detection is checked first. A fixation is assigned to the first detection whose bounding box contains its centre. Fixations that fall outside every detection are recorded as `unassigned`. This smallest-containing-box rule ensures that a fixation landing on a child held by an adult is attributed to the child detection, that a fixation on a rifle held by a shooter is attributed to the rifle detection, and that a fixation on a person standing inside a crowd region is attributed to

²⁰The 0.05 threshold is calibrated to the empirical spatial accuracy of webcam-based gaze estimates and the foveal radius implied by typical viewing distances. A tighter value of 0.025 was tested during pipeline development and produced visibly fragmented fixations in trials where genuine fixations spanned more than 2.5% of the screen due to webcam noise. The relaxed 0.05 value is the one used to produce the per-trial fixation lists analyzed in this paper; sensitivity of the fixation statistics to the minimum-duration threshold appears in appendix Table A.29. The 100 ms minimum follows the standard eye-tracking convention; shorter windows are dominated by tracker noise without indicating cognitively meaningful dwell. The 2000 ms maximum catches sustained looking on a single feature while excluding tracker dropouts where consecutive samples collapse to the same coordinate for several seconds.

the person detection. Generously drawn outer boxes do not absorb attention that semantically belongs to a nested object (Hessels et al. 2016). Crowd dwell, in particular, is the residual after individually-detected persons inside the crowd region have already claimed their fixations under the smallest-box rule; no explicit subtraction step is required.

Object-level dwell. For each (participant, stimulus, detection) triple we sum the durations of all fixations assigned to the detection. The resulting per-object dwell, measured in milliseconds, is the unit on which downstream analysis operates. This dwell-per-AOI construction is the standard measure of attention allocation in eye-tracking research (Holmqvist et al. 2011; Hessels et al. 2016; Vansteenkiste et al. 2015), and is the appropriate operationalization under the eye-mind assumption that cognitive processing of a region is reflected in gaze time on that region (Just and Carpenter 1980).

The next step takes the per-object dwell from Stage D and aggregates it into the eight theoretically motivated visual frames described in Table A.26. Each detection enters the aggregation with the theory category assigned by Gemini in Stage B. For each (participant, stimulus) trial, the frame-level dwell is the sum of object-level dwells across all detections whose category belongs to the frame. The dwell share for each frame is the frame dwell divided by total dwell across all categorized objects in that trial. Hypothesis tests in the main text operate on these participant-by-stimulus frame dwell shares, with participants nested in stimulus images and stimulus images nested in topics.

Table A.26: Topical visual frames, constituent Stage B categories, and theoretical motivation. Frames equal the sum of object-level dwells across the listed categories within each topic. The table reports immigration and gun control frames. January 6 and LGBT rights appear in Table A.27.

Topic	Frame	Categories	Theoretical motivation
Immigration	Humanizing	children; vulnerable or distressed adults; personal belongings (bags, clothing, possessions); hand-held protest signs	Children, vulnerable adults, belongings, and protest signs make migrants visible as identifiable persons. These cues fit work on iconic victims, refugee visual dehumanization, and media-induced dehumanization, where individualized suffering can interrupt anonymous mass threat imagery (Slovic et al. 2017; Bleiker et al. 2013; Esses et al. 2013). The frame should draw greater attention from liberal viewers.
Immigration	Depersonalizing	large crowds; border walls, fences, and checkpoints; breach moments where people force entry across a barrier; foreign national flags; tactical or paramilitary gear (helmets, body armor)	Crowds, border barriers, breach cues, foreign flags, and tactical gear make immigration legible through threat, control, and sovereignty. This frame matches research on refugee dehumanization and U.S. anti-immigration discourse, especially the recurring invasion and border-crisis repertoire (Bleiker et al. 2013; Esses et al. 2013; Chavez 2013; Santa Ana 2002; Flores 2003).
Gun control	Victim	children; vulnerable or distressed adults; firearms in active use (aimed or being fired); hand-held protest signs	Child vulnerability, bodily harm, active weapons, and protest signs anchor gun politics within a victim-centered regulatory frame. Haider-Markel and Joslyn show experimentally that public-safety frames invoking school-age children shift gun-policy support, with effects conditional on partisanship and political knowledge (Haider-Markel and Joslyn 2001). Wozniak's post-Sandy Hook survey adds that gun-control preferences are organized most consistently by political beliefs and affiliations (Wozniak 2017), making these victim cues plausible sites of partisan divergence.
Gun control	Weapons	displayed firearms (carried, holstered, or shown without active use); shooting-range safety gear (ear protection, shooting glasses); partisan political clothing (MAGA hats, candidate shirts, party-branded apparel); tactical or paramilitary gear	Displayed firearms, range safety gear, partisan clothing, and tactical gear index ownership practice, armed citizenship, and gun-rights identity. This frame follows work on NRA identity building, gun culture, and the shift toward self-defense-centered Gun Culture 2.0 (Lacombe 2019; Melzer 2009; Yamane 2023). Conservatives may read these cues as identity affirming, liberals as public-safety cues.

Note. The table's frame labels are topic-specific names for visual repertoires that travel across topics. The humanizing repertoire — children, personal belongings, and visibly distressed adults — appears in every topic. It is named the Humanizing frame in immigration, the Victim frame in gun control, the Identity-affirming frame in LGBT rights, and contributes the visible-injury component of the Insurrection symbols frame in January 6. The threat or disorder repertoire — large crowds, tactical or paramilitary gear, and breach acts — appears as the Depersonalizing frame in immigration and the Insurrection symbols frame in January 6. The partisan-identity repertoire — partisan political clothing and displayed weapons — appears as the Weapons frame in gun control and contributes to the Insurrection symbols frame in January 6. A category carries the same kind of visual meaning wherever it occurs; the topic supplies the referent.

Table A.27: Topical visual frames continued from Table A.26. The table reports January 6 and LGBT rights frames.

Topic	Frame	Categories	Theoretical motivation
January 6	Insurrection symbols	breach moments where rioters force entry into the Capitol; breach aftermath (damaged property, broken windows, scattered debris); tactical or paramilitary gear; Confederate battle flags; Trump campaign flags and banners; partisan political clothing (MAGA hats, party apparel); visibly injured or distressed adults	Breach, damage, tactical gear, Trump and Confederate symbols, partisan clothing, and visible injury anchor the insurrection reading of January 6. These cues match scholarship on definitional conflict after the attack and evidence on violent pro-Trump mobilization and militant ties (Zulli et al. 2023; Craig and Albertson 2024; Pape and Ruby 2022).
January 6	Active rioters	large crowds; American flags; hand-held protest signs	Crowds, American flags, and protest signs foreground mass action and patriotic protest symbolism. These are protest-paradigm cues that can support competing riot, protest, and insurrection labels by making different elements salient to different viewers (Kilgo and Mourão 2021; Entman 1993; Zulli et al. 2023).
LGBT rights	Identity-affirming	LGBT visibility symbols (Pride flags, rainbow flags, transgender flag); drag performers in performance settings; children; hand-held protest signs	Pride, rainbow, and trans symbols mark LGBT visibility and rights claims. Drag performance and child presence add the contested post-2020 layer, where parental-rights, drag-story-hour, and groomer narratives recode identity-affirming imagery as child threat. This frame follows work on gay-rights value conflict and recent anti-drag moral panic (Brewer 2003; Marwick et al. 2024).
LGBT rights	Bystanders	generic adults shown without LGBT visibility symbols, drag-performance cues, or partisan markers	Generic persons provide the operational baseline for images without explicit LGBT-coded symbols or performance cues. The contrast with the identity-affirming frame reflects research showing that LGBT-rights attitudes shift through value frames and identity-linked persuasion (Brewer 2003; Harrison and Michelson 2017).

Note. The table's frame labels are topic-specific names for visual repertoires that travel across topics. The humanizing repertoire — children, personal belongings, and visibly distressed adults — appears in every topic. It is named the Humanizing frame in immigration, the Victim frame in gun control, the Identity-affirming frame in LGBT rights, and contributes the visible-injury component of the Insurrection symbols frame in January 6. The threat or disorder repertoire — large crowds, tactical or paramilitary gear, and breach acts — appears as the Depersonalizing frame in immigration and the Insurrection symbols frame in January 6. The partisan-identity repertoire — partisan political clothing and displayed weapons — appears as the Weapons frame in gun control and contributes to the Insurrection symbols frame in January 6. A category carries the same kind of visual meaning wherever it occurs; the topic supplies the referent.

Table A.28: Gemini-Assigned Object Categories (Stage B) and Frame Mapping

Category	Frame assignment	N_{det}	N_{fix}
neutral_person	LGBT bystanders	629	60,664
personal_belongings	Immig. humanizing	140	15,471
child	Immig./Gun/LGBT	103	16,595
unmapped	(not assigned)	104	—
vulnerable_adult	Immig./Gun/Jan6	47	16,380
crowd_mass	Immig./Jan6	38	12,923
tactical_militant_gear	Immig./Gun/Jan6	30	3,196
breach_act	Immig./Jan6	30	6,782
political_clothing_partisan	Gun/Jan6	19	438
partisan_flag_trump	Jan6	15	249
border_infrastructure	Immig.	13	6,558
lgbt_symbol	LGBT	13	257
weapon_active	Gun	10	428
american_flag	Jan6	8	183
safety_gear_gun	Gun	8	199
foreign_flag	Immig.	7	639
police_uniform	(not in frames)	7	1,100
breach_aftermath	Jan6	6	160
weapon_passive	Gun	6	306
protest_sign	Immig./Gun/Jan6/LGBT	5	158
drag_performance	LGBT	4	287
capitol_architecture	(not in frames)	4	119
confederate_flag ^a	Jan6	0	0

Note. N_{det} is the number of unique detected objects assigned to each category across all 79 images (1,246 total detections). N_{fix} is the number of participant-object dwell records in the categorized fixation file (143,092 total). Categories are assigned by Gemini 2.5 Flash from the full image with Stage A bounding boxes overlaid. Each category maps to one or more theoretically motivated frames per topic (see Table A.26). “Unmapped” detections (8.3% of detections) are objects that Gemini could not assign to any predefined category. “Police_uniform” and “capitol_architecture” are recognized categories not used in any frame.

^a confederate_flag is a category in the Stage B closed list (Table A.25) that was detected in no image.

Table A.29: Fixation Statistics at Different Minimum Duration Thresholds

Min. duration	Mean N_{fix}	Mean AFD (ms)	Mean TFD (ms)	N_{valid}
60 ms	13.60	325.3	3,745	50,426
80 ms	11.69	343.2	3,619	50,375
90 ms	11.04	350.3	3,569	50,270
100 ms (default)	10.03	365.4	3,484	50,056
110 ms	9.61	372.5	3,452	49,801

Note. Each row applies a different minimum fixation duration threshold to the platform’s velocity-threshold (I-VT) fixation detection algorithm used for image-level measures (AFD, MFD, TFD). Higher thresholds exclude shorter fixations, reducing the fixation count and increasing mean duration. Total fixation duration decreases because excluded short fixations are not counted. The number of valid trials decreases slightly at higher thresholds because some trials lose all fixations. The main analysis uses the default 100 ms threshold. The object-level pipeline (Stage C) uses a separate dispersion-threshold (I-DT) algorithm with its own parameters documented in Appendix A.1.

Table A.30: Frame Coverage Across Participant-Image Pairs

Topic	Frame	N_{pairs}	Nonzero dwell	Coverage (%)
Immigration	Humanizing	39,673	22,625	57.0
Immigration	Depersonalizing	39,673	17,861	45.0
Gun Control	Victim	955	392	41.0
Gun Control	Weapons	955	363	38.0
January 6	Insurrection symbols	1,037	731	70.5
January 6	Active rioters	1,037	250	24.1
LGBT	Identity-affirming	1,165	626	53.7
LGBT	Bystanders	1,165	942	80.9

Note. N_{pairs} is the number of participant-image observations with at least one categorized object detection. “Nonzero dwell” is the number of pairs where at least one fixation landed inside a bounding box assigned to that frame. Coverage below 100% means that in some trials, participants did not fixate on any object belonging to that frame (dwell = 0 ms for those pairs). All pairs, including those with zero dwell, enter the regression and tercile analyses.

B Full Questionnaire

Consent Form I

The purpose of this research study is to explore how individuals visually perceive images that depict political issues. You are being invited to participate in this research study based on the following criteria: you are over 18 years old, reside in the United States, and possess fluency in written and spoken English.

Participation in this research study is completely voluntary. You have the option to decline participation if you choose to do so. Alternatively, you may initially agree to participate and later change your mind. Additionally, you have the flexibility to discontinue filling out the questionnaire at any time. Please note that if you decide to withdraw from the survey, none of your answers will be recorded. You retain the right to withdraw from participation even after providing consent.

If you agree to take part in this research, you will be requested to complete an online survey. Following that, you will be redirected to an eye-tracking platform where you will view several images and respond to related questions. Your involvement in this study is estimated to require approximately 15 minutes of your time.

You have the option to decline to answer any question that you do not wish to respond to. Moreover, you may choose to stop participating in the survey at any point. Please be aware that only individuals aged 18 or older are eligible to participate. If you are under 18 years old, we kindly ask you to discontinue your participation at this time.

The potential risks associated with your participation in this research primarily involve a breach of confidentiality. To safeguard participants against the risk of breach of confidentiality, all opinions expressed during the survey will be used in a generalized form and only after undergoing computer processing. All information will be stored in an anonymized format without any identifying details.

To protect your confidentiality, no questions will require you to provide information that

could be used to identify you. The risk of disclosing personal information is minimal and is no greater than when using the Internet. Anonymous responses will be securely stored on a password-protected computer in a password-protected location. No identifiable information will be collected, and the research data will not be associated with your name. Furthermore, the researcher(s) will not share your information with any third party. Your name or any other private information will not be utilized in any publications pertaining to this research.

If you have any questions or concerns about your rights as a research subject, please reach out to XXX (project principal investigator) at XXX or XXX. This study has received approval from XXX. If you have any inquiries or concerns regarding your rights as a participant, please contact the IRB office. The IRB administration can be reached at XXX or XXX.

You will be able to participate in the survey if you use a desktop device such as a laptop/PC/Macbook or a tablet!

The IRB Exempt XXX.

Consent Form II on Eye Tracking

There will be an eye-tracking component to this study, which will be conducted on a separate platform.

You will be redirected to the eye-tracking platform after completing the survey part.

The eye-tracking platform doesn't record any visual data (videos, images, sound) while the webcam is on. The platform doesn't need to send any video or images to external servers, as everything happens in a participant's web browser. The calibration process system learns what a participant's face looks like while she is looking at certain parts of the screen. The algorithms that the platform uses can predict gaze point position just by observing your face, and it all happens live in the web browser. No data is being sent anywhere else.

Since we are testing a new eye-tracking tool, it is important for participants to adhere strictly to the platform usability requirements to ensure accurate gazing results.

In this study, we'll ask you to participate in an eye-tracking session using your webcam.

Please:

- use a **desktop device** like a **laptop/PC/Mac(book)** or a **Tablet**,
- use **Chrome web browser**,
- make sure you have a **good quality, working webcam plugged in**,
- make a **webcam very steady** and below your eyes level,
- make sure you're in **good lighting conditions** - the light source is in front of you and illuminates your face evenly,
- don't rush through the calibration process,
- if you wear glasses, make sure there are no reflections on them,
- try to **keep your head still** during the whole test,
- if you're using a laptop, make sure it's **plugged in**.
- (for website testing) please, **share your screen or browser window** (NOT a tab) when asked.

It is crucial to follow the platform's requirements, otherwise, your results will not be valid and you will not be paid.

By clicking "I consent", you confirm that you agree with the conditions and are ready to proceed.

Questionnaire

1. What is your current age?

- Under 18 (excluded)
- 18 - 24

- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 - 84
- 85 or older

2. In which state of the United States do you live? [Drop-down list with all the US states]

3. Generally speaking, do you usually think of yourself as a DEMOCRAT, a REPUBLICAN, an INDEPENDENT, or other?

- Democrat
- Republican
- Independent
- Other party (specify)

4. In the survey flow, if DEMOCRAT: Would you call yourself a STRONG Democrat or a NOT VERY STRONG Democrat?

- Strong Democrat
- Not very strong Democrat

5. In the survey flow, if REPUBLICAN: Would you call yourself a STRONG Republican or a NOT VERY STRONG Republican?

- Strong Republican
- Not very strong Republican

6. In the survey flow, if INDEPENDENT: Would you call yourself a LEANING DEMOCRAT, LEANING REPUBLICAN or INDEPENDENT?

- Leaning Democrat
- Leaning Republican
- Independent

7. How would you describe your gender?

- Female
- Male
- Other

8. Please check one or more categories below to indicate what race(s) you consider yourself to be

- White
- Black or African American
- American Indian or Alaska Native
- Asian/Pacific Islander
- Multi-racial
- Other

9. Please indicate your annual household income

- Less than \$10,000
- \$10,000 - \$19,999
- \$20,000 - \$29,999
- \$30,000 - \$39,999

- \$40,000 - \$49,999
- \$50,000 - \$59,999
- \$60,000 - \$69,999
- \$70,000 - \$79,999
- \$80,000 - \$89,999
- \$90,000 - \$99,999
- \$100,000 or \$149,000
- More than \$150,000

10. Are you of Spanish or Hispanic origin or descent?

- Yes
- No
- Don't know

11. What is the highest level of school you have completed or the highest degree you have received?

- Less than high school
- High school graduate (high school diploma or equivalent including GED)
- Some college but no degree
- Associate's degree in college (2-year)
- Bachelor's degree in college (4-year)
- Master's degree
- Doctoral degree

12. How interested are you in what is going on in the government and politics?

- Not interested at all
- Somewhat not interested
- Indifferent
- Somewhat interested
- Strongly interested

13. We'd like you to rate how you feel towards both Democrats and Republicans on a scale of 0 to 100, which we call a "feeling thermometer." On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other.

- Democrats - sliding scale from 0 to 100
- Republicans - sliding scale from 0 to 100

14. When it comes to politics, would you describe yourself as liberal, conservative, or neither liberal nor conservative?

- Very conservative
- Somewhat conservative
- Slightly conservative
- Moderate; middle of the road
- Slightly liberal
- Somewhat liberal
- Very liberal

15. Do you oppose or support [the following statements]... ? (grid question with a 5-point: Strongly oppose, Somewhat oppose, Neither support not oppose, Somewhat support, Strongly support):

- Sending unauthorized immigrants now living in the United States back to their home country
- Race-based affirmative action programs designed to increase the number of African American and Hispanic students on college
- Same-sex marriage
- When the government restricts right to an abortion
- More state-sponsored programs that improve access to college education

Attention Check

Please click NEXT to proceed to the next part of the survey.

We will now show you a series of images that were used in different news stories. We ask you to answer a few questions regarding your perceptions of each of those photos.

Now we will show you some pictures. For each picture, you need to select ALL the objects shown in it. [Three test images follow here]

Redirect to the Eye Tracking Session Prompt

Please click on the link to proceed to the next part of the survey: [CONTINUE SURVEY](#)

We will now show you a series of images that were used in different news stories about immigration. We ask you to answer a few questions regarding your perceptions of each of those photos.

Eye Tracking Task

[Each Respondent receives 16/31/15 (depending on the wave of the study) images. Each respondent is given 5 seconds to gaze on the image, followed by 1 second break and redirect to a questionnaire about that image. Each image evaluation has the same structure and is presented in the following way]

1. What is your attitude towards the subject(s) that are portrayed in this picture?

- 1 - Extremely Negative
- 2
- 3
- 4 - Neither positive nor negative
- 5
- 6
- 7 - Extremely positive
- Don't know/ Not applicable