When Facts Take Sides

Elite Conflict and the Partisan Capture of Economic Beliefs

2026-07-20

In 1996 the General Social Survey asked whether immigrants take jobs; the parties were indistinguishable: 46 percent of Republicans agreed, 49 percent of Democrats. In 2024 the identical words divide them by over thirty points — a factual claim as party-sorted as gun control, sharpest among those best equipped to know better. Beliefs about unverifiable economic facts, we argue, run in two regimes: uncontested claims meet a shared folk intuition education corrects in both parties; contested ones seat party in its place, and knowledge accelerates rather than checks the displacement. Across five decades and eighteen elections, the informed gap proves no trait of the informed: it reverses with the president's party thirteen of thirteen times and switches on from zero at each issue's onset. It is opening, on schedule, over AI and jobs; our forecast is committed in advance. That gap is a manufactured premium: elite conflict, not ignorance, organizes economic belief.

1 Introduction

In 1996 the General Social Survey asked whether immigrants take jobs away from people born in America, and the parties could not be told apart: 46 percent of Republicans agreed, and 49 percent of Democrats. In 2024 the identical words returned a 33-point gap. The fact now divides the parties about as sharply as gun control — uncorrected for expressive responding, though at the most hostile expressive share the experimental record supports it remains the survey's largest fact gap (SI-B) — and it got there by two dated partisan movements: two decades of Democratic convergence toward the expert direction, then a cue-timed Republican return to the folk pole (Figure 1; Figure 12). Nothing about the claim's checkability changed. What changed is that elites made immigration a party question.

The claim's structure is what made that possible. Whether immigrants take jobs is a factual question — it has an answer — but the answer cannot be observed from ordinary life. The counterfactual is invisible, the causes are diffuse, and the verdict arrives, if at all, only after years. The citizen is a listener weighing a stranger's claim to be a time traveler: nothing available to the listener can check the report, so belief turns not on what is said but on who says it. We call this the Time-Traveler's Problem. Claims with this structure — AI and unemployment, trade and the middle class, immigration and jobs — leave a vacuum, and everything depends on which prior fills it.

The stakes are democratic competence. If beliefs about unobservable consequences track evidence, deliberation can discipline policy. If they track identity, then on exactly the claims experts are surest of, public belief mirrors partisan conflict rather than checking it. We show the latter is increasingly the case, and that it is most true of the citizens best equipped to know better. That inverts the instinctive remedy: on a captured fact the information cure points the wrong way, because the divide is widest exactly where knowledge is greatest — knowledge carries the out-group toward the experts while the in-group holds.

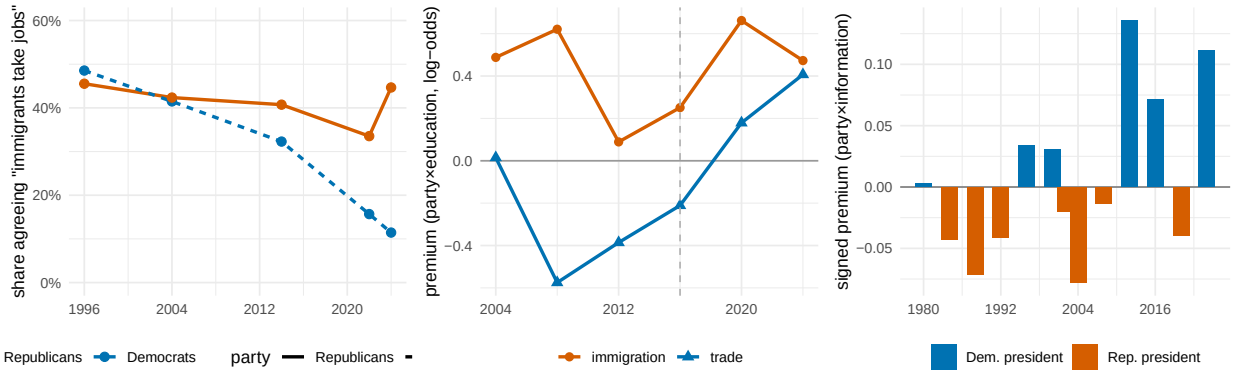
The paper argues that economic belief under unverifiability runs in two regimes. Where elites do not contest an issue, citizens fall back on a shared folk prior — the anti-foreign and anti-market biases of Caplan's taxonomy, which education corrects (Boyer & Petersen, 2018; Caplan, 2007). Where elites contest it, partisan identity displaces the folk prior, and the belief sorts by party. Our central empirical contribution is to watch a single belief cross from the first regime to the second. We hold the survey wording fixed, and we measure the issue's contestation[1] from elite behavior — a century of congressional speech — independently of the mass belief it explains.

Two parts of the argument stand on different footing, and we keep them apart. Unverifiability is the *scope condition*: it makes a fact capturable, and every claim we study sits at the unverifiable end of a coded index (SI-A; §7 estimates the cross-claim margin). Elite contestation is the *engine*: measured from elite behavior, it is the object of every identified result. The Time-Traveler's Problem names the boundary; the empirical work identifies the mechanism operating inside it.

The empirical anchor is the disagreement between the American public and professional economists, which is largest precisely on the claims economists are surest of (Johnston & Ballard, 2016; Sapienza & Zingales, 2013). On unverifiable, high-salience claims the gap is enormous (Table 52). But it is not uniform ignorance: it is sorted, and sorted asymmetrically. As elites contest an issue, the party for whom the expert-consistent position is congenial moves toward it, and the other party does not (Figure 1; §4 details the immigration benchmark).

Three findings make this more than a redescription of polarization. First, a factual belief came to sort like a preference: near zero through 2004, the partisan gap on "immigrants take jobs" reaches roughly 30 points by 2024, as wide as the gap on gun control (Figure 1; Figure 12). Second, the sorting is sharpest among the most sophisticated. The gap widens with political knowledge — the reverse of what an information-deficit account predicts — and it widens one-sidedly: the better-informed out-group moves toward the experts while the in-group holds (Figure 4). Third, the belief travels with consequential commitments: it predicts presidential vote and border-wall support net of party, ideology, and knowledge (Table 28). The qualifications these findings invite are consolidated in Table 1 rather than interleaved.

[1] Throughout, *elite contestation* names the degree to which elite partisan conflict has organized an issue — the mixing weight T of §3. It is elite issue-position divergence, measured from elite behavior (party platforms, congressional speech), and is distinct from affective polarization and from mass partisan sorting, which are its downstream symptoms; every operationalization in §4 is an elite-behavior measure.



(a) the switch-on, by party (b) two issues, two clocks (c) the manufactured premium

Figure 1: Three headline results. (a) The switch-on: shares agreeing 'immigrants take jobs' by party (GSS), from a −3-point Rep−Dem gap in 1996 to +33 by 2024. (b) Two issues, two clocks: the sophistication premium (party×education) switches on at each issue's own elite onset — immigration from the mid-2000s, trade only after the 2016 party-pole flip, with opposite signs in 2008–2012 (descriptive; four survey-years per issue). (c) The manufactured premium, the paper's powered test: the signed party×information premium on economic perception is positive under every Democratic president and negative under every Republican one, 13 of 13 measurable years (estimates and per-year detail in §7 and SI-B).

The object this paper isolates is a sophistication premium with no standing existence of its own, and the contribution is sharpest against the three results it will most be mistaken for. Jerit & Barabas (2012) established that partisan factual gaps are largest among the most knowledgeable, and located the cause in the information environment; what their design observes is the premium at a moment. We track the premium itself as elites assemble and reassemble it: absent before an issue is contested, switched on at each issue's own elite clock, signed and re-signed by the configuration in thirteen of thirteen measurable years, and null where no elite owns a pole — the trajectory that converts "polarization is worst among the informed" from a trait of citizens into a product of elites, and one their environment account can absorb only by making the environment itself a function of the configuration, which is our thesis. Against Zaller (1992): if a single awareness dimension enacted elite cues, it could not send one sophistication measure to both parties and another to only one. Yet that is what we find: education corrects the folk belief in *both* parties, while political knowledge moves only the out-group (Figure 2). Against Bartels (2002): what reverses with the incumbent is not only the perception gap but the sophistication premium itself. One equation delivers both movements. The quantity that does the work is one none of the three accounts predicts, and one issue evolution — a theory of preferences (Carmines & Stimson, 1989) — never needed: a premium that elite contestation itself creates (§3). The same move reconciles two classics. Converse (1964) found that mass publics lack self-generated constraint; Kahan (2013) found that the sophisticated are the most biased. On a captured fact both are right: the constraint is real but imposed from above, and it binds most tightly on those best able to receive the elite signal. The thesis is channel-agnostic. Whether transmission runs through identity-protective cognition or through exposure sorted by the configuration, both are capture, and both defeat the information-deficit remedy (§8). We name the object **the manufactured premium** — manufactured by the elite configuration, a claim about the premium's source, not about the psychology that transmits it.

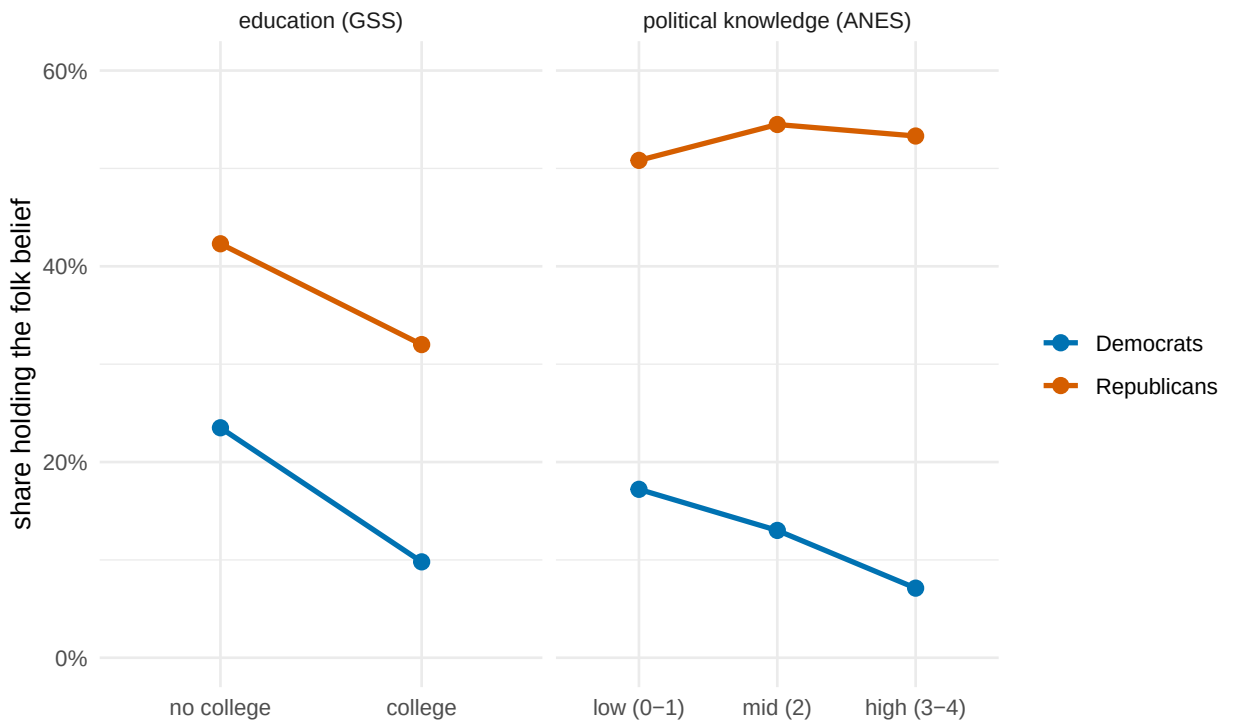


Figure 2: Two channels, two signatures. Left: education (GSS, 2014–2024) corrects the folk belief in both parties — the folk prior is shared. Right: political knowledge (ANES 2024) moves only the out-group — the identity channel is not. A single-dimension reception account predicts neither the pair nor the contrast.

The scope is a claim's *structure*, not its subject: truth-apt, hard to verify, elite-contested. Economics is the cleanest case, but we keep two levels of generality apart. The descriptive pattern travels — it recurs on disbelief in evolution and across most of twenty-six democracies. The sophistication-amplified form is clean only on the two economic facts; the evolution version dissolves under religiosity controls. One boundary inside the window deserves flagging now. The clean one-sided convergence is the law of the issue's *emergence*, 1996–2020. In 2024 the Republican share rebounded to its 1996 level, rewidening the party gap. But the rebound is flat across knowledge and sits on the first net-restrictive Republican platform in the Manifesto series — the signature of a supplied cue, not of capacity (SI-C). The theory also makes a forward prediction: beliefs about AI and employment, now weakly sorted, will partisanize as elite contestation of that issue rises — and its first within-person returns, from a linked Pew panel the two AI cross-sections turn out to contain, already run in the predicted direction with the predicted asymmetry (SI-D.7). The paper develops the model (§3), the design (§4–5), the evidence (§6), and the sophistication gradient and its generalization (§7), before turning to what it all means for convincing citizens of things they cannot check (§8).

The claim we stake the paper on is this: *a factual economic belief acquires partisan structure as elites contest its issue, and does so most sharply among the best-informed — whose partisan gap is set by the elite configuration, not by any standing trait of the informed.* Each half of the evidence supplies what the other cannot. The perception series shows, at full power, that the premium has no standing direction. The immigration series shows that the premium can be entirely absent before elites engage. Only together do they establish manufacture. Table 1 assigns each component claim

its design, its status, and its disconfirmer. Sincerity is bounded against the published incentive record rather than deferred (SI-B); the mechanism claim is channel-agnostic (§3); what remains open is registered, with named falsifiers (SI-E).

Table 1: Claims and evidence: each component claim with its design, data, status, and stated disconfirmer. (Sensitivity of the mechanism claim to unmeasured confounding is quantified in §7 and SI-B.)

Claim	Design	Data	Status	Disconfirmed by
A fixed-wording fact partisanized to value-cleavage magnitude	Within-item, across-era; three instruments, two houses	GSS + ANES, 1996–2024 (N = 33,031)	Replicates across instruments and houses (wild cluster bootstrap p = .001; all design forks positive)	Failure to replicate across instruments
Capture is content-neutral	Gap flips sign with incumbency	ANES perception, 18 elections	Holds in 18 of 18 elections (r = .84)	A fixed-sign gap
The manufactured premium (powered form)	Signed premium × incumbency, signs stated in advance	13 information measure years, 1980–2024	Holds 13/13 (exact p < .001; +0.054/SD, R^2 = .69)	Premium failing to reverse with the configuration
The switch-on form	Premium from zero on one stable belief; staggered event study	ANES CDF 2004–2016; 7 issues	Mass-gap form powered (DiD +0.34, placebo RI p = .003); premium form triangulated (2004 null + opposite-sign clocks + GM placebo, four survey-years)	Premium present before elite contestation
Within-person direction (party→belief)	Latent CLPM, reliability estimated	VOTER panel, 2011–2020	Identified against the reliability artifact (+0.55 vs +0.08); lead–lag, not causation	A verbatim-fact panel with belief leading
The converse case	Mirrored premium where the expert pole is Republican-congenial	GSS nuclear / GM / climate battery	Holds on nuclear (+0.10, p = .0002; GM null placebo)	Unmirrored convergence

Table 1: Claims and evidence: each component claim with its design, data, status, and stated disconfirmer. (Sensitivity of the mechanism claim to unmeasured confounding is quantified in §7 and SI-B.)

Claim	Design	Data	Status	Disconfirmed by
Sincerity	Partial identification in the expressive share λ (calibration bound)	Incentive-experiment record	Set-identified: switch-on survives every $\lambda \leq 0.8$; residual 15–17 pts	Knowledge-graded rebound; persistence after cue withdrawal
Comparative generality	Structural replication	Evolution (GSS); ISSP, 26 democracies	Descriptive replication; sophistication form a bounded non-test with ISSP instruments (SI-D.3)	Gap tracking fragmentation, not folk-pole ownership
The forward prediction (E.1)	Linked within-person panel on the arriving claim; forecast priced in advance	Pew ATP W27→W152 (n = 535), 2017–2024	First returns: party leads the arriving AI belief (+0.121, p = .002; reverse half the size); within-panel gap −1 → +12; forecast committed (25 pts at onset+4, 37 at onset+8, ±8 absent onset)	Onset without gap/premium movement; or the gap exceeding the band with no onset

2 Related literature

The puzzle — why citizens hold party-sorted beliefs about economic *facts* they cannot check, sharpest among the best-informed — sits at the intersection of five literatures, each owning a piece and none assembling the whole. That the public diverges from economists is established (Gordon & Dahl, 2013; Johnston & Ballard, 2016; Sapienza & Zingales, 2013), as is the within-profession agreement that makes the benchmark meaningful (Van Gunten et al., 2016), and that partisans inhabit dueling factual worlds more broadly (Marietta & Barker, 2019); what that literature does not ask is how a gap's *partisan* structure changes over time on fixed wording. Folk economics (Boyer & Petersen, 2018; Caplan, 2007) explains why the public errs in a shared folk-economic direction — not why the same bias is shared on automation and partisan on immigration; the two-regime model supplies the moderator.

Issue evolution and the partisan sort (Carmines & Stimson, 1986, 1989; Layman & Carsey, 2002; Levendusky, 2009) are the closest relatives; the seam is that preferences carry no truth value, and that issue evolution is silent on sophistication. When we say elites "supply an owner for one pole" of a claim, the ownership language is Petrocik's (Petrocik, 1996), transposed from issues a party is trusted to handle to poles of a factual claim a party comes to occupy. Party cues and motivated

reasoning supply the mechanism (Bisgaard, 2019; Bullock, 2011; Cohen, 2003; Gaines et al., 2007; Kahan, 2013; Lenz, 2012) — with the caveat, taken seriously in §5, that sophistication interactions cannot by themselves license cognitive readings (Druckman & McGrath, 2019; Tappin et al., 2020). Three prior findings bound the claim most tightly. Jerit & Barabas (2012) show partisan factual gaps *largest among the most knowledgeable*; we add the *origin* — their premium is a standing property of the environment at a moment, ours switches on at elite onset and is absent before. Shani (2006) shows that political knowledge fails to correct, and can amplify, partisan bias in political perceptions; and Gerber & Huber (2010) show that partisan economic assessments move with political control and carry behavioral weight. Because perception bias flips with the incumbent (Bartels, 2002), knowledge-amplified bias flipping alongside it is implicit in that lineage; what no prior design delivers is the *signed premium itself as the tracked object* — measured over four decades, absent before contestation, reversing 13 of 13 times with the configuration, its two newest points under sign predictions stated in advance (§7).

Climate is the domain precedent — a factual gap widening with asymmetric education amplification (McCright & Dunlap, 2011), elite supply conditioning which facts polarize (Merkley & Stecula, 2021; Tesler, 2018), politicization itself degrading the uptake of the science (Bolsen & Druckman, 2015) — a level-and-trend finding this design completes with fixed wording, an expert benchmark, a pre-contestation null, and the elite clock. A rival engine in that domain, polarized trust in science itself (Gauchat, 2012; Merkley, 2020), is answered here by design rather than assumption: the nuclear-power cell of §7 has educated Republicans converging *on* the expert position, which generalized expert-distrust cannot produce. The sincerity debate (Berinsky, 2018; Bullock et al., 2015; Graham, 2023; Peterson & Iyengar, 2021; Prior et al., 2015) is met with partial identification (SI-B); the persuasion work (Meyers et al., 2020; Stantcheva, 2021) tells us what moves belief in the folk regime and, by its silence on contested claims, what is at stake once a fact is captured.

What survives subtracting all five is the cross-derivative: a sophistication premium that is not a standing trait of the knowledgeable but manufactured by the elite configuration — absent before elites contest an issue, reversing where the configuration flips (§7) — a second-order dynamic neither a static knowledge-gap nor a divergent-diets account predicts.

3 Theory: belief under unverifiability, capture by elite contestation

Consider a citizen evaluating a truth-apt economic claim with verifiability $v \in [0, 1]$ — how readily its truth could be checked from ordinary experience. The citizen never weighs the truth τ directly; under unverifiability there is nothing to weigh. She weighs a signal whose precision rises with v. Optimal stated belief is then a precision-weighted mixture, $b^* = \alpha(v)\,\mathbb{E}[\tau \mid x] + (1-\alpha(v))$ prior, with $\alpha(0) = 0$ and $\alpha'(v) > 0$ (assumptions A1–A2 and the derivation in SI-C). What fails at low v is not the taste for truth but the instrument for finding it. The substantive content is the corner: as $v \to 0$ the signal conveys nothing, and belief equals the prior — the time-traveler's predicament made formal. Everything turns on *which* prior fills the vacuum.

The prior is not a single object but a regime-switching mixture of two dispositions. Let f be the **folk prior** — the shared anti-foreign, anti-market disposition that schooling corrects regardless of party (Boyer & Petersen, 2018; Caplan, 2007) — and p the **identity prior**, the position of one's party. The mixing weight is the issue's **elite contestation** T, the degree to which elite partisan

conflict has organized it:

$$\text{prior} \;=\; (1 - g(T))\, f \;+\; g(T)\, p, \qquad g'(T) > 0,\; g(0) = 0.$$

At low T the citizen falls back on the folk prior; as T rises, the identity prior progressively displaces it. The monotonicity $g'(T) > 0$ is not merely assumed. Elite polarization causally increases the weight citizens place on party cues (Druckman et al., 2013), so the model's central moving part has an experimental microfoundation the observational design could not provide on its own. The substitution of p for f as T rises on a claim of fixed wording is the paper's central dynamic.

One further assumption drives the results. Fidelity to the identity prior rises with political sophistication s: capacity lets a citizen recognize the party's position and marshal arguments for it where evidence cannot settle the matter (Kahan et al., 2012; Kahan, 2013). Write the identity term as $p(s)$ with $\partial p/\partial s > 0$, and let capacity *correct* the folk term rather than amplify it. Sophistication then plays opposite roles in the two regimes — amplifying partisan structure where T is high, attenuating the folk belief where it is low. This single asymmetry generates the central mechanism result, its falsification test, and a prediction of which party moves. Where a party's pole already coincides with f, capacity has nothing to add, so the out-party alone bears the amplification — an asymmetry across parties, not the symmetric sharpening of classic identity-protective cognition (§7).

Which party plays which role is fixed *ex ante*: folk-economic theory marks anti-immigration intuition as a shared, pre-political bias (Boyer & Petersen, 2018; Caplan, 2007), placing the Republican pole at f — Republicans anchor, Democrats converge. (The assignment rule generalizes: for every claim the paper studies, the folk pole is the response predicted by the documented pre-political heuristic of its domain — anti-foreign, make-work, and pessimistic bias for the economic claims, dread for nuclear and GM, distance-discounting for climate — codified and validated against the disengaged baseline in SI-A.5, before any partisan series is examined.) The data confirm it. Through 2004 the belief showed no partisan signal (46 versus 49 percent in 1996); as the cleavage opened, Republicans stayed while Democrats left, and pure independents left with them (§6) — the distinctively partisan act is the staying. Identifying the anchor as "whichever party stayed put" would be circular. The non-circular test is a fact whose folk pole is Democratic-congenial and expert pole Republican-congenial, on which Republicans should become the converging out-group. Nuclear-power risk supplies exactly that configuration, and the prediction holds (§7).

The formalism's contribution is narrow but real: the convex combination is standard, but making the three channels (α, g, s) *separable and multiplicative* in one gap equation converts the empirical asymmetry from description into a derived prediction (conditional on $p_R \approx f$), and implies the gap and the premium are the *same object scaled*, both $\propto g(T)$ — the structural signature §7 tests. The model proper is the reception-integrated $g(T, s)$ (§3.1); strict separability $g(T)$ is its post-diffusion limit. Timing results below use $g(T, s)$; level results use the limit.

These features combine into a single expression for the partisan gap. For a citizen of party $j \in \{R, D\}$, substituting the prior into $b^* = \alpha(v)\,\tau + (1 - \alpha(v))\,\text{prior}$ gives

$$b_j^* \;=\; \alpha(v)\,\tau \;+\; (1 - \alpha(v))\big[(1 - g(T))\, f(s) + g(T)\, p_j(s)\big],$$

where $f(s)$ is the folk prior that schooling moves toward the expert target ($\partial\,|f(s) - \tau|/\partial s < 0$) and $p_j(s)$ the party pole that sophistication sharpens. The *partisan gap* — the paper's object — is clean, because the shared folk term cancels:

$$\text{Gap}(s) \;=\; b_R^* - b_D^* \;=\; (1 - \alpha(v))\, g(T)\, (p_R(s) - p_D(s)).$$

The three factors divide the theory's labor: $\alpha(v)$ governs detachment from evidence (P1), $g(T)$ the rise of partisan loading as an issue becomes contested (P2–P4), and s the sophistication premium $\partial\,\mathrm{Gap}/\partial s = (1-\alpha)\,g(T)\,\partial(p_R - p_D)/\partial s$, positive only where T is high (P5). The §7 asymmetry is the special case $p_R(s) \approx f(s)$: when the in-party pole sits at the folk prior, only the out-party travels toward τ as s rises, so the widening is one-sided.

The comparative statics read off the gap equation as five numbered propositions; assumptions A1–A5 and complete derivations are in SI-C, and each proposition names the exhibit that tests it.

Proposition 1 (the gap). *As* $v \to 0$, belief detaches from evidence, producing large public–expert gaps on unverifiable claims. (Table 52.)

Proposition 2 (folk regime). *Where* T is low, divergence from the experts is universal, education-corrigible, and party-free. (Table 53.)

Proposition 3 (identity regime). *Where* T is high, party predicts divergence, and the out-group moves toward the expert-consistent pole. (Table 54, Figure 1.)

Proposition 4 (capture). *As* T rises on a fixed claim, the fact acquires partisan loading and comes to sort like the matched preference. (Figure 22.)

Proposition 5 (mechanism, conditional). *Where* T is high, the partisan gap on the fact widens with sophistication; where T is low it does not — positive at high T, non-positive at low T (zero in the model, mildly negative wherever schooling still corrects the folk belief). This conditional is the model's falsification test. (Table 2, Table 42.)

These five are the comparative statics the data test. The model carries one further implication we hold apart because current data cannot identify it — a **verifiability half-life**: as v rises on a fixed claim, the gap should shrink. Lacking within-claim variation in v, we register it as a prospective test (H4) in the AI study, where deployment makes the same claim progressively checkable (§8).

One refinement is stated now because the perception evidence (§7) will force it. Verifiability has two margins: *experiential* — checkability from one's own life — and *institutional*, checkability through trust-mediated intermediaries. The institutional margin is partially *endogenous to contestation*: as T rises the adjudicating intermediaries are absorbed into the configuration, and effective v collapses toward the experiential margin. A fact published monthly by a statistical agency can therefore behave as a low-v claim for a captured public — economic perception is the paradigm case (§7). What protects a claim is experiential verifiability, not publication.

P5 discriminates the model from the simplest rivals, though not all: a deficit account predicts the gap *shrinks* as people learn and naïve-Bayesian updating predicts convergence, so a gap that *widens* with knowledge contradicts both (Kahan, 2013; Taber & Lodge, 2006). What it does not separate is motivated reasoning from configuration-sorted exposure (Jerit & Barabas, 2012) — the channel question §5 takes up.

The model's relationship to issue evolution is now exact. Issue evolution (Carmines & Stimson, 1989) is the operation of $g(T)$ on *preferences*; we show the same $g(T)$ operating on truth-apt *facts*, and we add the sophistication term $p(s)$ that the theory of preference sorting never required and that distinguishes capture from a knowledge deficit. The scope is then a boundary, not a hedge. The theory concerns truth-apt, low-v, high-T economic claims. It is silent on verifiable facts (where $\alpha(v) \to 1$ and evidence binds), on pure preferences (which carry no truth value to be captured), and on uncontested claims (where f, not p, governs). Within its scope it is generative: it predicts

which facts are *capturable* — the unverifiable and the newly contested — and names the next, AI-and-employment, as that issue rises on T.

3.1 Relation to reception–acceptance

The mechanism lineage is Zaller's. In reception–acceptance (Zaller, 1992), when elites polarize, the aware polarize first, hardest, and in their side's direction — a premium tracking the configuration is what RAS trained readers to expect. Three things separate our account, each evidenced. First, an accuracy stake: RAS beliefs have no right answer, while ours sit against an expert benchmark, so capture exacts an epistemic cost RAS never priced. Second, the folk floor: $f(s)$ generates the education-corrigible, party-free low-T regime (Table 53) and the one-sidedness itself, where RAS predicts two-sided polarization of the aware. Third, two channels: education corrects both parties while knowledge moves only the out-group (Figure 2), a split a single awareness dimension cannot produce. Reception is a component of the model, not a rival to it. Writing $g(T, s)$ with $\partial^2 g / \partial T \, \partial s > 0$ places Zaller's reception function inside the mixing weight. Two RAS-flavored patterns in the data then arrive as derived predictions rather than anomalies: the premium *leads* the mass gap during switch-on, and the premium-to-gap ratio *declines* as diffusion completes (derivations and diffusion evidence in SI-C).

3.2 What would disconfirm the model

Exposure to disconfirmation is collected rather than scattered: six conditions — symmetric sophistication slopes while $p_R \approx f$; the gap emerging first among the *least* informed; a premium failing to track the elite configuration; an unmirrored converse case; a knowledge-graded rebound or persistence after cue withdrawal; AI failing to partisanize as its T rises — would each break the model. SI-E states each with where it was, or will be, put at risk; Table 1 carries the per-claim disconfirmers.

Because a model in which p_j tracks elite supply can be accused of explaining any in-group path after the fact, we state the finite menu of in-group behaviors it permits, and what each requires. The in-group may *anchor* (where its elite pole sits at the folk default, $p_R \approx f$ — the 1996–2020 record); it may *move with a measured cue* (where its elite pole departs from f, the movement knowledge-flat and dated to the cue — the 2022–2024 record); and it may *drift with the common baseline* (where the shared signal moves everyone — what independents index). The model forbids the complements: in-group movement *without* a measured elite cue; knowledge-*graded* in-group movement while $p_R \approx f$; and out-group movement *against* its own elite pole. Every observed path in the paper draws from the permitted menu, and each forbidden cell is where the account would have failed.

4 Data

The argument requires an expert benchmark, public belief on matched claims measured repeatedly over time, and a measure of elite contestation that is not itself a function of public opinion. The expert benchmark is the Kent A. Clark Center (IGM) US Economic Experts Panel (Kent A. Clark Center for Global Markets, 2024), reshaped to one record per economist per sub-question (753

records, seventeen claims) and collapsed to agree/not-agree — the sign of the consensus, not its intensity, is what the comparison requires. Public belief comes from the GSS 1972–2024 cumulative file (NORC at the University of Chicago, 2025); the Pew ATP automation (2017) and AI (2024) waves (Pew Research Center, 2024), the low-contestation comparisons; and the 2024 ANES (American National Election Studies, 2025), which adds a behavioral knowledge battery and consequential outcomes. The CES supplies a high-powered robustness check (Cooperative Election Study, 2025) (SI-B); the within-person dynamic rests on the VOTER panel, 2011–2020 (Democracy Fund Voter Study Group, 2020) (Section 6.5). Coding and verbatim wordings are in SI-A, analytic samples in SI-C (Table 50); mnemonics beginning `VCF` are Cumulative Data File variables, `V24…` 2024 Time Series variables.

4.1 Measuring elite contestation

The primary contestation series is external to the survey: the Democrat–Republican gap in the tone of congressional immigration speech across 140 years (Card et al., 2022a, 2022b) — generated by elite behavior, predetermined relative to the public belief it explains, and able to test whether elite polarization *precedes* mass sorting (Figure 18). No single index can carry a construct this broad, so T is operationalized five ways, each attached to the one result it powers — T cannot be whatever a section needs. Speech tone dates immigration's rise (near-collinear with the calendar; §5). An ordinal cross-claim ranking powers the gradient (P2–P3). On economic perception T is constant and incumbency identifies. The Manifesto platform divergence (Lehmann et al., 2025) measures both issues' signals continuously and dates every onset (§7). And a salient anti-immigration-party dummy powers the cross-national test, flagged as plausibly endogenous. None is clean; they converge, and none carries another's claim.

4.2 Matched pairs, truth-aptness, and availability

Folk claims are paired with the experts' verdict on the same causal question as *conceptual* matches; the inferential weight sits on the public items' dynamics, not the public–expert levels. For immigration no expert poll asks the verbatim question, so the benchmark is set-valued rather than a point: we anchor only the direction of the National Academies consensus (National Academies of Sciences, Engineering, and Medicine, 2017), a direction that is common ground between the field's opposing poles (Borjas, 2003; Card, 2005). Wherever inside the expert set the true effect lies, the folk belief sits outside it. Some respondents may read the claim distributionally. The unambiguously aggregate `immameco` item answers that worry: it partisanizes on the identical trajectory ($r = .95$; Figure 10). That the item is a truth-apt belief rather than a disguised preference is defended behaviorally. Fielded beside its matched preference in ANES 2024, the fact is *more* partisan ($+0.28$, $p < 10^{-7}$) and more knowledge-graded ($+0.12$, $p = .03$); a disguised preference should behave like the preference. Immigration's benchmark is direction-only, its referent arguable at the distributional margin. Economic perception (§7) has the unambiguous benchmark — realized GDP growth and unemployment, against which the gap is demonstrably bias (SI-B). The paper's normative payload therefore rests on the case with the cleanest target, while immigration supplies the switch-on that the constant-benchmark perception series cannot. The Claim Verifiability Index (SI-A; Krippendorff's $\alpha = 0.88$) certifies the unverifiability scope condition, whose cross-claim margin §7 estimates. No restricted microdata are redistributed; every survey is analyzed under its own native weights (SI-F).

5 Methods and identification

The estimand is a *structural change* — how the partisan organization of a fixed, truth-apt belief evolves as elites contest its issue — and contestation cannot be randomized. In its place, three features of the data strip out what a cross-issue comparison cannot: constant item wording, a contestation measure derived from elite behavior, and repeated cross-sections spanning the politicization.

5.1 The within-item, across-era design

Holding a single item's wording fixed differences away the wording confounds of cross-issue comparison: if the same question sorts respondents by party in 2024 but not in 2004, neither the question nor any stable trait of respondents can explain it. The within-item design establishes *that* the belief partisanized; attributing the change to elite conflict rests on the external contestation series — not strictly exogenous (elites anticipate voters), but *predetermined* and no mechanical echo of the belief studied (Gentzkow et al., 2019), with the direction of influence confronted in the precedence analysis. All estimates are survey-weighted logistic regressions with design-based standard errors (Lumley, 2004); the central interaction is reported across the full specification curve (Simonsohn et al., 2020) (SI-B).

5.2 Inference discipline

One rule governs every over-time claim: the unit of replication is the survey-year, not the respondent. The primary estimate is **+0.86 log-odds per decade** (survey-weighted logit of the folk belief on party × year, survey fixed effect, eleven survey-year clusters, $p < .0001$), +0.17 per decade as an average marginal effect (Mood, 2010) (Table 20). Because eleven clusters are too few for asymptotics, the stacked three-instrument, two-house series (33,031 respondents, sixteen clusters) gives **+0.71**, wild cluster bootstrap-t $p = .001$, permutation $p = .026$ (Figure 3). Robustness is uniform: both surveys separately, every leave-one-out, all forty-eight coding forks, instruments co-moving at $r = .95$ (SI-B; SI-C, Table 51).

One inference the design does *not* license: regressing the gap on contemporaneous T (+0.53, $p = .02$; SI-C, Figure 21) cannot separate contestation from time — the two are near-collinear over the window ($r = .91$). The case that the rise is contestation rests on what a bare trend cannot supply, led by the strongest: the perception gap flips sign with the incumbent across eighteen elections (§7), which no monotone trend can do. Precedence, the gradient, and the cross-issue timing corroborate on immigration.

5.3 The discriminating test of the mechanism

Sorting alone does not say *why*. Political sophistication separates the rivals: motivated reasoning predicts the partisan gap *intensifies* with sophistication, while information-deficit and naïve-Bayesian accounts predict it shrinks (Kahan et al., 2012; Taber & Lodge, 2006). What the interaction cannot exclude is *divergent information environments* (Jerit & Barabas, 2012), nor its rational-choice twin: a Bayesian whose configuration has sorted her *trust* — different priors, different source credences, no motivated cognition anywhere — reproduces every cross-sectional signature, which is why sophistication interactions are weak ground for cognitive claims (Druckman & McGrath, 2019;

Tappin et al., 2020). The thesis does not need to win that argument, and we state the position once: a Bayesian whose priors and trust network are supplied by the elite configuration is not a rescue of the information-deficit view but a mechanism of capture — the configuration has annexed the inputs — and both channels defeat the information-deficit remedy. The observational design disciplines the rival where it can (the informed-independents gradient, §7; the content-of-diet control and sensitivity benchmark, SI-B); the accounts separate fully only under randomized signals, which is the registered secondary question (SI-E, E.7). We estimate party×sophistication with a behavioral knowledge battery (ANES) and education (GSS), net of party×education and party×ideology (SI-B). The two measures are not interchangeable proxies for one "sophistication," and their roles are not assigned by fiat. The knowledge battery is, in substantial part, knowledge *of the elite configuration itself* — two of its four items ask which party controls a chamber — so capacity measured this way is definitionally the capacity to locate the party line; its assignment to the identity term $p(s)$ is analytic, not psychological. Education carries no information about the configuration, and its role is calibrated where the model says identity is switched off: in every low-contestation claim we field — automation, AI, GM crops — education moves belief toward the expert pole in both parties while party structure is absent or weak (Table 53, Table 54, Table 3). The roles thus fixed out of regime, the contested claims test the assignment rather than assume it: education is the predicted-weaker, conditional test (Table 42); knowledge is the magnitude test (§7). A third measure closes the triangle: general verbal ability (GSS `wordsum`), a cognitive capacity with no political content, must on this logic behave like education — folk-corrective in both parties, weakly sorting, stable across eras — and not like political knowledge; SI-D.4 (Table 59) reports exactly that pattern, including the sharpest cell: in the contested era verbal ability still moves *Republicans* toward the experts (-0.29, $p < .01$) while political knowledge leaves them flat.

5.4 Composition, replacement, and other threats

The across-era shift is not composition — party×era survives education×era and the full demographic battery, and it is *non-college* Democrats who moved most (Table 35; SI-B) — and not generational replacement: every Democratic ten-year birth cohort declines within itself, by 24 to 40 points (Figure 8), with 98 percent of the decline loading on period. (Replacement does distinctive work at one margin only — the *entry* margin: cohorts entering the electorate after elite onset arrive already sorted, at +25 points against −2 for pre-onset entrants, so capture compounds through configured entry as well as conversion; SI-B, Table 7.) The remaining threats — referent drift, survey mode, local labor markets, knowledge coding, reverse causation — are each closed or bounded in Appendices B and D; none can manufacture a party-*differential* trend on a fixed item.

5.5 Expressive responding

The gravest objection is that we measure expression, not belief (Bullock et al., 2015; Peterson & Iyengar, 2021; Prior et al., 2015). The belief predicts presidential vote and border-wall support net of party and knowledge, and the vote at an odds ratio near six net of seven-point party, ideology, and knowledge together (Table 28; SI-B) — consequence, though not yet sincerity, since a vote report is also a self-report. Rather than defer the question, we treat the expressive share λ as a sensitivity parameter and partially identify the sincere trajectory (SI-B). At every λ the incentive record supports, up to the paid-don't-know ceiling of 0.8, the worst-case sincere gap in 2024 (5.6 points) still exceeds the best-case ceiling in 1996 (3.4): the switch-on itself, not merely the endpoint,

is identified. The accuracy-incentive consensus leaves a residual of 15–17 points, larger than the entire observed gap at any point before 2014. Cheap talk cannot reproduce the structure regardless. Expressive accounts predict flat-or-reversed knowledge gradients and momentary responding, against the observed monotone gradient and durable within-person revision (SI-B). A rebound that became knowledge-graded, or a gap persisting after elite cue withdrawal, would disconfirm sincere capture.

6 Results

The evidence proceeds in the order that builds the argument: the gap exists (P1); it is folk where uncontested and identity where contested (P2–P3); a fact came to be captured, on the elite clock, and came to sort like a preference (P4); and — §7 — the capture intensifies with sophistication. SI-C (Table 49) collects the headline estimates in one table.

6.1 The gap exists (P1)

Where a folk belief can be set against the economists' verdict, the divergence is large and one-directional (SI-C, Table 52): not one panelist endorsed the claim that the 2016 protectionist plan would help the middle class, against a 36-point public endorsement, and on AI and unemployment the public runs seventy-seven percent against thirty-one. Immigration's cell in Table 52 is blank by design. Its benchmark is set-valued (§4): no expert poll asks the verbatim claim, so we record no number, anchor the direction on the National Academies consensus (National Academies of Sciences, Engineering, and Medicine, 2017), and set it against a folk belief held at forty-two percent. The gaps are descriptive context — every claim unverifiable from ordinary experience (P1); the inferential weight rests on the dynamics, which require no benchmark at all.

6.2 Folk where uncontested, identity where contested (P2–P3)

Where an issue is not yet contested, divergence bears the folk signature: on the AI-and-jobs claim (Pew 2024), education does the work party does elsewhere — college odds ratio 0.82 against a modest party 1.38 (SI-C, Table 53) — a shared error that knowledge erodes (Boyer & Petersen, 2018; Caplan, 2007), not yet a partisan possession. Across three claims ordered by elite contestation — automation, AI, immigration — the party odds ratio rises monotonically, $1.04 \rightarrow 1.70 \rightarrow 3.79$, while education corrects throughout (Table 54): identity switching on atop a folk belief schooling continues to erode. The three claims differ in house and period, so the gradient is the suggestive face of the argument; the load-bearing evidence is the fixed-instrument trajectory below, and composition is closed there (party×era OR 3.38 with education×era null; CES n $\approx$ 37,000 battery; SI-C, SI-D.5).

6.3 The partisanization of a fact (P4)

We now hold a single factual claim fixed and watch it cross from one regime to the other. The General Social Survey has asked, in identical wording across two decades, whether immigrants take jobs away from people born in America. The claim is truth-apt, it sits outside the expert set at either pole of the labor-economics debate (§4), and it is a paradigm time-traveler's report: its labor-market counterfactual is unobservable, so the citizen believes or disbelieves it on the strength

of who is asserting it. (The anchored expert direction concerns the aggregate, long-run claim — the reading the unambiguously aggregate good-for-the-economy item isolates, partisanizing on the identical trajectory at $r = .95$ — so the argument requires only that the folk belief overstate net job loss; §4, SI-A.) In 1996 the belief showed no partisan signal: Republicans and Democrats agreed at forty-six and forty-nine percent, a gap of −3 points (Figure 1; SI-C, Table 55). The gap is +1 in 2004, +8 in 2014, +18 in 2022, and +33 in 2024 (Republicans forty-five percent, Democrats eleven). A factual belief that was bipartisan within living memory has become a strongly party-sorted one, by 2024 nearly as sorted as the immigration-policy preference beside it (the convergence result below). It got there by two dated partisan movements — the long Democratic convergence, then the 2022–2024 Republican return to the folk pole — each on the elite instrument's clock (SI-C).

The timing runs elite-first, as issue evolution predicts (Carmines & Stimson, 1989). Elite polarization in congressional immigration speech (Card et al., 2022a) rose roughly a decade before the public's division (contemporaneous correlation 0.66; ten-year lag 0.92; Granger-style coefficient 0.72, $p = .063$, randomization $p \approx .10$; SI-C, Figure 18; SI-B). We report this as corroborative ordering only. Where the argument needs elite-first timing, it rests on the event study's flat pre-onset profile and the subnational clock (§7), not on this correlation; the novelty is in any case what is being reorganized — and, as §7 shows, the order in which the public falls into line. Two artifacts are ruled out directly. The decline is not social-desirability suppression: it is steeper, not weaker, in private web interviewing, and the gap is identical across modes (SI-C, Table 44). Nor is it referent drift, which moves everyone under fixed wording and cannot manufacture a party-differential trend: pure independents decline in lockstep with Democrats (−0.136 vs −0.134 per decade, $p = .88$) while Republicans hold flat against that common baseline (+0.123 per decade net, $p < .001$; SI-B, Figure 9). The common baseline itself is not unexplained noise, either: it tracks the *mean* of the elite signal — congressional immigration tone grew more positive overall while polarizing — at $r = -.86$, while the party gap tracks the *divergence* (SI-B, Table 13), both moments of mass belief moving with the corresponding, measured moments of elite speech. The partisan signature — the in-group anchored at the folk pole, exactly where the model pins p_R — survives whatever moves the baseline.

Capture also compounds through a channel conversion cannot see. Dating each respondent's political entry as birth year plus eighteen, the party gap at *first observation* tracks the elite contestation prevailing at entry: cohorts entering before 1996 arrive at −2 points, 1996–2011 entrants at +7, and entrants from 2012 onward at +25 — sorted on arrival, at nearly the full contemporaneous magnitude — while the 18–29 entry window shows the young already sorted in 2004 (+13 points) when the gap among their elders still sat at −2 (SI-B, Table 7). Conversion of those who once held the shared folk belief and configured entry of those who never did are distinct mechanisms with distinct futures: cue withdrawal can reverse the first but cannot un-socialize the second, an asymmetric-decay prediction registered with E.2.

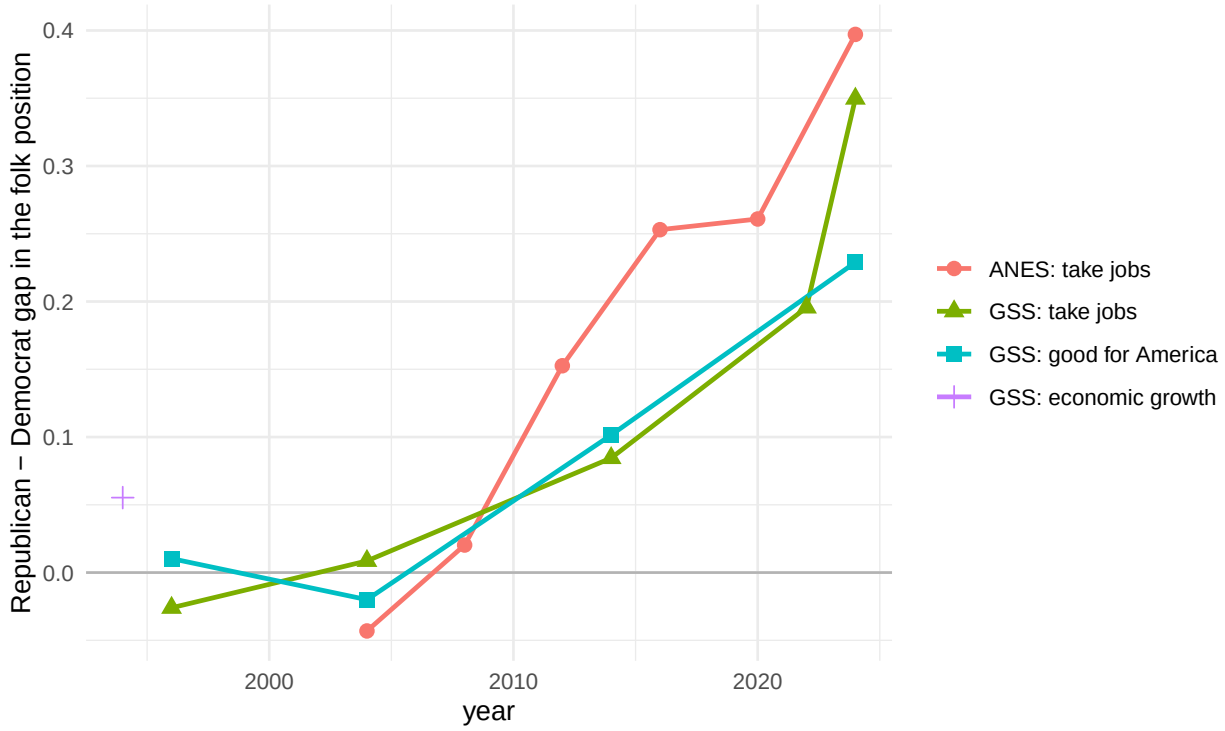


Figure 3: The switch-on replicates: the Rep−Dem gap in the folk position on three independent fixed-wording instruments across two survey houses — every instrument opens after ~2012 (pooled party×year +0.71 log-odds/decade; wild cluster bootstrap-t p = .001; 16 survey-year clusters, N = 33,031).

The trajectory is not sampling noise, and it does not rest on the GSS alone (Figure 3). Design-based intervals include zero in 1996–2004 and sit at 28–39 points by 2024; the ANES's separately fielded item traces the same crossing on an independent sample, mode, and weighting (SI-C, Figure 19, Figure 20). The widening is 98 percent Democratic in its arithmetic through 2022, but the independents baseline locates the partisan act. Pure independents decline in lockstep with Democrats, so the distinctively partisan movement is the Republican anchor at the folk pole against the corrective flow (+0.12 per decade net; SI-B, Figure 9) — the model's own geometry. The Republican share then rebounded to 45 percent in 2022–2024. The rebound replicates in the ANES, is not sophistication-graded, and sits on the measured 2024 platform cue (SI-C, Table 48, Table 46).

6.4 A fact comes to sort like a preference (P4)

The thesis made literal (SI-C, Figure 22): in 2004 the fact showed essentially no partisan gap while the matched reduce-immigration preference showed nine points; by 2024 the fact's gap had risen to roughly seventy percent of the still-larger preference gap — convergence in levels, not a faster-partisanization claim (SI-B). And the fact–preference coupling is flat throughout (~0.4; Figure 29). The two did not grow more correlated; the factual item acquired the partisan loading the preference had all along. A fact did not fuse with a preference; it *became* one. (This corrects an earlier within-person reading that failed the 2004 split-ballot structure; SI-D.6 documents the

correction, and every headline estimate is now recomputed from the committed aggregates on each build as a hard-failure check, SI-F.)

6.5 Within-person belief revision toward party, on a broader attitude (P4)

Cross-sections cannot show a belief changing inside a person. The VOTER panel (Democracy Fund Voter Study Group, 2020) re-interviewed a national sample in 2011 — before Trump — and in 2016, 2018, and 2020 on whether immigrants "mostly make a contribution" or are "mostly a drain": a broad immigration-impact belief, wider than the verbatim fact, which bounds what the panel can show. Capture reappears inside the same people: the party gap on the "drain" belief widened from 47 to 62 points (SI-C, Figure 24), by the familiar asymmetry — Democrats abandoned the belief (35% to 10%) while Republicans held (82% to 72%).

On the *direction*, three within-person estimators place party→belief above the reverse (change-score 0.12; cross-lag 0.33 vs 0.20; random-intercept CLPM (Hamaker et al., 2015) 0.40 vs 0.12, bootstrap difference 0.28 [0.19, 0.38]; Table 33). The standard worry — party measured more reliably, so the reliable construct spuriously leads — is answered twice. A two-dimensional reliability bound shows the lead reverses only where belief is nearly stable across nine years, exactly what the panel refutes. And a latent cross-lagged model with partisanship's reliability estimated rather than assumed returns +0.55 against +0.08 (four-indicator variant +0.66 vs +0.01), with differential attrition null ($p = .43$; SI-B). The direction is identified against the reliability artifact — as lead–lag, not causation. We claim no more than that. Recent reanalysis shows panel identification of causal direction is fragile under violations of the parallel-paths assumption these models require (Chiu et al., 2025). The verbatim-fact panel and the cue-event design (SI-E) are therefore registered as what would move the direction from lead–lag to shock-response.

The sorting concentrates exactly where capture predicts: two-thirds of respondents whose 2011 belief was out of step with their party had brought it into line by 2020 — one-directional travel by the people who began on the "wrong" side, extending party-following from policy attitudes (Slothuus & Bisgaard, 2021) to a truth-apt belief, where it carries an accuracy cost. The stated disconfirmer stands: a verbatim-fact panel in which belief led party would overturn the reading. And the panel that fields the exact item exists: the ANES re-interview components — 2016 respondents re-interviewed in 2020, 2020 respondents in 2024 — carry the verbatim take-jobs fact (the item the Cumulative File harmonizes as `VCF9223`) together with party identification at each wave, giving a two-to-three-wave verbatim-fact panel with behavioral knowledge measures. The analysis is scripted and run-ready (`R/66_verbatim_panel.R`, which stops with instructions until the wave files and panel linkage are placed); it, not the broad-item panel above, is where the disconfirmer is put at genuine risk. The UAS item bank and Nationscape's weekly density (2019–2021) remain the complements — the latter for the cue-event design.

In absolute terms the 2024 gap on the fact (33 points) matches the canonical value cleavages — gun permits (30), gay rights (33), the death penalty (35) — approaching abortion (45) and the matched preference (46; Figure 12): a truth-apt claim now divides the parties about as sharply as the values that organize American politics (SI-C). Nor is the magnitude the artifact of one wave: the gap already stood at 18 points in 2022, and the final widening — eleven of its fifteen points the Republican rebound — arrives on the largest Republican-platform cue spike in the Manifesto series (SI-C). The value-cleavage magnitude is reached by two partisan movements, each dated by the

elite instrument: two decades of Democratic convergence, then a cue-timed Republican return to the folk pole.

7 The sophistication gradient: the divide widens, not narrows

Is the capture the residue of ignorance, or something that strengthens with sophistication? What follows establishes the comparative static: the gap grows with knowledge, not shrinks, ruling out the deficit and naïve-Bayesian readings. The section's map is worth stating before its evidence, because "not a standing trait of the informed" admits three readings and no single exhibit eliminates all three. *(i)* The informed could have a standing *direction* — systematically more optimistic or expert-deferential; the perception sign-reversal (§7.6) kills that. *(ii)* Sophistication could be a standing *amplifier gain* — a fixed multiplier on whatever the party line is; only the switch-on-from-zero, the two-issue clocks, and the uncontested placebo kill that, because a constant gain multiplied into a rotating line reverses too. *(iii)* What survives both is a gain that elite contestation itself creates. The exhibits below are ordered as that division of labor requires, and §7.7 adjudicates the three readings in full once the evidence is on the table. Which surviving channel conveys the signal — cognition or exposure — is the registered secondary question (§3, §5; SI-E).

7.1 Emergence among the educated (GSS)

In the GSS the partisan gap emerges first among the college-educated — twenty-one points among graduates by 2014, near zero among non-graduates — and diffuses to the rest only later. Capture begins at the top, the signature of the reception-integrated model; we report it as a temporal ordering because the pooled interaction is positive but imprecise on few survey-years. College Democrats move first; non-college Democrats start higher, move later, travel farther; both Republican groups barely move — the $g(T, s)$ diffusion in cross-section (SI-C, Figure 15).

7.2 Confirmation in a second survey (ANES 2024)

The clean test comes from a different survey, a behavioral measure of sophistication, and many sampling units. In the 2024 ANES, the partisan gap on whether immigration is likely to take jobs widens monotonically with political knowledge — from thirty-four points among the least knowledgeable to forty-six among the most — and the design-based party-by-knowledge interaction is +0.51 ($p < .001$; Table 2, Figure 4). The most informed Democrats are the *most* likely to reject the folk belief; informed Republicans do not move. An information-deficit account predicts the gap should shrink with knowledge; it does the reverse. Because two of the four knowledge items ask which party controls a chamber, partisans might answer own-side questions flatteringly, smuggling party into the moderator. The battery survives the audit. The interaction runs +0.30 to +0.52 across every scoring — +0.51 sum score, +0.52 IRT, +0.46 dropping the flagged item, and +0.30 on the two civics items alone, the conservative floor we treat as co-primary since those items carry no partisan content — all significant; the DIF that exists runs against the flattery story, and partisan-identity strength — the one trait that plausibly drives both knowledge and the gap — leaves the estimate at +0.52 when entered interacted with both party and knowledge (SI-B, Table 22, Table 23).

Table 2: Partisan gap on 'immigrants take jobs' by political-knowledge level (ANES 2024).

knowledge	dem_share	rep_share	gap	n
low (0-1)	0.172	0.508	0.336	1630
mid (2)	0.130	0.545	0.415	1286
high (3-4)	0.071	0.533	0.462	1486

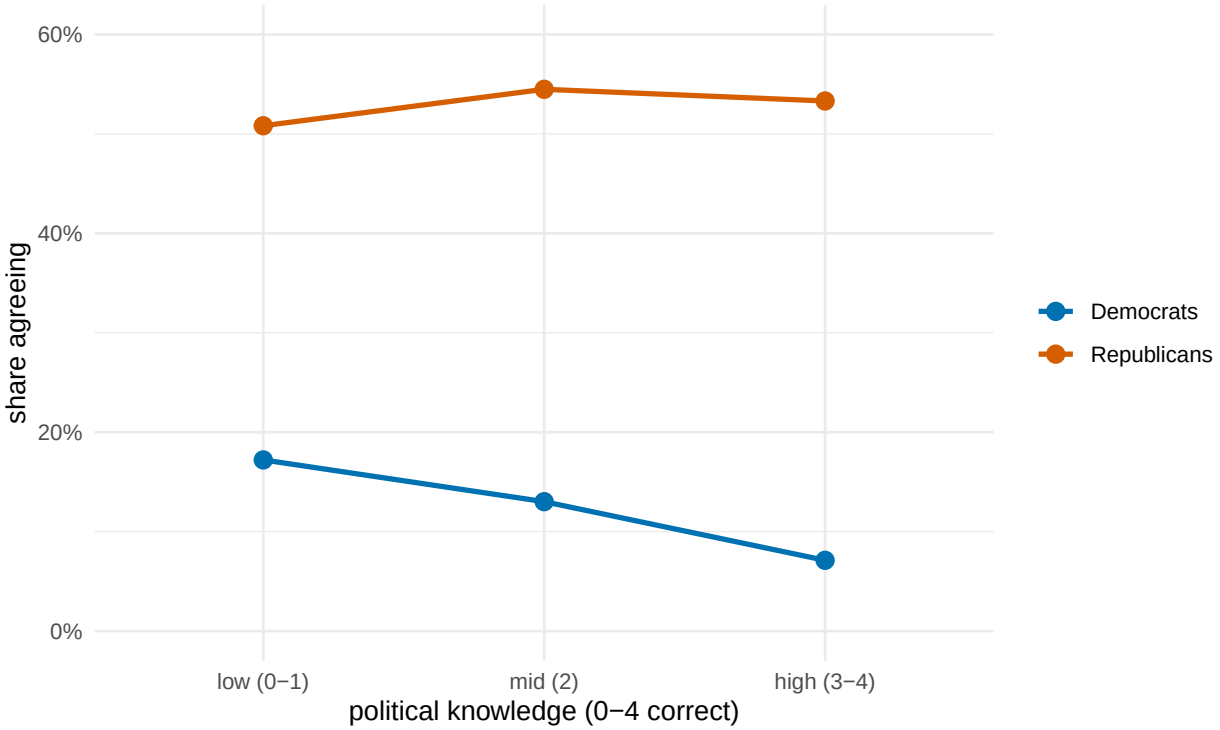


Figure 4: The gap widens with political knowledge — the opposite of an information-deficit prediction, and consistent with asymmetric motivated reasoning (party×knowledge +0.51, p < .001).

The amplification is one-sided: knowledge drives Democrats toward the expert-consistent pole (−0.49, $p < .001$) while leaving Republicans flat (+0.04, n.s.) — contradicting symmetric identity-protective cognition (Kahan, 2013) and naïve truth-tracking alike, and exactly what sophistication-enacting-the-identity-prior predicts.

Pure independents complete the design, converting a claim the model asserts into a measured result. If knowledge were generic truth-tracking, informed independents should converge on the expert-consistent pole as informed Democrats do. They do not: the knowledge slope among pure independents is flat (−0.08, n.s., n = 300), significantly shallower than the Democratic slope (difference +0.40, $p = .02$), with independents holding the folk belief at 35–41 percent across every knowledge level and Republicans and independents statistically indistinguishable in slope (SI-B, Table 14). Knowledge does not carry citizens toward the experts; it carries them toward their party's pole — and independents have no pole to be carried to. This also narrows the exposure

rival a further notch: the ambient information environment reaches independents too, yet does not move their informed members off the folk belief.

The party×knowledge interaction is positive in all twenty-four multiverse specifications (Figure 13), survives dropping weights, survives *dropping the functional form itself* — honest-forest heterogeneity estimation returns the premium at +0.05 per SD with knowledge carrying the projection and education carrying none, in both covariate specifications (SI-C, Table 39) — and replicates out of sample in the Cumulative File on a different sophistication measure — +0.41 pooled ($p \approx 10^{-10}$), present once immigration is contested (2008–2016), absent before (2004). On the probability scale, immune to logistic rescaling (Mood, 2010), the gap widens from 34 points at low knowledge to 46 at high, second difference +12.6 ([5.5, 19.6]; SI-B, Table 20); full decompositions in SI-C.

The two sophistication channels, set side by side, show directly that this is not reception-and-acceptance with an expert benchmark bolted on (Figure 2). Education — general human capital — lowers the folk belief in *both* parties (Democrats 24 to 10 percent across the college line, Republicans 42 to 32). It acts on the folk prior $f(s)$, which the parties share, and so sorts only weakly. Political knowledge — the identity-loaded channel — leaves Republicans flat and moves only Democrats, widening the gap from 34 to 46 points. It acts on $p(s)$, which departs from the folk prior only for the out-group. A reception–acceptance account in which sophistication is a single awareness dimension has no natural way to send one measure to both parties and the other to one; the two-channel split is exactly what the gap equation writes down, and a third capacity measure with no political content — verbal ability — patterns with education on every margin, including the in-party slope political knowledge cannot move (SI-D.4, Table 59).

7.3 Scope: sophistication amplifies only in the identity regime

The mechanism is bounded: sophistication does not make people polarize on everything. Using education as a common measure across the three claims, the party×education interaction runs null → negative → positive in the predicted order — correcting the folk belief where party is disengaged, defending the party line where it is engaged — though the immigration cell alone is imprecise (+0.34, $p = .27$; SI-C, Table 42). The weight rests on the design-clean ANES knowledge test above, which is robust to education, ideology, racial-resentment, and media-trust controls (Table 24, Table 25, Table 26). A negative-control battery of *verifiable* GSS science facts grounds the same boundary: the party×education interaction is null on identity-neutral items against +0.59 on the captured fact — suggestive rather than definitive, the battery not being pre-registered (SI-C, Table 43).

The two-regime model makes a prediction ordinary issue-evolution does not: the sophistication premium should itself *switch on* as an issue becomes contested — $\partial^2 \text{Gap}/\partial s\, \partial T > 0$ — rather than stand as a fixed trait of the informed. Two descriptive tests bear this out: on immigration the mass gap and premium rise together from a clean 2004 null; and the immigration and trade premiums switch on at *different* calendar times, carrying opposite signs in 2008–2012 — which no single secular trend can produce (SI-C). Both rest on few survey-years; we now measure the elite signal continuously, then at full power.

7.4 A measured elite-divergence instrument

From the Manifesto Project's coding of every U.S. platform (Lehmann et al., 2025) we construct, per issue and election year, the signed divergence in the parties' net restrictive emphasis — a

contestation index generated entirely by elite behavior. Across sixteen issue-survey-years the mass gap scales with the platform divergence it faces ($r = 0.63$; Figure 6a; within-issue permutation $p = .023$), and a pooled model with issue and year fixed effects gives $\partial\text{Gap}/\partial T = +0.10$. The trade gap is flat or reversed while the platforms agree (-0.07, 1988–2012) and switches on only as they diverge ($+0.35$, 2016–2024). This is $\partial\text{Gap}/\partial T > 0$ measured content-neutrally from elite behavior alone. What it does not deliver is the second derivative: the education-based premium is not separately identified against continuous divergence, because education is a weak proxy for the awareness that enters $g(T, s)$ directly (Zaller, 1992). The cross-derivative at full strength is the perception premium's sign-reversal, after the switch-on is dated.

The dose–response can also be seen whole. Pooling every issue-year the fixed-item record supplies — seven issues, ninety-one issue-years, five decades — Figure 5 plots each year's party-gap magnitude against the platform divergence it faced: the paper's thesis as one picture, the mass gap climbing the elite divergence across issues as different as welfare and trade. Two honesty notes travel with it. The pooled line is descriptive — the identified versions of this slope are the fixed-effects estimate above and the event study's dated onsets — and the curve's apparent concavity (a hinge fit places a flattening near the divergence level the long-captured issues occupy, slopes $+0.09$ then ~ 0) does not reach significance at ninety-one observations ($F \approx 2.3$, $p \approx .11$): capture appears to *complete* rather than compound indefinitely, but that is a hint the data license, not a law they establish.

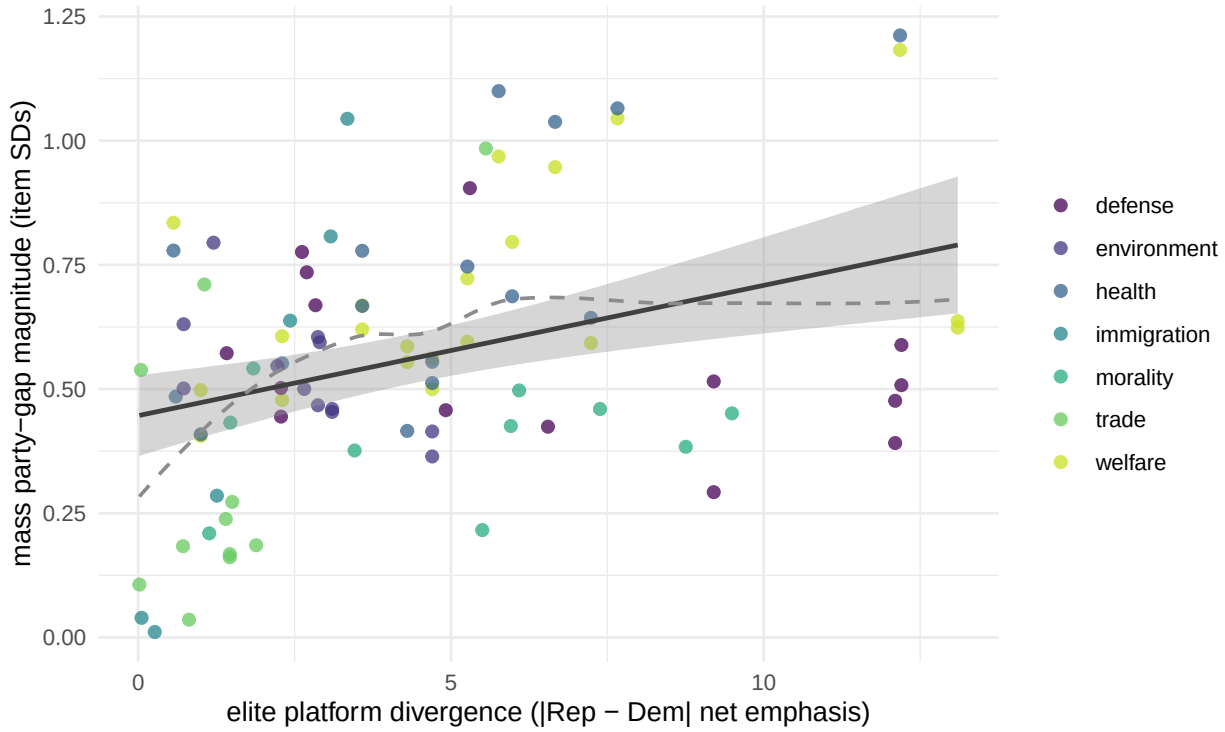
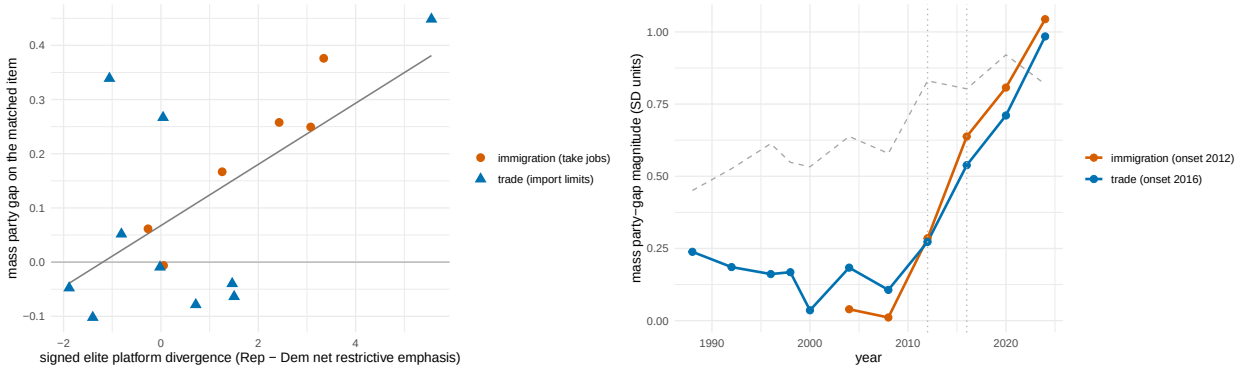


Figure 5: The capture curve: party-gap magnitude on each issue's fixed ANES item (item-SD units) against the Manifesto platform divergence it faced — 7 issues, 91 issue-years, 1970–2024. The line is a descriptive pooled fit (the identified slope is the FE estimate in the text, +0.10); the flattening at high divergence is suggestive (hinge F $\approx$ 2.3, p $\approx$ .11), not established.



(a) Gap on measured elite divergence

(b) Mass gaps switch on at their own elite clock

Figure 6: The elite clock. Left: the mass party gap on each matched ANES item against the signed Manifesto platform divergence it faces (one point per issue-survey-year; r = 0.63, within-issue permutation p = .023). Right: staggered event study — party-gap magnitude (|Rep−Dem|, standardized within item) for the two issues whose elite contestation begins within the window, against five long-partisan comparison issues (dashed); each treated issue is flat until its own elite onset (dotted) and rises only after, common shocks absorbed by calendar-year fixed effects.

7.5 A staggered event study: mass gaps switch on at their own elite clock

The two-clocks logic is a difference-in-differences claim, estimated as one. The issue universe is set by rule, not selection: every ANES Cumulative File item with a fixed-wording series spanning the window and a Manifesto issue category — seven issues — of which the treated are the two whose platform divergence has a dated within-window onset (immigration 2012, trade 2016), with the five long-divided, never-switching remainder (welfare, health, morality, environment, defense) as the comparison group. The outcome is the party-gap magnitude on each issue's fixed ANES item ($|\text{Rep} - \text{Dem}|$, standardized within item), regressed on event time with issue and calendar-year fixed effects; the year effects absorb any common "everyone polarized in the 2010s" surge (Figure 6b). The lead–lag profile does the work (Figure 7). Every pre-onset coefficient sits within noise of zero — and each lead is equivalence-bounded below the DiD itself (TOST $p \leq .010$; SI-B, Table 21) — while the profile jumps at onset and climbs thereafter (+0.38 at four years, +0.66 at eight-plus; DiD +0.34, $p < .001$). Mass division follows each issue's own elite clock, not the calendar. The estimates survive the staggered-DiD critique: a Callaway–Sant'Anna-style re-estimation returns a larger +0.44, so the TWFE figure errs conservative, and placebo-onset randomization gives $p = .003$ (SI-B, Table 15). Because "never-treated" status for the five comparison issues is an assertion, we also drop each control in turn; the DiD spans +0.28 to +0.37, every $p < .004$ (Table 16). Three hostile forks are met in SI-B (Table 17): shifting both onsets by up to four years in either direction leaves the DiD at +0.20 to +0.44 — earlier dating can only dilute the post period, and does; dating immigration by the *rival* elite clock, the speech series (2008), returns the same answer (+0.28, $p = .017$), so the two elite series' onset disagreement changes nothing; and re-running on the *signed* gap enlarges the estimate to +0.57, so the absolute-value outcome's noise floor is a conservative choice, not a crutch. The long-divided controls, finally, are not at ceiling: their gap magnitudes range over 0.3–1.2 item-SDs within issue (SI-B, Table 18), so the counterfactual trend is estimated from series with room to move. And the design is completed against its one open assumption: bias-aware

Rambachan–Roth bounds put the long-run coefficient's breakdown at two to four times the largest pre-period deviation, a Roth power analysis shows a violation big enough to fabricate the DiD would have been detected with near-certain power, and the modern estimator triad (imputation +0.38; Callaway–Sant'Anna +0.44; TWFE +0.34) agrees (SI-B, Table 19). With two cleanly late-contested issues this corroborates, rather than replaces, the perception sign-reversal.

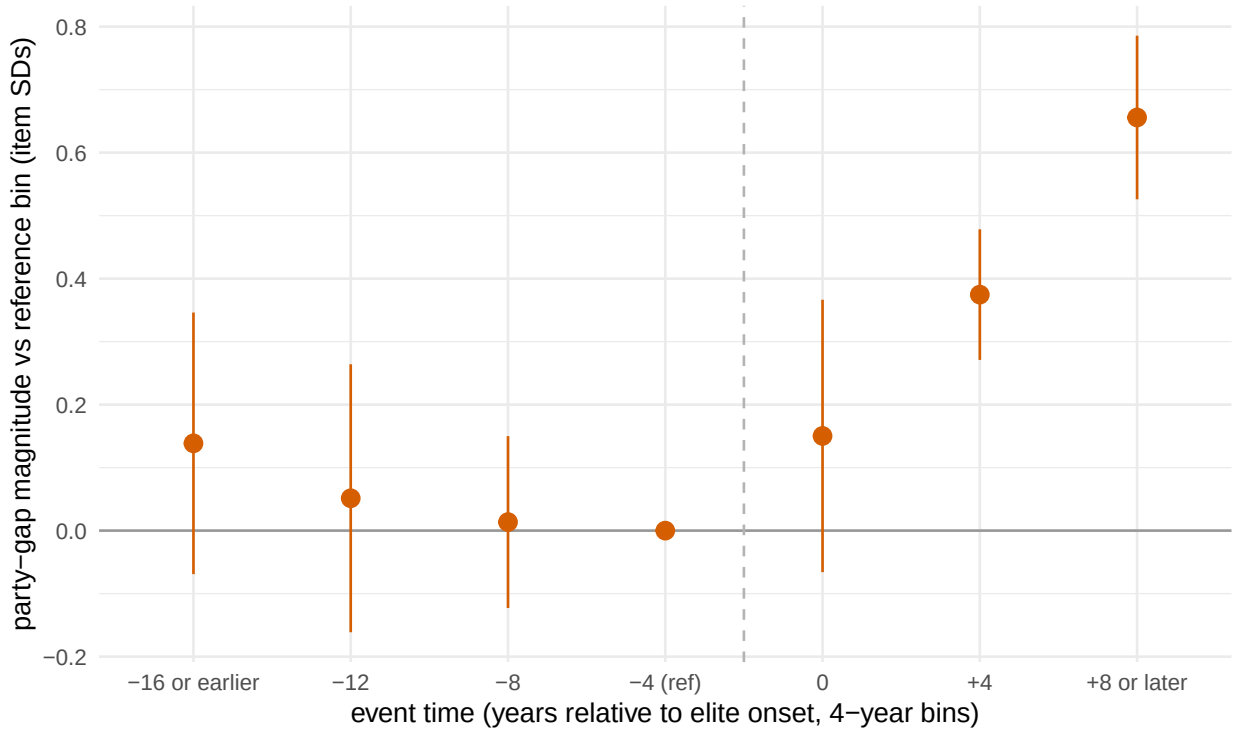


Figure 7: Event-time coefficients for the staggered elite-onset design: the mass party-gap magnitude (item SD units) by four-year bins relative to each treated issue's platform-dated onset, with issue and calendar-year fixed effects; never-treated issues are the comparison group. Bars are 95% CIs from issue-clustered SEs with t(6) critical values; reference period is the bin before onset. Pre-onset leads are flat; the gap opens only after the elite clock strikes.

The same logic staggers across places within the immigration issue itself, because the elite corpus records who is speaking. We pair each census division's congressional-delegation tone divergence (trailing decade, Card et al. corpus) with the division's own mass gap on the GSS immigration-economy items — 63 division-year-item cells, nine divisions, 1996–2022 — with division and calendar-year fixed effects. The national clock is absorbed by construction; identification comes only from *where* the gap opened earlier. It did open where the local delegation diverged earlier: $\partial \text{gap}/\partial D_{rt} = +0.66$ (division-clustered $p = .067$; exact randomization inference permuting the elite series across divisions within year, $p = .016$), sign-stable across items and design forks. The nine-cluster count and the corpus's 2020 end are caveats that bias against finding local structure (SI-B, Table 30). The elite clock moved at different speeds in different places, and the public's sorting kept local time. What the design cannot deliver is the sophistication-interacted version $\partial^2 \text{Gap}/\partial s\, \partial D_{rt}$. The triple interaction party×education×D_{rt} is null (-0.01, $p = .31$; college-binary variant -0.04, randomization $p = .48$). The reason is the measure: the only sophistication proxy the GSS fields with the fixed fact at the division level is education, which acts on the shared folk

prior — its slope scales with D_{rt} near-identically in both parties (Dem +0.030, Rep +0.022, both $p < .01$) — not on the identity term. And the design is powered only for the level leg (minimum detectable triple ≈ 0.11 against a plausible ≈ 0.03; replication files, R/57). The manufactured premium is therefore identified by the perception sign-reversal and the switch-on triangulation, not subnationally; a division-level political-knowledge series would be required, and none is fielded.

7.6 The manufactured premium

One further economic fact settles content-neutrality at once, on eighteen elections and roughly forty-eight thousand respondents: *retrospective economic perception* (`VCF0870`, 1980–2024), whether the national economy has gotten better or worse — truth-apt, the classic case of belief bending to identity (Bartels, 2002), and one on which knowledge is known to amplify rather than correct the bending (Shani, 2006).

The first result is content neutrality, and the data state it directly. The partisan gap in economic pessimism has no fixed sign: it flips with the party of the incumbent president (SI-C, Figure 23) — +0.84 under Democratic presidents (Republicans the pessimists), −0.76 under Republican ones, correlating with incumbency at $r = +0.84$ across the eighteen elections (presidential and midterm studies alike). The captured fact is owned by whichever party sits opposite the incumbent, so "one party simply defers to the experts" is impossible; under a Republican president it is Republicans who hold the rosy view. And the flip answers the identification worry that haunts the immigration series (§5) directly: a partisan gap that reverses sign on the electoral clock is not, and cannot be, a monotone trend in calendar time. We treat it as the load-bearing identification result, not a generality check.

The premium is itself a product of that configuration — the manufactured premium that gives the section its title. Within each year the party×information premium tracks the partisan gap almost perfectly ($r = +0.94$): under a Republican president informed Republicans are the most optimistic, and under a Democratic one informed Democrats are. The decisive test is what the signed premium does over time. It reverses with the incumbent president: positive under every Democratic administration, negative under every Republican one, in all thirteen years with an information measure. The pattern is exact (13/13, binomial $p < .001$), the magnitude substantial (+0.054 per SD of information, $R^2 = .69$), and the 1980–2016 interviewer-rating series delivers it on its own (+0.19, $p = .002$, $R^2 = .66$; Figure 1). The two newest points were added under sign predictions stated before estimation — a Republican incumbent forces 2020 negative, a Democratic incumbent forces 2024 positive — measured by behavioral knowledge items no interviewer impression can contaminate. Both landed as predicted. Randomization inference needing no asymptotics agrees: of all 1,716 possible assignments of party labels to the thirteen years, the observed one is the most extreme (exact $p = .0006$). Because incumbency arrives in consecutive runs — years within an administration share shocks — we also permute at the block level rather than treat years as exchangeable: over the eight presidencies the observed assignment is again the most extreme of all 70 (sign-match $p = .014$), and over the seven runs of consecutive same-party incumbency — the coarsest defensible exchangeability, where seven blocks permit a minimum one-sided p of .029 — the observed assignment attains exactly that minimum (SI-B, Table 9). The reversal also survives scoring the two near-zero years as ties (12/12 and 10/10, binomial $p = .0002$ and .001) and dropping 2020, the one year whose behavioral instrument is a single item (12/12, exact $p = .002$). A premium that reverses sign thirteen times on the electoral clock is not a function of standing sophistication or of calendar time. It is a function of the elite configuration — the content-neutral

analog of $\partial^2 \text{Gap}/\partial s\, \partial T$. Immigration shows the premium switching on from zero; perception shows it switching sign with the configuration. These are two expressions of one claim, and against the secular-trend threat the perception version bears the weight.

Two alternative readings remain, and both fail. *Measurement*: the rating-based years cannot be interviewer impressions of demeanor, because articulacy does not flip sign with the party in the White House — and the behavioral 2020/2024 points remove even that residue. *Accuracy*: the pessimists were not sometimes simply right. With election-year real GDP growth and unemployment change in the model, incumbency is essentially unchanged (gap $+0.80 \rightarrow +0.76$; premium $+0.054 \rightarrow +0.050$) while the fundamentals contribute nothing (SI-B, Table 10). And the same conclusion holds where accuracy is *measured at the individual level*, on the seven years no reasonable coding disputes: knowledge reduces a respondent's error against the realized economy only when the correction is party-congenial (-0.042 per SD) and not at all when it is not (-0.004, n.s.; difference $p = .007$; SI-B, Table 11) — with the pooled individual-level premium at $+0.056$ per SD ($p = .0005$, n = 29,238). The gap and its premium are bias with respect to a measurable benchmark — and the checkability that fails to protect this claim is exactly the institutional kind §3 marks as endogenous to contestation, a margin whose reach extends even to the pocketbook: the *personal*-finances retrospective carries its own incumbency-signed premium at half the national magnitude (12 of 13 years; SI-B, Table 12).

Immigration and perception are not two phenomena yoked together — slow partisanization here, valence cheerleading there (Bartels, 2002). They are the same equation with different terms in motion: immigration is $g(T)$ rising from zero; perception is a constant-$g(T)$ fact on which the sign of $(p_R - p_D)$ flips as incumbency reassigns the in-group. Immigration supplies the switch-on from zero that perception (flipping, not rising) cannot; perception supplies the powered cross-derivative that immigration cannot.

7.7 Three rival readings of the premium, and which test eliminates which

"Not a standing trait of the informed" admits three readings, and the tests above do not eliminate them interchangeably, so we name each rival and the test that separates it. *(i) Standing accuracy or direction:* the informed are systematically more optimistic, more expert-deferential — a fixed direction attached to sophistication. The perception series rules this out: a premium that reverses sign with the incumbent in all thirteen measurable years has no fixed direction (Figure 1, right panel). *(ii) Standing amplifier gain:* sophistication is a fixed multiplier on whatever the party line happens to be — the informed simply better, always and everywhere, at enacting the current line. This rival *survives* the perception reversal: if the gain is constant while the line rotates, the premium reverses with the incumbent exactly as observed, and the within-year premium-to-gap coupling ($r = +0.94$) is fully consistent with a constant gain multiplying a rotating gap. What kills (ii) is that the gain itself is sometimes *zero*: the immigration premium is absent in 2004, before elite onset — and "absent" is bounded, not asserted: the 2004 estimate is statistically smaller than every post-onset yearly estimate (TOST $p = .012$) and the pre-to-post change is significant in its own right ($+0.36$, one-sided $p = .019$; SI-B) — emerging only as the issue becomes contested; the two premiums switch on at two different calendar times, carrying opposite signs in 2008–2012; and where no elite party owns the folk pole — GM crops — the gain is null ($+0.01$, $p = .71$; equivalence-bounded below the mirrored nuclear premium, TOST $p = .005$, SI-B). A standing amplifier has no off position. *(iii) Configuration-manufactured gain:* the gain itself is created by elite contestation — absent at low T, switched on by elite onset, signed by the configuration. Only (iii) survives the tests

jointly. The division of labor should be read accordingly: the perception reversal is the paper's most powered result, and what it establishes at that power is *content-neutrality* — the elimination of (i); *manufacture* — the elimination of (ii) — rests on the switch-on-from-zero, the two clocks, and the uncontested placebo. The two together, and only together, deliver the headline claim.

Which claims get captured at all admits the same cross-claim test, estimated in full in SI-B: across every truth-apt claim the paper fields with party identification (ten fixed-wording claims, 44 claim-years, 1993–2024), the partisan gap is 0.13 SD where no elite party owns a pole and 0.63 SD where one does — a five-fold contestation margin (WLS +0.81, claim-clustered $p = .001$; randomization inference over contestation schedules, $p = .016$; SI-B, Table 31). The verifiability leg is not identified within the fielded range: every claim surveys field scores 6 or higher on the 0–12 index, because publics argue about claims they cannot check. That cell is registered as E.5. On the fielded record, what separates the claims citizens sort on from the claims they do not is whether elites supply an owner for one pole.

7.8 Generality: beyond immigration

The capture is not special to immigration. Disbelief in evolution partisanizes on the same one-sided trajectory (GSS 2006–2018); trade establishes content-neutrality through the 2016 party-pole flip, the converging out-group being whichever party departs from the folk pole; and cross-nationally (ISSP 2013, twenty-six democracies) the left–right gap on the same fact tracks whether a salient anti-immigration party owns the folk pole, not party-system fragmentation. All three are reported in full in the Supplementary Materials (SI-D).

The derived prediction, moreover, can be tested directly: the GSS environment rotations field three expert-consensus danger beliefs on one five-point scale, spanning the three configurations the model distinguishes (Table 3). On nuclear power — expert pole Republican-congenial (Markandya & Wilkinson, 2007; Schlömer et al., 2014; United Nations Scientific Committee on the Effects of Atomic Radiation, 2021), the folk pole (dread) owned by the Democratic elite side — it is *educated Republicans* who converge (education slope +0.16 vs +0.06; party×education +0.10, $p < .001$), the mirror image of immigration. On GM crops, where no elite party owns the folk pole (National Academies of Sciences, Engineering, and Medicine, 2016), education corrects both parties and the premium is null (+0.01, $p = .71$) — exactly the folk-regime placebo P2 predicts. On climate, expert pole Democratic-congenial, the asymmetry runs the familiar way (−0.15, $p < .001$), recovering McCright & Dunlap (2011) as one cell of a three-cell design. Same survey, same scale, items not selected after seeing the answer. The party that education moves toward the experts flips with the folk-pole assignment. That closes the circularity flagged in §3 and rules out the rival that Democrats simply defer to experts — including its strongest form, an asymmetric polarization of trust in science itself (Gauchat, 2012; Merkley, 2020), since a standing Republican distrust of experts cannot put educated Republicans *on* the experts' nuclear position: on nuclear power it is Democrats who hold the folk pole against them (caveats in SI-D.2b).

Table 3: The converse case, executed. Party×college interaction on the expert-consistent belief for three GSS danger items on one five-point scale (weighted LPM, year FE, HC1), with per-party education slopes. The converging party flips with the configuration: Republicans on nuclear power, both parties on GM crops, Democrats on climate.

issue (danger item)	expert pole congenial to	party×college	p	educ slope, Dem	educ slope, Rep	n
nuclear power (danger)	Republican	0.098	0.000	0.057	0.155	6000
GM crops (danger)	neither (untribalized)	0.013	0.705	0.092	0.107	4300
climate change (danger)	Democratic	-0.146	0.000	0.098	-0.018	7516

One reading of these convergences the model itself blocks: it never predicts truth-seeking. Under the gap equation the out-group converges on its *elite pole*, and in every case studied — including nuclear power, where Republican elites are pro-nuclear — that pole happens to coincide with the expert-consistent position. The flat independent gradient (§7, Table 14) is this point measured: convergence on the expert pole appears only where a party supplies the destination. Expert-consistency is therefore epiphenomenal to the mechanism: the benchmark does not cause the direction of anyone's movement; it prices the movement's epistemic consequence, telling us which party's capture exacts an accuracy cost and how large. This makes the democratic-competence stake exact rather than rhetorical. Capture is not a failure to reach a truth that better information would fix; it is a mechanism indifferent to truth whose cost the benchmark renders measurable. Were elite poles ever to sit opposite the experts on both sides, the model predicts convergence on error with equal confidence.

The converse table's natural companion is the timing falsifier: the same premium, tracked across two issues, switches on at two different calendar times, each on its own elite clock (Figure 1, center panel) — in 2008–2012 the two premiums carry *opposite signs*, which no single secular "sophisticates polarize more over time" trend can produce (design and caveats in SI-C).

8 Discussion

The organizing claim is that economic belief under unverifiability runs in two regimes, and that an issue can cross between them within a generation. Where a claim is uncontested, citizens hold a shared folk belief that information and explanation can move toward the experts (Meyers et al., 2020; Meyers et al., 2023). Where it is contested, information moves the out-group toward the experts while the in-group holds, so more of it widens the gap — the obstacle is not a deficit of knowledge but a surplus of identity (Kahan et al., 2012; Sapienza & Zingales, 2013). The VOTER panel shows the transition inside individuals: the same people, no better able to check the claim, revised a broad immigration-impact belief into partisan line as the issue contested (§6).

The two-regime model is an organizing lens, not a validated structural object — a bound we estimate rather than assert: the strict structural forms, fit by minimum distance to both series, are rejected exactly where the design's own logic says they must be (SI-C, Table 41). Its diffusion dynamics

show the difference. Strict separability implies a constant premium-to-gap ratio; instead the ratio falls as immigration becomes contested, the premium near its ceiling in 2008 while the mass gap is still near zero. Under the reception-integrated $g(T, s)$ (§3), both facts arrive as predictions rather than anomalies: the aware receive the elite signal earlier, so the premium leads the mass gap during switch-on and the ratio declines as diffusion completes (SI-C). The predictions that separate the model from a generic "sophisticates polarize more" account (Kahan, 2013) divide the labor stated in §7. The sign-reversal (13/13, at power) eliminates any standing direction of the informed. The switch-on from zero — the four-point immigration series, the cross-issue timing test, and the GM-crops placebo — eliminates a standing amplifier gain, with the converse corner shown on nuclear power (§7; the market-fact form stays registered, SI-E). The contribution is the descriptive capture, its content-neutrality, and the manufactured premium — whichever channel conveys it.

This is the answer to the question we opened with — how do you convince someone of a claim they cannot check? It depends on the regime. On an uncontested claim, supply the mechanism: explanation moves belief toward the experts. On a contested one, information is the wrong tool; the levers the theory implies are the withdrawal of elite contestation or a trusted in-group messenger — a conjecture the theory licenses, whose observable form is registered: if elite cue supply withdraws, the gap should close (SI-E) — faster among converts than among the cohorts configured at entry, who never held the pre-capture belief (§6; SI-B, Table 7) — with partisan media acting as the high-gain amplifier of the elite signal, not a competing cause.

The argument reaches beyond economics, but at two separable levels. The *descriptive* capture travels — it reproduces on disbelief in evolution and across most of twenty-six democracies (§7) — and should recur wherever a truth-apt claim is unverifiable and elite-contested (climate, public-health, new-technology safety are untested candidates; SI-E carries a scored watchlist, with verifiable ACA-type claims as the one cell where the framework forbids capture outright). The *sophistication-amplified* form is shown only on the economic facts. We name the broader pattern a prediction the framework licenses, not a regularity shown everywhere. And the named next case is no longer only a prediction: in the linked Pew panel, the same people who showed no party structure on the machine-jobs belief in 2017 show a twelve-point gap on the AI belief in 2024, with 2017 party leading the arriving belief and the reverse path half the size (SI-D.7) — the capture beginning within persons, on schedule — and the prediction is priced in advance, at roughly 25 points within four years of a platform-dated onset and 37 at eight-plus, with a $\pm$8-point band should no onset come (SI-E, Table 63).

The finding cautions the civic-knowledge tradition at the level the design licenses — about *information*, not cognition. The remedy for a captured fact cannot simply be more information, whichever engine carries the widening — a conclusion that converges with the correction literature on this very domain, where immigration misperceptions resist factual correction and corrected innumeracy fails to move downstream attitudes (Hopkins et al., 2019; Kuklinski et al., 2000). The causal wedge the observational design lacks exists experimentally on the input side: elite polarization increases reliance on party cues and dampens substantive evaluation (Druckman et al., 2013) — the manufactured premium is the field signature of that laboratory result. This is not the backfire effect (Guess & Coppock, 2020; Nyhan & Reifler, 2010; Wood & Porter, 2019): information does not make individuals more wrong; the between-party *gap* is wider among the better-informed because sophistication moves the out-group toward the experts while the in-group sits at the folk pole (Achen & Bartels, 2016; Flynn et al., 2017).

One experimental finding appears to run the other way: one-shot informational treatments move partisans in *parallel* (Coppock, 2023). The two results measure different objects — persuasion-in-parallel is the *impulse response* to a single dose; capture is the *equilibrium* under sustained, configuration-sorted cueing. A treatment can nudge both parties toward the experts on Tuesday while the configuration re-sorts them by November; the accounts part ways observably at cue withdrawal (SI-E).

The experiment this design lacks is, we think, the field's next test rather than this paper's missing piece, and the registered commitments (SI-E) are written to be run. The decisive one is E.6, which identifies the cross-derivative *causally* where the observational design can only triangulate it: a novel truth-apt claim, elite contestation manipulated between subjects, sophistication measured, the triple interaction party×sophistication×contestation the estimand — the manufactured premium under random assignment, on a single claim, with the secular-trend and selective-exposure confounds designed out. E.7 adapts the Fake-News-Effect design (Thaler, 2024) to separate motivated cognition from Bayesian updating on the verbatim fact, converting the mechanism from narrowed to identified; E.8 measures the expressive share on this item rather than transporting it. The observational registrations complete the frame: E.1 is a panel on AI-and-employment as its elite clock starts; E.3 fields the market-fact converse with a knowledge battery; E.5 fields matched verifiable and unverifiable claims; E.2's cue-withdrawal test will run itself the next time elite supply moves. Each names its disconfirmer in advance; the nine registrations of SI-E are offered as the agenda.

9 Limitations and scope

What did not hold. The record of this design is not uniformly confirmatory, and the misses are informative, so we collect them before the boundaries. Strict separability — the model's own simple form — is rejected by the premium-to-gap ratio test, and the estimated structural forms extend the bound: scalar rotation on perception and the one-path level-tracking form on immigration are both rejected by minimum-distance over-identification tests (SI-C, Table 41), which is why the model enters as an organizing lens with reception integrated (§8; replication files, R/37, R/75). The two-party scope conjecture failed outright and is revised below rather than asserted. The evolution sophistication signature dies under religiosity controls, narrowing that case to the descriptive claim (SI-D.1). The subnational triple interaction is a bounded non-test (§7); the verifiability leg of the gap equation is unidentified in the fielded range (§7; E.5) — and its within-survey probe runs in the predicted direction without separating (the personal-finances premium sits at half the national magnitude, ratio 1.8:1, quadruple $p = .21$; SI-B, Table 12); the GSS education interaction is scale-sensitive on the probability scale (SI-B); two items in the negative-control battery misbehave and are reported without a rescue (SI-C); the two-wave cross-national difference-in-differences is not significant (SI-D.3), and the cross-national *premium* is a bounded non-test with ISSP's instruments (SI-D.3); the strict-fixed-effects dynamic panel is uninformative for want of usable within-person variance, supporting neither direction (SI-B); and the model did not predict the timing of the 2024 rebound (SI-C). Each miss is reported where it bites, none is repaired after the fact, and the claims map (Table 1) grades what remains.

The design's boundaries. Precedence is estimated on the immigration *preference*, so elite-led timing *on the fact* is an inference from the preference's lead and the fact's emergence first among the sophisticated (§6). The mechanism is *consistent with* asymmetric motivated reasoning, not identified against divergent exposure. The interaction rules out the deficit and Bayesian accounts, survives

five party×confound controls and a measured content-of-diet control, and would take a confounder roughly eleven times the strongest measured proxy to overturn (RV ≈ 5%; §7). But no observational bound fully separates cognition from exposure; the thesis does not require it (both channels are capture, §3), and the channel question is registered (SI-E).

Verifiability does no within-sample identifying work: every studied claim is low in *experiential* verifiability, so verifiability enters only as a scope condition. That the most institutionally checkable claims are among the most captured is the prediction of §3's split, the perception series demonstrably orthogonal to its fundamentals (SI-B); the *experiential* half-life is itself registered — AI deployment should shrink the gap as it raises experiential v (SI-E). The design is observational: no causal claim about the fact–preference direction, the within-person direction identified only against the reliability artifact, as lead–lag (§6).

Generality runs at two levels: the *descriptive* capture is general (evolution, twenty-six democracies), while the *sophistication-amplified* form is clean only for immigration and economic perception, the evolution signature dying under religiosity controls. The converse prediction is confirmed on nuclear power with GM crops the placebo (§7); its market-fact form (minimum wage, rent control) lacks an off-the-shelf instrument and stays registered (SI-E). The within-economics longitudinal evidence rests on two framings of one immigration-economy belief, so economic generality leans on the gradient and the perception swings (Bartels, 2002).

Two further boundaries. Sincerity is met by partial identification rather than our own experiment (§5, SI-B), with the disconfirmer stated. And the two-party scope condition we conjectured is *not* supported: the cross-national gap is uncorrelated with party-system fragmentation — capture tracks whether an elite party owns the folk pole, not the number of parties — so we revise the condition rather than assert it. Each open question is restated as a registered directional prediction with a named falsifier (SI-E).

10 Conclusion

A truth-apt belief about the economy, its wording held fixed, came to sort by party nearly as strongly as a taste — and did so fastest among the citizens best equipped to know better. Followed as individuals across a decade, the same citizens revised a broad immigration-impact belief into partisan line, the party→belief direction identified against the reliability artifact as lead–lag (§6); and the cohorts who entered politics after the elites had configured the issue never held the shared belief at all, arriving sorted at twenty-five points where their pre-onset predecessors arrived at zero (§6). When a claim cannot be checked, belief tracks the identity of its source rather than the content of the message: the Time-Traveler's Problem. The next fact to sort is not merely predicted but already moving: within a linked panel, the same people unsorted on the machine-jobs belief in 2017 are twelve points apart on the AI belief in 2024, prior party leading the arriving belief (SI-D.7), and the forecast is committed in advance — magnitudes, dates, and the band that would falsify it (SI-E). Whether the discipline can carry unverifiable facts across the partisan filter — through mechanism, through messenger, through the withdrawal of elite contestation — is the question this diagnosis leaves for the work to come.

References

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists: Why elections do not produce responsive government*. Princeton University Press.

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius*, *3*. https://doi.org/10.1177/2378023117710578

American National Election Studies. (2025). *ANES 2024 time series study*. University of Michigan and Stanford University.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117–150. https://doi.org/10.1023/A:1021226224601

Benoit, K., De Marchi, S., Laver, C., Laver, M., & Ma, J. (2026). Using large language models to analyze political texts through natural language understanding. *American Journal of Political Science*. https://doi.org/10.1111/ajps.70050

Berinsky, A. J. (2018). Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. *The Journal of Politics*, *80*(1), 211–224. https://doi.org/10.1086/694258

Bisgaard, M. (2019). How getting the facts right can fuel partisan-motivated reasoning. *American Journal of Political Science*, *63*(4), 824–839. https://doi.org/10.1111/ajps.12432

Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, *65*(5), 745–769. https://doi.org/10.1111/jcom.12171

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics*, *118*(4), 1335–1374. https://doi.org/10.1162/003355303322552810

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, *91*(6), 3253–3285. https://doi.org/10.1093/restud/rdae007

Boyer, P., & Petersen, M. B. (2018). Folk-economic beliefs: An evolutionary cognitive model. *Behavioral and Brain Sciences*, *41*, e158. https://doi.org/10.1017/S0140525X17001960

Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review*, *105*(3), 496–515.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, *10*(4), 519–578. https://doi.org/10.1561/100.00014074

Burnham, M., Kahn, K., Wang, R. Y., & Peng, R. X. (2025). Political DEBATE: Efficient zero-shot and few-shot classifiers for political text. *Political Analysis*. https://doi.org/10.1017/pan.2025.10028

Caplan, B. (2007). *The myth of the rational voter: Why democracies choose bad policies*. Princeton University Press.

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., Abramitzky, R., & Jurafsky, D. (2022a). Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, *119*(31), e2120510119. https://doi.org/10.1073/pnas.2120510119

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., Abramitzky, R., & Jurafsky, D. (2022b). *Replication data for: Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration*. GitHub repository,

https://github.com/dallascard/us-immigration-speeches.
Card, D. (2005). Is the new immigration really so bad? *The Economic Journal*, *115*(507), F300–F323. https://doi.org/10.1111/j.1468-0297.2005.01037.x
Carmines, E. G., & Stimson, J. A. (1986). On the structure and sequence of issue evolution. *American Political Science Review*, *80*(3), 901–920.
Carmines, E. G., & Stimson, J. A. (1989). *Issue evolution: Race and the transformation of american politics*. Princeton University Press.
Chaisemartin, C. de, & D'Haultfœuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_01414
Chiu, A., Lan, X., Liu, Z., & Xu, Y. (2025). Causal panel analysis under parallel trends: Lessons from a large reanalysis study. *American Political Science Review*.
Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B*, *82*(1), 39–67. https://doi.org/10.1111/rssb.12348
Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822. https://doi.org/10.1037/0022-3514.85.5.808
Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206–261). Free Press.
Cooperative Election Study. (2025). *CES common content, 2022 and 2024*. Harvard Dataverse.
Coppock, A. (2023). *Persuasion in parallel: How information changes minds about politics*. University of Chicago Press.
Democracy Fund Voter Study Group. (2020). *Views of the electorate research (VOTER) survey, 2011–2020*. Democracy Fund Voter Study Group.
Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05
Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111–119. https://doi.org/10.1038/s41558-018-0360-1
Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*(1), 57–79. https://doi.org/10.1017/S0003055412000500
Egami, N., Hinck, M., Stewart, B. M., & Wei, H. (2023). Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models. *Advances in Neural Information Processing Systems 36*.
Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, *38*(S1), 127–150. https://doi.org/10.1111/pops.12394
Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., & Verkuilen, J. (2007). Same facts, different interpretations: Partisan motivation and opinion on iraq. *The Journal of Politics*, *69*(4), 957–974. https://doi.org/10.1111/j.1468-2508.2007.00601.x
Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the united states, 1974 to 2010. *American Sociological Review*, *77*(2), 167–187. https://doi.org/10.1177/0003122412438225
Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, *87*(4), 1307–1340.

https://doi.org/10.3982/ECTA14237
Gerber, A. S., & Huber, G. A. (2010). Partisanship, political control, and economic assessments. *American Journal of Political Science*, *54*(1), 153–173.
Gordon, R., & Dahl, G. B. (2013). Views among economists: Professional consensus or point-counterpoint? *American Economic Review*, *103*(3), 629–635. https://doi.org/10.1257/aer.103.3.629
Graham, M. H. (2023). Measuring misperceptions? *American Political Science Review*, *117*(1), 80–102. https://doi.org/10.1017/S0003055422000387
Guess, A., & Coppock, A. (2020). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, *50*(4), 1497–1515. https://doi.org/10.1017/S0007123418000327
Hainmueller, J., & Hopkins, D. J. (2014). Public attitudes toward immigration. *Annual Review of Political Science*, *17*, 225–249. https://doi.org/10.1146/annurev-polisci-102512-194818
Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889
Hopkins, D. J., Sides, J., & Citrin, J. (2019). The muted consequences of correct information about immigration. *The Journal of Politics*, *81*(1), 315–320. https://doi.org/10.1086/699914
ISSP Research Group. (2015). *International social survey programme: National identity III - ISSP 2013*. ZA5950 Data file Version 2.0.0; GESIS Data Archive, Cologne. https://doi.org/10.4232/1.12312
Jerit, J., & Barabas, J. (2012). Partisan perceptual bias and the information environment. *The Journal of Politics*, *74*(3), 672–684. https://doi.org/10.1017/S0022381612000187
Johnston, C. D., & Ballard, A. O. (2016). Economists and public opinion: Expert consensus and economic policy judgments. *The Journal of Politics*, *78*(2), 443–456. https://doi.org/10.1086/684629
Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424.
Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*, 732–735. https://doi.org/10.1038/nclimate1547
Kent A. Clark Center for Global Markets. (2024). *US economic experts panel (IGM forum)*. University of Chicago Booth School of Business.
Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.
Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *The Journal of Politics*, *62*(3), 790–816. https://doi.org/10.1111/0022-3816.00033
Layman, G. C., & Carsey, T. M. (2002). Party polarization and "conflict extension" in the american electorate. *American Journal of Political Science*, *46*(4), 786–802. https://doi.org/10.2307/3088419
Le Mens, G., & Gallego, A. (2025). Positioning political texts with large language models by asking and averaging. *Political Analysis*, *33*(3). https://doi.org/10.1017/pan.2024.29
Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x
Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Regel, S., Riethmüller, F., Weßels, B., & Zehnter, L. (2025). *The manifesto data collection. Manifesto project (MRG/CMP/MARPOR). Version 2025a*. Wissenschaftszentrum Berlin für Sozialforschung (WZB); Göttingen: Institut für Demokratieforschung (IfDem). https://doi.org/10.25522/

manifesto.mpds.2025a
Lenz, G. S. (2012). *Follow the leader? How voters respond to politicians' policies and performance*. University of Chicago Press.
Levendusky, M. (2009). *The partisan sort: How liberals became democrats and conservatives became republicans*. University of Chicago Press.
Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(8), 1–19. https://doi.org/10.18637/jss.v009.i08
Marietta, M., & Barker, D. C. (2019). *One nation, two realities: Dueling facts in american democracy*. Oxford University Press.
Markandya, A., & Wilkinson, P. (2007). Electricity generation and health. *The Lancet*, *370*(9591), 979–990. https://doi.org/10.1016/S0140-6736(07)61253-7
McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *The Sociological Quarterly*, *52*(2), 155–194. https://doi.org/10.1111/j.1533-8525.2011.01198.x
Merkley, E. (2020). Anti-intellectualism, populism, and motivated resistance to expert consensus. *Public Opinion Quarterly*, *84*(1), 24–48. https://doi.org/10.1093/poq/nfz053
Merkley, E., & Stecula, D. A. (2021). Party cues in the news: Democratic elites, republican backlash, and the dynamics of climate skepticism. *British Journal of Political Science*, *51*(4), 1439–1456. https://doi.org/10.1017/S0007123420000113
Meyers, E. A., Gretton, J. D., Budge, J. R. C., Fugelsang, J. A., & Koehler, D. J. (2023). Broad effects of shallow understanding: Explaining an unrelated phenomenon exposes the illusion of explanatory depth. *Judgment and Decision Making*, *18*, e24. https://doi.org/10.1017/jdm.2023.24
Meyers, E. A., Turpin, M. H., Bialek, M., Fugelsang, J. A., & Koehler, D. J. (2020). Inducing feelings of ignorance makes people more receptive to expert (economist) opinion. *Judgment and Decision Making*, *15*(6), 909–925.
Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82. https://doi.org/10.1093/esr/jcp006
National Academies of Sciences, Engineering, and Medicine. (2016). *Genetically engineered crops: Experiences and prospects*. The National Academies Press. https://doi.org/10.17226/23395
National Academies of Sciences, Engineering, and Medicine. (2017). *The economic and fiscal consequences of immigration*. The National Academies Press. https://doi.org/10.17226/23550
NORC at the University of Chicago. (2025). *General social survey, 1972–2024 cumulative data file*. NORC, Chicago.
Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2
Peterson, E., & Iyengar, S. (2021). Partisan gaps in political information and information-seeking behavior: Motivated reasoning or cheerleading? *American Journal of Political Science*, *65*(1), 133–147. https://doi.org/10.1111/ajps.12535
Petrocik, J. R. (1996). Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, *40*(3), 825–850. https://doi.org/10.2307/2111797
Pew Research Center. (2024). *American trends panel, waves 27 (2017) and 152 (2024)*. Pew Research Center, Washington, DC.
Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science*, *306*(5695), 462–466. https://doi.org/10.1126/science.1102081
Prior, M., Sood, G., & Khanna, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, *10*(4), 489–518. https://doi.org/10.1561/100.00014127

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, *90*(5), 2555–2591. https://doi.org/10.1093/restud/rdad018

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322. https://doi.org/10.1257/aeri.20210236

Sapienza, P., & Zingales, L. (2013). Economic experts versus average americans. *American Economic Review*, *103*(3), 636–642. https://doi.org/10.1257/aer.103.3.636

Schlömer, S., Bruckner, T., Fulton, L., Hertwich, E., McKinnon, A., Perczyk, D., Roy, J., Schaeffer, R., Sims, R., Smith, P., & Wiser, R. (2014). Annex III: Technology-specific cost and performance parameters. In *Climate change 2014: Mitigation of climate change. Contribution of working group III to the fifth assessment report of the IPCC* (pp. 1329–1356). Cambridge University Press.

Semenova, V., & Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, *24*(2), 264–289. https://doi.org/10.1093/ectj/utaa027

Shani, D. (2006). *Knowing your colors: Can knowledge correct for partisan bias in political perceptions?* [Unpublished manuscript].

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Slothuus, R., & Bisgaard, M. (2021). How political parties shape public opinion in the real world. *American Journal of Political Science*, *65*(4), 896–911. https://doi.org/10.1111/ajps.12550

Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *The Quarterly Journal of Economics*, *136*(4), 2309–2369. https://doi.org/10.1093/qje/qjab033

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences*, *34*, 81–87. https://doi.org/10.1016/j.cobeha.2020.01.003

Tesler, M. (2018). Elite domination of public doubts about climate change (but not evolution). *Political Communication*, *35*(2), 306–326. https://doi.org/10.1080/10584609.2017.1380092

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, *16*(2), 1–38. https://doi.org/10.1257/mic.20220146

United Nations Scientific Committee on the Effects of Atomic Radiation. (2021). *Sources, effects and risks of ionizing radiation: UNSCEAR 2020/2021 report, volume II*. United Nations.

Van Gunten, T. S., Martin, J. L., & Teplitskiy, M. (2016). Consensus, polarization, and alignment in the economics profession. *Sociological Science*, *3*, 1028–1052. https://doi.org/10.15195/v3.a45

Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change*, *77*(1–2), 103–120. https://doi.org/10.1007/s10584-006-9060-3

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Wu, P. Y., Nagler, J., Tucker, J. A., & Messing, S. (2023). *Large language models can be used to*

estimate the latent positions of politicians.
Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., & Wager, S. (2025). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *Journal of the American Statistical Association*, *120*(549), 38–51. https://doi.org/10.1080/01621459.2024.2393466
Zaller, J. R. (1992). *The nature and origins of mass opinion.* Cambridge University Press.

Supplementary Materials

The Supplementary Materials are organized as SI-A Measurement, SI-B Robustness, SI-C Supplementary results and exhibits, SI-D Generality, SI-E Registered commitments, and SI-F Reproducibility; they are excluded from the manuscript word count.

SI-A: Measurement — variables, coding, question wording, and the verifiability index

SI-A.1 Matched pairs and question wording

Each folk claim pairs a public-opinion item with the economic experts' verdict on the same causal claim, framed so that public and expert agreement point the same way. The pairs are *conceptual* matches, not identical wordings: the expert items use precise technical language and the public items are deliberately broad. The AI pair is the closest — a belief about the same outcome on both sides — while the trade and immigration pairs match the public folk belief to the experts' verdict on a closely related policy-causal claim. We treat the trade and immigration pairs as illustrative context and place the inferential weight on the individual-level and over-time analyses of the public items themselves. Verbatim wordings:

Pair	Expert claim (IGM / Kent A. Clark Center, verbatim)	Public item
AI → jobs	"Use of artificial intelligence over the next ten years will lead to a substantial increase in the unemployment rates in advanced countries."	Pew W152 `AIJOBS_W152`: "Thinking about the U.S. over the next 20 years, do you think that artificial intelligence (AI) will lead to [more / fewer / about the same number of] jobs?" (folk = "fewer")
Trade → jobs	"If all of the 'Seven actions to protect American workers' in President-elect Trump's 100-day plan are enacted, it will more likely than not improve the economic prospects of middle-class Americans over the next decade."	GSS `moretrde`: "(More) international trade leads to fewer jobs in the USA" (folk = agree)

Pair	Expert claim (IGM / Kent A. Clark Center, verbatim)	Public item
Immigration → jobs	*No direct IGM / Kent A. Clark Center item asks the verbatim claim; no expert point estimate is used.* Direction only, anchored on the National Academies (2017) consensus that immigration does not reduce native employment on net.	GSS `immjobs`: "Immigrants take jobs away from people born in America" (folk = agree); ANES `V242228` (see SI-A.3)

The immigration benchmark differs from the AI and trade pairs in one respect we state plainly and consistently across the paper: there is no expert *point estimate* for it, and we construct none. No IGM/Kent A. Clark Center item asks economists the verbatim "immigrants take jobs" claim. We anchor only the *direction* of the expert consensus, on the National Academies' finding that immigration does not reduce native employment on net (National Academies of Sciences, Engineering, and Medicine, 2017). Accordingly Table 52 leaves the immigration expert benchmark and gap blank rather than imputed, the reproducibility ledger records `immigration expert benchmark = NA` (script 04), and every immigration claim in the paper rests on the over-time and individual-level analyses of the public `immjobs` item set against the *direction* of consensus, never on a numeric benchmark. An earlier draft inverted the panel's agreement on a green-card item into an implied folk-agreement figure; we removed it as a non-sequitur (§4) and no such constructed estimate appears anywhere in the analysis.

SI-A.2 Variable coding

Survey	Variable	Item label	Coding
GSS	`immjobs`	immigrants take jobs away (5-pt agree–disagree)	folk = 1 if (strongly) agree (≤ 2)
GSS	`letin1`	number of immigrants should be (5-pt)	restrictionist = 1 if reduce (≥ 4)
GSS	`letin1a`	number of immigrants nowadays should be (5-pt)	restrictionist = 1 if reduce (≥ 4)
GSS	`moretrde`	more trade, fewer US jobs	folk = 1 if agree (≤ 2)
GSS	`partyid`	political party affiliation (0–7)	Democrat 0–2, Republican 4–6; 7 ("other party") excluded
GSS	`degree`	highest degree (0–4)	college = 1 if bachelor+ (≥ 3)

Survey	Variable	Item label	Coding
GSS	`wtssall` / `wtssps`	person (post-stratification) weight	analysis weight
Pew W152	`AIJOBS_W152`	AI → more / fewer / same jobs	folk = 1 if "fewer"
Pew W27	`ROBJOB4B_W27`	automation → people struggle to fill their time	folk = 1 if "likely"
Pew	`F_PARTYSUM_FINAL`	party summary	Republican vs Democrat
Pew	`F_EDUCCAT` / `F_EDUCCAT_FINAL`	education category	college = 1 if highest category
ANES	`V242228`	immigration likely to take jobs (4-pt)	fact = 1 if extremely/very likely
ANES	`V241227x`	party ID summary (1–7)	Democrat 1–3, Republican 5–7

One coding correction is disclosed rather than absorbed silently. Earlier builds coded GSS Republicans as `partyid` ≥ 4, which sweeps category 7 — "other party" — into the Republican group (up to ten percent of that group in recent waves). The scripts now exclude 7, and the correction runs in the paper's favor: on the capstone item the 2024 gap widens from 33.2 to 35.0 points and the 2022 gap from 17.8 to 19.6, with earlier waves essentially unchanged (deltas ≤ 0.4 points through 2014). Headline numbers quoted in this draft are the conservative included-7 values from the committed aggregates; the full pipeline re-run regenerates every aggregate under the corrected coding, and the R/56 consistency audit flags each quoted number that moves.

SI-A.3 ANES 2024 knowledge battery

The dependent variable for the cross-survey mechanism test is `V242228` (post-election): "How likely is it that recent immigration levels will take jobs away from people already here?", coded 1 if "extremely" or "very" likely, 0 if "somewhat" or "not at all" likely. The political-knowledge index sums four items, each scored correct = 1:

Variable	Item	Correct answer
`V241612`	Years in a full U.S. Senate term	6
`V241613`	Category of least federal spending	foreign aid
`V241614`	Party with most House members before the 2024 election	Republicans
`V241615`	Party with most Senate members before the 2024 election	Democrats

The two majority items are scored to the **pre-November-2024 Congress** (Republican House, Democratic Senate). Estimation uses the post-election design: weight `V240107b`, primary sampling unit `V240107c`, stratum `V240107d`. ANES includes a "catch" knowledge item, and some respondents may look answers up; both work *against* detecting a knowledge effect, so neither threatens the positive interaction reported.

SI-A.4 The Claim Verifiability Index

Each studied claim is scored 0–3 on four dimensions and summed to a 0–12 index; the anchors for each level are:

Dimension	0	1	2	3
Temporal lag	resolves in < 1 yr	1–3 yr	3–10 yr	> 10 yr
Counterfactual dependence	direct comparison available	weak counterfactual	strong counterfactual required	inherently unobservable counterfactual
Attributional diffuseness	single clear cause	few competing causes	many competing causes	causally over-determined
Personal unobservability	verifiable from own experience	partly observable	rarely observable	never personally observable

The index certifies that the studied claims are genuinely difficult to verify — the premise of the design — and the studied claims cluster at the high end (mean 9.5 on the 0–12 scale; Table 4). It is also used once as a regressor: §7 estimates the claim-level margin of the partisan gap on the CVI and elite contestation across every fielded claim, where the contested range of the index (6–12) permits a bounded test. Following standard practice for content-analytic measures (Krippendorff, 2004), all seventeen claims were scored independently by a second coder applying the same rubric. Agreement is high: across the sixty-eight dimension-level ratings the two coders agree exactly 88 percent of the time and never differ by more than one point, giving Krippendorff's $\alpha = 0.88$ on the ordinal dimensions (and $\alpha = 0.98$ on the summed index) — comfortably above the pre-registered $\alpha \geq 0.67$ threshold for adequate agreement (the two codings are in `data/cvi_two_coders.csv`; the instrument is `data/cvi_coding_sheet.md`). Within the seventeen expert claims the CVI lacks the variation a moderator would need; the cross-claim margin in §7 supplies what variation the fielded record contains (all fielded claims score ≥ 6), and the verifiability half-life is held apart from the numbered predictions in the theory section.

Table 4: Claim Verifiability Index scores (two independent coders; unit of analysis = 68 dimension-level ratings; Krippendorff's $\alpha = 0.88$ ordinal, 0.98 on totals).

question_id	lag	cf	diff	pers	CVI
2012_02_07_rent_control_A	3	3	3	3	12
2015_09_22_15_minimum_wage_A	2	3	2	3	10
2015_09_22_15_minimum_wage_B	2	3	3	3	11
2016_11_16_100_day_plan_A	3	3	3	3	12
2016_11_16_100_day_plan_B	3	3	3	3	12

Table 4: Claim Verifiability Index scores (two independent coders; unit of analysis = 68 dimension-level ratings; Krippendorff's $\alpha = 0.88$ ordinal, 0.98 on totals).

question_id	lag	cf	diff	pers	CVI
2021_01_13_after_brexit_A	3	3	3	3	12
2021_01_13_after_brexit_B	3	3	3	3	12
2026_01_29_aca_subsidies_A	1	2	1	1	5
2026_01_29_aca_subsidies_B	2	3	3	2	10
2026_01_29_aca_subsidies_C	2	1	2	3	8
2026_05_13_ai_work_and_education_A	3	3	3	2	11
2026_05_13_ai_work_and_education_B	3	3	3	2	11
2026_05_13_ai_work_and_education_C	3	2	3	3	11
2026_05_22_america_versus_europe_A	1	1	1	2	5
2026_05_22_america_versus_europe_B	1	1	1	2	5
2026_06_03_permanent_residency_rules_A	1	2	1	2	6
2026_06_03_permanent_residency_rules_B	2	2	2	2	8

SI-A.5 The folk-pole rubric, and its validation against the disengaged baseline

Every prediction that turns on which pole of a claim is the *folk* pole — the one-sidedness of the immigration convergence, the mirrored nuclear case, the GM placebo, the climate cell — requires that the assignment be fixed independently of the partisan data it explains. The rule is therefore codified: **the folk pole of a claim is the response predicted by the documented pre-political heuristic governing its domain**, taken from the cognitive literature and fixed before any partisan series is examined. The assignments and their sources:

Domain	Heuristic	Source	Folk pole
immigration, trade	anti-foreign bias	Caplan (2007); Boyer & Petersen (2018)	immigrants/trade take jobs
AI, automation	make-work bias (lump of labor)	Caplan (2007)	machines destroy work
economic perception	pessimistic bias	Caplan (2007)	economy worsening
nuclear power, GM crops	dread and contamination risk	Slovic (1987)	dangerous
climate	psychologically distant risk, discounted	Weber (2006)	danger underweighted
evolution	psychological essentialism	Boyer & Petersen (2018)	humans did not evolve

Two features do the disciplining. First, the rubric is *directional in both senses*: dread places the folk pole **above** the expert risk assessment (nuclear, GM), distance places it **below** (climate) — so a generic "publics distrust experts" story, which predicts one-directional deviation, cannot mimic it. Second, the rubric is validated behaviorally against respondents who have no elite pole to enact:

pure independents in each claim's *earliest* fielded wave (Table 5). On every claim the disengaged deviate from the expert benchmark in the rubric's direction, and on five of the seven they hold the folk response as an outright majority (immigration 53 percent agreeing, 69 percent of the decided; nuclear 86 percent on the danger side; GM 70 percent; AI 77 percent; automation 65 percent). The two exceptions are stated rather than argued around: on climate the criterion is directional — 48 percent of disengaged respondents in 1993 sit below the "very dangerous" pole the experts occupy, with the single largest cell at the midpoint — because distance discounts the danger without reversing it; and on trade the first fielding is late (2008), where disengaged agreement (39 percent, against an expert pole at zero) deviates toward the folk pole without commanding a decided majority. No claim's disengaged baseline sits *on* the expert pole, and no deviation runs against the rubric.

Table 5: Folk-pole validation against the disengaged baseline: pure independents in each claim's earliest fielded wave, weighted. The folk-side share uses each item's coding from SI-A.2; the decided share excludes the agree–disagree midpoint. On every claim the disengaged deviate from the expert benchmark in the rubric's direction.

Claim	Heuristic	Disengaged baseline	Folk-side response	Folk share	Decided folk share	n	Majority criterion
immigrants take jobs	anti-foreign bias (Caplan 2007; Boyer & Petersen 2018)	GSS 1996	agree	0.528	0.686	185	yes
trade -> fewer jobs	anti-foreign / protectionist bias (Caplan 2007)	GSS 2008 (first fielding)	agree	0.390	0.477	121	vs experts only (see text)
AI -> fewer jobs	make-work bias (Caplan 2007)	Pew 2024	fewer	0.771	NA	146	yes
automation -> insufficient work	make-work bias (Caplan 2007)	Pew 2017	likely	0.650	NA	83	yes
nuclear power danger	dread risk (Slovic 1987)	GSS 1993	dangerous (1-3)	0.858	NA	171	yes
GM crops danger	dread / contamination (Slovic 1987)	GSS 2000	dangerous (1-3)	0.704	NA	184	yes
climate danger	psychological distance discounting (Weber 2006)	GSS 1993	below expert pole (3-5)	0.484	NA	157	directional (see text)

SI-B: Robustness

The headline mechanism result is robust to specification, weighting, and coding.

Table 6: Full specification curve: the party×knowledge interaction (log-odds) and p-value across all 24 specifications.

belief	party	controls	weighting	coef	p
very	leaners	none	weighted	0.608	0.0e+00
very	leaners	none	unweighted	0.642	0.0e+00
very	leaners	age	weighted	0.559	0.0e+00
very	leaners	age	unweighted	0.623	0.0e+00
very	leaners	agesq	weighted	0.558	0.0e+00
very	leaners	agesq	unweighted	0.621	0.0e+00
extreme	leaners	none	weighted	0.554	4.0e-07
extreme	leaners	none	unweighted	0.696	0.0e+00
extreme	leaners	age	weighted	0.510	4.2e-06
extreme	leaners	age	unweighted	0.695	0.0e+00
extreme	leaners	agesq	weighted	0.510	5.7e-06
extreme	leaners	agesq	unweighted	0.696	0.0e+00
very	strong	none	weighted	0.923	0.0e+00
very	strong	none	unweighted	0.885	0.0e+00
very	strong	age	weighted	0.855	1.0e-07
very	strong	age	unweighted	0.866	0.0e+00
very	strong	agesq	weighted	0.846	2.0e-07
very	strong	agesq	unweighted	0.861	0.0e+00
extreme	strong	none	weighted	0.646	1.0e-03
extreme	strong	none	unweighted	0.805	1.2e-06
extreme	strong	age	weighted	0.640	1.9e-03
extreme	strong	age	unweighted	0.846	7.0e-07
extreme	strong	agesq	weighted	0.639	2.1e-03
extreme	strong	agesq	unweighted	0.844	8.0e-07

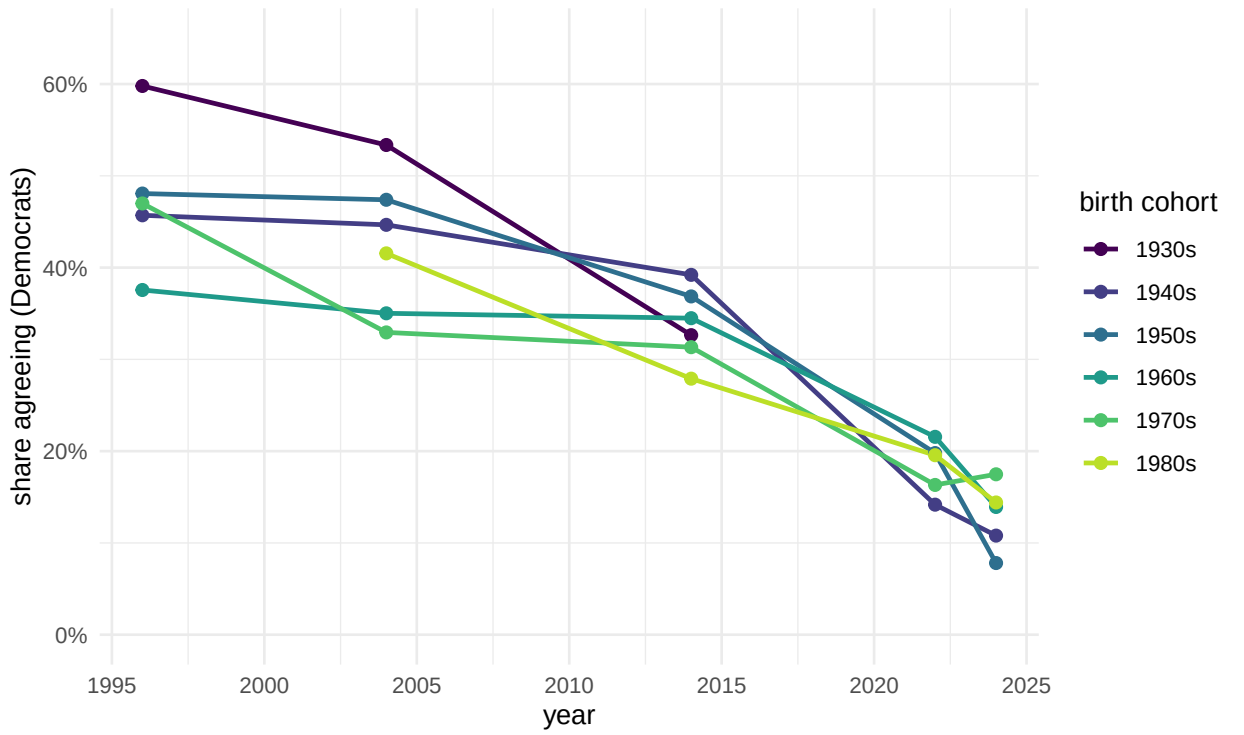


Figure 8: The generational-replacement alternative, answered without an APC identifying assumption: share of Democrats agreeing 'immigrants take jobs' by year, within ten-year birth cohorts (GSS, cells n $\geq$ 40, cohorts observed in $\geq$ 3 waves). Every cohort declines within itself (-24 to -40 points); the aggregate decline is attitude change, not composition.

Configured at entry: the imprinting margin. The within-cohort exhibit shows conversion; its complement is what capture does to citizens who were never converts because they entered the electorate after the elites had already configured the issue. Dating each respondent's political entry as birth year + 18, the first-observed party gap by socialization cohort tracks the elite contestation prevailing at entry (Table 7; `R/73`): cohorts entering before 1996 show -2 points at first observation; 1996–2011 entrants show $+7$; entrants from 2012 onward arrive at **+25 points** — sorted on arrival, at nearly the full contemporaneous magnitude. The wave-by-wave entry window (respondents 18–29) tells the same story with a detail worth keeping: the young were already sorted in 2004 ($+13.4$ points) while the gap among their elders was still -2 — capture reached *entering* cohorts before it converted incumbents, the socialization-side twin of the gap's emergence among the educated (§7.1). The pooled contrasts are disjoint: $+3.4$ (± 9.9) for the 1996–2004 entry window against $+24.6$ (± 7.6) for 2022–2024. Two mechanisms, then, compound: conversion of those who once held the shared folk belief, and configured entry of those who never did. The distinction is not bookkeeping — it prices the future. Cue withdrawal (E.2) can reverse conversion; it cannot un-socialize configured entrants; so if elite supply withdraws, the model now predicts the gap closes faster among cohorts who once held the pre-capture belief than among those who never did — an asymmetric-decay prediction the imprinting margin makes testable, registered with E.2.

Table 7: Configured at entry: the party gap on the fact at first observation, by political-socialization cohort (entry = birth year + 18), and within the 18–29 entry window by wave (GSS; cells with fewer than 25 respondents per party not estimated).

Block	Cohort / wave	Rep − Dem gap
gap at entry, by socialization cohort	entered pre-1996, first observed 1996	−0.022 (n=1064)
gap at entry, by socialization cohort	entered 1996–2011, first observed 2014	+0.068 (n=216)
gap at entry, by socialization cohort	entered 2012+, first observed 2022/24	+0.247 (n=412)
entry window (18–29) vs 30+, by wave	1996	young −0.055 (n=211) | 30+ −0.019 (n=855)
entry window (18–29) vs 30+, by wave	2004	young +0.134 (n=193) | 30+ −0.020 (n=794)
entry window (18–29) vs 30+, by wave	2014	young +0.157 (n=126) | 30+ +0.068 (n=826)
entry window (18–29) vs 30+, by wave	2022	young +0.211 (n=246) | 30+ +0.184 (n=1342)
entry window (18–29) vs 30+, by wave	2024	young +0.296 (n=187) | 30+ +0.361 (n=1356)
pooled entry-window contrasts	1996+2004	+0.034 ± 0.099 (n=404)
pooled entry-window contrasts	2022+2024	+0.246 ± 0.076 (n=433)

The premium series, year by year

Table 8: The signed sophistication premium by year: incumbent party, premium per SD of the year's information measure, native-scale premium, analytic n, and the information instrument (interviewer rating 1980–2016; behavioral items 2020/2024, signs predicted before estimation).

year	incumbent	premium (per SD)	premium (native)	n	instrument
1980	Democratic	0.003	0.013	1336	interviewer rating
1984	Republican	-0.042	-0.165	1874	interviewer rating
1988	Republican	-0.072	-0.263	1713	interviewer rating
1992	Republican	-0.041	-0.164	2134	interviewer rating
1996	Democratic	0.034	0.131	1494	interviewer rating

Table 8: The signed sophistication premium by year: incumbent party, premium per SD of the year's information measure, native-scale premium, analytic n, and the information instrument (interviewer rating 1980–2016; behavioral items 2020/2024, signs predicted before estimation).

year	incumbent	premium (per SD)	premium (native)	n	instrument
2000	Democratic	0.030	0.111	1551	interviewer rating
2002	Republican	-0.020	-0.085	1359	interviewer rating
2004	Republican	-0.078	-0.288	1037	interviewer rating
2008	Republican	-0.013	-0.050	2001	interviewer rating
2012	Democratic	0.136	0.521	1785	interviewer rating
2016	Democratic	0.072	0.270	973	interviewer rating
2020	Republican	-0.040	-0.087	7104	House-majority recall (behavioral)
2024	Democratic	0.112	0.094	4877	civics battery, 4 items (behavioral)

Block-structure inference for the premium reversal

The 1,716-assignment exact test treats the thirteen years as exchangeable. They are not quite: incumbency arrives in consecutive runs, and years within an administration share shocks (a recession spans a term), so free year-permutation could overstate the evidence. Table 9 re-runs the randomization at every defensible level of exchangeability, with the incumbency coefficient and the sign-match count as joint statistics. At the presidency level (eight blocks, 70 assignments) the observed assignment is the most extreme possible on both statistics. At the coarsest level — seven runs of consecutive same-party incumbency, where only 35 assignments exist and the minimum attainable one-sided p is $1/35 = .029$ — the observed assignment attains exactly that minimum; the two-sided coefficient p is .057, which we report rather than suppress, noting that the direction was stated in advance for every year (the theory signs the premium by the incumbent, so the one-sided test is the registered one). Two further forks discipline the sign count itself: scoring the near-zero years as ties (dropping $|\text{premium}| < 0.01$, then < 0.02) leaves 12/12 and 10/10 (binomial $p = .0002$, .001); and dropping 2020 — the one year whose behavioral instrument is a single item — leaves the coefficient unchanged (+0.054) at 12/12, exact $p = .002$.

On the 2020 instrument: the CDF carries no valid Senate-majority recall for 2020 (`VCF0730` is empty that year), so House recall is the only behavioral item available, and it is the item type our own 2024 audit flags for partisan DIF (Table 22). The direction of any such contamination is, however, conservative here: the pre-2020 House majority was Democratic, so own-side flattery misclassifies knowledgeable *Republicans* as low-knowledge, attenuating the Republican optimism gradient under

the Republican incumbent — biasing the 2020 premium toward zero, against the predicted negative sign it nonetheless shows.

Table 9: Inference for the premium sign-reversal at every level of exchangeability. Exact permutation tests at the year, presidency, and party-run level (the last permits a minimum one-sided p of .029, which the observed assignment attains), plus ties and drop-2020 forks.

Test	Units	Assignments	Statistic	p
year-level exact (all 1,716 label assignments)	13	1716	incumbency coefficient +0.054 / sign-match 13 of 13	0.0006 (two-sided; observed most extreme)
presidency-block exact (8 blocks: Carter, Reagan, Bush 41, Clinton, Bush 43, Obama, Trump, Biden)	8	70	incumbency coefficient / sign-match	coef 0.029 two-sided; sign 0.014 (observed most extreme)
party-run-block exact (7 runs of consecutive same-party incumbency)	7	35	incumbency coefficient / sign-match	coef 0.057 two-sided, 0.029 one-sided (directional prediction stated in advance); sign 0.029 — observed most extreme; minimum attainable at 7 blocks
near-zero years as ties: drop $\|\text{premium}\| < 0.01$ (1980)	12	NA	sign-match 12 of 12	0.00024 (one-sided binomial)
near-zero years as ties: drop $\|\text{premium}\| < 0.02$ (1980, 2008, 2002)	10	NA	sign-match 10 of 10	0.00098 (one-sided binomial)
drop 2020 (single-item behavioral instrument year)	12	924	incumbency coefficient +0.054 / sign-match 12 of 12	0.0022 (exact two-sided)

Fundamentals anchor for the perception series

The economic-perception gap and premium are identified against incumbency, but the perception item has a measurable referent, so we anchor both series to realized fundamentals: election-year real GDP growth (BEA) and the change in the annual-average unemployment rate (BLS). Incumbency survives untouched and the fundamentals do no work (Table 10) — alone, they explain a fifth of the variance with neither significant, against incumbency's seven-tenths. The partisan structure of the perception is bias against the measured economy, not a report of genuinely different economies.

Table 10: Anchoring the perception series to realized fundamentals (election-year real GDP growth, unemployment-rate change). The incumbency coefficient is essentially unchanged with fundamentals in the model, and the fundamentals themselves are not significant.

Quantity	Estimate	p
partisan gap (18 elections): incumbency, alone	0.801	0.0000
partisan gap (18 elections): incumbency, + fundamentals	0.756	0.0001
partisan gap (18 elections): real GDP growth	-0.179	0.1539
partisan gap (18 elections): unemployment change	-0.281	0.1916
signed premium (13 years): incumbency, alone	0.054	0.0004
signed premium (13 years): incumbency, + fundamentals	0.050	0.0041
signed premium (13 years): real GDP growth	-0.009	0.4682
signed premium (13 years): unemployment change	-0.012	0.5237
gap ~ fundamentals only: R^2	0.194	NA
premium ~ fundamentals only: R^2	0.246	NA

The year-level anchoring can be discounted as an eighteen-observation regression, so the same claims are re-established at the individual level (Table 11; `R/69`). First, the premium as one powered estimate: pessimism regressed on party × knowledge × incumbency with year fixed effects across all thirteen information-measure years gives +0.056 per SD (year-clustered $t(12)$ $p = .0005$; n = 29,238) — the micro-level twin of the year-level +0.054, and the magnitude companion to the sign test. Second, *measured error*, with accuracy coding kept incontestable by restricting to years any reasonable mapping classifies alike — implied "better" only where election-year growth ≥ 3.5 percent *and* unemployment fell (1984, 1988, 1996, 2004), implied "worse" only where growth ≤ 0 *and* unemployment rose a point or more (1980, 2008, 2020). On those seven years (n = 16,559), a truth that favors the other side raises a respondent's error by +0.26 on the 0–1 scale; political knowledge reduces error among the truth-*congenial* (−0.042 per SD, $p = .014$) and does nothing among the truth-*uncongenial* (−0.004, $p = .75$; difference +0.034, $p = .007$). Knowledge corrects error only when the correction is party-congenial — the accuracy-scale statement of the manufactured premium, and the individual-level answer to any residual worry that the partisan structure of perception reflects genuinely different economies.

Table 11: The perception series at the individual level: the pooled three-way premium (13 years, year FE), and measured error against unambiguous fundamentals (7 years where any reasonable accuracy mapping agrees). Knowledge corrects error only where the correction is party-congenial.

Quantity	Estimate	SE	p	n
pooled individual-level manufactured premium: rep × knowledge × incumbency (13 years, year FE)	0.0555	0.0118 (year-clustered)	.0005 (t(12)); HC1 p < 1e-26	29238
error, truth-uncongenial main effect (7 unambiguous years)	0.2602	—	—	16559
knowledge slope on error, truth-congenial respondents	-0.0416	0.0122 (year-clustered)	.014 (t(6))	8617

Table 11: The perception series at the individual level: the pooled three-way premium (13 years, year FE), and measured error against unambiguous fundamentals (7 years where any reasonable accuracy mapping agrees). Knowledge corrects error only where the correction is party-congenial.

Quantity	Estimate	SE	p	n
knowledge slope on error, truth-uncongenial respondents	-0.0039	0.0115 (year-clustered)	.75 (t(6))	7942
difference: knowledge × uncongenial on error	0.0335	0.0083 (year-clustered)	.007 (t(6)); HC1 p = 1.4e-07	16559

Even the wallet: the premium on the personal retrospective. The CDF fields the *personal* twin of the perception item — whether the respondent is better or worse off than a year ago (`VCF0880`) — in the same years, to the same respondents, under the same information measures: the closest thing in the design to an experientially checkable economic claim. It is not exempt (Table 12). The personal item carries its own configuration-signed premium: the party×information interaction matches the incumbent's party in twelve of thirteen years (the single miss, 1992, is +0.0006 — exactly zero), with an incumbency slope of +0.029 ($p = .001$, $R^2 = .64$) and a pooled three-way of +0.032 (year-clustered $p = .0001$) — informed partisans report even their own household's year through the configuration, at roughly half the national magnitude. We draw the discipline before the moral: the national-to-personal premium ratio (1.8:1) runs in the direction the verifiability term predicts, but the same-respondent difference is *not* separable at this power (item×incumbency×party×knowledge = +0.017, year-clustered $p = .21$), so this is not an identification of $\alpha(v)$ and is not presented as one — the designed contrast remains E.5's. What it does establish is the reach of the institutional margin §3 describes: a retrospection about one's *own* finances is evaluative and attributionally diffuse enough for the configuration to sign its premium too (Gerber & Huber, 2010), which is why the scope condition is written in terms of experiential verifiability of the *claim*, not mere proximity of its referent.

Table 12: The premium on the personal-finances retrospective (VCF0880), beside the national item: configuration-signed at half the national magnitude (12/13), with the same-respondent difference bounded but not separable (p = .21) — reported as reach of the institutional margin, not as an identification of the verifiability term.

Quantity	Estimate	Note
national item: signed premium ~ incumbency (13 years)	0.0540	$R^2 = .69$; sign-match 13/13
personal item: signed premium ~ incumbency (13 years)	0.0290	p = .001, $R^2 = .64$; sign-match 12/13 (miss: 1992, premium +0.0006)
pooled 3-way premium, national (inc×rep×know, year FE)	0.0572	year-clustered p = .0005, n = 29,238
pooled 3-way premium, personal	0.0318	year-clustered p = .0001, n = 29,266
ratio national : personal	1.8000	predicted direction (national less experientially checkable)

Table 12: The premium on the personal-finances retrospective (VCF0880), beside the national item: configuration-signed at half the national magnitude (12/13), with the same-respondent difference bounded but not separable (p = .21) — reported as reach of the institutional margin, not as an identification of the verifiability term.

Quantity	Esti-mate	Note
same-respondent difference (item × inc × rep × know)	0.0170	year-clustered p = .21 — NOT separable; reported as bounded

The independents placebo: bounding referent drift

Fixed wording is not automatically a fixed stimulus: the composition of "immigrants" and the salience of legal status differ between 1996 and 2024, and the `immameco` co-movement cannot rule this out because it shares the referent. The test that can is the *pure-independent* series (Figure 9): referent drift is common to all respondents, so it moves independents as it moves partisans, while capture moves partisans apart around the independent baseline. Independents track Democrats almost exactly (−0.136 vs −0.134 per decade; difference-in-trends +0.003, $p = .88$), while Republicans diverge from the common baseline by +0.123 per decade ($p < .001$). Two implications, stated plainly. First, the partisan gap cannot be a referent-drift artifact: drift has no mechanism for holding one party flat against a baseline it moves. Second, the decomposition is informative about *where* the partisan act sits over time: the Democratic decline travels with the ambient baseline, while the distinctively partisan movement is the Republican anchor at the folk pole — which is precisely where the model places the in-group ($p_R = f$), and consistent with the cross-sectional gradient evidence that knowledge does its distinctive work within whichever party the configuration assigns.

What, then, *is* the ambient corrective flow the placebo conditions on? Not a free parameter: the elite signal has a first moment as well as a second, and the corpus's own headline finding is that congressional immigration tone grew more positive overall *while* polarizing (**?**). Pairing the two elite moments with the two mass moments across the five capstone waves (Table 13; `R/68`), the independent baseline tracks the elite *mean* tone at $r = -0.86$ — as the shared signal warmed, the disengaged left the folk belief — while the party gap tracks the *divergence* at $r = +0.78$. We flag the limit as plainly as the fit: at five waves the two elite moments themselves co-move too strongly for each mass moment to be assigned its own elite moment (the cross-correlations are comparable), so this is a named, measured candidate for the ambient flow — the shared component of elite speech — not a separated decomposition; the separation is what the unified continuous index (E.9) is for. What the pairing already rules out is the reading on which the baseline's decline is unexplained noise: both moments of mass belief move with the corresponding, measured moments of the elite signal.

Table 13: Two moments of the elite signal against two moments of mass belief (five capstone waves; elite series are trailing five-year means of congressional immigration tone, truncated at the corpus's 2020 boundary for the 2022/2024 waves). The independent baseline tracks the elite mean (r = −0.86); the party gap tracks the divergence (r = +0.78); at five waves the elite moments co-move too strongly for a separated assignment (see text).

GSS wave	Elite series through	Elite mean tone	Elite divergence	Independent folk share	Party gap
1996	1996	0.995	0.300	0.528	-0.026
2004	2004	1.140	0.302	0.528	0.009
2014	2014	1.155	0.686	0.323	0.084
2022	2020	1.243	0.683	0.185	0.196
2024	2020	1.243	0.683	0.204	0.350

The cross-sectional companion to the trend placebo is the independents' *knowledge gradient* (Table 14; `R/61_independents_knowledge.R`). Scoring pure independents (`V241227x` = 4) on the identical battery and scale as the partisans, the knowledge slope on the fact is −0.47 among Democrats ($p < .001$), +0.04 among Republicans (n.s.), and −0.08 among pure independents (n.s., n = 300): informed independents do not leave the folk belief, and the Democratic slope is significantly steeper than the independents' (difference +0.40, $p = .02$) while the Republican and independent slopes are indistinguishable (+0.11, $p = .45$). The two margins divide the labor: over *time*, independents move with Democrats (the ambient corrective flow) and Republicans anchor; across *capacity*, independents sit still and informed Democrats converge on their party's pole. Each margin isolates one party's distinctively partisan act, and both are capture. The gradient also disciplines the truth-tracking and exposure rivals at once: an information environment that carried the expert signal to the informed as such would move informed independents too. It does not. (Point estimates replicate the pipeline's survey-weighted logits; the small independent sample bounds the precision, stated on the table. The flat independent slope is also equivalence-bounded below the Democratic slope it is contrasted with — TOST $p = .004$ at the full magnitude — though a cell of n = 300 cannot bound it near zero, $\text{MDE}_{80} = 0.42$; Table 21.)

Table 14: The independents placebo, cross-sectional form: share holding the folk belief by political knowledge (ANES 2024, weighted). Knowledge moves Democrats only; pure independents (n = 300) are flat at every level, indistinguishable from Republicans in slope.

knowledge	Dem share	Ind share	Rep share	n Dem	n Ind	n Rep
low (0-1)	0.172	0.351	0.508	797	169	833
mid (2)	0.130	0.409	0.545	653	82	633
high (3-4)	0.071	0.345	0.533	873	49	613

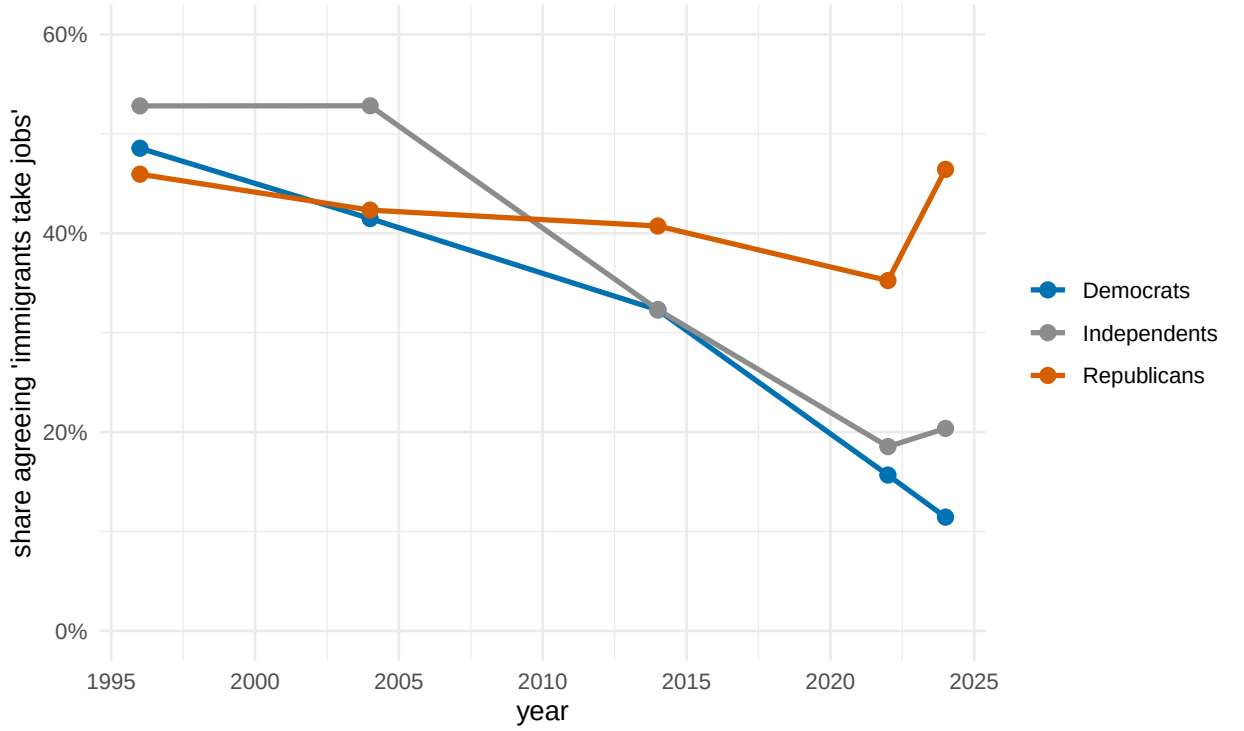


Figure 9: The independents placebo. Share holding the folk belief ('immigrants take jobs'), GSS 1996–2024, by party with pure independents as the drift baseline. Independents decline in lockstep with Democrats (difference-in-trends +0.003, p = .88); Republicans hold flat against the common baseline (+0.123 per decade net, p < .001). Referent drift moves everyone and so cannot produce the party-differential trend.

Heterogeneity-robust event study and placebo onsets

Staggered onsets estimated by two-way fixed effects invite the Goodman–Bacon / negative-weights objection. With two treated cohorts and a five-issue never-treated pool the scope for contamination is small, and re-estimation confirms it (Table 15): the Callaway–Sant'Anna-style group-time effects (cohort base = last pre-onset observation, never-treated controls) aggregate to +0.44, *above* the TWFE difference-in-differences of +0.34, so the main-text figure errs conservative. Placebo-onset randomization — 5,000 reassignments of which two issues are treated and when, requiring at least two pre- and post-onset observations — places the observed effect past the 99th percentile of the placebo distribution.

Table 15: Event-study robustness: Callaway–Sant'Anna-style estimation with never-treated controls, and placebo-onset randomization inference (5,000 draws reassigning treated issues and onset years).

Quantity	Value
TWFE DiD (post, issue+year FE)	+0.335
CS-style ATT, immigration cohort (base 2008, never-treated controls)	+0.419
CS-style ATT, trade cohort (base 2012, never-treated controls)	+0.455

Table 15: Event-study robustness: Callaway–Sant'Anna-style estimation with never-treated controls, and placebo-onset randomization inference (5,000 draws reassigning treated issues and onset years).

Quantity	Value
CS-style ATT, overall (equal cohort weights)	+0.437
placebo-onset randomization: draws	5000
placebo-onset randomization: one-sided p (DiD >= observed)	0.0028
placebo-onset randomization: two-sided p (|DiD| >= observed)	0.0060
placebo distribution 95th / 99th percentile	+0.176 / +0.255

Table 16: Leave-one-comparison-issue-out: the TWFE DiD re-estimated dropping each never-treated issue in turn. The estimate is stable (+0.28 to +0.37, all p < .004): no single comparison issue supplies the counterfactual.

Dropped control issue	DiD	p
welfare	0.349	0.0015
health	0.366	0.0002
morality	0.278	0.0013
environment	0.344	0.0004
defense	0.310	0.0039
none (full panel)	0.335	0.0002

Onset dating, the rival clock, and outcome-form forks

The design's remaining soft spots are the onset dates (the paper's two elite series disagree on immigration's rise by several years), the absolute-value outcome (bounded below, so near-zero true gaps carry a positive noise floor), and the possibility that long-divided controls sit at ceiling. All three are tested (Table 17, Table 18; `R/67`; inference is issue-clustered with $t(6)$ critical values, more conservative than the main-text CR p-values). Shifting both onsets -4 to $+4$ years bounds the DiD between +0.20 and +0.44: earlier dating dilutes the post period with pre years, exactly as a true onset implies, and never erases the effect. Dating immigration by the speech clock (2008) instead of the platform clock returns +0.28 ($p = .017$) — the two elite series disagree about *when* precisely, not about *whether*, and either dating delivers the result. Fixing each issue's direction by its mean signed gap and re-running on the signed outcome gives +0.57 ($p = .001$), larger than the published +0.34, so the |gap| floor biases the headline *down*. And the control issues have ample headroom: their gap magnitudes average 0.40–0.73 item-SDs with within-issue ranges up to 0.8 SD, against treated pre-onset levels averaging 0.14 — the counterfactual is supplied by series that demonstrably move.

Table 17: Event-study forks: onset shifts, the rival (speech) elite clock, and the signed-gap outcome (TWFE DiD; issue-clustered CRV1 with t(6) critical values — more conservative than the main-text CR inference).

Fork	DiD	p	Note
baseline (platform onsets 2012/2016)	0.335	.007 (issue-clustered, t(6))	replicates R/33; published CR p = .0002
both onsets shifted −4 years	0.201	.081	earlier dating can only dilute post with pre years
both onsets shifted −2 years	0.335	.007	no fielded year enters the window
both onsets shifted +2 years	0.437	.005	NA
both onsets shifted +4 years	0.437	.005	NA
immigration dated by the speech clock (2008)	0.283	.017	the rival elite series gives the same answer
signed-gap outcome (direction fixed by issue mean)	0.573	.001	the |gap| floor concern is conservative: signing enlarges the estimate

Table 18: Control-issue headroom: distribution of the party-gap magnitude (item-SD units) for the five never-treated issues. None sits at a ceiling; each moves across its series.

Issue	Mean	SD	Min	Max	Years
defense	0.550	0.161	0.292	0.904	15
environment	0.528	0.116	0.364	0.795	12
health	0.728	0.254	0.409	1.212	16
morality	0.399	0.110	0.210	0.541	10
welfare	0.687	0.203	0.406	1.183	21

Sensitivity to parallel-trends violations, and the power of the pre-test

Testing pre-trends and equivalence-bounding them (Table 21) still leaves the modern question open: what if parallel trends were violated by amounts the pre-period cannot rule out (Rambachan & Roth, 2023), and could a violation that matters have slipped past a pre-test of limited power (Roth, 2022)? Both are answered on the estimated event-time system (Table 19; `R/76`). First, bias-aware bounds: allowing post-onset violations as large as $\bar{M}$ times the largest pre-period deviation $\bar{B}$ — taken both at its point value (0.139) and at the ultra-conservative 95-percent upper envelope of the noisiest pre-bin (0.305) — the DiD's breakdown value is $\bar{M}^* = 1.23$ (point) and 0.56 (envelope), the +4 coefficient's 2.10 and 0.96, and the long-run +8 coefficient's **3.98 and 1.81**: the effect the theory cares about survives violations two to four times anything the pre-period exhibits, and clears even the envelope bound with $\bar{M} = 1.8$. Exact Rambachan–Roth relative-magnitude confidence sets are computed beside these closed-form bounds on the pipeline run. Second, the power of the pre-test itself: the joint Wald test on the three pre-bins detects linear violations of slope ≥ 0.091 per four-year bin with 80-percent power; a trend at exactly that detection threshold could bias the

DiD by at most +0.21 of its +0.335 — a real residual we state rather than round away — while **a trend large enough to fabricate the entire DiD (slope 0.147) would have been detected with essentially certain power**. Third, the estimator triad completes: the efficient imputation estimator (Borusyak et al., 2024) gives +0.384 (issue-jackknife se 0.103), sitting between the TWFE +0.335 and the Callaway–Sant'Anna +0.437, and the de Chaisemartin–D'Haultfœuille dynamic estimator (Chaisemartin & D'Haultfœuille, 2024) — robust to *non-absorbing* treatment, the right check where platform divergence can recede — runs alongside in `R/76`.

Table 19: The event study against its one open assumption: bias-aware breakdown values in the Rambachan–Roth relative-magnitudes framework (closed-form, with the exact CIs computed beside them in R/76), the Roth power analysis of the pre-test, and the modern estimator triad.

Quantity	Value	Note
bias-aware bound, DiD (+0.335): breakdown $\bar{M}^*$ (point $\bar{B} = 0.139$)	1.23	survives violations up to 1.2× the largest pre-deviation; lower bound at $\bar{M} = 1$: +0.03
bias-aware bound, DiD: breakdown $\bar{M}^*$ (95%-upper $\bar{B} = 0.305$)	0.56	ultra-conservative envelope (upper CI of the noisiest pre-bin)
bias-aware bound, event-time +4 (+0.375): breakdown $\bar{M}^*$	2.10 / 0.96	point / 95%-upper $\bar{B}$
bias-aware bound, event-time +8 (+0.656): breakdown $\bar{M}^*$	3.98 / 1.81	the long-run effect survives even the ultra-conservative envelope at $\bar{M} = 1.8$; lower bound at $\bar{M} = 1$: +0.41 / +0.25
pre-test power (Roth): slope detectable with 80% power	0.091 per 4-yr bin	joint Wald on the three pre-bins, Σ from the issue-clustered fit
largest undetected-trend bias on the DiD	+0.21	a trend at exactly the 80%-power threshold biases the DiD by at most 0.21 of its 0.335
slope needed to fabricate the entire DiD	0.147 — detected with ≈100% power	a violation large enough to produce the whole effect would not have slipped past the pre-test
BJS imputation estimator (efficient; untreated-cell FE imputation)	+0.384 (issue-jackknife se 0.103)	triad: TWFE +0.335, CS +0.437, BJS +0.384
de Chaisemartin–D'Haultfœuille dynamic estimator	computed on the pipeline run (R/76)	non-absorbing-treatment robustness; package row fills on author's machine
exact Rambachan–Roth relative-magnitude CIs	computed on the pipeline run (R/76)	HonestDiD conditional/hybrid intervals appended beside the closed-form bounds above

Probability-scale versions of the headline interactions

Because log-odds interaction coefficients are not comparable across samples, years, or model specifications — unobserved heterogeneity rescales the logistic coefficients (Mood, 2010) — every

headline interaction is also reported as an average marginal effect on the probability scale, with the interaction itself as a *second difference* (Table 20). The knowledge-based primary survives the translation cleanly: the design-based party gap on the fact widens from 34 points among low-knowledge respondents to 46 among high, a second difference of +12.6 points (95% CI [5.5, 19.6], PSU-clustered). The over-time widening is +17 points per decade on the probability scale ([11, 24], survey-year-clustered), matching the random-intercept specification in §5. The GSS education version, by contrast, is scale-sensitive: +0.55 in log-odds ($p = .02$) but only +3 points ([−6, +11]) as a second difference — consistent with education's pre-assigned role as the weaker, conditional proxy (education also corrects the folk error in both parties, compressing the probability-scale contrast), and a caution we report plainly rather than bury. The signed premium series needs no translation: it is estimated by linear regression on the 0–1 pessimism scale throughout.

Table 20: Headline interactions as average marginal effects on the probability scale (delta-method 95% CIs; ANES 2024 clustered by PSU, pooled series by survey-year, GSS by HC1). The interaction is reported as a second difference in the predicted-probability party gap.

Quantity	Estimate	CI lower	CI upper
ANES24 fact: party gap, low knowledge (0-1)	0.336	0.279	0.393
ANES24 fact: party gap, mid knowledge (2)	0.415	0.362	0.467
ANES24 fact: party gap, high knowledge (3-4)	0.462	0.425	0.499
ANES24 fact: second difference (high - low)	0.126	0.055	0.196
pooled: widening per decade, prob scale (AME)	0.174	0.112	0.235
GSS immjobs 2014-24: party gap, non-college	0.210	0.160	0.260
GSS immjobs 2014-24: party gap, college	0.235	0.173	0.298
GSS immjobs: second difference (college - non-college)	0.025	-0.055	0.106

The procedure follows specification-curve and multiverse practice (Simonsohn et al., 2020; Steegen et al., 2016): rather than defend a single model, we report the full distribution of estimates across defensible analytic choices. Three points follow. First, every specification yields a positive, significant interaction (Figure 13 and the table above), so the result is not an artifact of a particular control set or coding. Second, the estimate is essentially unchanged with the survey weights removed, so it is not an artifact of the weighting. Third, the fact–preference coupling holds for both the older levels item (`letin1`) and the newer reduce item (`letin1a`) (Figure 29), so the stable-coupling result is not an artifact of item wording. Throughout, each survey is analyzed with its own native design weights; no single weight is pooled across surveys.

Two further caveats. The precedence test rests on eleven biennial observations; its lead–lag correlation and Granger-style coefficient are reported as suggestive temporal ordering, not as a powered time-series result, and we do not subject so short a series to unit-root or cointegration testing, which it cannot support. In place of asymptotic inference we use a time-shift randomization that respects autocorrelation: we correlate the public gap against the elite series displaced by every lag in a wide window and rank the observed ten-year alignment within that distribution. The lag-ten correlation falls near the ninetieth percentile ($p \approx .10$) — consistent with elites leading, but conservative enough that we claim only ordering, not a precise lag or an effect size. This randomization is the appropriate small-sample replacement for the asymptotic Granger p, and it is deliberately the weaker of the two figures we report. Second, the `immjobs` item was fielded in 1996, 2004, 2014, 2022, and 2024 — not in 2021 — so the GSS's partial shift to web administration in 2021 does not enter the capstone series.

The one notable irregularity, the 2022 GSS education split, is visible in Figure 15 and does not affect the design-based interaction or the cross-survey ANES confirmation. Across the five capstone waves the `immjobs` item was administered to comparable, randomly assigned ballot subsamples, with analytic Republican and Democratic counts per wave reported in Table 55 (ranging from roughly 480 to 960 per party); the trend is not an artifact of the irregular spacing of fielding years, and — as §6 notes — the split-ballot structure is precisely why we retire the earlier within-person "fusion" reading in favor of the partisan-loading result.

How absent is absent? Equivalence bounds for the load-bearing nulls

Several of the paper's pivotal quantities are *absences* — the pre-onset premium, the GM-crops placebo, the flat independents gradient, the education×era null, the event-study pre-trends, the 2004 mass gap — and an absence asserted from a non-significant coefficient is only as good as the power behind it. We therefore treat every null that carries argumentative load to the same discipline (Table 21): the exact design-based (sandwich) standard error, the 90% confidence interval, two one-sided tests (TOST) against a full-magnitude bound fixed by the theoretically relevant comparison and against half that bound, and the minimum detectable effect at 80% power. The bounds are stated before the tests: for the 2004 premium, the smallest post-onset yearly estimate (+0.39) and half the pooled post-onset estimate; for the GM placebo, the mirrored nuclear premium; for the independents slope, the Democratic slope it is compared against; for education×era, the party×era estimate; for the event-study leads, the DiD; for the 2004 mass gap, the first post-onset gap (2014).

Three results and one confession. First, the switch-on's pre-period survives: the 2004 premium is statistically smaller than every post-onset yearly estimate (TOST $p = .012$), and — because a single year of n = 921 cannot exclude intermediate values ($\text{MDE}_{80} = 0.42$) — we also test the pre-to-post *change* directly: the pooled 2008–2016 premium exceeds the 2004 estimate by +0.36 (one-sided $p = .019$). "Absent, then present" is a tested difference, not a glance at two p-values. Second, the placebos hold at full magnitude: the GM premium, the independents slope, the education×era term, and all three event-study leads are each equivalence-bounded below their full comparison magnitudes (all TOST $p \leq .010$, the −16 bin exactly at .010). Third, the 2004 mass gap is bounded below the first post-onset opening in the GSS ($p = .017$) and in the two-survey precision pool ($p = .003$); the ANES 2004 point alone is imprecise, and its sign runs *opposite* the eventual gap. The confession: the strict half-magnitude bound is met by the well-powered cells (education×era, the −8 lead) and not by the small-n cells (the 2004 premium, GM, independents, the earliest lead bin), whose single-cell MDEs sit near 0.1–0.4 log-odds. We report that plainly: the nulls are bounded below the effects they are contrasted with, which is what the argument requires; they are not bounded near zero, which no honest reading of these sample sizes could claim.

Table 21: Equivalence bounds for the load-bearing nulls. Each null is reported with its design-based (sandwich) SE, 90% CI, TOST p-values against a pre-stated full-magnitude bound and half of it, and the 80%-power minimum detectable effect. The premium-change row is a direct difference test (one-sided), not a TOST.

Quantity	Estimate	SE	90% lo	90% hi	Bound (full)	TOST p	Bound (half)	TOST p (half)	MDE80
premium, 2004 (pre-onset; ANES CDF)	0.054	0.149	-0.192	0.299	0.390	0.012	0.205	0.155	0.418
premium change, pooled 2008-2016 minus 2004	0.355	0.171	0.074	0.636	NA	NA	NA	NA	NA
premium, GM crops placebo	0.013	0.033	-0.041	0.067	0.098	0.005	0.049	0.138	0.092
knowledge slope, pure independents	-0.076	0.149	-0.321	0.169	0.471	0.004	0.236	0.142	0.417
education x era (log-odds)	0.129	0.174	-0.157	0.415	1.218	0.000	0.609	0.003	0.487
event-study lead, -16 or earlier	0.139	0.085	0.000	0.278	0.335	0.010	0.168	0.366	0.237
event-study lead, -12	0.051	0.087	-0.092	0.194	0.335	0.000	0.168	0.089	0.243
event-study lead, -8	0.014	0.056	-0.078	0.106	0.335	0.000	0.168	0.003	0.156
mass gap 2004, GSS (design-based)	0.009	0.036	-0.050	0.067	0.084	0.017	0.042	0.176	0.100
mass gap 2004, ANES CDF	-0.040	0.036	-0.100	0.020	0.084	0.114	0.042	0.480	0.102
mass gap 2004, precision-pooled (GSS+ANES)	-0.015	0.025	-0.057	0.027	0.084	0.003	0.042	0.146	0.071

Wording robustness: a second immigration-economy fact

A natural worry is that the partisanization is an artifact of the specific phrase "immigrants take jobs." It is not. The GSS fields a second, differently framed claim about the same underlying fact — whether immigrants are, on balance, good for the American economy (`immameco`) — in 1996, 2004, 2014, and 2024. Its partisan gap follows an all-but-identical path (Figure 10): flat through 2004, opening by 2014, and reaching twenty-two points by 2024, driven by the same one-sided movement — the share of Democrats calling immigrants good for the economy rises from 68 to 94 percent while Republicans barely move. The two fact-item gap trajectories correlate at $r = .95$ across the four common years. The capture is a property of the belief about immigration and the economy, not of any one item's wording. A direct measurement of wording *stability* — diachronic contextual embeddings of the claim's usage contexts in elite and news text, benchmarked against known-drift and known-stable control phrases — is specified in the replication repository (`tools/semantic_drift_protocol.md`); a stability curve inside the control band would convert

the fixed-meaning premise from asserted to measured, and a break dated to the elite clock would itself be informative.

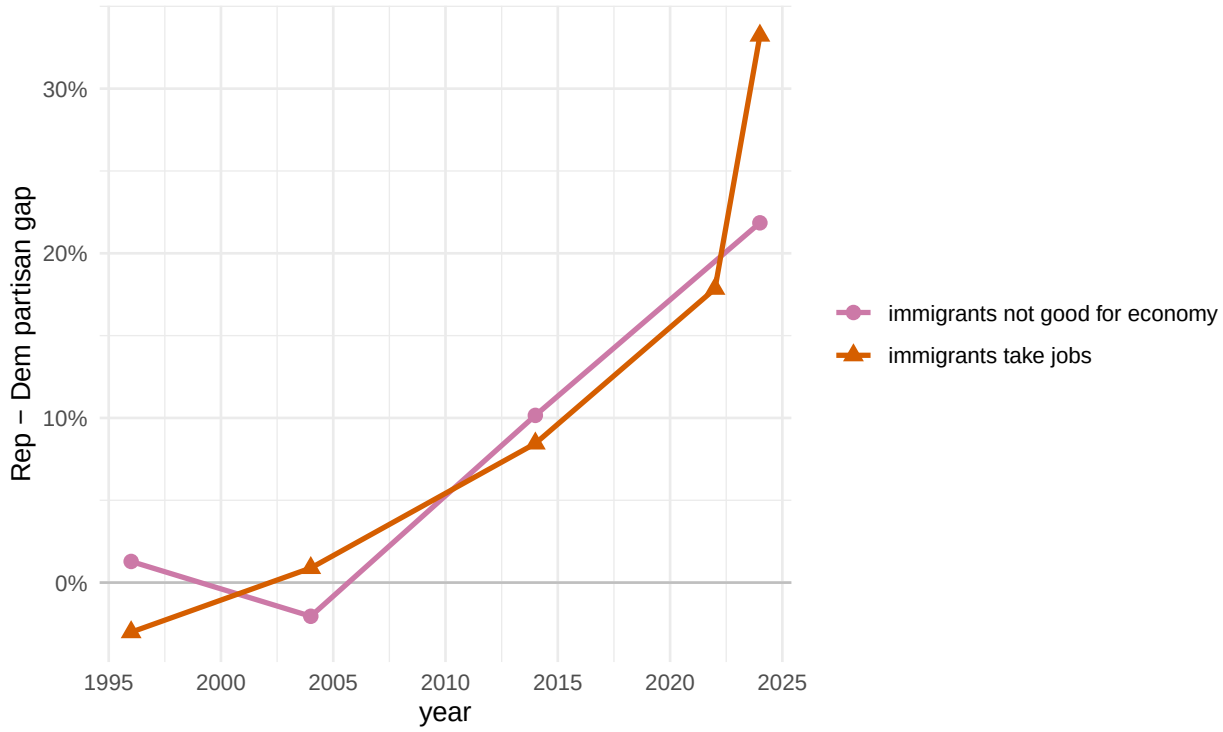


Figure 10: Wording robustness: the partisan gap on a differently worded immigration-economy fact (whether immigrants are *not* good for the economy, immameco) tracks the gap on 'immigrants take jobs' (r = .95 across the four common years; GSS).

Mechanism specificity: knowledge, not its correlates

A natural objection to the mechanism is that "political knowledge" is merely a proxy for education or ideology, so the knowledge×party interaction would vanish once those interactions are included. It does not. The interaction retains essentially its full magnitude and high significance controlling for party×education and party×ideology, while both of those interactions are themselves null — it is specifically political sophistication, not its correlates, that amplifies the sorting.

A second door is measurement: two of the four knowledge items ask which party controls the House or Senate, and partisans might answer own-side-control questions asymmetrically, smuggling partisanship into the moderator. We audit the battery three ways (Table 22). Each item is tested for party-DIF matched on the rest-score: at this sample size every item shows statistically detectable uniform DIF, but only the House-control item crosses the ETS non-negligible boundary (−0.45 log-odds), and the signs are the opposite of flattery — on three of four items, including House control, *Republicans* are less likely to answer correctly at a given rest-score, so any contamination biases the interaction *downward*. Re-scoring by a 2PL IRT model and re-estimating returns +0.52; the two civics items alone (Senate term length, foreign-aid spending — no partisan content) return +0.30 ($p = .008$); dropping the flagged item returns +0.46. The headline interaction is invariant to how knowledge is scored, which closes the last measurement reading of the mechanism test.

Table 22: Party-DIF per knowledge item, matched on the rest-score (uniform = party main effect; non-uniform = party × rest-score; flag = Bonferroni-significant and ≥ 0.43 log-odds, the ETS non-negligible boundary). Negative = Republicans less likely correct at a given rest-score.

item	uniform DIF	p	non-uniform	flag
k_senterm	-0.30	2.8e-06	-0.18	FALSE
k_foraid	-0.37	6.0e-07	-0.16	FALSE
k_house	-0.45	0.0e+00	0.05	TRUE
k_senate	0.23	3.6e-04	-0.08	FALSE

Table 23: The party×knowledge interaction is scoring-invariant: sum score, 2PL IRT (EAP), the two civics items alone, and the battery excluding the DIF-flagged House-control item (design-based logit, ANES 2024).

scoring	party×knowledge	se	p
sum score 0-4 (headline)	0.51	0.11	8.1e-06
2PL IRT (EAP)	0.52	0.11	3.6e-06
civics items only (Senate term, foreign aid)	0.30	0.11	8.4e-03
excluding DIF-flagged (k_house)	0.46	0.12	7.0e-05
net of rep x strength and knz x strength	0.52	0.11	3.8e-06

Table 24: Party×knowledge interaction on the immigration fact, net of party×education and party×ideology (ANES 2024).

Model	party×knowledge	p	n
baseline	0.526	0.000	4402
+ party x education	0.467	0.001	4369
+ party x ideology	0.491	0.000	3920
+ both	0.483	0.001	3896

A sharper version of the objection names a specific rival engine. For evolution we reported, against our own interest, that the sophistication signature did not survive religiosity: the party×education interaction on disbelief fell to non-significance once church attendance and biblical literalism and their party interactions entered (§7), because disbelief in evolution is largely structured by religiosity. The immigration analog of that confound is racial resentment (or authoritarianism): each is correlated with both party and political sophistication and is the natural alternative driver of immigration sorting. We therefore subject the knowledge×party interaction on the fact to the same kind of test, controlling for party×racial-resentment and party×authoritarianism (the four-item ANES batteries; racial resentment runs, as expected, far higher among Republicans). Unlike the evolution case, it survives: the interaction falls only from a log-odds of 0.53 to 0.46 with party×racial-resentment, to 0.41 with party×authoritarianism, and to 0.38 with both, remaining significant at $p < .01$ throughout (Table 25). The amplification of immigration-fact sorting by political knowledge is not

racial resentment, or authoritarianism, wearing a sophistication mask — which is the most important thing the contrast with evolution establishes, and the reason we ran the test that evolution failed.

Table 25: The party×knowledge interaction on the immigration fact survives party×racial-resentment and party×authoritarianism (ANES 2024; survey-weighted logit, cluster-robust by PSU) — unlike the evolution sophistication effect, which did not survive its religiosity analog.

Model	party×knowledge (log-odds)	p	n
baseline	0.526	0.000	4402
+ party x racial resentment	0.455	0.000	4385
+ party x authoritarianism	0.405	0.001	4277
+ party x resentment + authoritarianism	0.381	0.001	4271

The divergent-information-environments account is the harder rival: it predicts the same one-sided amplification through *exposure* rather than cognition (§5, §7). We cannot identify against it observationally, but we can ask whether the readiest content-relevant proxies absorb the interaction. Party×trust-in-media — mainstream-news trust is sharply party-stratified, the natural channel for divergent diets — leaves party×knowledge essentially intact (0.53 to 0.49), and party×political-attention does no more. A sharper proxy goes further: from the ANES 2024 media battery — the named television, radio, website, and newspaper sources each respondent reports using — we build party×exposure to a respondent's *own-party-leaning* outlets (Fox and conservative talk radio for Republicans; MSNBC, CNN, NPR for Democrats). It narrows the interaction to 0.44 but leaves it strongly significant ($p < .001$; Table 26). Read as a mediation, pro-attitudinal exposure carries about a quarter of the amplification (mediated share $\approx 25\%$) and leaves roughly three-quarters as a direct, non-exposure effect that remains significant under the natural direct/indirect decomposition. Even the measured slant of what people consume does not absorb the amplification — and the sharpest possible proxy, the immigration-frame tone of each outlet's actual coverage scored by an elite-tone classifier (Card et al., 2022a), is the one refinement left to future data. No observational proxy captures message *content* fully, so the account is narrowed hard, not closed — and the thesis holds either way, since exposure sorted by the elite configuration is capture too (§3). The registered falsifier for the channel question is premium movement under measured content shifts (SI-E).

The formal sensitivity analysis makes the same point on a common scale, and we benchmark it honestly against our *strongest* measured confounder rather than a weak one. With party×racial-resentment and party×media-trust both in the model, the interaction's Cinelli–Hazlett robustness value (Cinelli & Hazlett, 2020) is $RV_{q=1} = 4.9\%$: an unobserved confounder would have to explain 4.9% of the residual variance in *both* the interaction and the fact to reduce the estimate to zero. Modest in the abstract — but modest *relative to what*? The natural benchmark is not the weak media-trust proxy but party×racial-resentment, a substantively powerful, well-measured confounder that itself survives as a real predictor of the fact. Its partial R^2 is about 0.44% (geometric mean across the treatment and outcome sides). To reach the robustness value, an unmeasured confounder would have to be roughly *eleven times* as strongly associated with both partisanship-knowledge and the immigration fact as racial resentment — a strong, real thing — is; benchmarked against the weaker party×media-trust proxy the multiple is about sixty. No confounder we can name approaches even the racial-resentment bar. So the exposure rival is narrowed considerably further than the raw figure suggests — though, absent a measure of consumption content, still not closed.

Table 26: The party×knowledge interaction on the immigration fact survives controlling for content-relevant exposure proxies — party×media-trust and, in the last row, party×exposure to a respondent's own-party-leaning named outlets (ANES 2024; survey-weighted logit, cluster-robust by PSU). It narrows (to 0.44 under the diet measure) but does not close the divergent-diets account.

Model	party×knowledge (log-odds)	p	n
baseline	0.526	0	4402
+ party x trust in news media	0.493	0	4395
+ party x trust in traditional media	0.521	0	4380
+ party x both media-trust measures	0.492	0	4373
+ party x pro-attitudinal outlet exposure	0.445	0	4402

A related quantity at scale: the preference, by education (CES)

The asymmetric amplification is not an artifact of the ANES or of modest samples. In the Cooperative Election Study it replicates on the matched immigration *preference* at roughly thirty-eight thousand respondents per survey, across two years and three differently worded restrictionist items (Table 27). On the cleanest item — support for reducing legal immigration by half, where neither party sits near a floor or a ceiling — education pulls Democrats toward openness about twice as steeply as Republicans (Democrats from 34 to 18 percent restrictionist across the college line, Republicans from 68 to 58), so the partisan gap *widens* with schooling and the party-by-education interaction is +0.38 ($p < 10^{-15}$). The 2024 items repeat the pattern (interactions +0.52 and +0.80), the only caveat being that Republican support for a border wall is near its ceiling, which mechanically limits in-group movement there. Two points follow. First, the comparative static the paper rests on — the partisan gap *widening* with capacity, the out-group moving while the in-group holds — is real and large beyond any sampling doubt (its cognitive reading, capacity recruited to enact the party line, remains the candidate §5 cannot yet separate from exposure). Second, the captured fact, having converged on the preference's partisanship (§6), is now driven by the very same machinery that governs the preference: the assimilation is complete in mechanism as well as in level. The boundary of this check: it uses *education* on a *preference*, not the behavioral knowledge battery on the fact, so it confirms a related quantity rather than the headline one. Its scale, however, lets us take one observational step on the cognition-versus-exposure question of §5, which smaller samples cannot. If the amplification were merely party-correlated *exposure* — the more attentive of each side absorbing more of their own diet — then controlling for political attention should absorb it. It does not. Adding a party-by-political-attention interaction *halves* the party-by-education interaction but leaves it firmly significant (from +0.36 to +0.18 in the same respondents, $p <$ .001): education amplifies the partisan gap beyond what differential attention explains. This does not isolate cognition completely, but it is more than a concession that exposure and capacity are indistinguishable — here, at n ≈ 38,000, they partly separate, and the part that survives is not absorbed by differential attention, the readiest exposure proxy though not the whole of exposure; what remains of the split is the registered channel question (SI-E). The same control run on the *headline* pairing — political knowledge and the fact, in the ANES — tells the same story (§7): the party-by-knowledge interaction falls only from +0.51 to +0.40 ($p = .001$) once party-by-attention is included. The CES contributes scale; the ANES contributes directness, applying the test to the very quantity the mechanism claim is about.

Table 27: The mechanism at scale (CES 2022–2024, n ≈ 38,000 each): the partisan gap on immigration restriction widens with education, driven by the out-group, across three items. Shares are the restrictionist share; the interaction is party×college (log-odds).

Year	Item	Dem (no coll → coll)	Rep (no coll → coll)	party×college	p	n
2022	reduce legal immigration 50%	0.337 → 0.181	0.679 → 0.575	0.384	0.00000	38464
2024	build a border wall	0.272 → 0.115	0.921 → 0.9	0.803	0.00000	38894
2024	increase border patrols	0.717 → 0.624	0.945 → 0.949	0.516	0.00078	38894

Behavior, against expressive responding

Expressive responding (Bullock et al., 2015; Peterson & Iyengar, 2021; Prior et al., 2015) holds that partisans may report beliefs they do not hold to cheer their side. The observational design cannot fully exclude this — the calibration bound below quantifies how much it could explain — but the factual belief carries independent political weight that pure cheerleading on a single survey item would not produce, in line with the precedent that partisan economic assessments carry behavioral consequence (Gerber & Huber, 2010). In ANES 2024 it predicts consequential behavior, 2024 presidential vote and border-wall support, *net of party identification and political knowledge*:

Table 28: The factual belief predicts behavior net of party and knowledge (ANES 2024; odds ratios from survey-weighted logits).

Outcome	Fact OR	p	n
Trump vote (vs Harris)	7.56	0	3630
Favor border wall	5.52	0	3511

The sharpest form of the objection is that "net of party" under-models identity when party is only a three-category measure. The result does not depend on that coarseness. Re-estimating the fact's effect on presidential vote on a common sample while progressively adding seven-point party identification, then seven-point ideology, then knowledge, the odds ratio falls from roughly nine but stabilizes just above six and stays overwhelmingly significant (Table 29): the belief predicts the vote at an odds ratio of about 6.1 net of a full identity battery. A report that tracked nothing but partisan identity should be largely absorbed by these controls; it is not.

Table 29: The factual belief predicts presidential vote net of progressively richer identity controls (ANES 2024; odds ratio on the fact, survey-weighted logit).

Identity controls	Fact OR	p	n
Party (3-cat)	8.71	0	3327
+ 7-pt party ID	6.87	0	3327
+ 7-pt ID + ideology	6.38	0	3327
+ 7-pt ID + ideology + knowledge	6.10	0	3327

A belief that predicts the vote net of a rich identity battery shows that it travels with consequential partisan commitments — but, since a vote report is itself a self-report open to the same expressive pressures as the belief report, this does not by itself establish sincerity. What settles how much expressive responding *could* explain is the calibration bound, next.

The subnational elite clock: models and forks

The main text reports the headline (+0.66; division-clustered $p = .067$; randomization $p = .016$). The design: for each census division and GSS fielding year, the delegation's Republican-minus-Democratic difference in net restrictionist tone across all immigration speeches in the trailing decade (Card et al. corpus), paired with the division's mass partisan gap on the two GSS immigration-economy items in item SD units — 63 division-year-item cells, nine divisions, 1996–2022 — with division, calendar-year, and item fixed effects, so the national clock is absorbed by construction. The estimate is sign-stable across both items separately, a minimum-cell-size fork, and a fork ending the panel in 2014; the 2022 elite window truncates at the corpus's 2020 end and division-level cue environments are permeable to national media, both biasing *against* local structure, since misattributed national exposure loads onto the absorbed year effects.

Table 30: The subnational elite clock. Mass partisan gap on the GSS immigration-economy items (division-year-item cells, item-SD units) regressed on the division's congressional-delegation tone divergence in the trailing decade (Card et al. corpus), with division, calendar-year, and item fixed effects; WLS by cell n. Randomization inference permutes the elite series across divisions within year (5,000 draws).

quantity	value
dGap/dD_rt (region+year+item FE, WLS)	0.6572
cluster SE (9 divisions)	0.3507
cluster p	0.0669
RI p (elite series permuted within year)	0.0156
cells	63
divisions	9
years	4
fork: immjobs only	0.671 (p = 0.133, cells = 36)
fork: immameco only	1.114 (p = 0.115, cells = 27)
fork: cells n >= 40	0.667 (p = 0.064, cells = 61)
fork: pre-2020 elite windows only (1996-2014)	1.006 (p = 0.077, cells = 54)

The verifiability margin across claims

The scope condition — capture operates on claims citizens cannot verify — is certified by the CVI (SI-A) and admits a first cross-claim estimate. Every truth-apt claim the paper analyzes that was fielded to a mass sample with party identification — ten fixed-wording claims across the GSS, ANES, and two Pew panels; forty-four claim-years, 1993–2024 — receives its CVI score and an elite-contestation indicator dated from elite behavior alone (platform onsets as in the event study; the SI-D.2b configuration codings for the science claims). The outcome is the partisan gap in item

standard deviations. Under the gap equation the two legs separate: $g(T)$ predicts a contestation margin; $\alpha(v)$ predicts a CVI slope *within* the contested regime.

The contestation leg is unambiguous: the weighted mean gap is 0.13 SD where no elite party owns a pole and 0.63 SD where one does — a five-fold margin across ten claims, two houses, three decades (WLS +0.81, claim-clustered $p = .001$; randomization inference reassigning entire contestation *schedules* across claims, $p = .016$; Table 31). What separates the claims citizens sort on from the claims they do not is not topic, scale, or house but whether elites supply an owner for one pole.

The verifiability leg is not identified within the fielded range, and we report that plainly: CVI × contested = −0.04 (randomization $p = .38$), the sign flipping (+0.09, n.s.) under the fielded-wording recoding. The reason is design, not theory: every claim fielded to the public scores 6 or higher on the 0–12 index — survey houses field the questions publics argue about, and publics argue about claims they cannot check — so the low-CVI anchors that would identify $\alpha(v)$ have no public twins. The margin estimates one leg of the gap equation and brackets the other; the missing cell is registered (E.5): field matched verifiable and unverifiable claims under contestation, and the model stakes itself on the gap opening only where checking fails. (Extension-item CVI codings are single-coder pending the second-coder pass; all forks are in the replication files.)

Table 31: The claim-level margin. Standardized partisan gap regressed on elite contestation and the Claim Verifiability Index across 44 claim-years (10 fixed-wording claims, GSS/ANES/Pew, 1993–2024). WLS weighted by analytic n; claim-clustered SEs; randomization inference permutes assignments across claims (5,000 draws).

specification	term	estimate	se	p
main (repo-matched CVI)	CVI	0.017	0.013	0.211
main (repo-matched CVI)	contested	0.806	0.233	0.001
main (repo-matched CVI)	CVI:contested	-0.037	0.037	0.326
fielded-wording CVI	CVI:contested	0.087	0.135	0.522
randomization inference	CVI:contested	-0.037	NA	0.378
randomization inference	contested	0.501	NA	0.016

The calibration bound: how much could expressive responding explain?

Rather than defer sincerity to a future experiment, we bound the expressive share against the published experimental record. The incentive experiments agree on the magnitude: monetary accuracy incentives cut partisan factual gaps roughly in half — about 55 percent in Bullock et al. (2015), with the same range on economic retrospections (Prior et al., 2015) and on political information (Peterson & Iyengar, 2021) — rising to roughly 80 percent in Bullock et al. (2015)'s upper bound only when a *paid "don't know"* is added — a design that works partly by moving low-confidence respondents to "don't know" rather than by revealing insincerity, which is Graham (2023)'s point that much apparent misperception is guessing, not held falsehood. List experiments, which remove the expressive incentive without paying for accuracy, find limited expressive responding even on partisan rumors (Berinsky, 2018). Table 32 applies each published share, taken at face value, to the 33-point 2024 gap on the fact:

Table 32: The expressive-responding calibration bound. Each row applies a published expressive share, at face value, to the 33-point 2024 partisan gap on the immigration fact. The sincere residual is compared with the 1996 point estimate (≈3 points, itself statistically indistinguishable from zero).

Scenario	Expressive share	Sincere residual (pts)	× the 1996 estimate
Accuracy incentives (Prior et al. 2015; Peterson & Iyengar 2021)	50%	16.5	5.5
Accuracy incentives (Bullock et al. 2015)	55%	14.8	4.9
Accuracy + paid don't-know (Bullock et al. 2015, upper bound)	80%	6.6	2.2

Applying published incentive-experiment shares to this item assumes transportability across items and settings — the shares were estimated on economic retrospections and political-information items, not immigration facts — and the identified-set construction is designed to absorb exactly that uncertainty, since it holds at every share up to a ceiling above anything the record reports for any item, under adversarial year-varying paths. The bound formalizes as partial identification. Let $\lambda_t \in [0, 0.8]$ be the expressive share of the observed gap in year t — the ceiling being the most hostile value the record supports (the paid-don't-know upper bound). The sincere gap is then set-identified, $g_t^* \in [(1 - 0.8)\,\underline{g}_t,\ \overline{g}_t]$, composing the worst-case λ with the sampling interval, and Figure 11 plots the identified band around the whole 1996–2024 trajectory rather than the 2024 endpoint alone. The exercise's teeth: the *floor* of the 2024 band (5.6 points — $\lambda = .8$ against the CI lower bound) exceeds the *ceiling* of the 1996 band (3.4 points — $\lambda = 0$ against the CI upper bound). The bands are disjoint, so the switch-on is identified under every expressive-share path the literature supports, including adversarial paths in which λ_t varies freely across years — and a λ_t that grows from nothing on exactly the elite clock is capture in expressive clothing, not an alternative to it. The knowledge gradient survives every knowledge-homogeneous λ (uniform scaling preserves order); flattening it adversarially requires the top tercile's expressive share to exceed the bottom's by at least 26 points with the bottom at zero, a differential no incentive experiment reports. And the value-cleavage benchmark is untouched by any uniform λ (all gaps shrink alike), while a fact-*only* λ of one-half leaves a 16.5-point residual — still larger than the fact's entire pre-2014 trajectory, though below the uncorrected value benchmarks, which we state rather than argue around.

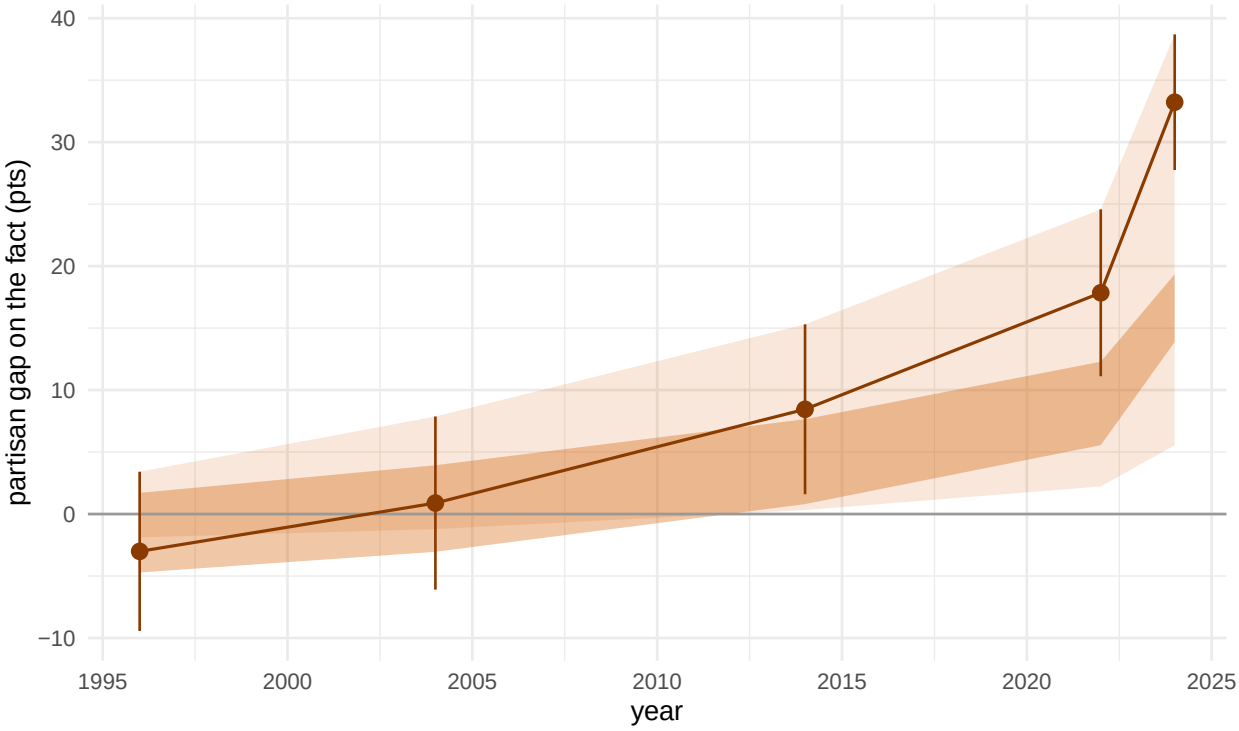


Figure 11: Partial identification of the sincere partisan gap under expressive responding. Points: observed GSS gap with 95% CIs. Dark band: sincere gap identified set at $\lambda = 0.5$ (the accuracy-incentive consensus). Light band: full identified set with λ free in $[0, 0.8]$ (paid-don't-know ceiling), composed with sampling error. The 2024 floor (5.6 pts) exceeds the 1996 ceiling (3.4 pts): the switch-on survives every expressive share the published record supports.

Three things follow. First, the *level*: at the top of the accuracy-incentive range the sincere residual is 15–17 points — roughly five times the 1996 point estimate, larger than the entire observed gap at any point before 2014, and, since a uniform expressive correction shrinks every gap alike, still the largest fact gap in the survey. Even the paid-don't-know extreme leaves ~7 points, more than double the 1996 estimate. Second, the *trend*: expressive responding is an account of levels, not of a twenty-eight-year trajectory; for cheerleading to produce the trend, the expressive share itself would have to grow from nothing on exactly the elite clock the event study dates — at which point "expression tracking elite configuration" is capture under another name. Third, the *structure*: expressive accounts predict flat-or-reversed knowledge gradients (cheering requires no sophistication) and momentary responding; we observe a monotone knowledge gradient (34→42→46 points across terciles) and durable nine-year within-person revision, and the one segment that *is* knowledge-flat — the 2024 Republican rebound — is precisely the segment already quarantined as cue-shaped rather than capacity-driven (SI-C). The expressive share consistent with the published experimental record cannot account for the level, the trend, the gradient, or the within-person durability. The reading stays falsifiable: a rebound that became knowledge-graded, or a gap that persisted after elite cue withdrawal, would disconfirm sincere capture.

The within-person panel (VOTER): cross-lagged direction and change-score

The within-person capture result (Section 6.5) rests on the Views of the Electorate Research (VOTER) panel, which re-interviewed the same national sample across 2011–2020; the belief item asks whether immigrants "mostly make a contribution to American society" or are "mostly a drain," coded so that higher values are the anti-immigrant (folk) direction, and party is the seven-point identification at each wave. Both variables are standardized within the analysis sample.

The headline estimate is the random-intercept cross-lagged panel model (RI-CLPM) of Hamaker et al. (2015), fit by maximum likelihood on the complete four-wave sample (2011, 2016, 2018, 2020; n = 2,008 full-participation respondents; RMSEA = 0.067). A pair of random intercepts absorbs each respondent's stable trait level in party and in belief, so the cross-lagged paths describe *within-person* deviations rather than between-person differences — the correction the two-wave CLPM cannot make. Cross-lags are constrained equal across waves (a conservative simplification given the unequal 5/2/2-year spacing). The within-person party→belief path (0.40) exceeds the belief→party path (0.12), and a nonparametric bootstrap of the difference (300 resamples) places it at 0.28 with a 95% interval of [0.19, 0.38] that excludes zero in every resample (Table 33). Two simpler estimators agree: the ordinary two-wave standardized CLPM (2011→2020) gives party→belief 0.33 versus belief→party 0.20, and a change-score regression of the 2011→2020 belief change on 2011 party — which differences out each respondent's baseline — gives $\beta = 0.12$ ($p < .001$). That the asymmetry is *larger*, not smaller, once trait variance is partialled out is the reverse of what an artifact-of-stable-dispositions reading predicts.

Table 33: Random-intercept cross-lagged panel model on the VOTER panel (four waves 2011–2020, standardized; RI-CLPM separates within-person dynamics from stable between-person traits). The within-person path from party to the later belief is more than triple the reverse path, and the difference is significant by bootstrap.

Parameter	Estimate	SE	p	Note
within: party_t -> belief_t+1 (CL_p2b)	0.403	0.039	<.001	cross-lag, constrained equal across waves
within: belief_t -> party_t+1 (CL_b2p)	0.120	0.016	<.001	cross-lag, constrained equal across waves
within: belief stability (AR_b)	0.231	0.025	<.001	autoregressive
within: party stability (AR_p)	0.828	0.024	<.001	autoregressive
path difference (party-leads minus belief-leads)	0.283	NA	<.001	bootstrap 95% CI [0.189, 0.379]; 300/300 resamples > 0
N (complete four-wave respondents)	2008.000	NA	NA	full-participation panel; RMSEA=0.067

The complementary descriptive trajectory (Figure 24) and the sorting summary — party–belief alignment rising from 74% to 83%, with two-thirds of the 2011-misaligned converging by 2020 — together with the two-wave CLPM and change-score robustness are written by `R/17_voter_panel.R` to `voter_within_person_gap.csv`, `voter_cross_lagged.csv`, `voter_riclpm.csv`, and `voter_sorting.csv`. The RI-CLPM is reproducible in R via `lavaan` (model syntax in `R/17_voter_panel.R`); the committed `voter_riclpm.csv` records the estimates this manuscript renders.

Reliability sensitivity (the direction's conditional identification). Because party is measured more reliably than the single belief item, we disattenuate the two-wave cross-lags across a grid of assumed party and belief reliabilities and ask where the party→belief lead reverses (Table 34; `R/29_reliability_grid.R`). The lead holds across the whole plausible range and reverses *only* in the corner where party reliability $\gtrsim 0.90$ *and* belief reliability $\lesssim 0.70$ — at the belief's noise floor of 0.64 (its full nine-year stability). Belief reliability that low requires the underlying belief to have been near-stable across the nine years (true-score autocorrelation $\gtrsim 0.89$), which the panel refutes (alignment 74→83%, two-thirds of out-of-step respondents revised). Under any belief reliability consistent with a belief that *moved* ($\gtrsim 0.75$), the lead survives even at the top of the range estimated for party identification. The direction is therefore identified against the reliability artifact conditional on the documented belief change.

Three estimator families, and what each can identify. We say plainly what the strictest estimator finds, because a reader of the reproducibility ledger will find it anyway. An ML-SEM dynamic panel with individual fixed effects, lagged outcomes, and predetermined regressors (Allison et al., 2017) — the treatment that protects against time-invariant confounding under the *least* charitable assumptions — attenuates **both** cross-lags to zero on this item (party→belief −0.02, belief→party +0.00, both n.s.; `voter_mlsem_dpm.csv`, `R/59`), with fit statistics that reject the model outright (CFI 0.38, RMSEA 0.50). This is exactly what the reliability grid above predicts: with fixed effects and lagged outcomes absorbing all stable variance, the only identifying variation left is within-person *transient* change in a three-category item across four unevenly spaced waves — variance the grid shows is noise-dominated — so the estimator has nothing to work with and attenuates everything, the Nickell-type logic of dynamic panels at T = 4 compounded by single-item measurement error. We therefore read the three families as a bracket, not a horse race: the RI-CLPM and the latent CLPM with estimated reliability, which retain usable variance, put party→belief well above belief→party (0.40 vs 0.12; 0.55 vs 0.08); the strict-FE treatment, which retains almost none, finds neither path — and, in particular, supplies *no* support for the reverse path under the assumptions most hostile to our reading. What no estimator in any family produces, anywhere, is a belief→party lead. The bracket's honest summary: the direction is identified against the reliability artifact and against random-intercept confounding, is uninformative under strict fixed effects for want of usable variance, and is disconfirmable by the verbatim-fact panel (below), not by re-estimation of this one. One member remains to be added on the pipeline run: the equal-cross-lag constraint across the panel's unequal 5/2/2-year gaps — confessed above as a conservative simplification — is removed entirely by continuous-time SEM (Driver et al., 2017), which estimates the drift matrix of the underlying process and makes unequal spacing identifying variation rather than a violated assumption; the run-ready script (`R/79`) pre-states the reading rule (a belief→party drift exceeding party→belief would count against §6.5) and pre-registers the same model for the verbatim panel.

Table 34: Two-dimensional reliability sensitivity bound (VOTER). Whether the party→belief lead survives, by assumed party reliability (columns) and belief reliability (rows). It reverses only at party ≳0.90 and belief ≲0.70 — the corner that assumes the belief was near-stable over nine years, which the documented within-person change refutes.

rho_belief	party rel 0.8	party rel 0.85	party rel 0.9	party rel 0.95
0.64	lead	lead	reverses	reverses
0.70	lead	lead	lead	reverses
0.75	lead	lead	lead	lead
0.80	lead	lead	lead	lead

Table 34: Two-dimensional reliability sensitivity bound (VOTER). Whether the party→belief lead survives, by assumed party reliability (columns) and belief reliability (rows). It reverses only at party $\gtrsim$0.90 and belief $\lesssim$0.70 — the corner that assumes the belief was near-stable over nine years, which the documented within-person change refutes.

rho_belief	party rel 0.8	party rel 0.85	party rel 0.9	party rel 0.95
0.85	lead	lead	lead	lead

SI-C: Supplementary results

This appendix collects secondary results the main text summarizes at the point each bears on the argument: the composition robustness of the sorting, the benchmark of the fact against canonical value cleavages, the one-sided sophistication asymmetry with its multiverse and out-of-sample replication, and the descriptive switch-on signature (structural separability and cross-issue timing).

Sorting, not composition

A natural alternative is composition: perhaps the parties' belief gap widened because their make-up changed, not because anyone's belief sorted. It did not (Table 35). In a pooled model interacting both party and education with era, the party×era term is large and significant (odds ratio 3.38) while the education×era term is null (1.14, $p = .46$). The across-era divergence is identity sorting, not the diploma divide — and, directly against a composition account, it is non-college Democrats who moved most.

Education is not the composition that matters most for immigration: the parties also diverged on race, ethnicity, and nativity. Re-estimating the party×era interaction with race/ethnicity/nativity/age/region×era controls leaves the coefficient undisturbed (if anything larger), though the subsampled items cut the analytic n by a third and the estimate to marginal significance (Table 60); a better-powered cross-sectional check on the matched preference in the CES (n $\approx$ 37,000) confirms the party gap survives ideology and the full demographic battery (+1.64 $\rightarrow$ +0.91 log-odds, $p < 10^{-150}$; Table 61). Compositional change does not visibly account for the sorting — though over-time ideological re-sorting is a residual the cross-sections cannot close, met only partially by the panel (§6). Full results are in Supplementary Materials SI-D.5.

Table 35: Pooled interaction model: party×era survives; education×era is null.

term	estimate	std.error	statistic	p.value	conf.low	conf.high
rep	1.12	0.08	1.41	0.16	0.96	1.32
post	0.23	0.13	-10.98	0.00	0.17	0.30
college	0.45	0.09	-8.56	0.00	0.37	0.54
rep:post	3.38	0.16	7.58	0.00	2.47	4.63
post:college	1.14	0.17	0.74	0.46	0.81	1.60

A fact as sorted as a value

How party-sorted is that, in absolute terms? The natural yardstick is the set of issues we already treat as *defining* the partisan divide. In 2024 the Republican–Democrat gap on the factual belief that immigrants take jobs (33 points) is as large as the gaps on the canonical moral and cultural cleavages — gun permits (30), gay rights (33), the death penalty (35) — and approaches abortion (45) and the matched immigration-policy preference (46) (Figure 12). A truth-apt empirical claim now divides the parties about as sharply as the value questions that organize American politics. That is the substantive meaning of capture: not merely that the gap grew, but that it grew into the company of the deepest cleavages — while remaining, unlike them, a question with a right answer, in the sense that the professional consensus on its aggregate version is settled (the National Academies' review finds little to no long-run effect of immigration on native employment or wages; National Academies of Sciences, Engineering, and Medicine (2017)). We invoke that benchmark only to fix the *direction* of the divergence, not a precise distance, and we lean on it most lightly for immigration — where, unlike trade and AI, no direct expert poll exists — precisely because the public–expert gap there is anchored on review consensus rather than a point estimate.

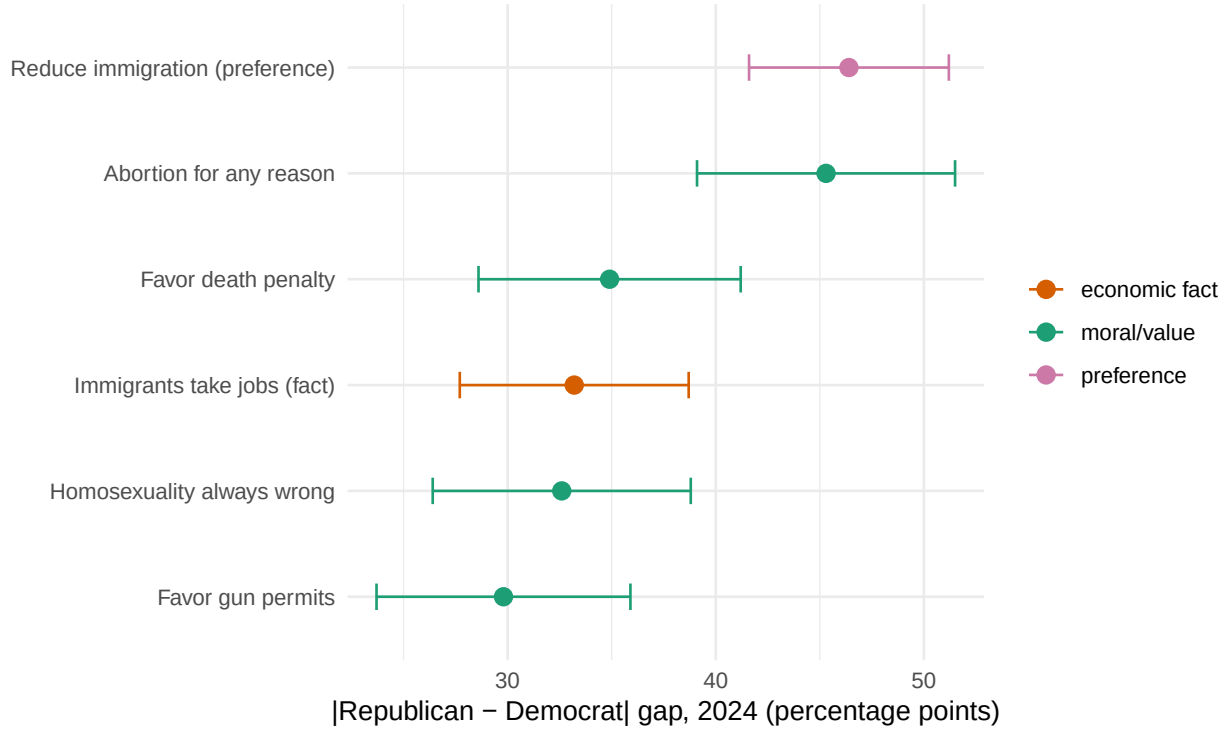


Figure 12: The 2024 partisan gap on the immigration fact in context: as large as canonical value cleavages, approaching the matched policy preference (GSS 2024; |Rep − Dem|, 95% CI).

The amplification is borne by the out-group

The amplification is not symmetric across parties, and the asymmetry is itself informative. Decomposing the knowledge effect within each party (Table 36), the slope of the folk belief on political

knowledge is sharply negative among Democrats (log-odds −0.49, $p < .001$) — the most knowledgeable Democrats hold the belief at seven percent against seventeen among the least — while among Republicans it is flat and insignificant (+0.04, $p = .36$). The two slopes differ decisively: a direct test of the difference (the party×knowledge interaction) returns +0.53, $p < .001$, so the asymmetry is a tested quantity, not an eyeballed contrast. This bears on two rivals at once. Classic identity-protective cognition is *symmetric* — it predicts sophistication sharpens both poles (Kahan et al., 2012; Kahan, 2013) — and a naïve account in which knowledge tracks the truth predicts both sides move toward the expert-consistent pole; the one-sidedness contradicts each. It is, instead, what the present model predicts: sophistication amplifies enactment of the identity prior $p(s)$, and for Republicans the party line coincides with the folk default, so capacity has nothing to add, while for Democrats the line departs from it, so knowledge carries them toward the experts.

What the one-sidedness does *not* establish: it rules out symmetric cheerleading and naïve truth-tracking, but not *divergent information environments*, in which the expert signal sits in the Democratic media diet (Jerit & Barabas, 2012). Three controls narrow the exposure channel without closing it — the interaction survives party×political-attention (+0.51 → +0.40), party×media-trust (+0.53 → +0.49), and party×exposure to own-party-leaning named outlets (+0.53 → +0.44; Table 26) — but even the diet measure indexes *which* outlets a partisan consumes, not the *content* of the messages, so we read "knowledge enacts the party line" as a description of the pattern, not an identified cognitive claim (§5).

The aggregate counterpart is the capstone's one-sided movement (§6). General *education*, by contrast, erodes the folk error on *both* sides (consistent with P2) — which is the point: it is *political knowledge*, the capacity to apply the party line, that moves the out-group alone. The same asymmetry governs the matched immigration preference in the Cooperative Election Study at ≈ 38,000 respondents per survey (party×education +0.38 to +0.80; SI-B), so the machinery is not a small-sample artifact.

Table 36: Within-party decomposition: political knowledge moves Democrats toward the expert-consistent pole but leaves Republicans flat (ANES 2024; survey-weighted logit of the folk belief on standardized knowledge, within party).

Party	Belief, low knowledge	Belief, high knowledge	Knowledge (log-odds)	p	n
Democrats	0.172	0.071	-0.489	0.00	2323
Republicans	0.508	0.533	0.037	0.36	2079

The result does not hinge on the knowledge battery: it recurs in the GSS with *education*, but through a different channel (Table 58; Supplementary Materials SI-D.4). Education corrects the folk error in *both* parties (consistent with P2), about twice as steeply for Democrats, so the gap widens only modestly; political knowledge moves the out-group alone, widening it from 34 to 46 points. The two differ exactly as the two-channel model predicts — education acts on the shared folk prior $f(s)$, political knowledge on the identity term $p(s)$ — which is why the knowledge interaction, not the education one, is the decisive test.

Robustness: multiverse and heterogeneity

Across 24 specifications the party×knowledge interaction is positive in 100% and significant (p<.05) in 100% (coefficients 0.51–0.92).

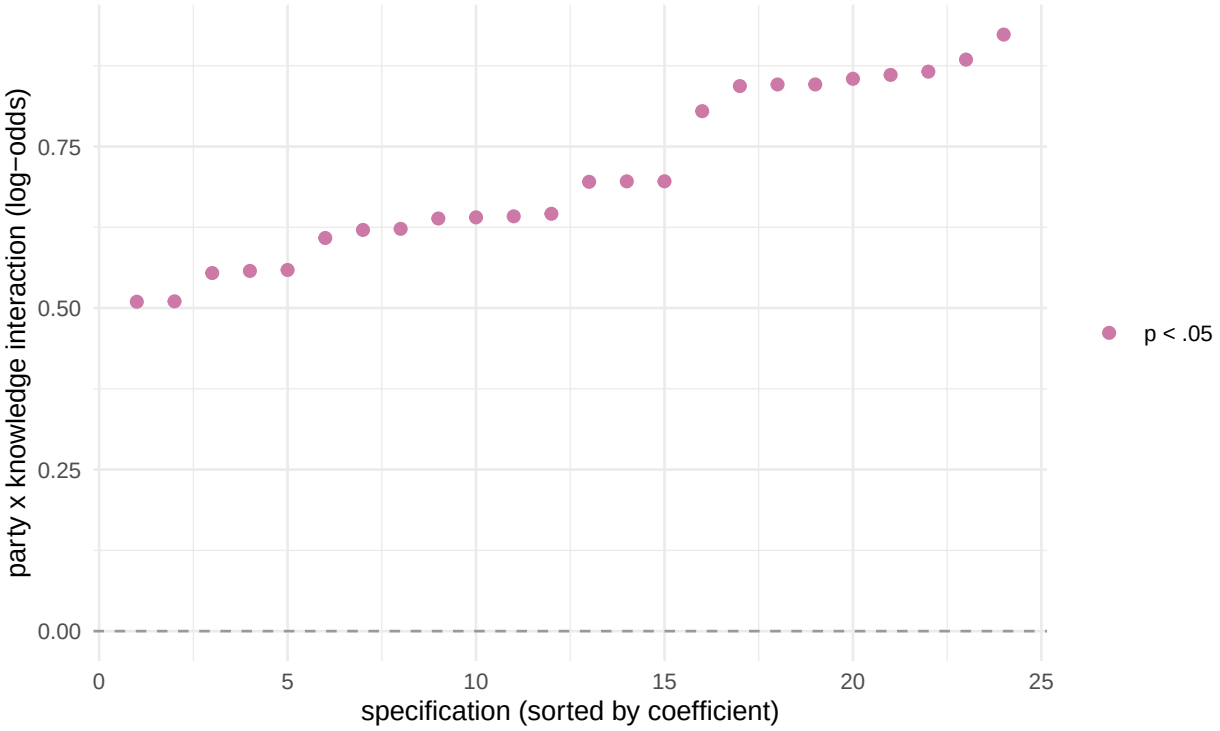


Figure 13: Specification curve: the party×knowledge interaction across 24 specifications (sorted; purple = p < .05).

Table 37: Heterogeneity: Rep–Dem gap on the fact by knowledge level, within partisan strength (ANES 2024).

Partisan strength	Low-knowledge gap	High-knowledge gap
strong	0.36	0.63
weak/lean	0.22	0.33

Table 38: Heterogeneity: party×knowledge interaction by age cohort (ANES 2024).

Age cohort	Interaction (log-odds)	n
<40	0.429	1085
40-64	0.563	1826
65+	0.710	1497

The premium, functional-form-free: honest-forest heterogeneity

The headline premium is a parametric interaction, so it inherits the interaction's functional form. The current frontier for heterogeneity drops that inheritance: estimate the conditional partisan–belief association $\tau(x)$ nonparametrically with honest sample-splitting (Athey et al., 2019), then summarize it two ways no interaction coefficient can (Table 39; `R/78`; cross-fitted AIPW scores, five folds, seed fixed). The framing is stated before the numbers: party is not randomized, so the estimand is the design-honest heterogeneous *association* — exactly the logit's epistemic status, with the doubly-robust machinery buying functional-form freedom, not causality. Under the paper's full control battery, the best linear projection of $\tau(x)$ on knowledge (Semenova & Chernozhukov, 2021) returns the premium at **+0.052 per SD** ($p = .042$) — the logit's own average-marginal scale, recovered by machinery that assumed no interaction — and the multivariable projection retains knowledge (+0.051) while assigning education **zero** (−0.015, $p = .61$): the two-channel split, reproduced with no channel built in. The rank-weighted average treatment effect (Yadlowsky et al., 2025) answers a question the interaction cannot pose: knowledge *ranks* the machine-learned heterogeneity (AUTOC +0.044, $p = .050$) and education does not (−0.013). Two honesty notes carry the exhibit. Overlap: with ideology and resentment in x, party is near-deterministic for half the sample (54.6 percent of propensities clipped) — which is why this specification mirrors the logit-with-controls estimand — and the overlap-clean specification (dropping both) recovers the raw-scale association (+0.341) with the knowledge projection attenuated on the DR-score scale (+0.027, n.s.; +0.048 multivariable), forest smoothing damping a four-item battery's gradient absent the partial-out. The design-based interaction therefore remains the headline; what the forest contributes is triangulation of *form* — in both specifications, knowledge carries the projection and education never does.

Table 39: The premium under honest-forest heterogeneity estimation (cross-fitted AIPW; canonical grf version in R/78): best linear projections of (x), rank-weighted AUTOC, and DR-score gaps by knowledge tercile, under the full-battery and overlap-clean covariate sets — with clipping and attenuation stated, not smoothed.

Quantity	Value	Note
specification A — full covariate set (the paper's control battery: knowledge, education, ideology, resentment, age, media trust)	NA	NA
average partisan association (AIPW)	+0.221 (se 0.021)	residual association net of the battery; propensity clipping 54.6% — party is near-deterministic in ideology+resentment, stated, and why the estimand mirrors the logit-with-controls
BLP of ˆ(x) on knowledge	+0.052 (se 0.026, p = .042)	the premium, functional-form-free, at the logit's own AME scale (+0.05/SD)

Table 39: The premium under honest-forest heterogeneity estimation (cross-fitted AIPW; canonical grf version in R/78): best linear projections of (x), rank-weighted AUTOC, and DR-score gaps by knowledge tercile, under the full-battery and overlap-clean covariate sets — with clipping and attenuation stated, not smoothed.

Quantity	Value	Note
multivariable BLP: knowledge / education	+0.051 (p = .07) / −0.015 (p = .61)	knowledge carries the projection; education carries none — the two-channel split, machinery-free
RATE/AUTOC, knowledge ranking	+0.044 (boot se 0.023, p = .050)	knowledge ranks the machine-learned heterogeneity; education AUTOC −0.013
DR-score gap by knowledge tercile	0.148 / 0.263 / 0.287	covariate-adjusted scale (nets the battery), gradient preserved
specification B — overlap-clean covariate set (drops ideology and resentment from x)	NA	NA
average partisan association (AIPW)	+0.341 (se 0.022)	clipping 5.4%; matches the raw-scale gap
BLP knowledge (single / multivariable)	+0.027 (p = .38) / +0.048 (p = .13)	attenuated on the DR-score scale: without the ideology/resentment partial-out, forest smoothing damps the discrete battery's gradient — the design-based interaction (+0.51 log-odds; second difference +12.6 [5.5, 19.6]) remains the headline; this exhibit triangulates form, not magnitude
education BLP / AUTOC	−0.051 / −0.046	education never carries the projection, in either specification

Out-of-sample replication

The multiverse establishes that the 2024 interaction is robust *within* its survey; it cannot establish that the result is not peculiar to that survey or to the four-item battery. The ANES Time Series Cumulative File lets us check both at once. It carries the identical "take jobs" fact, harmonized across years (`VCF9223`), together with a *different* sophistication measure — the interviewer's rating of the respondent's general level of political information — fielded in the earlier, fully in-person waves. Re-estimating the party×sophistication interaction on the fact thus replicates the result out

of sample in two senses simultaneously: a different operationalization of sophistication, in different survey-years. It reproduces cleanly (Table 40): the interaction is +0.39 in 2008, +0.40 in 2012, and +0.52 in 2016 — the last essentially the 2024 magnitude — and +0.41 pooled ($p \approx 10^{-10}$). It is, tellingly, *absent* in 2004 (+0.05, n.s.): before immigration had hardened along party lines, sophistication did not amplify the gap, exactly the conditional law the theory predicts (P2/P3) — amplification appears only once the issue becomes contested. The 2004 null is held to equivalence discipline rather than read off a p-value: the year alone cannot exclude premiums up to about +0.30 ($\text{MDE}_{80} = 0.42$), but it *is* statistically smaller than every post-onset yearly estimate (TOST against the smallest, +0.39: $p = .012$), and the pooled post-onset premium exceeds it by +0.36 (one-sided $p = .019$) — the switch-on is a tested difference, not a contrast of significance stars (SI-B, Table 21). The headline mechanism is therefore neither a one-survey nor a one-instrument artifact; it recurs across four surveys and two distinct sophistication measures, and switches off in the pre-elite contestation year as the model requires.

Table 40: Out-of-sample replication (ANES Cumulative File): party×political-information interaction on the immigration 'take jobs' fact, using the interviewer's information rating in place of the behavioral battery. Present once the issue is contested (2008–2016), absent before (2004).

Year	Interaction (log-odds)	p	n
2004	0.054	0.69	921
2008	0.390	0.00	1803
2012	0.398	0.00	1662
2016	0.524	0.00	874
pooled 2008-2016 (yr FE)	0.409	0.00	4339

The structural signature: the sophistication premium is itself a product of elite contestation

The model's gap equation, $\text{Gap}(s) = (1-\alpha)\, g(T)\, (p_R(s) - p_D(s))$, is *multiplicatively separable*, and that separability makes a prediction sharper than anything tested so far. Two of its quantities are observable over time on the same fact: the mass partisan gap — the $g(T)$ term — and the *sophistication premium*, the party×sophistication interaction, which is $\partial\,\text{Gap}/\partial s$ and which the model makes proportional to $g(T)$ as well. Ordinary issue-evolution predicts the gap should grow as elites polarize an issue. The two-regime model predicts something issue-evolution does not: the sophistication premium should grow *with it*, from zero before contestation — the cross-derivative $\partial^2\text{Gap}/\partial s\,\partial T > 0$. The premium is not a standing trait of the sophisticated; it should switch on as the issue becomes contested.

It does. Taking the immigration "take jobs" fact across the years the elite-contestation series covers (ANES Cumulative, 2004–2016) and measuring sophistication by the interviewer's political-information rating, both quantities move together with elite contestation (Figure 14). In 2004, before immigration was contested, both are essentially zero — the mass gap is −0.04 and the sophistication premium is +0.05 (n.s.). As elite contestation rises across 2008–2016, both rise in step: the mass gap to +0.23 and the premium to +0.52, tracking the elite series at correlations of +0.86 and +0.91, and each other at +0.85. We are careful about which part of this is discriminating. That the premium and the gap co-scale is *not* itself a test of the model: it is the algebraic consequence of

the multiplicative form $\text{Gap}(s) = (1-\alpha)g(T)\Delta p(s)$ we assume, in which both are proportional to $g(T)$ by construction. The discriminating fact is the *switch-on from zero*: the premium is essentially absent before the issue is contested (the clean 2004 null, +0.05 n.s.) and *emerges* with elite contestation. A generic "sophisticates polarize more on partisan issues" account (Kahan, 2013) expects the premium wherever partisanship is present; it does not expect it to switch on as an issue becomes contested from a pre-cleavage baseline. It is that switch-on, anchored by the 2004 null, that the generic account does not predict — and it rests on four survey-years, which is why we hold it as a descriptive pattern, not a powered estimate. The timing *within* the window — the premium running ahead of the mass gap, near its ceiling while the gap is still climbing — is the reception-integrated $g(T, s)$ prediction (§3): the aware receive the elite signal first, and diffusion down the sophistication distribution completes later.

One further detail confirms the *channel* structure: the same premium measured by general *education* does not switch on — it is already large in 1996, before the cleavage — because education acts on the folk prior $f(s)$, corrigible in both regimes, while political information acts on the identity term $p(s)$, which enters only through $g(T)$. We hold the result to a modest standard (four survey-years, anchored by the 2004 null), and are clear about what carries what across the three results that follow: *this* immigration result is the cleanest single test of the switch-on but is thin and collinear with time; the next subsection strengthens it against that secular-trend confound by cross-issue differential timing; and the economic-perception result after it tests *content-neutrality*, not the switch-on. None stands in for another.

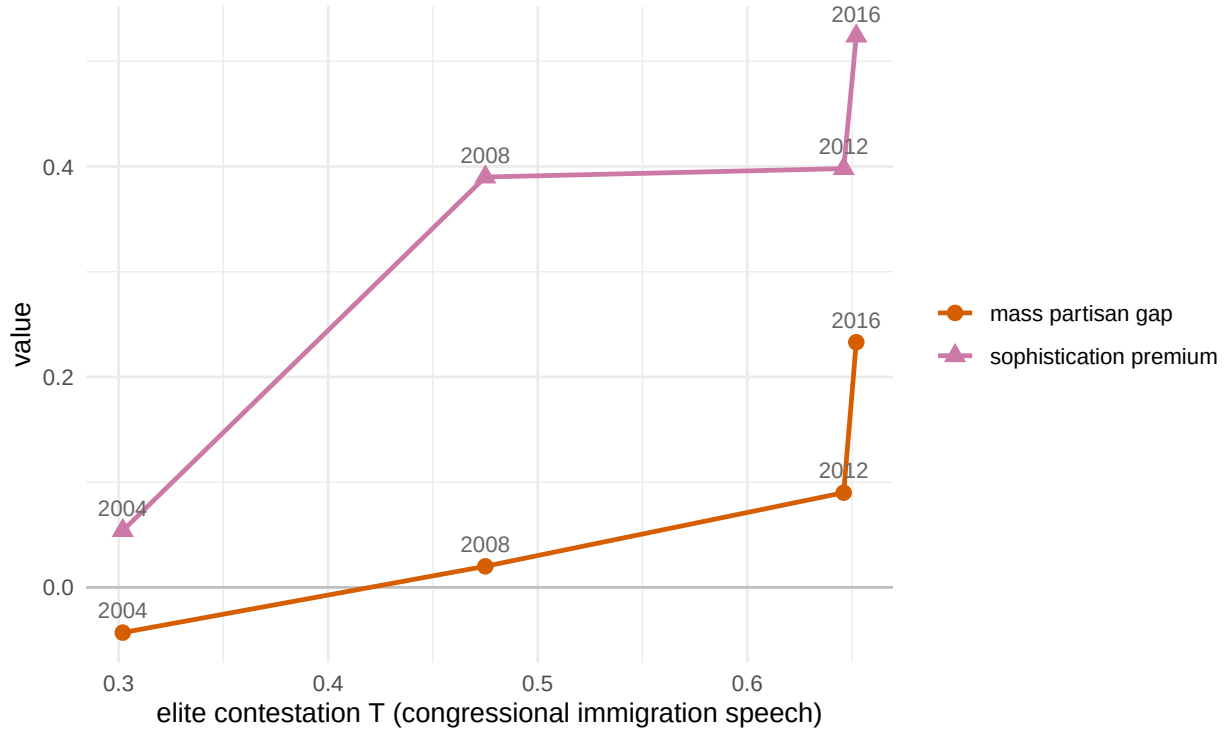


Figure 14: The multiplicative structure, made visible. On the immigration fact (ANES Cumulative 2004–2016), the mass partisan gap and the sophistication premium (party×political-information interaction) both rise from ~zero in step with elite contestation — the $g(T)$ term and its sophistication-derivative scaling together, as $\text{Gap} = (1-\alpha)\, g(T)\, \Delta p(s)$ predicts.

The gap equation, estimated: the structural bound

"Organizing lens, not a validated structural object" (§8) is priced rather than pleaded: we estimate the strict structural forms and tabulate where they break (Table 41; `R/75`), so the bound is a result — and a reviewer who attempts the fit finds it already run. On perception, the strictest reading of "same equation, poles rotating" is a scalar rotation, $\text{premium}_t = c \cdot \text{incumbency}_t$: weighted least squares over the thirteen per-year premiums gives $c = 0.046$, the rotation itself is exact (thirteen of thirteen signs), and the over-identification test *rejects* constancy decisively ($^2(12) = 67.3$, $p < .001$) — the configuration signs the premium while its gain scales with the year's own gap (the $r = +.94$ coupling of §7.6), which no single scalar carries. On immigration, the strictest one-path form — $\text{Gap}_t =\ \cdot \text{G}_t$ and $\text{Premium}_t = \lambda \cdot \text{G}_t^{\theta}$ on a single measured elite series — is rejected with either series as G: with speech tone, $^2(5) = 12.7$ ($p = .026$), the tone path too flat across 2012–2022 to produce a doubling gap and too high in the 1990s to produce a null one; with platforms, $^2(5) = 10.4$ ($p = .064$), failing the mirror-image way — the 2008 premium (+0.39) sits at platform-G of zero, the premium *preceding* the platform clock, which is the two-clock structure §7.5 dates, here quantified. The diffusion exponent is unidentified at these moments (95% profile CI [0.05, 3.0]). The reading is the one the paper's design already embodies: level-tracking structural forms fail exactly where the ratio test said they must, and identification therefore proceeds through dated onsets, sign reversals, and event-time profiles rather than through level-tracking of any single elite series. The unit-level joint model becomes estimable when E.9's common index puts both issues' contestation on one scale; until then the structural bound stands as a tabulated miss, in the ledger's genre.

Table 41: The gap equation, estimated: minimum-distance fits of the strict structural forms, and where each breaks. The rejections quantify why the model enters as an organizing lens with reception integrated, and why identification runs through dated onsets and sign reversals rather than level-tracking.

Quantity	Value
perception, strict scalar rotation: premium = c × incumbency	c = 0.046 (se 0.004); rotation exact (13/13 signs) but constancy REJECTED: chi2(12) = 67.3, p < .001 — the gain scales with the year's own gap (r = +.94, §7.6), not a constant
immigration, one latent path (speech-tone G): Gap = G, Premium = $\lambda\hat{\text{G}}$	= 0.106, λ = 0.518, = 1.50; chi2(5) = 12.7, p = .026 — REJECTED; unidentified (95% profile CI [0.05, 3.00]; = 1 not rejected, p = .57)
immigration, platform-G fork	chi2(5) = 10.4, p = .064 — rejected the mirror-image way: the 2008 premium (+0.39) sits at platform-G = 0; the premium precedes the platform clock (the two-clock structure, quantified)
reading	level-tracking structural forms fail where the design says they must (ratio test, R/37); identification proceeds through dated onsets, sign reversals, and event-time profiles, not level-tracking; the unit-level joint model awaits E.9's common index

Two issues, two clocks: dating the switch-on against a secular trend

The switch-on faces one threat the immigration series alone cannot dispatch: a secular trend. Perhaps the sophisticated simply polarized more on everything across these decades, for reasons unrelated to any issue's specific elite contestation, and immigration's premium rose with the calendar as a by-product. A single calendar-time trend can produce *one* issue's premium switching on with the calendar. It cannot produce *two* issues' premiums switching on at two *different* calendar times, each tracking its own elite-polarization clock. That is a test we can run without new data, and it is the strongest identification of the switch-on the design affords.

Trade contested on a different clock than immigration. Immigration's elite series and its premium move from the mid-2000s; trade's party poles did not flip until 2016, when the Trump-era Republican party abandoned the free-trade consensus for protectionism (Figure 26). Were the premium a secular property of the sophisticated, the trade premium would have risen alongside immigration's through the 2000s. It did not. Estimating the party×education interaction on the import-limits position across the ANES cumulative file (Figure 1, center panel), the trade premium is *negative or null* throughout the pre-flip era — through 2016 the educated of both parties move toward the free-trade experts, educated Republicans the most free-trade (the P2 regime; premium -0.21 to -0.58) — and turns *positive* only after the 2016 flip ($+0.18$ in 2020, $+0.41$ in 2024, $p = .006$). The immigration premium, on the same measure and years, is positive throughout (2004–2024). In 2008 and 2012 the two premiums carry *opposite signs in the same years* — immigration $+0.62$ and $+0.09$, trade -0.58 and -0.39 — which no common secular clock can produce.

The within-party slopes show the mechanism, and it is the immigration asymmetry reproduced on a second issue once the poles have flipped. Before 2016 the protectionist folk default leaned Democratic and schooling pulled both parties toward free trade; after 2016, with the Republican party now at the protectionist folk pole, educated Democrats — the new out-group — converge on the free-trade experts (protectionism falling from 0.35 to 0.25 across the college line in 2024) while educated Republicans hold the folk line (0.74 to 0.72). The premium did not merely appear; it *switched sign* as the party configuration reversed, exactly as a content-neutral, elite-configuration-driven account requires.

Two instrument substitutions in this test require flagging. The trade item is the import-limits *preference*, not a trade *fact* — no trade fact is fielded with a sophistication measure across the flip, so we use the preference and carry the fact–preference gap as elsewhere — and the sophistication measure is *education*, not the interviewer-information rating the immigration structural signature uses, because that rating was discontinued after 2016, before trade's switch-on could be observed. With those caveats the result does what immigration alone could not: it dates the premium's emergence to *each issue's own contestation clock*, so the switch-on is no longer a single calendar-collinear series, and the secular-trend confound — the one threat the immigration cross-derivative could not rule out — is answered by the data rather than by assertion. It remains descriptive (a preference, education, six survey-years), but it is the one observational strengthening of the cross-derivative available without experimental control of the elite signal, and it raises what the structural signature can bear: not one collinear series, but two issues switching on at two times.

Who moved, and when

This sequence reconciles two facts that can look opposed: the partisan gap *emerges* first among graduates (here), yet it is non-college Democrats who ultimately *move* most (§6). They are two stages of one process (Figure 15). College Democrats abandon the folk belief a decade before anyone else — opening the early gap among the educated — while non-college Democrats, starting from a far higher baseline (55 percent in 1996), peel away later but travel farther, to 18 percent by 2024 (a 37-point fall against the college Democrats' 24). Republicans of both education levels barely move throughout. The educated lead; the less-educated catch up and, from their higher start, cover the greater distance. First-mover and largest-mover are different roles, and Democrats of different educations play them in turn.

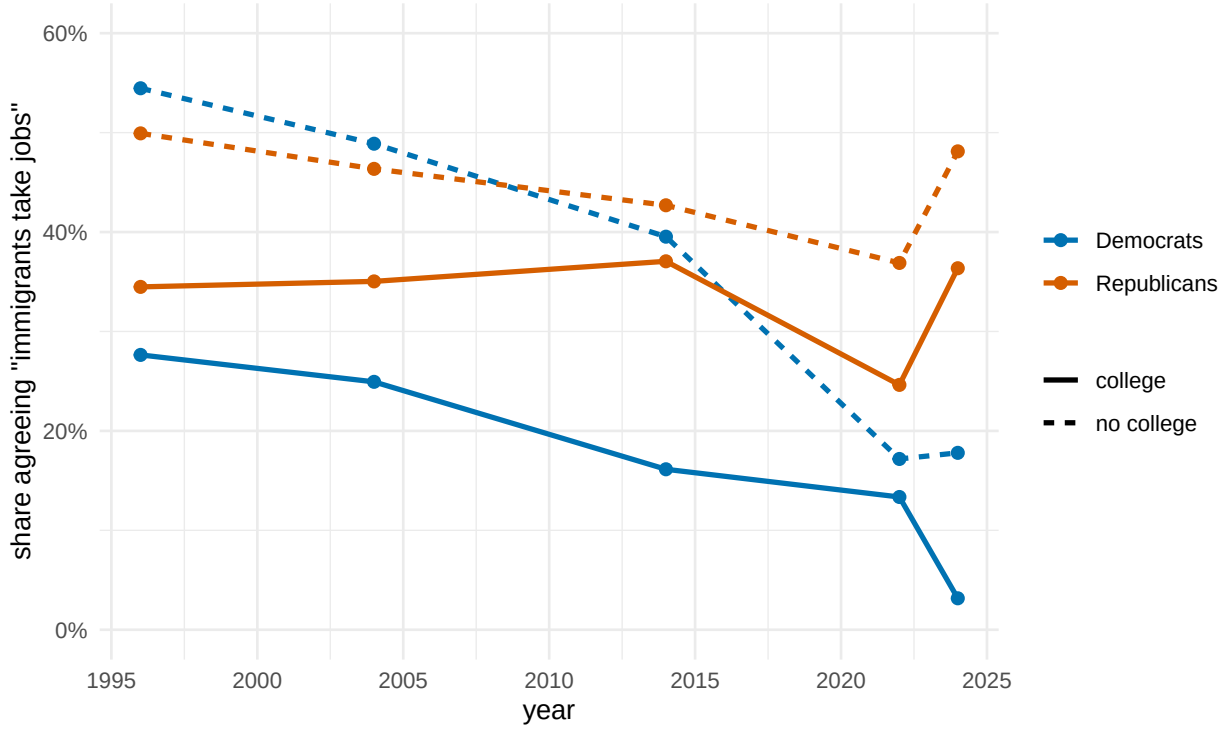


Figure 15: Who moved, and when: share agreeing 'immigrants take jobs', by party and education (GSS). College Democrats peel away first; non-college Democrats start higher, move later, and travel farther; both Republican groups barely move.

Scope table: party×education by elite contestation

The main text summarizes the three-claim education gradient (§7); the full table:

Table 42: Scope of the mechanism: party×education amplification of the folk belief by elite contestation. Null where uncontested, correcting at low-mid, amplifying at high; the immigration cell is positive but imprecise (+0.34, p = .27).

Claim	Elite contestation	party×education	p	n
Automation (2017)	low	0.056	0.780	4033
AI (2024)	low-mid	-0.352	0.034	4413
Immigration (2022-24)	high	0.337	0.273	3379

Negative-control battery

A negative-control battery grounds the scope condition empirically rather than by assertion, and directly answers the objection this section opened with — that the result might be nothing more than "the sophisticated polarize on everything." We estimate the same party×education interaction on a battery of *verifiable* GSS science facts (Table 43). On the identity-neutral ones — whether the Earth orbits the Sun, whether electrons are smaller than atoms, whether antibiotics kill viruses, how lasers work — the interaction is null (all $p > .15$, $|\text{coefficient}| \leq 0.2$), against +0.59 on the captured immigration fact. Sophistication does *not* widen partisan gaps on facts that carry no identity. Two items (whether all radioactivity is man-made, whether the Earth's core is hot) do show a positive interaction. We report these without a rescue: we did not pre-register which science items count as identity-neutral, so we will not retrofit a story for why two of them behave like captured facts, and we flag the battery as *suggestive* rather than a clean pre-specified control. What we can say is the modest thing the data support: across a battery of verifiable science facts the interaction is null on most and positive on the captured immigration fact and a minority of science items, so the amplification does not attach to verifiable facts *in general* — consistent with its tracking identity rather than sophistication-on-everything, but not a decisive negative control. A pre-registered identity-neutral battery is the version that would settle it.

Table 43: Negative-control battery: party×education interaction on verifiable GSS science facts. Null on most identity-neutral facts (vs +0.59 on the captured immigration fact), consistent with the amplification tracking identity rather than a general tendency of the sophisticated to polarize. Two science items show positive interactions and the battery was not pre-specified, so it is suggestive, not a decisive control (see text).

Verifiable fact	party×education	p	n
earthsun	0.147	0.366	6861
electron	0.034	0.797	5387
radioact	0.574	0.003	6424
lasers	0.198	0.169	5024
viruses	-0.008	0.942	8330
hotcore	0.521	0.043	6451

Social desirability and survey mode

The 38-point Democratic decline on the fact is not social-desirability suppression. One competing reading of that decline is that expressing "immigrants take jobs" grew less acceptable on the left, so the drop is partly reduced reporting rather than changed belief. The strongest observational lever, survey mode, runs against it. If the decline were suppression under social pressure, it should be *weaker* in self-administered (web) interviewing than interviewer-administered (in-person); the GSS fielded immjobs in both modes in 2022 and 2024 (Table 44), and it is not weaker but steeper — Democrats hold the folk belief *less* on the web (7 vs 15 percent in 2024; self-administration coefficient -0.70, $p < .001$), the reverse of suppression, and the partisan gap is identical across modes (34 vs 33 points in 2024). Mode is not randomly assigned, but this is the cleanest non-experimental evidence available — the calibration bound (SI-B) quantifies the residual the objection could claim — and it points away from social desirability (the within-person panel and the `immameco` co-movement, $r = .95$, point the same way).

Table 44: Social desirability vs. survey mode (GSS 2022, 2024). The folk-belief rate by party under interviewer-administered (in-person) and self-administered (web) interviewing. The Democratic rate is lower, not higher, on the web, and the partisan gap is essentially identical across modes — the opposite of what social-desirability suppression predicts. Mode is not randomly assigned (compositional caveat).

Year	Dem (in-person → web)	Rep (in-person → web)	gap in-person	gap web
2022	0.181 → 0.128	0.358 → 0.298	0.176	0.170
2024	0.154 → 0.073	0.483 → 0.411	0.329	0.338

The 2024 Republican rebound

The widening is, on net, overwhelmingly one-sided — ninety-eight percent of the net 1996–2024 gap change is Democratic movement toward the expert-consistent direction — but the Republican path is non-monotone (45 → 42 → 41 → 34 → 45), so we report it interval by interval (Table 47) rather than call the in-group "anchored." Through 2022 the change is 70–85 percent Democratic; the exception is 2022–2024, when the Republican share jumped from 34 to 45 percent, replicating independently in the ANES (a 39→52 rebound; Table 48). The rebound is the model's third permitted in-group path (§3.2), and it is identified by fingerprints stated there, not by accommodation here. In-group movement is permitted only *on a measured cue*, and cue-driven movement must be (i) dated to the cue, (ii) flat across sophistication, and (iii) indifferent to lived exposure. The 2022–2024 wave shows all three: it sits on the largest Republican-platform cue spike in the Manifesto series (below); the Republican knowledge slope is flat in 2024 (+0.04, n.s.; Table 45), a uniform cue effect rather than capacity-driven enactment; and it is uniform across state-level exposure change (SI-C, `R/58`). One confession, made once: the model permits this path but did not predict its *timing* — no account in the literature did — and what the framework buys instead is that the rebound's three fingerprints were checkable, and checked. The emergence-window asymmetry ($p_R \approx f$, 1996–2020) had its own disconfirmers — a symmetric sophistication slope while $p_R \approx f$, or the gap emerging first among the *least* informed — and we observe neither. The cue itself is measured, by the same instrument that dates every onset in the event study: the 2024 Republican platform is the first in the modern Manifesto series whose net multiculturalism emphasis turns *restrictive* — per608

− per607 = +0.48, against −0.87 in 2016 and 2020 and a previous series maximum of −0.06 — while the Democratic side stays deeply negative (−2.87), pushing the signed divergence to its series maximum (3.34; Table 46). The rebound therefore sits *on* a measured, Republican-side cue spike, dated in-sample by the platform clock, rather than beyond the end of the elite series. (The speech corpus cannot corroborate at the floor-tone margin: Republican congressional immigration tone is flat-to-*declining* over 2015–2020 — the anti share falls from about 0.53 to 0.39 — and the corpus ends in 2020; extending it through the 2021–2025 Congressional Record is pure labor with the published classifier, and a campaign-rhetoric corpus would capture the cue channel platforms proxy. Until then the platform series carries the dating alone.) We keep the contingency falsifiable with a registered forward prediction: if the rebound is an elite-cue effect ($p_R > f$ because Republican elites now supply an anti-immigration cue), it should *persist* only while that cue is supplied and *reverse* if elite supply withdraws, and it should remain uniform across sophistication rather than becoming knowledge-graded — the signature of cue-taking, not capacity-driven enactment. A rebound that instead grew knowledge-graded, or persisted after elite cue-withdrawal, would disconfirm the cue reading.

One rival the knowledge-flatness test cannot touch is *lived exposure*: the 2022–2024 migration surge was real, and experiential updating is as knowledge-flat as cue-taking, so the two readings separate only *geographically* — a national cue predicts a rebound uniform across local exposure, lived experience predicts one concentrated where exposure rose. Measured against the change in each state's foreign-born share (CES self-reports, n ≈ 60,000 per wave), the Republican rebound on the verbatim fact is uniform: rebound × exposure-change +0.004 (state-clustered $p = .89$; randomization inference permuting exposure across states, $p = .88$), the Democrats-as-within-state-control triple +0.018 ($p = .64$), the 2024-level fork −0.005 ($p = .80$). The matched *preferences* show at most a marginal exposure gradient (oppose-legal-status +0.023, $p = .004$; border patrols −0.005, n.s.) — lived exposure touches the preference margin while the fact's rebound tracks the national cue, one more instance of the fact behaving as configuration-driven where the preference admits experience. The cue reading survives its sharpest observational rival (replication files, R/58).

Table 45: The 2024 Republican rebound is not sophistication-driven: the within-party sophistication slope on the fact, by year. The Republican slope is flat throughout (including 2024), so the rebound is a uniform in-group shift, not knowledge-graded enactment.

Year	Measure	Republican slope	Democrat slope
2004	interviewer-info	-0.363	-0.418
2008	interviewer-info	-0.015	-0.425
2012	interviewer-info	-0.135	-0.551
2016	interviewer-info	-0.112	-0.649
2024	behavioral	0.037	-0.489

Table 46: The 2024 cue, measured: net restrictive multiculturalism emphasis (per608 − per607) by party platform, Manifesto Project 2025a. The 2024 Republican platform is the first in the modern series to turn net-restrictive; the Democratic platform stays deeply negative.

year	Democrats	Republicans	divergence
2008	-1.82	-2.08	-0.26

Table 46: The 2024 cue, measured: net restrictive multiculturalism emphasis (per608 − per607) by party platform, Manifesto Project 2025a. The 2024 Republican platform is the first in the modern series to turn net-restrictive; the Democratic platform stays deeply negative.

year	Democrats	Republicans	divergence
2012	-1.32	-0.06	1.26
2016	-3.30	-0.87	2.43
2020	-3.94	-0.87	3.08
2024	-2.87	0.48	3.34

Table 47: Interval-by-interval decomposition of the partisan-gap change on the fact (GSS immjobs). Most of the long-run widening is Democratic movement, but the 2022–2024 interval is Republican-driven — the in-group is not monotonically anchored.

Interval	Δ Dem (pts)	Δ Rep (pts)	Δ gap (pts)	% from Dem	% from Rep
1996-2004	-7.1	-3.2	3.9	69	31
2004-2014	-9.2	-1.6	7.6	85	15
2014-2022	-16.6	-7.2	9.4	70	30
2022-2024	-4.2	11.1	15.4	28	72

Table 48: The 2024 Republican rebound on 'immigrants take jobs' replicates across surveys (not a GSS-wave artifact): the Republican rate falls then rises into 2024 in both the GSS and the ANES, while Democrats decline monotonically.

Survey	Year	Dem share	Rep share	Gap
ANES	2016	0.244	0.497	0.253
ANES	2020	0.133	0.394	0.261
ANES	2024	0.123	0.520	0.397
GSS	2014	0.323	0.407	0.084
GSS	2022	0.157	0.335	0.178
GSS	2024	0.114	0.447	0.332

Supporting exhibits for the main text

Model assumptions and complete derivations

Assumptions. A1 (signal precision). The citizen holds a prior over the truth τ centered at prior_j with precision $\pi_0 > 0$ and observes one signal $x = \tau + \varepsilon$, $\varepsilon \sim N(0, 1/\pi(v))$, with $\pi(0) = 0$ and $\pi'(v) > 0$: experiential verifiability is signal precision. **A2 (loss).** Stated belief minimizes $L(b) = \mathbb{E}[(b-\tau)^2 \mid x] + \kappa\,(b - \text{prior}_j)^2$, $\kappa \geq 0$: posterior accuracy plus a congruence penalty for departing the prior. **A3 (prior mixture).** $\text{prior}_j = (1 - g(T))\,f(s) + g(T)\,p_j(s)$ with $g(0) = 0$, $g'(T) > 0$; in the reception-integrated variant $g = g(T, s)$ with $\partial^2 g/\partial T\,\partial s > 0$ and $g(T, s) \to g(T)$ as diffusion completes. **A4 (folk correction).** $f(s)$ is common to both parties and $\partial\,|f(s) - \tau|/\partial s < 0$.

A5 (identity sharpening; configuration). $\partial\,(p_R(s) - p_D(s))/\partial s > 0$ where T is high; the studied configuration places the in-party pole at the folk default, $p_R(s) \approx f(s)$, with the expert target τ on the far side of f from p_D.

Derivation of b^*. Under A1, $\mathbb{E}[\tau \mid x] = \frac{\pi(v)\,x + \pi_0\,\text{prior}_j}{\pi(v)+\pi_0}$. The first-order condition of A2 gives $b^* = \frac{\mathbb{E}[\tau|x] + \kappa\,\text{prior}_j}{1+\kappa}$. Substituting, $b^* = \alpha(v)\,x + (1-\alpha(v))\,\text{prior}_j$ with $\alpha(v) = \frac{\pi(v)}{(1+\kappa)(\pi(v)+\pi_0)}$, so $\alpha(0) = 0$, $\alpha'(v) > 0$, $\alpha < 1$, and in expectation over signals $\mathbb{E}\,b_j^* = \alpha(v)\,\tau + (1-\alpha(v))\,\text{prior}_j$ — the reduced form in §3, with the identity motive entering twice (as prior center and, if $\kappa > 0$, as congruence taste) and v entering only through precision.

The gap equation. Substituting A3 into $\mathbb{E}\,b_j^*$ and differencing across parties, the folk term cancels: $\text{Gap}(s) = \mathbb{E}\,b_R^* - \mathbb{E}\,b_D^* = (1-\alpha(v))\,g(T)\,(p_R(s) - p_D(s))$.

Proposition 1. As $v \to 0$, $\alpha \to 0$ and $\mathbb{E}\,b_j^* \to \text{prior}_j$, so $|\mathbb{E}\,b_j^* - \tau| \to |\text{prior}_j - \tau|$, which is invariant to the strength of the expert consensus: the public–expert gap survives arbitrarily strong consensus. □

Proposition 2. At $g(T) = 0$, $\text{prior}_j = f(s)$ for both j, so $\text{Gap}(s) = 0$ at every s (party is irrelevant), and $\partial\,|\mathbb{E}\,b^* - \tau|/\partial s = (1-\alpha)\,\partial\,|f(s) - \tau|/\partial s < 0$ by A4 (education corrects everyone). □

Proposition 3. At $g(T) > 0$, $\text{Gap}(s) = (1-\alpha)\,g(T)\,(p_R - p_D) \neq 0$. Under A5's configuration ($p_R \approx f$), $\mathbb{E}\,b_R^*$ is a mixture of f with itself and is flat in the identity channel, while $\mathbb{E}\,b_D^*$ moves with $p_D(s)$ — the out-group carries all movement, and since τ lies on p_D's side of f, that movement is toward the expert-consistent pole. □

Proposition 4. On a fixed claim (v, wording constant), $\partial\,\text{Gap}/\partial T = (1-\alpha)\,g'(T)\,(p_R - p_D) > 0$; as $g \to 1$ the fact's partisan loading approaches $(1-\alpha)(p_R - p_D)$, the loading structure of the matched preference. □

Proposition 5. $\partial\,\text{Gap}/\partial s = (1-\alpha)\,g(T)\,\partial(p_R - p_D)/\partial s$, which is zero when $g(T) = 0$ (by P2's cancellation) and strictly positive when $g(T) > 0$ (by A5): the premium exists only in the identity regime. *Remark (the mild low-T negative).* The model proper predicts exactly zero at $g = 0$; the small negative interactions observed on low-T claims arise when $f'(s)$ operates on parties at unequal baselines — a measurement fact about bounded scales, outside the model and noted where it appears. □

Timing implications of $g(T, s)$. Write $g(T, s) = g(T)\,r(T, s)$ with reception r increasing in s, $r \leq 1$, and $r \to 1$ (all strata reached) as diffusion completes. Then $\text{Gap}(s) = (1-\alpha)\,g(T)\,r(T, s)\,\Delta p(s)$ where $\Delta p = p_R - p_D$, and the premium is $\partial\,\text{Gap}/\partial s = (1-\alpha)\,g(T)\,[\,r_s\,\Delta p + r\,\Delta p_s\,]$. *(i) The premium leads the gap:* at onset, $g(T)$ is small and r is near zero except at high s where $r_s > 0$ is largest, so the s-averaged gap is indistinguishable from zero while the premium — carried by the $r_s\,\Delta p$ term concentrated at high s — is already positive. *(ii) The premium-to-gap ratio declines:* the ratio is $r_s/r + \Delta p_s/\Delta p$, and as diffusion completes $r \to 1$, $r_s \to 0$, so the ratio falls monotonically to the separable limit $\Delta p_s/\Delta p$ — the two-stage diffusion SI-C documents in cross-section, and the reason level results use the $g(T)$ limit while timing results use $g(T, s)$. □

Table 49: Summary of the main estimates.

Finding	Estimate	Inference
Public-expert gap (immigration)	+39 pts	public 42% vs expert 3%
Party OR – automation (low T)	1.04	[0.84, 1.30]
Party OR – AI (mid T)	1.70	[1.44, 2.01]
Party OR – immigration (high T)	3.79	[2.91, 4.93]
Composition: party x era OR	3.38	[2.47, 4.63]
Composition: education x era OR	1.14	p = .46 (n.s.)
Precedence (elite leads public)	lag-10 r = .92	Granger beta = .72, p = .063; randomization p ≈ .10
Capstone: partisan gap on the fact	-0.03 -> +0.33	trend r = .92, p = .03
Fact converges on preference	9% -> 72%	levels convergence 2004-2024
Party x knowledge (mechanism)	+0.51	p < .001
… net of education + ideology	+0.48	p < .001
Sophistication amplifies only at high T	immigration + / automation null / AI -	scope/falsification
Fact -> Trump vote (net of party)	OR 7.56	p < .001

Table 50: Analytic samples: size and composition by survey (weighted shares; % Republican among party identifiers, % college = bachelor's or higher).

Survey	Years	N	% Republican	% college
GSS (immjobs)	1996-2024	8177	48	29
Pew ATP W27 (automation)	2017	4116	44	30
Pew ATP W152 (AI)	2024	4566	52	35
ANES 2024	2024	4722	48	35

Table 51: Design-fork robustness of the party×year slope (survey-weighted logistic, cluster-robust by survey-year; log-odds per decade). Positive and significant across both surveys separately and every leave-one-survey-year-out.

Design fork	slope per decade	p
pooled	0.858	<.0001
GSS only	0.584	<.0001

Table 51: Design-fork robustness of the party×year slope (survey-weighted logistic, cluster-robust by survey-year; log-odds per decade). Positive and significant across both surveys separately and every leave-one-survey-year-out.

Design fork	slope per decade	p
ANES only	1.137	<.0001
drop ANES2004	0.843	<.0001
drop ANES2008	0.824	<.0001
drop ANES2012	0.856	<.0001
drop ANES2016	0.861	<.0001
drop ANES2020	0.843	<.0001
drop ANES2024	0.743	<.0001
drop GSS1996	1.018	<.0001
drop GSS2004	0.862	<.0001
drop GSS2014	0.862	<.0001
drop GSS2022	0.896	<.0001
drop GSS2024	0.846	<.0001

Table 52: Public–expert agreement on unverifiable economic claims (weighted shares; gap = public − expert; CVI from SI-A). Descriptive context only. Immigration carries no expert point estimate (no direct IGM item exists; its direction is anchored on the National Academies consensus, SI-A), so its benchmark and gap are left blank rather than imputed. The trade pair matches a broad public folk item to the experts' verdict on a bundled policy claim (SI-A.1), so its printed gap is illustrative, not a measured distance.

pair	CVI	public_agree	expert_agree	gap
AI -> unemployment	11	0.77	0.31	0.45
Trade -> job loss	12	0.36	0.00	0.36
Immigration -> job loss	6	0.42	NA	NA

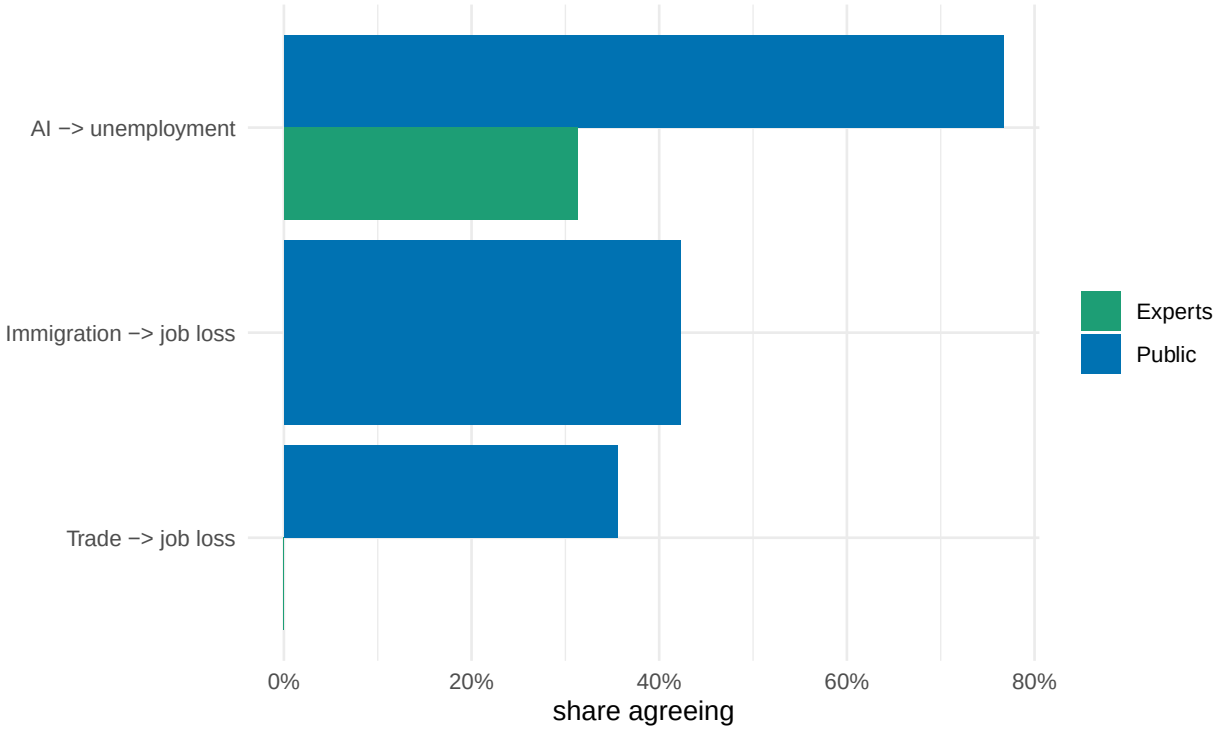


Figure 16: Public versus expert agreement on unverifiable economic claims.

Table 53: Individual-level divergence model (AI pair, Pew W152): odds ratios from a survey-weighted logit.

term	estimate	conf.low	conf.high	p.value
partyRep	1.38	1.10	1.74	0.01
college	0.82	0.69	0.98	0.03
ideo	1.09	0.97	1.21	0.14
partyRep:college	0.68	0.49	0.95	0.02

Table 54: Party odds ratios (Republican vs Democrat) on the folk belief, by tribalization.

claim	estimate	conf.low	conf.high
automation_2017	1.04	0.84	1.30
ai_2024	1.70	1.44	2.01
immigration_2022_24	3.79	2.91	4.93

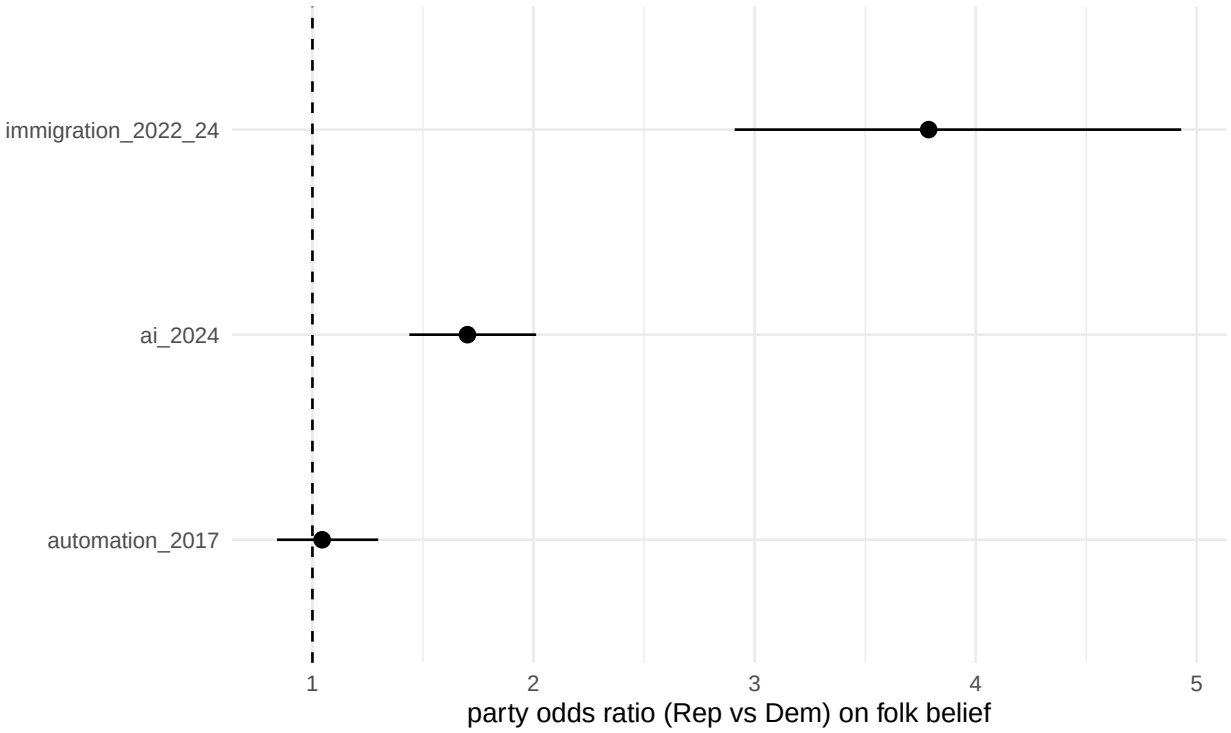


Figure 17: The identity effect rises with tribalization.

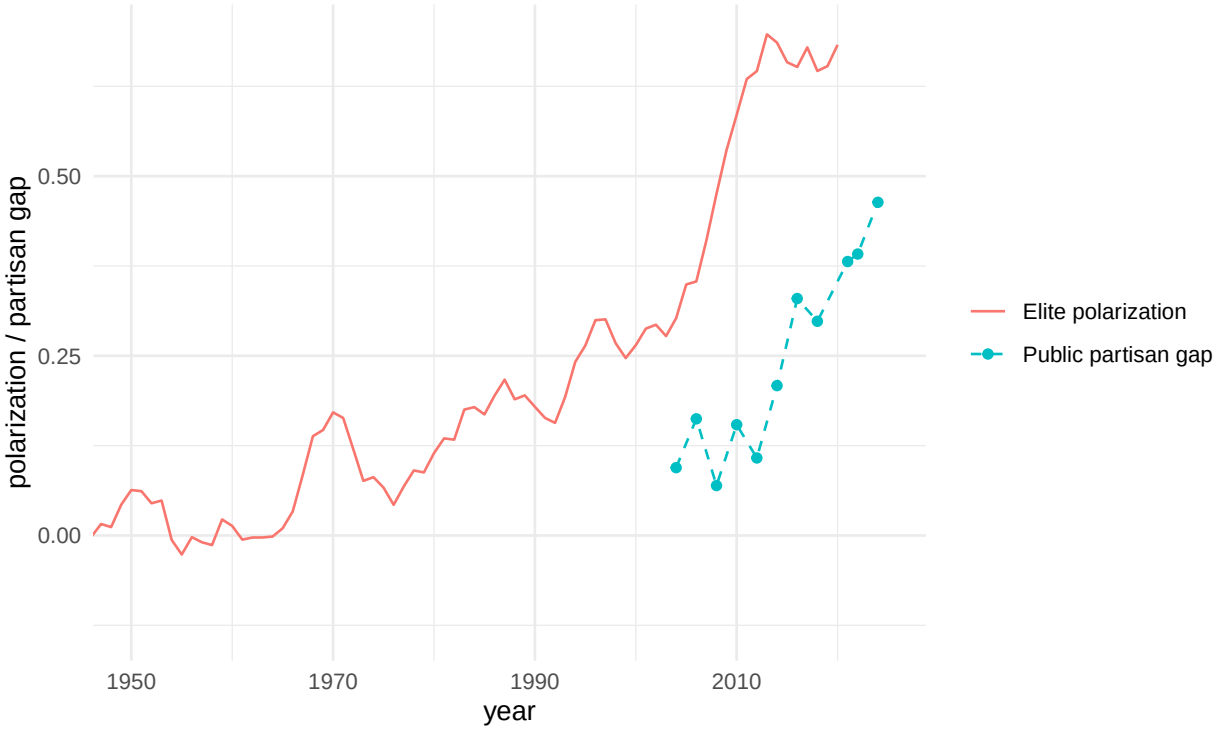


Figure 18: Elite polarization preceded the public partisan gap (lag-10 r = 0.92; Granger $\beta = 0.72$, p = .063; suggestive temporal ordering, not a causal estimate).

Table 55: Share agreeing 'immigrants take jobs', by party and survey year.

year	dem_share	rep_share	gap	n_dem	n_rep
1996	0.485	0.455	-0.030	610	484
2004	0.415	0.424	0.009	514	482
2014	0.323	0.407	0.084	569	418
2022	0.157	0.335	0.178	963	782
2024	0.114	0.447	0.332	833	802

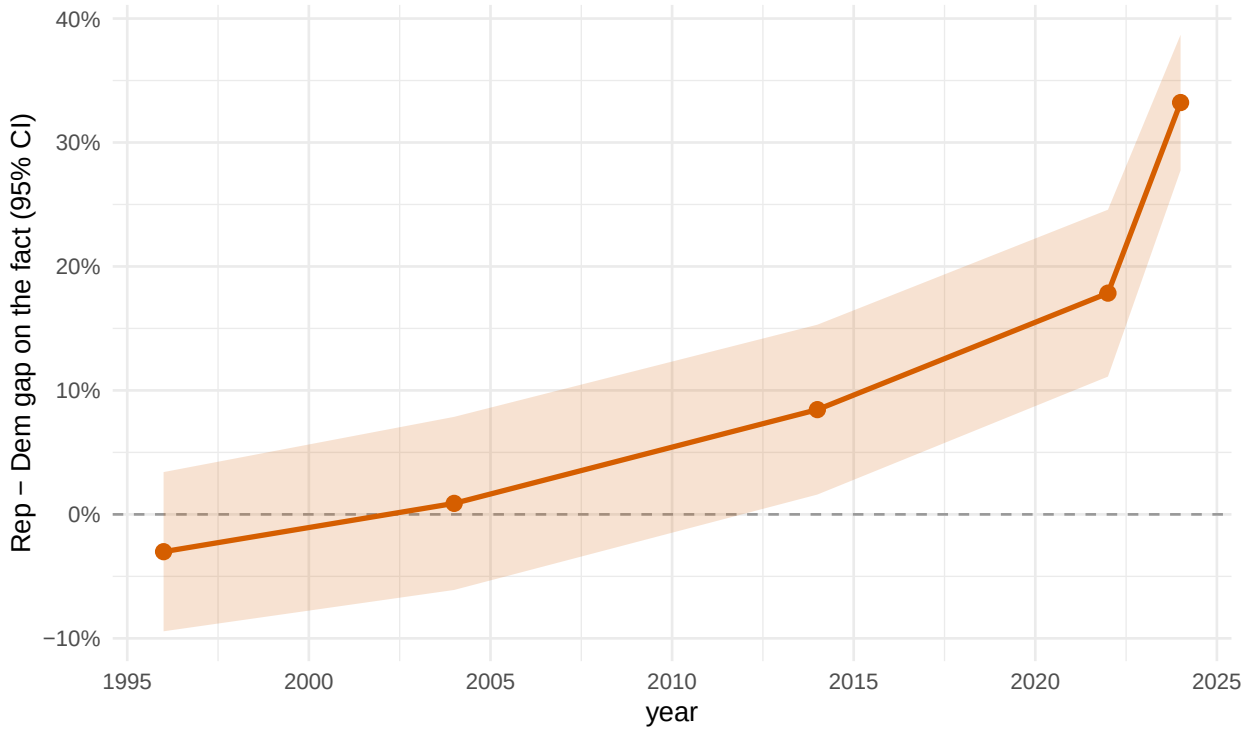


Figure 19: The partisan gap on the fact with design-based 95% confidence intervals: indistinguishable from zero through 2004, significantly positive and growing thereafter.

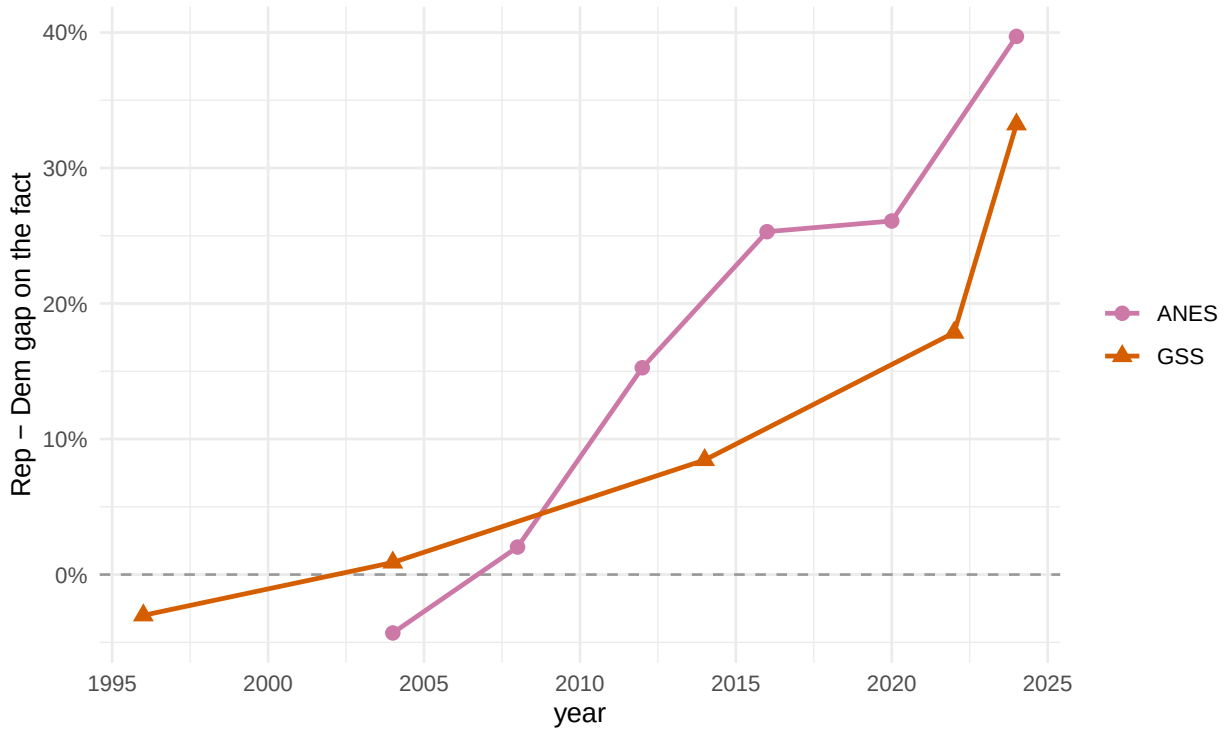


Figure 20: Two independent fixed-wording series. The partisan gap on 'immigrants take jobs' rises across both the GSS (immjobs) and the ANES (cumulative-file VCF9223), reproducing the same non-partisan-to-partisan crossing on two national surveys with independent samples and weighting.

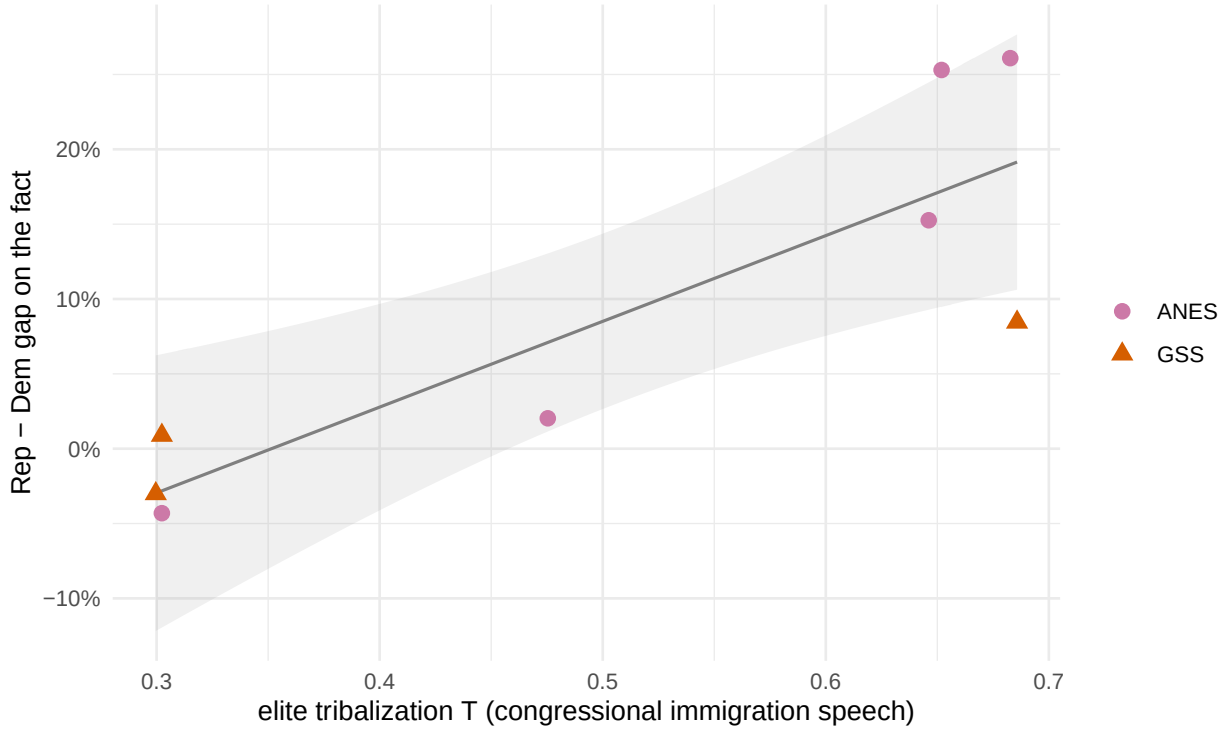


Figure 21: Gap on tribalization, not year. Pooling GSS and ANES survey-years for which the externally-measured tribalization index is observed, the partisan gap on the fact rises with contemporaneous elite tribalization (T coefficient +0.53, p = .02; raw r = +0.86).

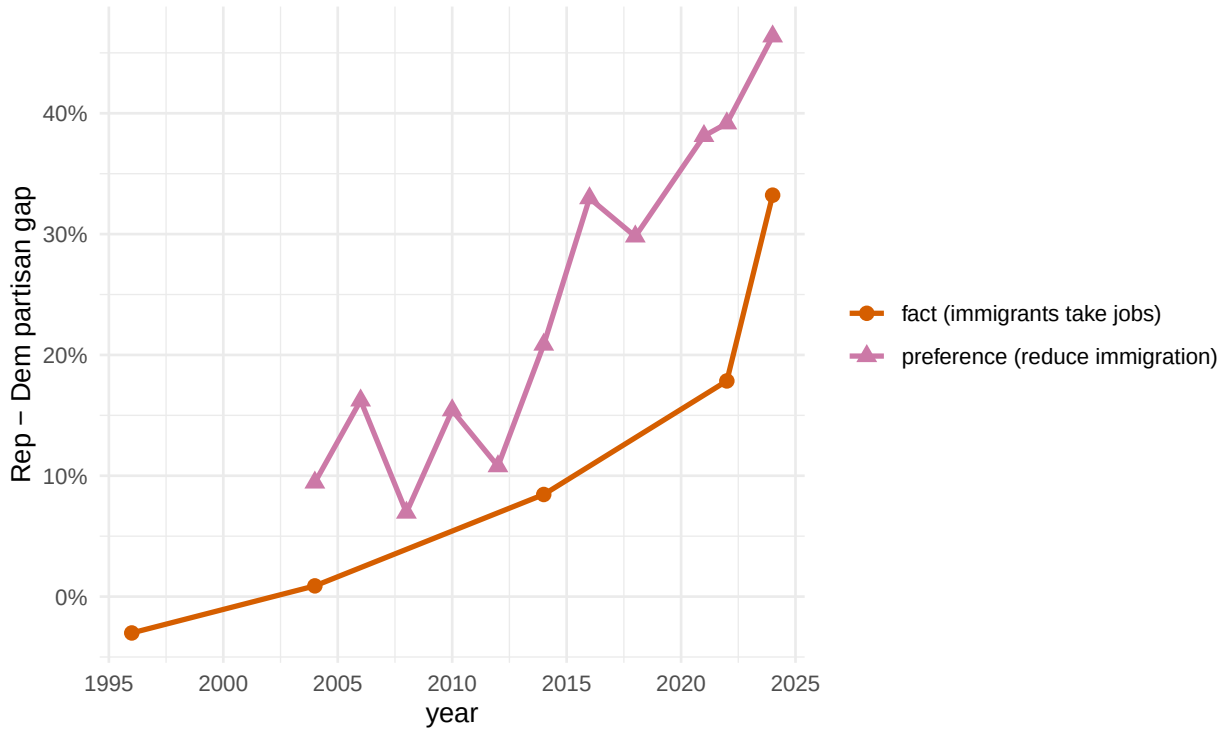


Figure 22: A fact comes to sort like a preference. The partisan gap on the factual belief climbs in levels to value-cleavage magnitude by 2024, reaching roughly 70% of the matched policy preference's gap (we rest on levels, not the share-of-preference ratio, which is unstable on the near-zero 2004 base; §6) — the fact migrates into the partisan structure the preference already had.

The perception series and the VOTER trajectory

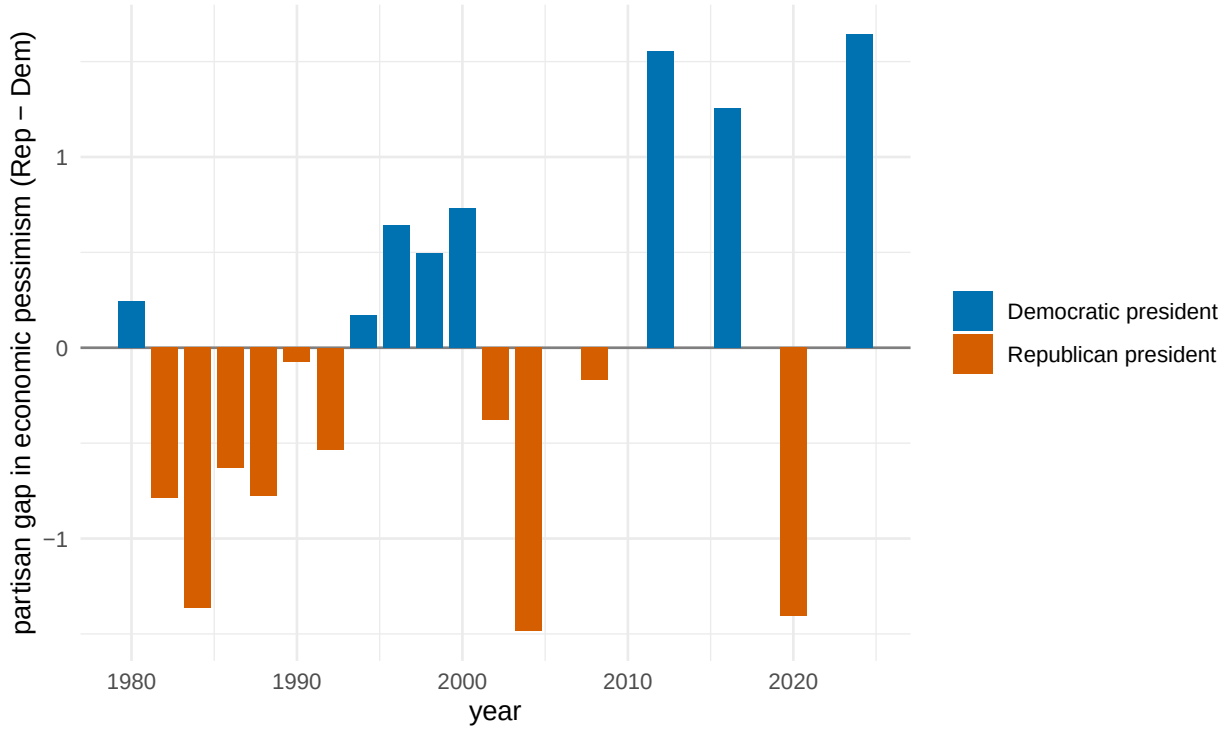


Figure 23: Content neutrality at power: the partisan gap in retrospective economic pessimism (Republican minus Democrat, log-odds), ANES 1980-2024. The gap is positive under every Democratic president and negative under every Republican one (r with incumbency = +0.84 across eighteen elections): the captured fact is owned by whichever party sits opposite the incumbent.

Table 56: The manufactured premium, year by year: signed party×information premium on economic perception, instrument, and incumbent party. 1980–2016 use the interviewer information rating; 2020/2024 use behavioral knowledge items with signs predicted before estimation.

year	incumbent	instrument	signed premium (/SD)	n
1980	Democratic	interviewer rating	0.003	1336
1984	Republican	interviewer rating	-0.042	1874
1988	Republican	interviewer rating	-0.072	1713
1992	Republican	interviewer rating	-0.041	2134
1996	Democratic	interviewer rating	0.034	1494
2000	Democratic	interviewer rating	0.030	1551
2002	Republican	interviewer rating	-0.020	1359
2004	Republican	interviewer rating	-0.078	1037
2008	Republican	interviewer rating	-0.013	2001
2012	Democratic	interviewer rating	0.136	1785
2016	Democratic	interviewer rating	0.072	973
2020	Republican	House-majority recall (behavioral)	-0.040	7104

Table 56: The manufactured premium, year by year: signed party×information premium on economic perception, instrument, and incumbent party. 1980–2016 use the interviewer information rating; 2020/2024 use behavioral knowledge items with signs predicted before estimation.

year	incumbent	instrument	signed premium (/SD)	n
2024	Democratic	civics battery, 4 items (behavioral)	0.112	4877

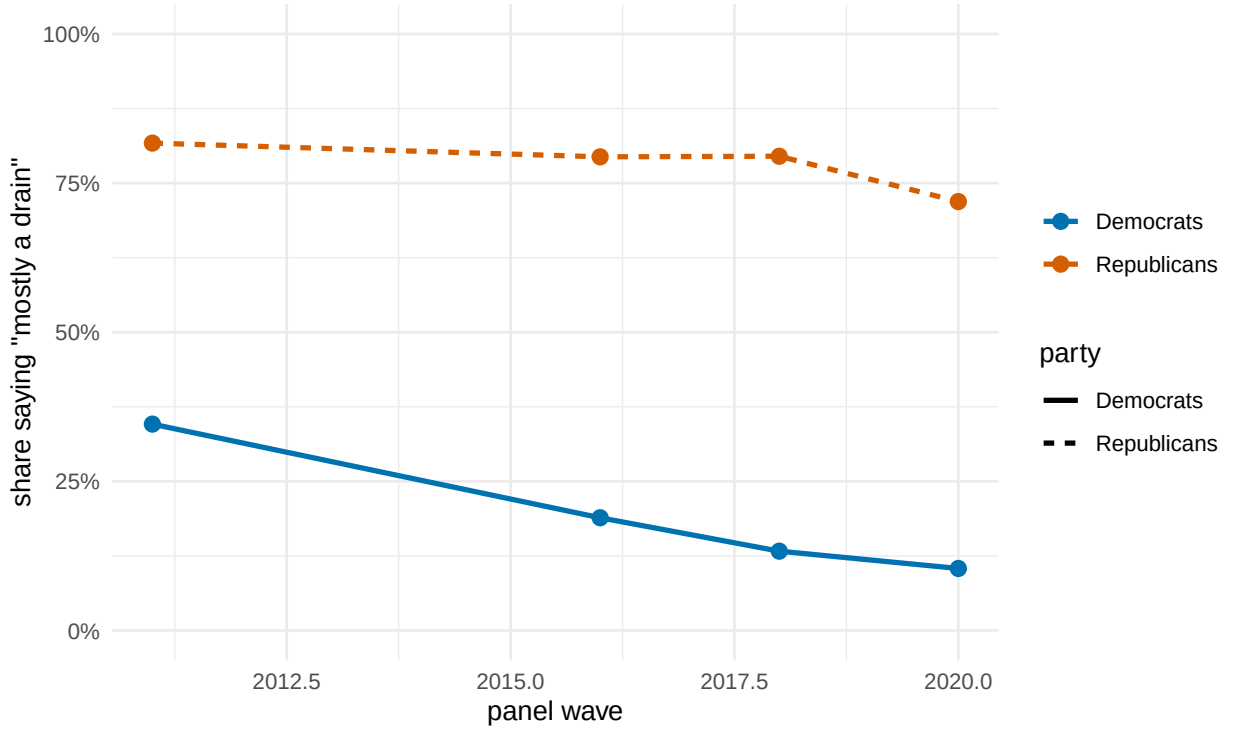


Figure 24: Within-person capture (VOTER panel, 2011-2020): share calling immigrants 'mostly a drain', by party, same respondents re-interviewed. The gap widens from 47 to 62 points by the familiar asymmetry - Democrats abandon the belief while Republicans hold.

SI-D: Generality, content neutrality, and cross-national scope

This appendix reports in full the three generality demonstrations summarized in §7.

SI-D.1 Beyond economics: disbelief in evolution

The mechanism is defined by the *structure* of a claim — truth-apt, unverifiable, elite-contested, with a folk default one party's pole shares — not by its content, so it should appear on any belief with that structure. Consider disbelief that humans evolved from earlier species. The scientific consensus is near-universal; the claim is unverifiable from everyday life; the folk default (that humans are not animals) is the kind of pre-political intuition the model places at f; and elites contest it along partisan and religious lines. The GSS science module fields it (`evolved`, 2006–2018),

and the immigration signature reappears (Figure 25): the Republican–Democrat gap in *disbelief* grows from 14 points (2006) to 24 (2018), one-sidedly — Democrats converge on the consensus (43 to 34 percent) while Republicans hold the folk default (57 to 58). A sophistication signature appears — the party×education interaction on disbelief is +0.59 ($p = .001$) — but does not survive the obvious confound: disbelief in evolution is heavily structured by religiosity, and once church attendance, biblical literalism, and their party interactions enter, the interaction falls to +0.21 and loses significance ($p = .15$). So evolution corroborates the *capture* — a non-economic fact acquiring partisan structure, asymmetrically, over two decades — but its sophistication-amplification is entangled with religiosity in a way the immigration result is not. We lean on evolution for the partisanization, not as independent evidence for the sophistication mechanism.

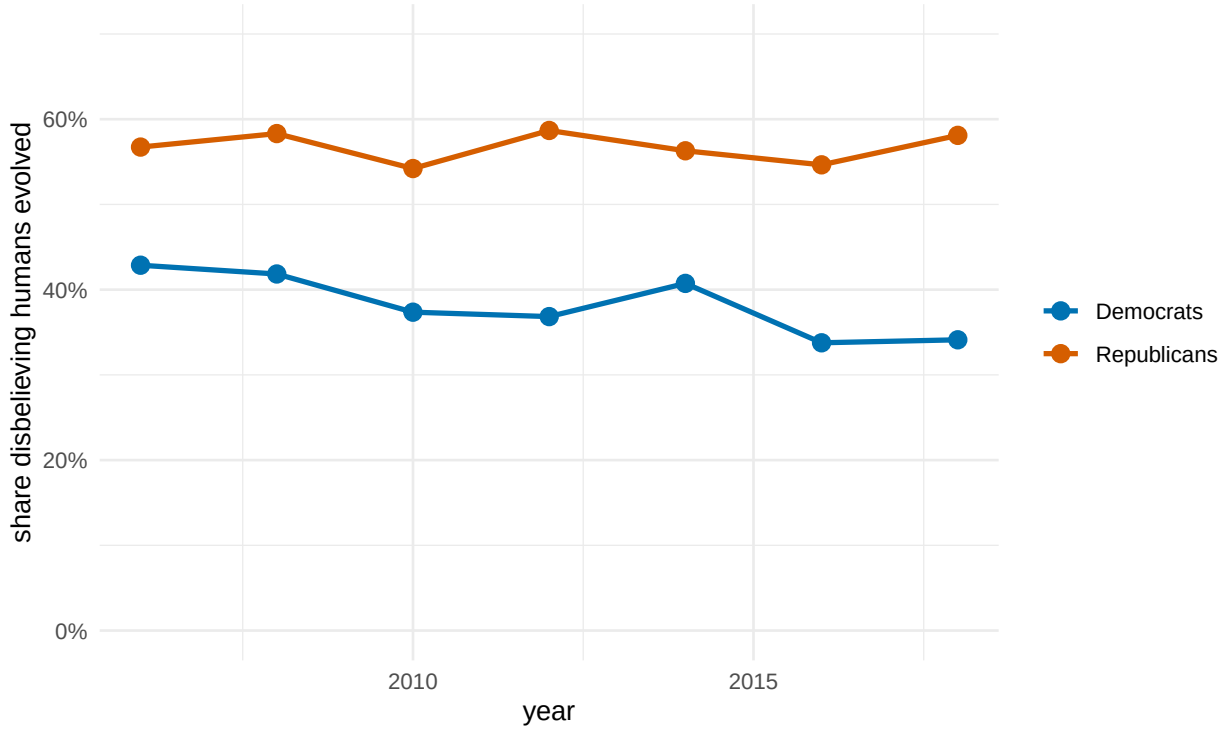


Figure 25: The same capture, a different domain: the partisan gap on whether humans evolved (GSS science module, 2006–2018). Disbelief sorts by party — Democrats converging on the scientific consensus, Republicans anchored at the folk default — exactly as on immigration.

SI-D.2 Content neutrality: the trade party-pole flip

Immigration and evolution both have the *same* party — the Democrats — moving toward the expert position, which is observationally consistent with a simpler claim than ours: that Democrats are merely more deferential to expert consensus. The model asserts something content-neutral and stronger — that the converging out-group is *whichever* party departs from the folk default. The decisive case has the folk default Democratic-congenial and the experts siding with Republicans; trade supplies it. Folk economics marks protectionism as the shared default (Caplan, 2007); the experts hold that trade raises aggregate welfare. On the ANES import-limits item the party at the

protectionist folk default *switches* (Figure 26): from 1988 through 2012 Democrats were the more protectionist party (Rep−Dem gap −3 to −11 points), so the expert-aligned, freer-trade side was Republican; by 2016 the gap had flipped to +13 and by 2024 to +42, as the Trump-era Republican party adopted protectionism and Democrats moved toward the economists. The party holding the folk error against the experts is set by elite conflict, not a standing disposition to defer. Its limit: because trade became contested only *after* the parties swapped poles, its low-elite contestation era is a P2 regime in which sophistication corrects both sides, so trade demonstrates content neutrality through the *direction and flip of the gap*, not through a sophistication-amplification test with Republicans converging — the test the nuclear-power item supplies (SI-D.2b; §7).

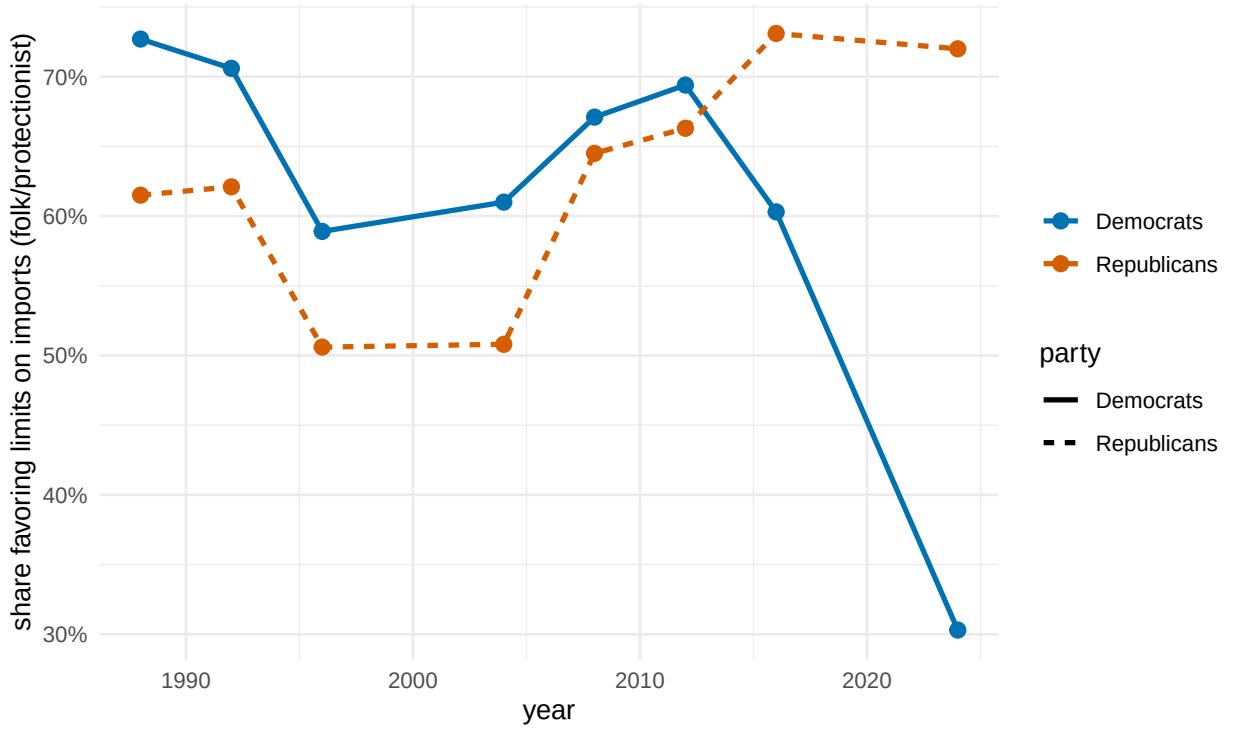


Figure 26: Content neutrality. On the ANES import-limits item, the protectionist (folk-economic) default is held by Democrats from 1988–2012 and by Republicans after 2016: the party that departs from the experts is set by elite conflict, not by a fixed disposition to defer. Share favoring limits on imports, by party.

SI-D.2b The converse case in one battery: nuclear power, GM crops, climate

The GSS environment rotations (1993–2021) field three expert-consensus danger beliefs on one five-point scale, jointly spanning the three elite configurations the model distinguishes. Nuclear power carries a strong expert consensus of low mortality and environmental risk per unit of energy generated — in the comparative mortality estimates (Markandya & Wilkinson, 2007), the UNSCEAR radiological assessments (United Nations Scientific Committee on the Effects of Atomic Radiation, 2021), and the IPCC lifecycle emissions figures, where nuclear sits with wind rather than with fossil fuels (Schlömer et al., 2014); it is experientially unverifiable; elites contested it for decades; and the folk default — dread — has historically been owned by the Democratic pole, making

the expert-consistent pole Republican-congenial: the mirror image of immigration's configuration. GM-crop risk carries a comparable consensus (National Academies of Sciences, Engineering, and Medicine, 2016) but *no* elite owner of the folk pole, so the model's P2 regime predicts education correcting both parties with no premium. Climate is the standard configuration, the expert pole Democratic-congenial (McCright & Dunlap, 2011). The predictions follow from the model, not the data: party×college on the expert-consistent belief positive for nuclear, null for GM crops, negative for climate, with the per-party education slopes asymmetric in mirrored directions on the two contested items.

All three hold (Table 3, main text; Figure 27). On nuclear the education slope among Republicans is +0.155 ($p < .001$) against +0.057 among Democrats — college Republicans reach "not dangerous" at 38 percent against 23 for non-college, while Democrats sit at 11–17: the educated wing of the folk-dread party barely moves. On GM crops both slopes are +0.09 to +0.11 (each $p < .001$) and the interaction is +0.013 ($p = .71$). On climate the Republican slope is flat (−0.018, $p = .39$) while the Democratic slope is +0.098 — the asymmetric education amplification documented by McCright & Dunlap (2011), recovered here as one cell of a three-cell design. Per-year estimates (Table 57) show the nuclear interaction positive in five of six waves, carried by the 1993–2010 rotations and attenuated in 2016–2021 — a fade consistent with the recent *bipartisan* softening of elite conflict over nuclear power (as elite contestation withdraws, the model expects the premium to decay), though we flag that reading as post hoc. Caveats: the items ask about danger "for the environment," a wording on which a referee could pit unresolved waste storage against the mortality and lifecycle records; we note it, note that the three anchors triangulate the same directional pole, and note above all that the prediction was fixed *ex ante* by the elite configuration — which party owned the folk pole, itself assigned by the SI-A.5 rubric (dread for nuclear and GM, distance-discounting for climate) and validated on the disengaged baseline there, not by any reading of the item — so contesting the benchmark contests the pricing of the convergence, not its direction; the benchmark is directional rather than a point estimate, exactly as for immigration; education is the paper's weaker sophistication proxy (§7); and the 2021 wave changed mode. None of these touches the design's point: the identity of the converging party is set by the configuration, not by a standing partisan disposition toward expertise.

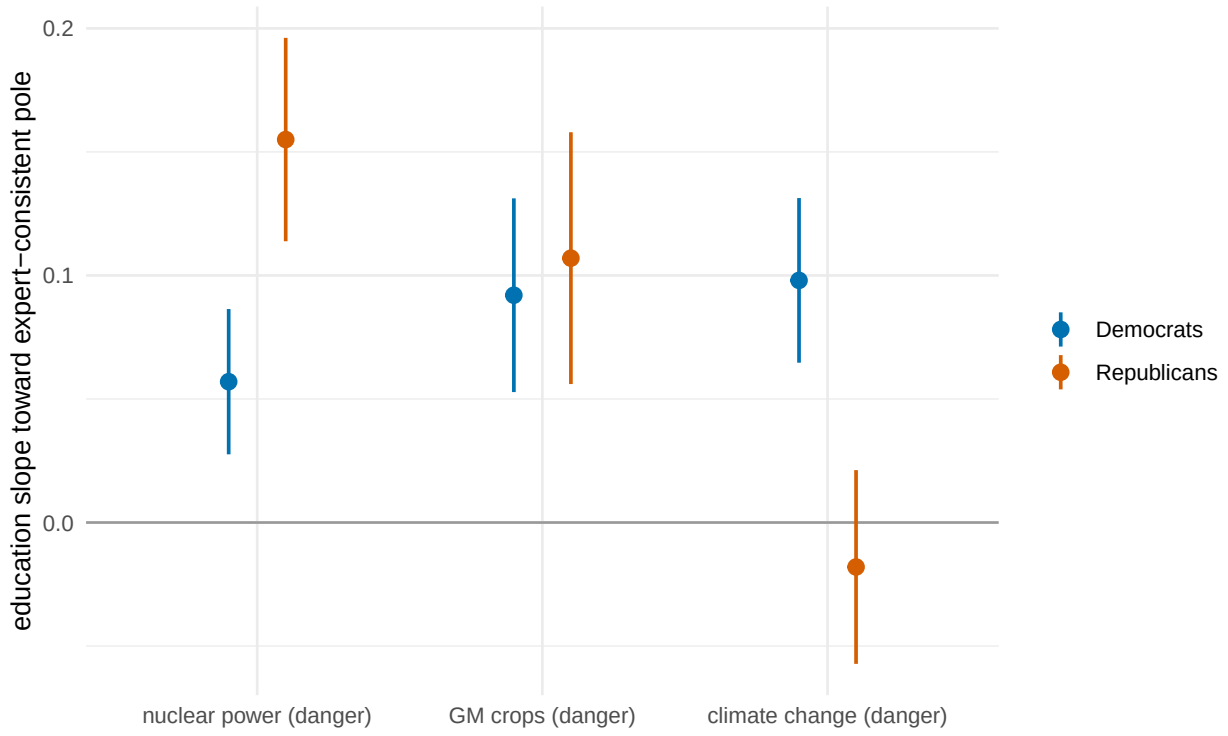


Figure 27: Education slopes toward the expert-consistent pole, by party, for the three GSS danger items (pooled years, weighted LPM with year fixed effects; bars are 95% CIs). The converging party flips with the elite configuration: Republicans on nuclear power, both parties on GM crops, Democrats on climate.

Table 57: Per-year party×college interactions on the expert-consistent belief (weighted LPM, HC1 SEs), by item.

issue	year	party×college	SE	p	n
nuclear power (danger)	1993	0.108	0.057	0.056	1208
nuclear power (danger)	1994	0.141	0.061	0.021	1056
nuclear power (danger)	2010	0.187	0.066	0.004	1073
nuclear power (danger)	2016	0.014	0.068	0.833	722
nuclear power (danger)	2018	0.115	0.069	0.097	592
nuclear power (danger)	2021	0.033	0.060	0.580	1349
GM crops (danger)	2000	0.023	0.082	0.781	766
GM crops (danger)	2010	0.107	0.074	0.148	950
GM crops (danger)	2016	-0.056	0.070	0.426	710
GM crops (danger)	2018	0.069	0.088	0.432	582
GM crops (danger)	2021	-0.056	0.062	0.362	1292
climate change (danger)	1993	-0.196	0.069	0.004	1182
climate change (danger)	1994	-0.092	0.073	0.210	999
climate change (danger)	2000	-0.057	0.080	0.476	846
climate change (danger)	2010	-0.138	0.071	0.050	1069
climate change (danger)	2016	-0.171	0.084	0.042	704

Table 57: Per-year party×college interactions on the expert-consistent belief (weighted LPM, HC1 SEs), by item.

issue	year	party×college	SE	p	n
climate change (danger)	2018	-0.141	0.099	0.154	583
climate change (danger)	2021	-0.081	0.055	0.142	1347
climate change (danger)	2024	-0.034	0.088	0.697	786

SI-D.3 Cross-national scope (ISSP 2013)

The mechanism is built on the two-party United States, raising a scope question: should capture be weaker where many parties diffuse the partisan signal? The ISSP 2013 National Identity module (ISSP Research Group, 2015) fields the *same* fact — whether "immigrants take jobs away from people born in [country]" — in twenty-six democracies, with a harmonized left–right measure (Figure 28). Three results. First, the capture generalizes: the right holds the folk belief more in 77 percent of countries. Second, the naive scope conjecture *fails*: the gap is uncorrelated with party-system fragmentation ($r = +0.16$, $p = .44$, the wrong sign). Third, what *does* track the gap is elite issue-supply: the left–right gap is four times larger where a salient anti-immigration party exists (0.13 vs 0.03; Welch $t = 2.96$, a cross-sectional description, surviving a party-number control, $p = .006$). The moderator is plausibly *endogenous* (ownership and sorting co-determined), so we read it as descriptive corroboration, and the single cross-section compares countries at different stages of their own contestation (the US mid-ranked in 2013, matching its GSS path). The identity mechanism is not confined to two-party systems and does not weaken with party number; its footprint is set by whether elites supply a party to own the claim.

We can partly address the endogeneity with a second wave. Stacking ISSP 2003 (National Identity II) with 2013 into a country-year panel of the seventeen democracies present in both, the mean left–right gap on the identical fact rose from 0.041 to 0.078, and countries with a salient radical-right party widened about twice as much as those without (+0.051 vs +0.025). But with only two waves and seventeen countries the within-country difference-in-differences is not significant (+0.025, $t = 0.69$), and continuous elite-supply measures from party platforms are too noisy to sharpen it — the Manifesto nationalist family code conflates governing nationalists with anti-immigration challengers. The two-wave panel is therefore *directionally consistent* with the cross-sectional reading — capture tracking elite supply, now within countries over time — but underpowered; identifying the within-country moderation cleanly would take a third wave and externally coded party positions (CHES). That third wave now exists: the ISSP 2023 National Identity and Citizenship integrated file (ZA10010, v1.0.0) is released at GESIS, fielding the identical item a decade further into most of these countries' radical-right trajectories; the replication code auto-detects the file and upgrades the two-wave difference to a three-wave country-fixed-effects test the moment it is placed in the data directory (R/38). We flag it as the highest-yield extension rather than claim it as done. The *premium* form of the comparative claim is also now tested to the record's limit, and the result is a bounded non-test reported in the ledger's genre (`R/72`; `issp_premium_nontest.csv`). Estimating right×education (harmonized degree, standardized within country) on the identical fact per country and wave, country-wave premiums are estimable but noisy — median standard error ≈ .04, so single-country intervals span roughly ±8 points — and the United States, the positive control where the premium is independently documented, shows +0.05 (2003) and +0.08 (2013), sitting exactly where the domestic education-based second difference sits (+0.03 [−0.06, +0.11], Table 20): on

this instrument — education, the paper's weak proxy, crossed with bloc-level left–right rather than party — the premium is small even where the knowledge-based premium is known to be large. The radical-right moderation of the premium *change* is null (DiD −0.010, Welch $p = .80$) amid cross-country heterogeneity too extreme to be substantive (−0.28 to +0.13). With these instruments the cross-national premium question cannot be answered in either direction; the party-resolved BES panel is the vehicle with the resolution the second-order quantity needs. A second, sharper out-of-sample commitment is drafted for separate registration: the British Election Study panel (waves 1–30, 2014–2024) fields a fixed-wording economic immigration belief within persons across Britain's fast, dated elite reconfiguration of the issue — including the prediction, inexpressible in a two-party system, that after 2016 capture follows the Leave–Remain configuration rather than party (`paper/preregistration_bes.md`).

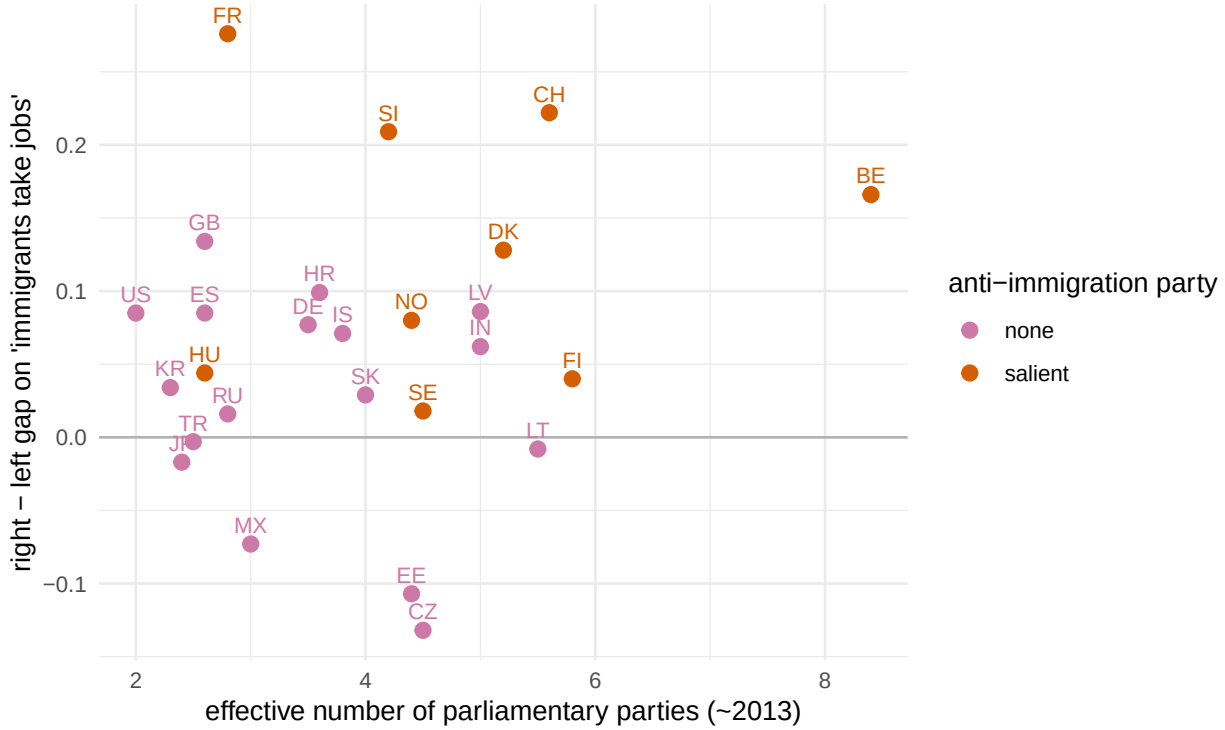


Figure 28: Cross-national scope (ISSP 2013, 26 democracies). Left–right gap on the 'immigrants take jobs' fact against party-system fragmentation (effective number of parliamentary parties). The gap is positive in most countries but is *uncorrelated* with the number of parties (r = +0.16, n.s.); it is larger where a salient anti-immigration party exists (orange). The US (two-party) is mid-ranked, consistent with its still-rising GSS trajectory in 2013.

SI-D.4 Knowledge versus education: the two-channel test

The result does not depend on the ANES knowledge battery; it recurs in the GSS with education, through a revealingly different door (main text, Figure 2; the underlying slopes in Table 58). Across the high-contestation years (2014–2024), education lowers the folk belief in *both* parties: Democrats from 24 to 10 percent across the college line, Republicans from 42 to 32. General schooling corrects

folk-economic error in everyone — but it corrects Democrats about twice as steeply (within-party log-odds −1.04 against −0.44), so the partisan gap widens only modestly with education (19 points among non-graduates to 22 among graduates). Political knowledge in the ANES does something sharper: it leaves Republicans flat and moves only Democrats, widening the gap from 34 to 46 points. The contrast is a prediction of the model: education acts on the folk prior $f(s)$, common to both parties; political knowledge acts on $p(s)$, which departs from folk intuition only for the out-group. A measure of general human capital corrects the shared error in everyone and sorts weakly; a measure that loads on the identity channel sorts sharply — which is why the knowledge interaction is the decisive test. What education's correction *consists of* — enlightenment, sociotropic values, cosmopolitan socialization — is contested in the immigration-attitudes literature (Hainmueller & Hopkins, 2014), and nothing here adjudicates it: education's role is calibrated on non-immigration claims (automation, AI, GM crops; §5), so whatever mixture schooling carries, it is the folk-corrective channel on every uncontested claim, and the two-channel result needs no more.

Table 58: GSS within-party education slopes on the immigration fact (2014–2024). Education corrects both parties — unlike ANES political knowledge, which moves only the out-group — but corrects Democrats about twice as steeply.

Party	No college	College	Education (log-odds)	n
Democrats	0.235	0.098	-1.037	2365
Republicans	0.423	0.320	-0.445	2001

A third capacity measure settles the channel assignment on content rather than outcome. The two-channel reading can be accused of circularity — education and knowledge assigned to $f(s)$ and $p(s)$ because of the very contrast they explain. General verbal ability (GSS `wordsum`, a ten-item vocabulary test) breaks the circle: it is cognitive capacity with zero political content, so its role is fixed by what it contains, in advance of any estimate — the model *requires* it to behave like education and *forbids* it from behaving like political knowledge. It obeys both (Table 59). Wordsum corrects the folk belief in both parties in both eras (Democrats −0.46 and −0.51, Republicans −0.29 in each era, all $p < .01$); the party×wordsum interaction is directionally positive but not significant in either era (+0.17, +0.22); and — the discriminating fact — it does *not* switch on across eras (change +0.05, $p = .78$), against political knowledge's +0.05 → +0.51 on the same fact. Sharpest of all: in the contested era the Republican *wordsum* slope is −0.29 ($p < .01$) where the Republican *knowledge* slope is +0.04 (n.s.). Capacity as such still carries the in-group toward the experts; only capacity to *locate the party line* leaves it anchored at the folk pole. Three measures, one rule: the channel is set by whether the measure contains the configuration, not by how much capacity it certifies.

Table 59: The third capacity measure: party×wordsum on the immigration fact, by era (GSS; survey-weighted logit, year FE within era, sandwich SEs; wordsum standardized within era). Verbal ability behaves like education — folk-corrective in both parties, weakly sorting, era-stable — and unlike political knowledge, which moves only the out-group and switches on at elite onset.

Era	n	party×wordsum	SE	p	Dem slope	SE	Rep slope	SE
pre-onset (1996, 2004)	1694	0.174	0.116	0.135	-0.456	0.076	-0.290	0.089
contested (2014, 2022, 2024)	2210	0.223	0.136	0.102	-0.508	0.098	-0.293	0.100

SI-D.5 Composition controls: race, nativity, and the CES check

Education is not the composition that matters most for immigration: the parties also diverged over these decades on race, ethnicity, and nativity, first-order predictors of immigration beliefs. Re-estimating the party×era interaction adding race, ethnicity, nativity, age, and region, each interacted with era (Table 60), the point estimate does not attenuate — it is, if anything, slightly larger (log-odds +0.27 → +0.37 with race/ethnicity/nativity×era, +0.33 with age and region). But because the GSS asks the nativity and ethnicity items of only a subsample, the analytic n falls by a third and the coefficient drops to marginal significance ($p \approx .07$–$.11$): its magnitude is undisturbed, but the test is underpowered there. The Cooperative Election Study supplies a properly powered version on the matched immigration *preference* at n ≈ 37,000 (Table 61): the party gap survives both ideology and the full demographic battery — the raw +1.64 log-odds falls to +0.75 controlling ideology and +0.91 with demographics added ($p < 10^{-150}$) — and the party×college interaction survives (+0.36 → +0.14, $p = .008$). The partisan structure is not reducible to ideological or demographic composition in the cross-section. One channel these controls do *not* close is over-time ideological self-selection — pro-immigration moderates re-sorting into the Democratic column, which demographic controls would miss; the within-person panel is the proper defense (§6); its item-breadth caveat is recorded once, in the claims map (Table 1).

Table 60: Party×era on the fact, adding race/ethnicity/nativity/age/region×era controls (GSS, survey-weighted logit). The coefficient is robust — even larger — to the demographic shifts that matter most for immigration; the wider p-values reflect the one-third sample loss from the subsampled nativity/ethnicity items, not attenuation.

Model	party×era (log-odds)	p	n
baseline	0.271	0.007	3077
+ education x era	0.262	0.010	3077
+ race, ethnicity, nativity x era	0.370	0.071	1979
+ age, region x era (full)	0.328	0.110	1976

Table 61: Better-powered composition test (CES 2022, n ≈ 37,000): the party gap on the matched immigration preference survives ideology and demographic (race/nativity/age) controls, and the party×college interaction survives too. Cross-sectional and on the preference — it addresses compositional confounding at a moment, not over-time re-sorting (the panel, §6, addresses that).

Quantity	estimate (log-odds)	p	n
party gap: baseline	1.638	0.000	37115
party gap: + ideology	0.747	0.000	37115
party gap: + ideology + race/nativity/age	0.913	0.000	37115
party x college: baseline	0.358	0.000	37115
party x college: + ideology + demographics	0.136	0.008	37115

SI-D.6 Fused, not converging: the split-ballot correction

The fact and the preference were always correlated, at roughly 0.4, and that coupling is essentially flat across the whole period among respondents asked both items (Figure 29). What changed is not that the two items grew more correlated but that the factual one acquired the partisan loading the preference had carried all along. This matters because an earlier reading of these data — a within-person fact–preference correlation rising from negative to positive — does not survive the survey's split-ballot structure: the two items were administered to disjoint samples in 2004, so no such correlation is even estimable that year, and the apparent negative value was an artifact of treating non-respondents as disbelievers. The corrected result is cleaner: a fact did not fuse with a preference, it *became* one, in the precise sense of acquiring its partisan signature. The correction is also the visible face of a general safeguard: the build now recomputes every quoted headline number from the committed aggregates and fails hard on any mismatch (SI-F, script 56), so an error of this class cannot recur silently. A scope note follows from the repeated-cross-section design: everything we observe is between-group — party *aggregates* moving across waves, not individuals crossing within their lifetimes — so in the cross-sections "capture" names a change in the partisan distribution of the belief, and the within-person dynamic requires the panel of §6.

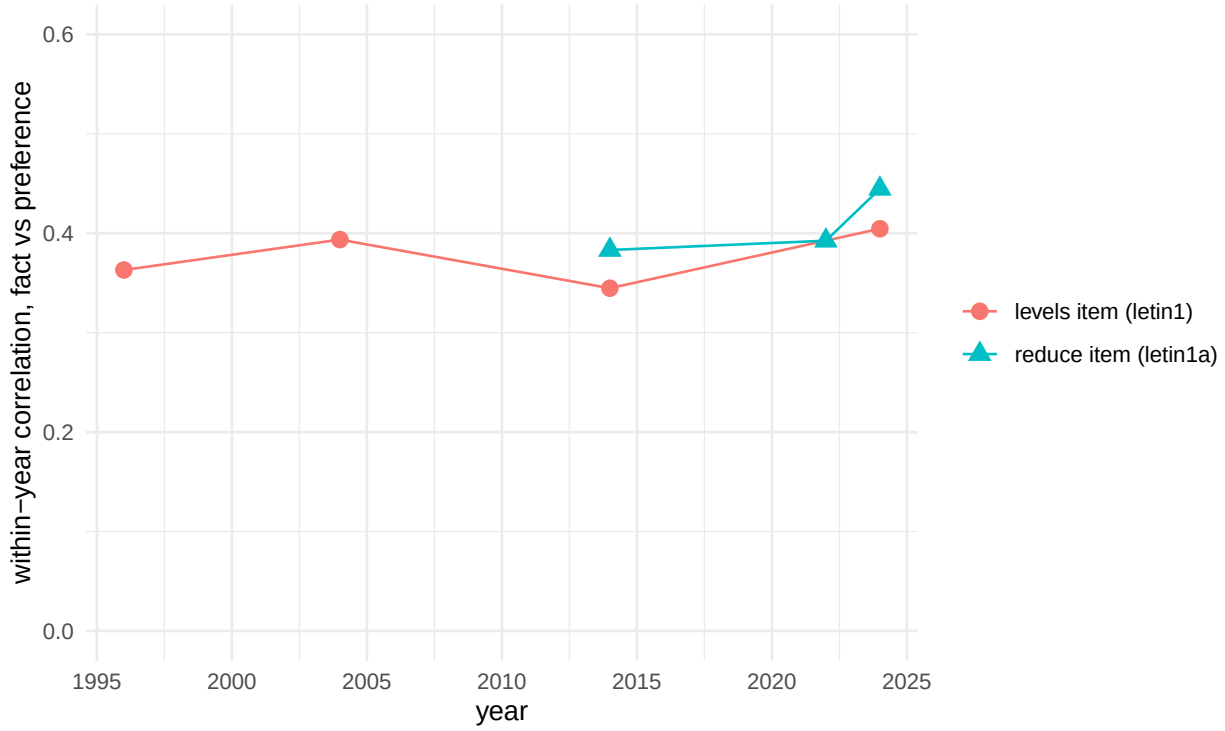


Figure 29: Fact and preference are stably coupled (true co-respondents), not converging.

SI-D.7 First returns on the forward prediction: the AI capture, within persons

The registered prediction E.1 — that AI-and-employment beliefs will partisanize as elite contestation of the issue rises — turns out to have a within-person test already running. The two Pew waves used as cross-sections in §6 share persistent panel identifiers: 535 respondents who were partisans in 2017 answered both the automation item (W27) and the AI-jobs item (W152), a seven-year window spanning exactly the transition the paper's gradient documents in cross-section — party structure absent on the machine-jobs belief in 2017 (odds ratio 1.04, Table 54), weakly present by 2024 (1.70). Within these same people, that transition is visible and *directional* (Table 62). In 2017 the linked panel's party gap on the automation belief is −1.2 points — indistinguishable from the full W27 partisan sample's +1.3, so panel retention did not select on the quantity of interest. By 2024 the same 535 people show a +12.0-point gap on the AI belief. The folk share rises in *both* parties — the ambient level moves, as the shared elite signal on AI turns anxious — but Republicans rise twenty-two points against the Democrats' nine: the partisan act is the differential, exactly as in the immigration decomposition, and the cross-lag isolates it. The cross-lag structure runs the way capture requires: 2017 party predicts the 2024 belief net of the 2017 belief (+0.121, $p = .002$; robust to either wave's weights, no weights, a college control, a logit form, and a change-score variant), while the reverse path — 2017 belief predicting 2024 party — is less than half the size and not significant (+0.056, $p = .08$; +0.046 scoring 2024 nonpartisans at one-half). The 2017-party × college interaction on the 2024 belief is +0.10 ($p = .21$): directionally the reception-first pattern, underpowered at this n, and reported as such.

Three caveats bound the reading, stated as plainly as the result. The panel is small (n = 535) and its retention is selective on education (college 26 versus 31 percent), which is why the college-controlled

fork is shown — and attrition is met with bounds, not only balance: under monotone selection, Lee trimming bounds (Lee, 2009) set-identify the within-panel 2024 gap at [+0.105, +0.291] (bootstrap lower limit +0.025; `R/77`, `atp_lee_bounds.csv`) — worst-case differential attrition can roughly halve or double the gap but cannot erase it. The item's referent shifts from automation to AI, so conditioning on the 2017 belief handles baseline levels, not construct identity — the estimand is new-belief formation on the arriving claim, predicted by prior party, which is what capture *means* at onset. And the elite-side dating of AI's onset is not yet in hand (the platform series runs through 2024; E.9's index will carry it), so this is the mass side of the prediction, with the elite clock still to be read. With those bounds: the paper's named next case has begun to move, within persons, in the predicted direction, with the predicted asymmetry, before publication — the forward prediction's first returns rather than its test.

Table 62: First returns on E.1: the linked Pew ATP panel (2017 automation → 2024 AI, 535 respondents partisan in 2017). Party leads the arriving belief within persons (+0.121, p = .002, all forks); the reverse path is less than half the size; the within-panel party gap moves from −1 to +12 points.

Quantity	Estimate	SE	p	n
2017 baseline, full W27 partisans: folk share / Rep share / college / folk gap (R−D) (W27 weights)	0.639 / 0.438 / 0.313 / +0.013	—	—	4033
2017 baseline, linked panel: folk share / Rep share / college / folk gap (R−D) (W27 weights)	0.665 / 0.448 / 0.256 / +0.001	—	—	535
forward: 2017 party → 2024 AI belief, net of 2017 belief (LPM, W152 weights)	+0.121	0.039	.002	535
…fork: W27 weights	+0.102	0.047	.031	535
…fork: unweighted	+0.138	0.035	.0001	535
…fork: + 2017 college control	+0.121	0.039	.002	535
…fork: survey-weighted logit (log-odds)	+0.755	0.249	.003	535
…fork: change-score Δbelief ~ 2017 party	+0.132	0.056	.0190	535
reverse: 2017 belief → 2024 party, net of 2017 party (partisans both waves)	+0.056	0.032	.084	526
…fork: 2024 nonpartisans scored 0.5	+0.046	0.032	.155	535
within-panel folk levels 2017 → 2024, Democrats (W152 weights)	0.636 → 0.725	—	—	NA
within-panel folk levels 2017 → 2024, Republicans (W152 weights)	0.624 → 0.845	—	—	NA
within-panel party gap 2017 → 2024 (W152 weights)	−0.012 → +0.120	—	—	NA
reception check: 2017 party × college on 2024 belief	+0.097	0.078	.21 (underpowered)	535

SI-E: Registered empirical commitments

Six model-level disconfirmers precede the forward registrations, each with its status. *Symmetric sophistication slopes while $p_R \approx f$* — both parties moving with sophistication during emergence would fail the one-sidedness derivation (*at risk 1996–2020; held*: the amplification is borne by the

out-group alone, §7). *The gap emerging first among the least informed* — deficit-style diffusion from the bottom would contradict $g(T, s)$ (*held*: it emerged first among the most informed, §6–7). *A premium failing to track the elite configuration* — constancy across configurations would return it to a standing trait (*at risk in thirteen years; held at power*: 13/13 reversals, §7). *The converse case failing in the mirrored direction* — Democrats converging on a fact whose expert pole is Republican-congenial would break content-neutrality (*held on nuclear power*, GM crops the uncontested placebo, §7). *A knowledge-graded rebound, or persistence after cue withdrawal* (*standing*, E.2). *AI failing to partisanize as T* rises (*standing*, E.1).

The paper's remaining open questions are stated here as directional predictions, registered in advance on the Open Science Framework with a third-party timestamp (anonymized view-only registration: [OSF LINK — insert before submission]; the registration text is also committed verbatim in the replication archive) so that the theory is tested prospectively rather than rationalized in hindsight. Each carries its disconfirmer on its face.

E.1 The next capture (AI and employment). AI-and-employment beliefs, weakly sorted in 2024 and education-*corrigible* (the party×education interaction is currently negative — schooling still corrects the folk error in both parties; Table 42), will partisanize as elite contestation of the issue rises: the mass gap will open on the elite clock, and the sophistication premium will switch on from its current null, in the direction of whichever party's pole the folk default comes to oppose. *First returns, in hand:* the linked ATP panel (SI-D.7) shows the within-person side beginning — party leading the arriving belief (+0.121, $p = .002$, reverse path half the size), a −1-to-+12-point gap opening inside the same 535 people, Republicans toward the folk pole. What remains registered is the elite-clock leg (dating AI's onset from the platform/speech series as it forms) and the premium leg (a knowledge battery beside a fielded AI item). *Disconfirmed if* elite contestation of the issue rises while the mass gap or the premium stays flat — capacity failing to enact a supplied elite cue.

E.1, priced. Directional registration is cheap, so the prediction is committed with magnitudes and dates (Table 63; `R/74`). From the 2024 baseline (a 9.7-point full-sample gap) and the event study's estimated capture path translated to the AI item's own scale, the model forecasts a gap of roughly **25 points [21, 30] within four years of a platform-dated AI-and-labor onset, and 37 [32, 43] at eight-plus** — bands the two realized capture paths bracket (immigration's binary gap stood at 25.3 four years after its onset; trade's ran 13 to 42 across eight). The forecast binds only upon a dated onset by the same instrument that dated immigration and trade; absent onset, the gap is predicted to stay within ±8 points of baseline. The falsifiers cut both ways: an onset with the gap still inside the no-onset band at onset+4 disconfirms capture, and a gap exceeding the band *without* an onset disconfirms the elite clock. Verification venues are named — the recurring Pew AI waves, the GSS from 2026, the ANES in 2028 — so the theory can be marked to market on a calendar it does not control.

Table 63: E.1, priced: the dated quantitative forecast for the AI-jobs party gap, committed in advance — baseline, conditioning onset, point forecasts with intervals, reference curves, the no-onset band, and two-sided falsifiers.

Quantity	Value
baseline: 2024 full-sample AI-jobs gap (Rep − Dem, W152)	9.7 pts (Rep 0.813, Dem 0.716; n = 4,420)

Table 63: E.1, priced: the dated quantitative forecast for the AI-jobs party gap, committed in advance — baseline, conditioning onset, point forecasts with intervals, reference curves, the no-onset band, and two-sided falsifiers.

Quantity	Value
onset condition (the forecast binds only if this occurs)	a platform-dated AI-and-labor onset: opposed net platform positions (MARPOR-successor / E.9 index), the instrument that dated immigration (2012) and trade (2016)
forecast, four years after a dated onset	gap $\approx$ 25 pts [21, 30]
forecast, eight-plus years after a dated onset	gap $\approx$ 37 pts [32, 43]
reference curves (realized capture paths)	immigration binary gap at onset+4: 25.3 (ANES 2016); trade preference gap onset to onset+8: 13 $\rightarrow$ 42
no-onset counterfactual	gap stays within $\pm$8 pts of 9.7 (control issues' within-issue variation)
premium leg	currently education-corrigible (party$\times$education -0.35); the sophistication premium switches on to positive within one electoral cycle of onset
verification venues	Pew ATP AI waves (recurring); GSS 2026+; ANES 2028
falsifier (capture side)	a dated onset with the gap still inside the no-onset band at onset+4
falsifier (elite-clock side)	the gap exceeding the band with NO dated onset — mass-first capture

The watchlist. The same commitments generalize to a short ranked list of currently uncapturable-looking claims the framework marks as next, each with its rubric-assigned folk pole (SI-A.5), its verifiability score (SI-A.4), and the condition under which it sorts — so the theory publishes not one falsification opportunity but a slate. **AI and employment** (CVI 11; folk pole: machines destroy work) — priced above. **Minimum-wage employment effects** and **rent control** (CVI 10–12; folk poles Democratic-congenial, expert poles Republican-congenial — the E.3 geometry): the model predicts both sort *only if* an elite party comes to own a pole, with educated *Republicans* as the converging out-group — the mirrored direction, so capture here would confirm content-neutrality a second time, and symmetric sorting without elite ownership would disconfirm. **ACA-type enrollment claims** (CVI 5–8) are the slate's control row: verifiable within a survey cycle, they are predicted *not* to open a knowledge-graded gap even under contestation (E.5) — the one cell where the framework forbids capture outright. Each row names its disconfirmer; none requires the paper to be right about which party engages first, only that sorting wait for ownership and skip the checkable.

E.2 The rebound under cue withdrawal. The 2024 Republican rebound is read as elite-cue-driven ($p_R > f$ under supplied cues), not capacity-driven: it should *persist only while the cue is supplied*, *reverse* if elite supply withdraws, and *remain flat across knowledge* throughout. *Disconfirmed if* the rebound becomes knowledge-graded, or persists after elite cue withdrawal — either would also revive the expressive and sincere-capture questions the calibration bound (SI-B) closes on current evidence. A third rival — lived exposure to the 2022–2024 migration surge, which knowledge-flatness cannot exclude — is already tested and fails: the rebound on the fact is uniform across state-level exposure change (SI-C).

E.3 The market-fact converse case. The model's mirrored-direction prediction is confirmed on nuclear-power risk (§7, SI-D.2b); what remains registered is its specifically *market-fact* form:

on a contested economic claim whose folk pole is Democratic-congenial and whose expert pole is Republican-congenial — minimum-wage or rent-control employment effects are the canonical candidates — *Republicans* should be the converging out-group and the sophistication premium should run in the mirrored direction. No off-the-shelf instrument currently fields such a fact with a sophistication measure (we searched the GSS, the ANES files, and the 2022/2024 CES); the prediction stands ready for the first survey that does. *Disconfirmed if* Democrats converge, or the premium runs unmirrored.

E.4 The experiential verifiability half-life. As AI deployment raises *experiential* v on the fixed employment claim — the outcome becoming observable in respondents' own labor markets — the gap should shrink on the experiential margin even where the issue stays elite-contested, separating the two verifiability margins §3 distinguishes. *Disconfirmed if* rising experiential verifiability leaves the gap untouched while elite contestation is held by the elite series.

E.5 The missing verifiability cell. The claim-level margin (§7) estimates the contestation leg of the gap equation across ten fielded claims but cannot identify the verifiability leg, because no claim scoring below 6 on the CVI has a public twin. The registered test: field matched claim pairs under elite contestation — one member verifiable within the survey cycle (CVI ≤ 5, of the ACA-enrollment kind), one unverifiable (CVI ≥ 10) — with party identification and a knowledge battery. The model predicts the partisan gap and its sophistication premium open on the unverifiable member only. *Disconfirmed if* gaps of equal size open on both — capture indifferent to checkability would break the gap equation's $\alpha(v)$ term, not just prune it.

The five above are prospective *observational* predictions. Three *experimental* designs follow, each written to be preregistered and fielded, because the observational record has a boundary the paper is explicit about: the manufactured-premium cross-derivative is identified by triangulation across the perception sign-reversal and the switch-on, not as a single estimated quantity on one claim — and the subnational attempt to close that gap fails for want of a division-level knowledge measure (SI-B). Random assignment closes it. The designs name their treatments, estimands, and disconfirmers; power targets and the full protocol are in the OSF registration.

E.6 The cross-derivative, experimentally: elite contestation × sophistication. This is the within-claim causal version of the paper's central mechanism. Respondents are shown a *novel, plausible, truth-apt* economic claim on which they can hold no prior — a low-salience or lightly-synthetic quantitative claim about the employment effect of an obscure policy — and randomly assigned to a *low-contestation* framing (the claim is described as under nonpartisan expert study, no party cue) or a *high-contestation* framing (the parties are described as sharply opposed, with the party-position pairing embedded so the folk-congenial pole is owned by one side). Political sophistication is measured pre-treatment with a standard knowledge battery; the outcome is stated belief in the claim. The estimand is the triple interaction *party* × *sophistication* × *contestation condition*. The model predicts it is *positive*: the party×sophistication interaction is null under low-contestation framing and switches on under high-contestation framing — $\partial^2 b / \partial s \, \partial T > 0$ recovered causally, on a single claim, with the secular-trend and selective-exposure confounds eliminated by random assignment and the absence of priors. Druckman, Peterson and Slothuus (2013) established the *level* leg experimentally (elite polarization raises cue reliance); the sophistication-interacted leg — the manufacture of the premium itself — is the novel quantity, and no prior experiment has isolated it. The design is powered for a triple interaction (a materially larger sample than a main-effect study; the OSF registration reports the simulation-based target), and analysis is preregistered as the triple-interaction coefficient with sophistication entered continuously and, secondarily, by tercile. *Disconfirmed if* the party×sophistication interaction is present under low-contestation

framing, or fails to switch on under high — either would show the premium is a standing trait, not manufactured.

E.7 Cognition versus exposure: a motivated-reasoning discriminant on the verbatim fact. The observational design narrows the divergent-exposure rival (RV $\approx 5\%$, benchmarked at eleven times party×racial-resentment; content-of-diet mediation $\approx 25\%$) but cannot separate motivated cognition from configuration-sorted priors. Thaler's Fake-News-Effect design (2024) identifies motivated reasoning from Bayesian updating on any numerical-answer question by eliciting the *median* of the belief distribution and comparing responses to congenial versus uncongenial signals — and validates on immigration among other issues. We adapt it to the verbatim immigration-employment quantity: present the claim, deliver a randomized signal, and test whether the *sophisticated* distort their assessment of the signal's veracity in the party-congenial direction (motivated cognition) rather than updating as Bayesians who merely began from different priors (exposure). The estimand is the sophistication × signal-congeniality interaction in the distortion of source trust. This is the current experimental standard for exactly the cognition-versus-Bayesian separation the paper registers as secondary; landing it would let the mechanism claim be made outright rather than narrowed. *Disconfirmed if* the sophisticated update symmetrically on congenial and uncongenial signals — the Bayesian-with-different-priors prediction. The observational counterpart is now buildable with validated tools: LLM measurement of political text matches expert and crowd coders on issue positions (Benoit et al., 2026; Le Mens & Gallego, 2025; Wu et al., 2023), so each ANES respondent's *named* outlets can be scored for the immigration frame content they actually carried during the field window, replacing outlet identity with measured message content in the party × content control — with the annotation layer run through design-based supervised learning (Egami et al., 2023), which corrects the bias that raw LLM labels induce in downstream estimates even at high accuracy, on an open-weight, political-domain-validated classifier (Burnham et al., 2025) (protocol in the replication repository, `tools/content_of_diet_protocol.md`); the knowledge interaction surviving *measured content* would narrow the exposure rival past any prior observational design.

E.8 Measuring the expressive share directly, and a real-stakes behavioral separation. The sincerity bound (SI-B) transports the expressive share λ from incentive experiments on *other* items. Two designs measure it for *this* one. First, a Bullock-style accuracy-incentive treatment (2015; 2015) and, in parallel, a Bayesian Truth Serum (2004) or list-experiment elicitation on the verbatim `immjobs` item, replacing the transportability assumption with a direct estimate of λ for the immigration fact. Second, a real-stakes separation of belief from cheap talk: an incentivized choice in which respondents can earn a small payment by matching the actual published native-employment statistic, or a costly-information-acquisition choice, which no vote report can proxy. Together these upgrade the behavioral evidence from "the belief predicts another self-report" to "the belief predicts a real-stakes act," and convert the calibration bound's assumed λ into a measured one. *Disconfirmed if* the partisan gap collapses to within the 1996 ceiling once the item is incentivized on its own respondents — expression, not belief.

E.9 One elite-contestation index, continuous, through the present. The paper operationalizes T five ways, each attached to one result; five measures can be read as five researcher choices. The committed upgrade replaces them with a single continuous, position-based index of elite issue divergence, scored directly from the Congressional Record — including 2021–2025, closing the corpus's 2020 end — with the now-validated LLM toolkit (Benoit et al., 2026; Le Mens & Gallego, 2025; Wu et al., 2023) — annotations passed through design-based supervised learning (Egami et al., 2023) so that downstream estimates and standard errors remain valid under arbitrary annotation error, with an open-weight political-NLI classifier (Burnham et al., 2025) as the scoring engine and a hand-coded

random subsample as the correction anchor — validated in the overlapping window against both the Manifesto platform series and the Card et al. tone series, and decomposable by state delegation to power the subnational designs (protocol: `tools/unified_contestation_index_protocol.md`). *Disconfirmed if* the index diverges from the two existing series where they overlap — the measures would not be measuring one construct — or if results held under the five proxies fail under the one index.

These nine — five observational predictions, three experimental designs, and one measurement commitment — with the falsifiers stated throughout the text (a knowledge-graded rebound; premium movement under measured content shifts; symmetric sophistication slopes while $p_R \approx f$; the gap emerging first among the least informed), are the claims on which the framework has agreed in advance to be judged. The experimental trio is written so that E.6 identifies the cross-derivative causally where the observational design can only triangulate it, E.7 converts the mechanism from narrowed to identified, and E.8 replaces a transported sincerity bound with a measured one — the three points at which a referee can most fairly say the observational paper stops.

SI-F: Reproducibility and data availability

The analysis is fully scripted and version-pinned. During review the repository is available as an anonymized view-only archive; the public repository is linked upon publication. All analysis code, the `renv` lockfile, and the non-identifiable derived aggregates will be deposited in the APSR Dataverse on acceptance. Raw microdata are not redistributed and are available from the original providers under their own terms: GSS (NORC), ANES, Pew ATP, CES (Harvard Dataverse), VOTER (Democracy Fund), ISSP (GESIS), the Manifesto Project, and the Card et al. speech corpus; none requires more than free registration. From the repository root, `source("setup.R")` restores the package versions recorded in `renv.lock`; `Rscript run_all.R` then runs the numbered scripts `R/01`–`R/79`, writing derived aggregates to `data/processed/`, closing with `R/56` — a hard-failure consistency block that recomputes every headline number quoted in the abstract, Figure 1, and the claims map from the committed aggregates and verifies that all cross-references in this manuscript's source resolve — followed by the figure script (`R/59` and `R/60` stop with instructions if their raw inputs are not yet in place). A fixed seed (`20260616`) is set in `run_all.R`. This manuscript renders its tables and figures directly from the committed aggregates in `data/processed/`, so the document and the analysis cannot diverge; the aggregates are non-identifiable, allowing every exhibit to be rebuilt without the restricted microdata. All estimation was performed in R 4.6.0; design-based inference uses the `survey` package (Lumley, 2004), with `tidyverse` for data manipulation and `haven` for survey-file import. Exact package versions are pinned in `renv.lock`; the public repository (with the commit hash recorded in the released version) holds the scripts, the lockfile, the committed aggregates, and this manuscript source. A clean run reproduces the following.

Script	Object	Expected value
01	cleaned expert panel	753 economist × sub-question rows
02	public–expert gap; micro model	AI gap +0.45; college OR 0.82
03	Claim Verifiability Index	range 5–12, mean 9.5; Krippendorff's $\alpha = 0.88$

Script	Object	Expected value
04	public–expert gaps	AI +0.45, trade +0.36; immigration expert benchmark = NA (no direct IGM item; direction anchored on NAS)
05	contestation gradient	party OR 1.04 → 1.70 → 3.79
06	composition control	party×era OR 3.38; education×era null
07	precedence	lag-10 r = 0.92; Granger β = 0.72, p = .063
08	partisanization (capstone)	gap −0.03 → +0.33; trend r = .92, p = .03
09	mechanism (GSS)	gap emerges among graduates; 98% Democratic convergence
10	mechanism (ANES)	party×knowledge +0.51, p < .001
11	robustness	multiverse 100%/100%; knowledge survives education+ideology controls
12	convergence + behavior	fact gap 9% → 72% of preference; fact predicts vote (OR 7.6)
13	conditional law + descriptives	amplification: immigration + / automation null / AI −
14	sorting benchmark	immigration fact gap 33 pts ≈ guns/gay rights/death penalty
15	generality (evolution)	disbelief gap +14 → +24 pts; party×education +0.59, p = .001
16	mechanism at scale (CES)	party×education +0.38 to +0.80 on immigration restriction, n ≈ 38k
17	within-person capture (VOTER panel)	drain-belief gap +47 → +62 pts; RI-CLPM within-person party→belief 0.40 > belief→party 0.12 (diff 0.28, boot CI [0.19, 0.38])
18	out-of-sample mechanism (ANES CDF)	party×info on take-jobs fact: +0.39/+0.40/+0.52 (2008/12/16), null 2004
19	content neutrality (ANES trade)	protectionist-default party flips: Rep−Dem gap −0.08 (pre-2012) → +0.42 (2024)
20	cross-national scope (ISSP 2013)	right-more-folk gap in 77% of 26 countries; corr(ENP, gap) = +0.16 (n.s.)

Script	Object	Expected value
21	structural signature (ANES CDF + elite series)	gap & sophistication premium both ~0 in 2004, co-rise with T; corr +0.86/+0.91
22	CES composition test (preference, n ≈ 37k)	party gap +1.64 → +0.91 net of ideology + race/nativity/age ($p < .001$); party×college +0.36 → +0.14
23	cross-issue differential timing (ANES CDF)	trade premium −0.58/−0.39 (2008/12) → +0.18/+0.41 (2020/24); immigration +0.62/+0.09 same years (opposite signs)
24	over-time multilevel inference (GSS+ANES)	party×year +0.86 log-odds/decade, cluster-robust $p < .0001$ (11 survey-year clusters); 14/14 design forks positive
25	mode-sensitivity / social desirability (GSS)	Dem folk lower on web than in-person (self-admin coef −0.70, $p < .001$); gap mode-invariant (33 vs 34 pts, 2024)
26	VOTER reliability bound (2-D grid)	party→belief lead reverses only at party reliability $\gtrsim$0.90 AND belief $\lesssim$0.70 (near-stable-belief corner the documented change refutes); else identified
27	perception premium sign-reversal (ANES CDF)	signed party×info premium on econ perception sign-matches incumbency 11/11; premium ~ incumbency +0.19, $p = .002$, $R^2 = .66$
28	small-cluster inference (wild bootstrap)	party×year slope wild cluster bootstrap-t $p = .005$ Rademacher / .003 Webb (11 clusters); permutation $p = .026$
29	VOTER reliability 2-D grid	lead survives all cells except party rel $\geq$0.90 × belief rel $\leq$0.70 (conditional identification)

Script	Object	Expected value
30	measured elite-divergence instrument (MARPOR + ANES CDF)	mass gap ~ signed platform divergence across 16 issue-years, r = 0.63, within-issue permutation p = .023; pooled ∂Gap/∂T = +0.10; trade gap −0.07 (pre-2016) → +0.35 (2016+)
31	RV benchmark (ANES 2024, Cinelli–Hazlett)	party×knowledge RV = 4.9%; killer confounder ≈ 11× party×racial-resentment (a strong, well-measured confounder; ≈60× party×media-trust)
32	pooled switch-on, 3 instruments/2 houses (GSS immjobs+immameco+immecon, ANES take-jobs)	party×year +0.71 log-odds/decade, 16 clusters, N=33,031; wild bootstrap p = .001 (Rademacher) / .001 (Webb); per-instrument +0.12/+0.09/+0.21 all sig
33	staggered event study (7 issues, ANES CDF, MARPOR-dated onsets)	flat pre-trend (+0.07, p = .40); post-onset +0.25 (0–7y) / +0.63 (≥8y), DiD +0.34 p < .001, net of issue + calendar-year FE
34	RV re-benchmark + exposure mediation (ANES 2024)	RV=4.9%; killer confounder ≈11× party×racial-resentment; pro-attitudinal outlet exposure mediates ≈25%, direct share ≈75% (sig)
35	latent CLPM, multi-indicator partisanship (VOTER)	party→belief +0.55 vs +0.08 (2-indicator, 2011→2020, CFI .98/TLI .94/RMSEA .13); +0.66 vs +0.01 (4-indicator); attrition party×belief null (p = .43)
36	age-period-cohort decomposition (GSS immjobs, Democrats)	period share of 1996–2024 decline = 98%; Dem period slope −0.134→−0.138/decade net of cohort+age; period R^2 2× cohort
37	premium-to-gap ratio test of strict separability	ratio not constant: falls 19.5 → 4.4 → 2.2 (2008–2016) — premium leads the gap; the reception-integrated $g(T, s)$ reading, reported rather than left as algebra

Script	Object	Expected value
38	cross-national country panel (ISSP 2003+2013, 17 countries)	mean L–R gap on fact 0.041→0.078; salient-RR countries widen 2× (DiD +0.025, n.s. at 2 waves); directional, underpowered
39	fact-vs-preference behavioral discriminant (ANES 2024)	fact more partisan than matched preference (+0.28, p < .001) and more knowledge-graded (+0.12, p = .03); immameco co-move r=.95
40	Republican elite tone trend (Card corpus, 2015–2020)	Republican anti-immigration tone flat-to-declining (~0.53→0.39, −0.03/yr); does not presage 2024 rebound; registered forward prediction stated
41	premium extension 2020/2024, behavioral knowledge (ANES CDF + 2024 study)	2020 −0.040/SD (R incumbent), 2024 +0.112/SD (D incumbent), both as predicted; series 13/13 (exact binomial p < .001); per-SD premium ~ incumbency +0.054, p < .001, R^2 = .69
42	converse-case battery (GSS environment rotations)	nuclear party×college +0.10 (p = .0002), educ slope Rep +0.16 vs Dem +0.06 (mirrored asymmetry); GM crops +0.01 (p = .71, both slopes positive — folk placebo); climate −0.15 (p < .001) — converging party flips with the folk-pole assignment
43	exact randomization inference, premium series	all 1,716 incumbency assignments enumerated: observed coefficient and 13/13 sign-match both the most extreme possible (exact p = .0006); binomial sign test p = .0002

Script	Object	Expected value
44	probability-scale AMEs for headline interactions	knowledge tercile gaps 0.34→0.42→0.46, second diff +0.126 [0.055, 0.196]; era widening +0.17/decade [0.11, 0.24]; GSS education second diff +0.03 [−0.06, +0.11] (scale-sensitive, reported plainly)
45	event-study robustness (CS + placebo onsets)	Callaway–Sant'Anna-style ATT +0.44 (cohorts +0.42/+0.46) > TWFE +0.34 (conservative); placebo-onset RI 5,000 draws: observed past 99th pctile, one-sided p = .003
46	independents placebo (GSS immjobs)	pure independents decline in lockstep with Democrats (−0.136 vs −0.134/decade, diff p = .88); Republicans hold flat against the common baseline (+0.123/decade net, p < .001) — referent drift cannot produce the party-differential trend
47	fundamentals anchor (perception series)	incumbency unchanged with realized fundamentals in the model: gap +0.80 → +0.76 (p = .0001), premium +0.054 → +0.050 (p = .004); GDP growth and unemployment change never significant — bias against the measured economy, not differential accuracy
48	within-cohort decline, assumption-free (GSS immjobs, Democrats)	every ten-year birth cohort declines within itself, −24 to −40 pts first-to-last observation — generational replacement excluded without APC assumptions

Script	Object	Expected value
49	expressive-responding calibration bound (published incentive effects)	at the top of the accuracy-incentive range (50–55% expressive) the sincere residual of the 33-pt 2024 gap is 15–17 pts (≈5× the 1996 estimate); paid-DK extreme (80%) leaves 6.6 pts; expressive accounts cannot produce the trend, the monotone knowledge gradient, or nine-year within-person durability
50	content-of-diet exposure control (ANES 2024)	party×knowledge +0.526 → +0.445 ($p < .001$, n = 4,402) net of party×own-party-leaning named-outlet exposure; narrows ~15%, does not close (consistent with the RV benchmark)
51	claim-level verifiability margin (44 claim-years, 10 fielded claims)	contested mean gap 0.63 SD vs 0.13 uncontested (WLS +0.81, claim-clustered $p = .001$; RI over contestation schedules $p = .016$); CVI×contested unidentified in the fielded range (−0.04, RI $p = .38$; fielded-wording recode +0.09 n.s.)
52	subnational elite clock (63 division-year-item cells, 9 divisions)	∂gap/∂D_rt = +0.66 (division-clustered $p = .067$; RI within year $p = .016$); sign-stable across item, cell-size, and pre-2020 forks
53	expressive share as sensitivity parameter (λ bands)	2024 sincere floor 5.6 pts > 1996 ceiling 3.4 pts — bands disjoint at every $\lambda \leq .8$; switch-on identified under adversarial year-varying paths
54	knowledge battery party-DIF audit + rescoring (ANES 2024)	only House-control item crosses the ETS boundary (−0.45, direction *against* flattery); interaction scoring-invariant +0.51 / +0.52 IRT / +0.30 civics-only / +0.46 pruned / +0.52 net of strength interactions

Script	Object	Expected value
55	event-time coefficient plot + leave-one-control-out	leads 0.14/0.05/0.01 (CIs cover zero); +0.38 at +4y, +0.66 at +8y; LOO DiD spans +0.28 to +0.37, all $p < .004$
56	consistency assertions (hard build failure)	20+ headline-number asserts + full crossref audit; ALL PASS on the committed aggregates
57	subnational triple (cross-derivative vs local clocks)	bounded non-test: −0.04 (RI $p = .48$), MDE 0.11 ≈ 4× any plausible premium modulation; educ×D slopes near-identical by party (+0.030/+0.022)
58	2024 rebound: cue vs lived exposure	Republican rebound on the fact uniform across state foreign-born exposure change (+0.004, RI $p = .88$); one of two CES preference items shows a small gradient (+0.023)
59	ML-SEM dynamic panel, third estimator family (VOTER)	pending re-provision of the raw panel file; salvage dry run: both cross-lags attenuate to zero under strict FE (bounded non-test; no reverse-path support)
60	close-elections instrument for the subnational clock	pending the MEDSL House 1976–2022 returns file; script run-ready (2SLS, first-stage F, Anderson–Rubin set)
61	independents knowledge gradient (ANES 2024)	knowledge slope on the fact: Dem −0.47 ($p < .001$), pure independents −0.08 (n.s., n = 300), Rep +0.04 (n.s.); Dem−Ind slope difference +0.40 ($p \approx .02$), Rep−Ind +0.11 (n.s.)

Script	Object	Expected value
62	equivalence bounds for the load-bearing nulls	2004 premium TOST vs smallest post-onset estimate p = .012; post-minus-2004 change +0.36, one-sided p = .019; GM placebo TOST p = .005; independents slope TOST p = .004; education×era TOST p < .001; event-study leads TOST p ≤ .010; 2004 mass gap pooled TOST p = .003; MDEs reported for every cell
63	premium reversal: block-structure inference	year-level exact p = .0006 (1,716); presidency-block most extreme of 70 (sign p = .014); party-run-block most extreme of 35 (one-sided p = .029, the attainable minimum); ties forks 12/12 and 10/10; drop-2020 coef +0.054, 12/12, exact p = .002
64	wordsum channel test (GSS)	party×wordsum +0.17 pre / +0.22 contested (both n.s.; era change p = .78 — no switch-on) while wordsum corrects both parties in both eras (Dem −0.46/−0.51, Rep −0.29/−0.29, all p < .01); contested-era Republican wordsum slope −0.29 vs knowledge slope +0.04
65	folk-pole rubric validation	disengaged (pure-independent, earliest-wave) baselines deviate from the expert benchmark in the rubric's direction on 7/7 claims; outright folk-side majorities on 5/7 (immigration 53%, nuclear 86%, GM 70%, AI 77%, automation 65%); climate directional (48% below the expert pole, modal cell at midpoint), trade vs-experts-only (39% agree against an expert pole at zero)

Script	Object	Expected value
66	verbatim-fact panel (ANES re-interviews)	run-ready (stops with instructions): 2016→2020 and 2020→2024 cross-lags + three-wave RI-CLPM on the verbatim take-jobs item; awaits the 2016/2020 Time Series files (free registration); disconfirmer stated — a fact→party lead on the verbatim item overturns §6.5
67	event-study onset/outcome forks	onset shifts −4..+4 bound the DiD at +0.20..+0.44; speech-clock dating of immigration gives +0.28 (p = .017); signed-gap outcome +0.57 (p = .001); control headroom 0.3–1.2 item-SD vs treated pre-onset 0.14
68	two moments of the elite signal	independent baseline ~ elite mean tone r = −0.86; party gap ~ divergence r = +0.78; elite moments co-move at five waves, so a separated assignment awaits the unified index (E.9)
69	individual-level accuracy recast	pooled rep×knowledge×incumbency premium +0.056/SD (year-clustered p = .0005, n = 29,238); measured error on 7 unambiguous years: knowledge slope −0.042 (congenial) vs −0.004 (uncongenial, n.s.), difference p = .007 — knowledge corrects only party-congenial error
70	ATP panel: first returns on E.1	linked W27→W152 panel (n = 535 2017-partisans): party→arriving AI belief +0.121 (p = .002, all forks); reverse +0.056 (p = .08); within-panel gap −1.2 → +12.0 pts; balance on 2017 folk gap (+0.001 vs +0.013 full sample)

Script	Object	Expected value
71	personal-finances premium (VCF0880)	incumbency-signed 12/13 (miss 1992 at +0.0006); slope +0.029 (p = .001, R^2 = .64); pooled 3-way +0.032 (p = .0001); national:personal ratio 1.8; same-respondent quadruple +0.017 (p = .21, bounded, not separable)
72	ISSP premium bounded non-test	country-wave premiums median SE .04; US positive control +0.05/+0.08 ≈ domestic education second difference; RR moderation of Δpremium −0.010 (p = .80); heterogeneity −0.28..+0.13 — instruments too blunt for the second-order quantity
73	entry-cohort imprinting	gap at first observation by socialization cohort: pre-1996 entrants −2 pts, 1996–2011 +7, 2012+ entrants +25 (sorted on arrival); 2004 young already +13 while 30+ still −2; pooled entry-window contrasts +3.4 (±9.9) vs +24.6 (±7.6) — conversion and configured entry compound; asymmetric-decay prediction registered with E.2
74	E.1 priced (AI forecast)	baseline 9.7 pts; forecast 25 [21, 30] at onset+4 and 37 [32, 43] at onset+8, conditional on a platform-dated onset; no-onset band ±8; two-sided falsifiers; venues Pew/GSS 2026+/ANES 2028
75	the gap equation, estimated	perception scalar rotation: c = 0.046, rotation exact but constancy rejected (chi2(12) = 67.3); immigration one-path form rejected with either elite series (chi2(5) = 12.7 / 10.4; 2008 premium at platform-G = 0 — the premium precedes the platform clock); theta unidentified [0.05, 3.0]; structural bound tabulated

Script	Object	Expected value
76	RR sensitivity + pre-test power + triad	breakdown $\bar{M}^*$: DiD 1.23/0.56, +4 2.10/0.96, +8 3.98/1.81 (point/envelope $\bar{B}$); = 0.091, DiD-fabricating slope detected with ≈100% power, max undetected bias +0.21; BJS +0.384 (jk se .103) between TWFE +0.335 and CS +0.437
77	Lee bounds (ATP attrition)	within-panel 2024 gap set-identified [+0.105, +0.291] under monotone selection; bootstrap lower limit +0.025 — attrition can halve or double the first-returns gap, not erase it
78	premium, functional-form-free	honest-forest BLP: knowledge +0.052/SD (p = .042), education −0.015 (multi-BLP); AUTOC knowledge +0.044 (p = .050) vs education −0.013; clipping 54.6% stated; overlap-clean fork attenuated (+0.027 n.s.) — form triangulated, logit remains headline
79	continuous-time SEM (VOTER)	run-ready: drift-matrix estimation removes the equal-cross-lag constraint on 5/2/2-year gaps; reading rule pre-stated; pre-registered for the verbatim panel