Political Autoimmunity:
Adverse-Interest Voting and the Functions That Decide It

PIATRA . INSTITUTE

June 2026

Abstract

To say a group voted against its interests is to run a calculation and hide most of its inputs. The claim presupposes one interest where a voter has several, reads a candidate's likely harm to a group as the whole of the ledger while ignoring whatever the candidate offers, treats a knowable harm and an unforeseeable one as the same act, and never says whether it is counting voters or counting votes. This paper turns the accusation into a measurement model that exposes each of those inputs and prices the verdict's dependence on them. Adverse policy risk for a group, candidate, and policy domain factors as a product of exposure, institutional dependence, candidate hostility, implementation probability, and magnitude; a foreseeability gate and a salience gate separate the risk a voter could know and weigh from raw exposure; and net alignment under a given definition of interest is the candidate's weighted protective benefit minus the weighted adverse risk, so that gross risk and net alignment are different quantities whose difference is the protective term the folk accusation drops. The model is run on three synthetic but anchored 2024 cases, support for Trump among LGBTQ, Muslim, and Latino voters, solved exactly and propagated through 40,000 seeded Monte-Carlo draws. Three buried parameters govern the answer. The interest function moves the same Latino vote from aligned at a net of 0.267 under a material reading to misaligned under a rights reading, and the net verdict changes sign for one of the three groups while holding negative for the other two. The counting frame inverts the most-misaligned group from LGBTQ, judged per supporter, to Latino, judged per the votes a bloc moves. The awareness gate removes between 0.431 and 0.584 of each group's gross risk, and because each input is propagated at its own measurement-quality concentration the leftover uncertainty is dominated by the quantities measured worst: exposure alone accounts for a 0.467 share of the variance in the sharpest case and the policy-hostility term that the accusation leans on for only 0.016. One configuration is robust, the rights-dependence reading of the LGBTQ case per voter, and it survives in 0.901 of draws while the headline ordering as a whole survives in 0.664. The construct's contribution is to make the hidden inputs explicit and report how much of any verdict each one drives. It does not issue the verdict, and the paper is explicit about why it cannot.

1. A Folk Accusation and Its Free Parameter

A version of the same sentence is written after every election. A group supported a candidate whose program will cost that group the protections it relies on, and the support is read as a mistake the group made about itself. The sentence has a long pedigree in popular political writing, most influentially in the argument that working-class voters in the American interior trade their economic interests for cultural ones and lose both (Frank, 2004). It has the shape of an empirical claim and almost none of the content, because the quantity it rests on, the group's interest, is never defined, and the most direct test of the popular version reverses it: across American voters the probability of voting Republican rises with income, so the picture of the poor voting against material interest while the rich vote for it does not survive contact with the income gradient (Gelman, 2008; Bartels, 2008). The accusation survives this, but only as a claim about a choice its author has left undisclosed.

The choice is the interest function: the map from a group and a candidate to a number that says how well the vote served the group. Fix one interest function and the accusation becomes testable. Fix a different one and the same vote can change verdict. The folk version smuggles a particular interest, usually material self-interest, into a sentence whose grammar implies there could be only one, and the smuggling is what gives the claim its air of obviousness. Decades of survey research remove even the presumption that material self-interest is the natural default, because self-interest, measured directly, predicts policy attitudes weakly and inconsistently, and symbolic and group attachments predict them better (Sears and Funk, 1991). Whatever a vote is tracking, it is usually not the voter's own balance sheet.

This paper takes the accusation seriously enough to build the instrument it lacks. The instrument is a decomposition that separates the things the folk claim fuses: the adverse policy risk a candidate poses to a group, the protection the same candidate offers, whether the risk was foreseeable, whether the affected issue was salient to the group, and which definition of interest is being applied. The name for the pattern the instrument measures is political autoimmunity, support from a group for a coalition whose likely policies weaken the legal, administrative, redistributive, and status protections that group disproportionately depends on. The word is chosen for the structure it points at, a system's protections turned against the system that maintains them, and not for any verdict about the voters, who may be trading the protection for something they value more and may be right to.

What the construct measures is narrower than several things it could be confused with, and the boundaries matter because each neighbor holds a charge the construct is built to avoid. False consciousness, in its Marxian sense, presumes the group has a true interest it has been deceived about, and the deception is the explanandum (Elster, 1985); the present model fixes no true interest and treats the choice among interests as the analyst's disclosed assumption rather than the voter's error. Cognitive dissonance names a psychological state and its reduction (Festinger, 1957), a mechanism inside a voter, where political autoimmunity is a relation among exposure, policy risk, awareness, and vote share, observable without any claim about what the voter feels. Intersectional scholarship has shown for thirty years that a group defined by one axis is cross-cut by others, so that the interest of women, or of Latino voters, or of the working class is never a single quantity to begin with (Crenshaw, 1991). The model inherits that lesson as a constraint: it computes per subgroup where it can, and it reports the instability of any group-level number rather than papering over it.

2. Seven Interests, Not One

The premise that a vote serves one interest is the premise the evidence least supports, and the alternatives are not a list of irrationalities but a list of well-studied utilities. The cleanest statement of the underlying distinction is Sen's: a person's choices answer to their welfare, to their wider preferences, and to their commitments, and these can come apart without any failure of reason, so that acting against one's own welfare ranking out of commitment is not a mistake but a different maximand (Sen, 1977). An interest function is a choice of maximand. The model holds seven of them, each with a literature that says it is real.

The material reading takes interest as income, employment, benefits, prices, and wealth, and it is the reading the folk accusation usually intends. The rights-dependence reading takes interest as the legal and administrative protections a group leans on, anti-discrimination enforcement, due process, bodily autonomy, immigration security, the neutrality of agencies, and it is the reading under which "autoimmunity" is most legible, because these are the protections a hostile administration can withdraw. The two readings already diverge for any group whose material position and whose rights position move in opposite directions under the same candidate.

The remaining five readings come from the study of why votes track something other than the voter's own welfare. The instrumental value of a single vote is negligible, since the probability of casting the decisive ballot is vanishingly small, and this is the oldest result in the economic theory of voting (Downs, 1957). Riker and Ordeshook absorbed the problem by adding a term to the calculus, a value the voter gets from the act of voting itself regardless of its effect (Riker and Ordeshook, 1968), and Brennan and Lomasky built the term into a full theory in which the ballot is expressive, a place to register identity and allegiance because it is nearly costless to do so (Brennan and Lomasky, 1993). Expressive voting is the strongest challenge to any external interest model, because it says that what an outside analyst books as self-harm the voter books as utility, and a model has to let that reading win where it fits. The subjective reading is the discipline that follows: take interest to be what voters themselves rank as important, measured rather than imputed.

Three further readings name specific contents the expressive ballot can bear. Party identification is a standing social identity, acquired early and revised rarely, and it predicts the vote more steadily than any season's issues (Campbell et al., 1960; Green, Palmquist and Schickler, 2002); the in-group attachment and out-group antagonism underneath it are the machinery that social identity theory described (Tajfel and Turner, 1979). As partisanship has sorted into line with race, religion, and place, the identities reinforce one another and the affect attached to the out-party has grown into a motive in its own right (Mason, 2018; Iyengar, Sood and Lelkes, 2012; Iyengar et al., 2019). System-justification theory documents a stranger pattern, that members of disadvantaged groups sometimes defend the very arrangements that disadvantage them, supplying a motive for supporting a dominant coalition from inside a group it subordinates (Jost and Banaji, 1994; Jost, Banaji and Nosek, 2004), a tendency social-dominance theory traces to a measurable preference over group hierarchy (Sidanius and Pratto, 1999) and that the weak and shifting attachment of minority voters to either party leaves room to act on (Hajnal and Lee, 2011). Status, as distinct from money, is the content the work on the recent realignments keeps finding: the sense of a group's position relative to others, whose perceived threat moved votes in 2016 on Mutz's contested but careful account (Mutz, 2018), and whose texture the ethnographies of resentment render directly (Cramer, 2016; Hochschild, 2016). These supply the expressive-status and coalition-entry readings, in which acceptance and standing are the interest the vote serves.

What the subjective reading must respect, and what an external reading discards at its peril, is that much of this content is structured commitment rather than residual ignorance. Moral intuitions are weighted differently across the spectrum, with loyalty, authority, and sanctity weighing on the right that an interest model built only from care and fairness cannot see (Graham, Haidt and Nosek,

2009); a disposition toward order and conformity, activated by perceived threat, organizes a coherent politics rather than a confused one (Stenner, 2005); ethnocentrism and racial resentment shape policy preference through channels whose very measurement remains contested (Kinder and Kam, 2009; Kinder and Sanders, 1996). To book any of these as an error is to assume the conclusion the model is built to hold open. The seventh reading is protest. A vote can be cast to punish a party rather than to select a government, and retrospective-voting theory treats this as rational: the electorate holds incumbents to account for outcomes, and punishment is the instrument (Key, 1966; Fiorina, 1981). A protest reading downweights the adverse risk of the punished alternative, because the voter's maximand is the signal sent rather than the policy received. Whether the protest was well-calculated is a separate question the model can pose but cannot, from vote choice alone, answer.

None of the seven is the right one. That is the claim being made: interest is a parameter, and the literatures above are seven defensible settings of it, each turning some apparent self-harm into religious commitment, expressive identity, status, protest, or a tradeoff the voter would make again. A model that fixed one setting would be doing in mathematics what the folk accusation does in prose. The model fixes none, and reports what changes as the setting changes.

3. The Decomposition

Let g index a group or subgroup, c a candidate or coalition, and j a policy domain. The adverse policy risk a candidate poses to a group in a domain is the product of five quantities, each scaled to the unit interval:

$$R_{gcj} = E_{gj}\, D_{gj}\, H_{gcj}\, P_{cj}\, M_{gj}.$$

E_{gj} is the group's exposure to the domain, the share of the group whose circumstances the domain reaches. D_{gj} is institutional dependence, how much the group's standing in the domain rests on protections an administration can alter rather than on private means it controls. H_{gcj} is the candidate's hostility, read from platform, record, appointments, and the commitments of the coalition, on a scale this paper takes from a signed policy score and folds to its adverse part. P_{cj} is the probability the candidate can and will implement the policy, which depends on unified government, the courts, and administrative capacity. M_{gj} is the magnitude of the harm if implemented. The product form encodes a claim worth stating plainly: a risk is large only when every factor is large, and a single small factor, an unexposed group or a candidate without the means to act, collapses the domain's contribution regardless of how alarming the other factors look. Multiplication is the model's way of refusing to be frightened by any one number.

Summed over a group's domains and weighted by the group's support for the candidate, the product gives a gross risk, $V_{gc} \sum_j R_{gcj}$, where V_{gc} is the group's vote share for c. This is the quantity nearest to what the folk accusation gestures at, and the model treats it as a starting point rather than a conclusion, because two corrections stand between it and any verdict of self-harm. A candidate is rarely hostile to a group on every axis at once. A platform that withdraws a protection in one

domain may extend a benefit in another, and that benefit is part of the vote's return, so reading the same signed policy score from its protective side gives B_{gcj}, the protection the candidate offers the group in the domain, on the same scale as the harm. Domains also differ in weight under a given interest, which is where the interest function of the previous section enters as W^m_{gj}, the importance interest model m assigns the domain, normalized across the group's domains.

Net alignment under interest model m is then the weighted protection minus the weighted risk,

$$\mathrm{NA}_m(g,c) = V_{gc} \sum_j W^m_{gj} \left(B_{gcj} - R_{gcj}\right),$$

and the autoimmunity score keeps only the domains where adverse risk clears the protection by more than a tolerance τ,

$$A_m(g,c) = V_{gc} \sum_j W^m_{gj} \max\!\left(0,\ R_{gcj} - B_{gcj} - \tau\right).$$

The tolerance, set at 0.02, keeps domains of negligible net adversity from accumulating into a verdict.

The relation between gross risk and net alignment is the model's first result, and it is elementary. Write the weighted gross risk under model m as $G_m = V_{gc} \sum_j W^m_{gj} R_{gcj}$ and the weighted protective offset as $\Pi_m = V_{gc} \sum_j W^m_{gj} B_{gcj}$. Then $\mathrm{NA}_m = \Pi_m - G_m$, so a group is net-misaligned, $\mathrm{NA}_m < 0$, if and only if $G_m > \Pi_m$. A judgment of self-harm is a claim that the weighted risk exceeds the weighted protection. The folk accusation observes a piece of G_m, the alarming domain, and concludes $\mathrm{NA}_m < 0$, and the inference holds only when Π_m is small, which is to say only when the candidate offers the group nothing the group wanted. That condition is sometimes met and usually assumed. Making Π_m a term rather than an omission is most of what the decomposition does, because it converts a verdict into a comparison and puts back the half of the comparison the accusation leaves out.

4. Foreseeability, Salience, and the Dissolution of "Irrational"

A risk no one could have seen is not evidence about a voter's judgment, and a risk in a domain the voter never cared about is not evidence about their priorities. The model separates both from raw risk with two gates, and the separation is where the charge of irrationality dissolves into questions that can be asked one at a time.

Foreseeability has a public side and a private side. The public side, F_{cj}, is whether the adverse policy was knowable before the vote from the candidate's prior record, explicit promises, party platform, media coverage, and the warnings of advocacy groups, the supply of available signal. The private side, a_{gcj}, is whether voters in the group reported awareness of the issue, the signal actually received. The distinction is not pedantic. The receive-accept-sample account of opinion makes the supply of elite signal the precondition for any mass perception of an issue (Zaller, 1992), and

the long-standing finding of thin issue constraint in mass publics warns that the supply often goes unreceived (Converse, 1964); voters who do receive it reason with it through cues and shortcuts rather than full information (Lupia and McCubbins, 1998), and received belief can be confident and wrong, a different failure from the simple absence of information (Kuklinski et al., 2000). A model that collapsed these would confuse a voter who was misled with one who was uninformed and with one the signal never reached. The gate combines the two sides as $K_{gcj} = \lambda F_{cj} + (1-\lambda)\, a_{gcj}$, with $\lambda = 0.5$ giving them equal weight, and multiplies the risk by K to give foreseeable risk.

Salience is the second gate. Group exposure to a domain does not become a voter's priority until the group is conscious of the domain as bearing on it, the condition the literature on group consciousness and linked fate identifies as the trigger that converts demographic position into political behavior (Dawson, 1994; Miller et al., 1981). The model holds S_{gj}, the salience of the domain to the group, and multiplies again to give priority risk, $R_{gcj}K_{gcj}S_{gj}$. The construction is deliberately deflationary. Each gate can only shrink the risk, never enlarge it, so the model moves from what a candidate could do to a group, through what was foreseeable, to what was both foreseeable and salient, and the gap between the first and the last is the room the folk accusation fills with an assumption of negligence.

The gates also supply a typology that replaces the binary of rational and irrational with categories that name the actual case. A vote facing high foreseeable, salient risk with low reported awareness is uninformed adverse-interest voting, a candidate for the information failure the folk claim imagines, and the only one of the categories where the imagined story holds. A vote facing the same risk with high awareness and an explicit higher-priority issue is informed tradeoff voting, a price knowingly paid rather than a mistake, the case the realist tradition insisted on when it argued the electorate is not a fool (Key, 1966), and which the model must recognize so that it does not label every tradeoff an error. Between them sit protest voting, where the adverse risk is accepted to send a signal, and its miscalculated variant, where the signal's cost was underestimated, separable only with a counterfactual the model can frame but not settle. Framing research is the reminder underneath all of this that salience and awareness are constructed by how a choice is presented and are not fixed properties of the voter (Tversky and Kahneman, 1981), so the gates measure a state of the information environment as much as a state of the voter. The realist literature presses hardest here, doubting that retrospective judgment is as considered as the rational tradition hopes (Achen and Bartels, 2016; Caplan, 2007); the model takes no side, and instead makes awareness an input whose value changes the verdict and whose measurement is therefore the thing to argue about.

5. Three Cases, Computed

The model is exhibited on three pairings from the 2024 United States presidential election, support for Trump among LGBTQ voters, Muslim voters, and Latino voters, chosen because each was offered in popular commentary as a case of a group voting against itself and because the three differ in the parameters the model isolates. Everything that follows is computed from a synthetic dataset whose exposure, dependence, hostility, implementation, magnitude, salience, and awareness val-

ues are stipulated to display the model's behavior. They are not estimates. The numbers are facts about a fully specified model, in the sense in which a worked example in mechanics is a fact about the example, and they are offered because the model is the simplest object in which the claimed effects appear. Only the vote shares point at reported figures, and they are anchored by sources of unequal quality: the Latino share at 0.46 to a validated-voter study that found Trump roughly even with Harris among Hispanic voters (Pew Research Center, 2025), the Muslim share at 0.21 to an advocacy-group exit poll, recorded as such and not as validated-voter evidence (Council on American-Islamic Relations, 2024), and the LGBTQ share at 0.12 to exit-poll summaries, the weakest anchor, used only for order of magnitude. The asymmetry is left visible because it is part of the object: the cases easiest to discuss in public are not the cases easiest to measure.

Each group is given three domains. The LGBTQ case spans federal civil-rights protection, transgender health and documentation, and hate-crime and discrimination enforcement, all rights-dependent and all facing a hostile administration; its per-cell gross risks are 0.459, 0.207, and 0.153, the highest single-cell risks in the study, because exposure, dependence, hostility, and implementation are jointly high. The Muslim case covers entry and travel restriction, the Gaza-driven foreign-policy grievance that motivated much of the defection, and religious-bias enforcement, with risks of 0.246, 0.086, and 0.142; the protest domain scores low on adverse risk because its hostility is lower and its content is expressive rather than protective. The Latino case takes in immigration enforcement, the economy, and civil-rights enforcement, at 0.252, 0.026, and 0.089, and the economy domain is the one place in the study where the candidate is read as protective rather than hostile, holding a benefit B no other cell does. Figure 1 shows the nine cells and what the gates do to them.

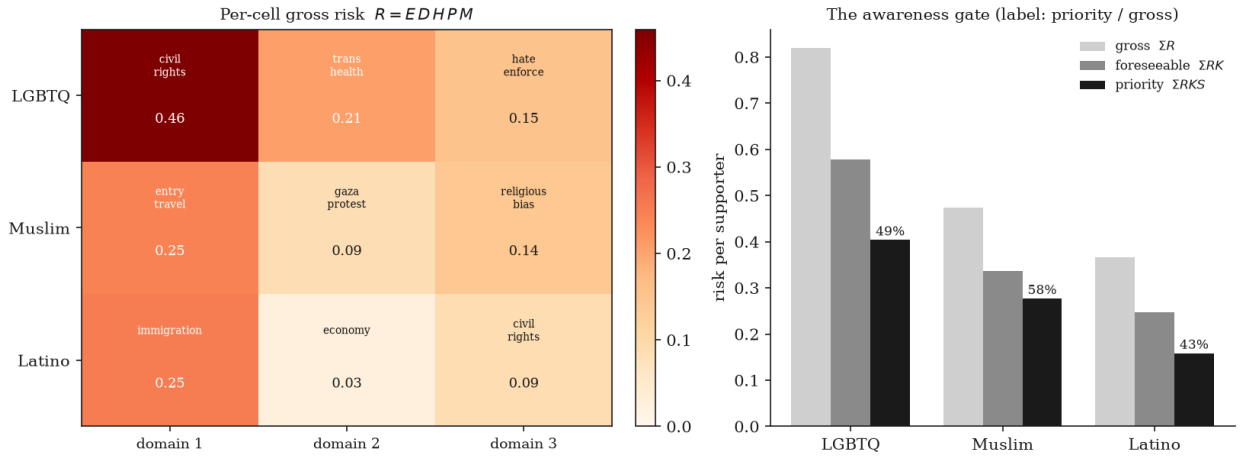

Figure 1: Left: per-cell gross risk $R = E\,D\,H\,P\,M$ for the nine group-domain cells. The LGBTQ civil-rights cell at 0.459 is the sharpest single risk in the study; the Latino economy cell at 0.026 is the mildest, because the candidate is read as protective there. Right: the awareness gate at work. For each group the three bars are gross risk $\sum R$, foreseeable risk $\sum RK$ after the foreseeability gate, and priority risk $\sum RKS$ after the salience gate; the percentage is priority over gross, the fraction of raw risk that survives as both foreseeable and salient. The gates remove between 43% and 58% of gross risk, and they remove different fractions from different groups.

The aggregate per-supporter risks tell the first part of the story. Gross risk per supporter is 0.819 for the LGBTQ case, 0.473 for the Muslim case, and 0.367 for the Latino case, an ordering that looks decisive. The awareness gate then removes a different share from each: priority risk falls to 0.404, 0.277, and 0.158, an attrition of 0.493, 0.584, and 0.431. The Muslim case retains the most because its dominant adverse domain, entry restriction, is among the most foreseeable in the study, openly promised and previously enacted, rather than because its risks are larger; the gate rewards a candidate's transparency about hostility by keeping more of it on the ledger. The LGBTQ and Latino cases lose roughly half their gross risk to the gates, the LGBTQ case because some of its sharpest exposure sits in a lower-salience domain, the Latino case because its dominant domain is somewhat less foreseeable and the economy domain, where the group is most conscious, holds little adverse risk to begin with.

6. The Verdict Is Not in the Data

The per-supporter ordering, LGBTQ then Muslim then Latino, is stable across all seven interest functions, and it is also the wrong unit for the question the accusation asks. A claim about a group voting against itself is a claim about a bloc, and a bloc's effect on an election scales with its size. Weighting each group's priority risk by its vote share reverses the order: population-weighted priority risk is 0.048 for the LGBTQ case, 0.058 for the Muslim case, and 0.073 for the Latino case. The group with the highest risk per supporter is the lowest-impact bloc, and the group with the lowest risk per supporter is the highest-impact bloc, because the LGBTQ share is small and the Latino share large. Both numbers are correct, and they answer different questions. Per supporter, the LGBTQ case is the sharpest instance of a vote at odds with a group's rights dependence. Per the votes a bloc actually moves, the Latino case is where the same pattern bears most on an outcome. The folk accusation never says which it means, and the two readings name different groups.

The interest function is the second buried parameter, and its effect is sharpest on net alignment, where the protective term of Section 3 does its work. Figure 2 traces net alignment for the three groups across the seven interest functions. The LGBTQ line sits near -0.273 and barely moves, because the case has no protective domain for any interest function to weight: every reading agrees the vote is misaligned, and they differ only in degree. The Muslim line stays negative throughout, between roughly -0.111 under the protest reading, which downweights the adverse domains in favor of the expressive one, and -0.189 under the coalition-entry reading. The Latino line crosses zero. Under the material reading, which weights the economy domain where the candidate is protective, Latino net alignment is $+0.267$, a vote that serves the group's interest; under the rights-dependence reading, which weights immigration and civil-rights enforcement, it is -0.061, a vote against it. The same vote, the same data, opposite verdicts, and the only thing that changed is the analyst's choice of what counts as interest. Of the three groups, exactly one has its verdict's sign decided by that choice, and that unevenness is itself the finding: assumption-sensitivity is not uniform, and the model says which cases are robust and which are manufactured.

The third parameter is the one the model is most pointed about, because it concerns not the choice of assumption but the measurement of the inputs the assumption acts on. Priority risk in a single

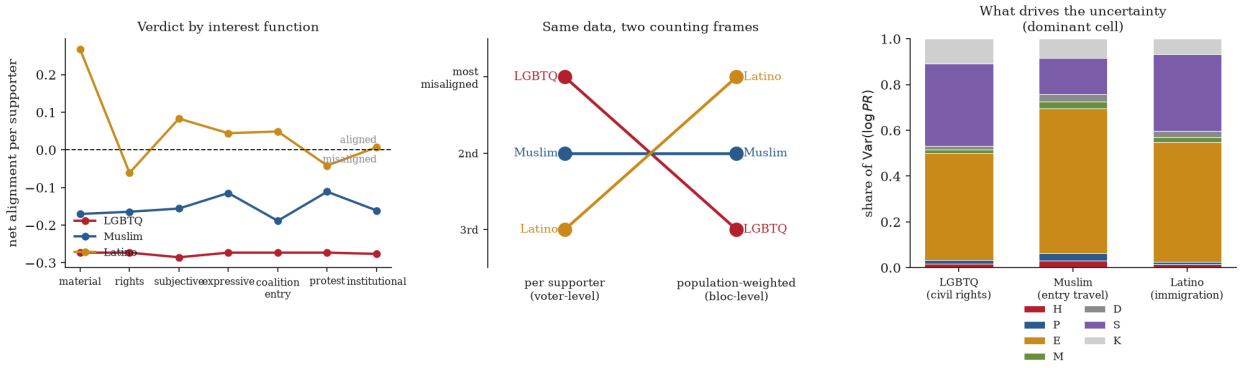

Figure 2: Left: net alignment per supporter for each group across the seven interest functions, with the zero line dividing aligned from misaligned. The LGBTQ case is robustly negative and nearly flat; the Muslim case is negative throughout; the Latino case crosses from aligned under the material, subjective, expressive, coalition-entry, and institutional readings to misaligned under rights dependence and protest. Center: the counting-frame inversion. Ranked per supporter, LGBTQ is the most-misaligned group and Latino the least; ranked by population-weighted priority risk, the order reverses, because vote share runs the other way. Right: the share of the variance in each group's dominant-cell priority risk attributable to each input, from an exact log-variance decomposition in which each input is drawn at its own measurement-quality concentration. Exposure E, the salience gate S, and the awareness gate K, the three survey-estimated inputs given wide priors, dominate; the policy-hostility term H, on which the folk accusation leans, is given a tight prior and contributes least.

domain is a product of independent factors, so the variance of its logarithm decomposes exactly into a sum of per-factor variances, and the share of each factor is a clean attribution of where the uncertainty in a verdict comes from. That attribution is meaningful only because the model assigns each input its own measurement-quality concentration rather than a single shared one. The three survey-estimated inputs, exposure, salience, and reported awareness, are drawn from wide Beta priors, and the four inputs an analyst reads off the public record, hostility, implementation probability, magnitude, and institutional dependence, are drawn from tight ones, so the decomposition reflects a stipulated difference in how well each input is measured and not merely the stipulated means. For the LGBTQ case's dominant cell, exposure takes a 0.467 share of the variance, the salience gate 0.36, and the awareness gate 0.107, while candidate hostility, implementation, magnitude, and dependence take 0.016 each. Exposure and the awareness gate together account for 0.574, and exposure and salience for 0.827. The pattern holds across the three cases: the verdict's uncertainty is dominated by exposure, salience, and awareness, the quantities an analyst estimates from surveys with the widest error, and is least sensitive to candidate hostility, the quantity an analyst can read off the public record with the most confidence. A one-at-a-time swing confirms it. Moving each input of the LGBTQ rights-model score across its uncertainty interval, the score swings by 1.269 of its value when exposure moves and by a fifth of that or less for any policy factor. The accusation rests its weight on the input that bears almost none of it.

Propagating every input as a Beta distribution at its own measurement-quality concentration and

drawing 40,000 seeded samples puts intervals on the verdict. The per-supporter priority risks have medians of 0.394, 0.27, and 0.148 for the three cases, with 90% intervals running from 0.206 to 0.636 for the LGBTQ case, 0.148 to 0.429 for the Muslim case, and 0.064 to 0.282 for the Latino case, intervals wide enough to overlap. The headline per-supporter ordering survives in 0.664 of the draws, so roughly one draw in three reorders the three groups under uncertainty alone. The single most robust statement the study contains, that the LGBTQ case is the most misaligned under the rights-dependence reading taken per supporter, holds in 0.901 of draws, and even it fails in about one in ten. The point is not that the model is too uncertain to use. The point is that the uncertainty is reportable, and that a responsible verdict is an interval with a probability attached rather than the flat declarative the genre prefers. The reflexive use of attribution here, decomposing the model's own output among its inputs in the manner of a value over the factors of a game (Shapley, 1953) and of variance-based global sensitivity analysis (Saltelli et al., 2008), is the model auditing itself: before it grades a vote, it reports which of its own assumptions the grade depends on.

7. The Standing to Name the Interest

The decomposition converts a folk accusation into a measurement with disclosed inputs, and the measurement's main result turns on the instrument rather than the electorate: for the cases public argument actually fights over, the verdict is governed by three analyst choices, the interest function, the counting frame, and the awareness gate, and under realistic uncertainty it is not robust. One configuration in the study holds up, the rights-dependence reading of the LGBTQ case taken per voter, and it can be trusted only because so many of the neighboring configurations do not. This inverts the rhetorical use the construct is usually put to. A measure of political autoimmunity, carefully built, is more often a reason to withhold the accusation than to make it.

Each of the three choices is a place where the arithmetic stops and a commitment begins, and naming them is the same as saying what the model refuses to do. The interest function is a normative claim about what a group ought to value, and no measurement discharges it; the model's service is to price it, to show that adopting the rights reading rather than the material one is the whole of what produces the misalignment verdict in the Latino case, so the argument can move to the ground it actually occupies, the defense of the reading. The counting frame is a choice of question, per voter or per bloc, and the two name different groups from the same data, so a verdict that does not declare its unit has not yet said what it is about. The awareness gate turns on a quantity, what a group knew and weighed, that surveys estimate worst, which is why the same gate that dissolves the charge of irrationality into the informed tradeoff and the uninformed miss cannot, from vote choice alone, tell the tradeoff from the miscalculated protest: the two differ in a counterfactual no cross-section observes. And below the group score lies the voter it is not, because exposure, salience, and awareness vary within every group the study names, and the cross-pressured voter pulled by several memberships at once is the normal case (Lazarsfeld, Berelson and Gaudet, 1944; Crenshaw, 1991); a population-weighted number is an ecological summary, and read as a per-voter verdict it is the ecological fallacy with a coat of arithmetic.

Under all three sits the commitment the construct shares with the accusation it was built to disci-

pline. To call a vote self-harm is to claim standing to define a group's interest against the group's own expressed choice, and that standing is exactly what the expressive and subjective readings deny. The model does not settle the claim; it keeps the subjective reading, interest as what voters say they value, among the seven, and measures how far each external reading departs from it. Where an external reading and the subjective one agree, the verdict stands on firm ground. Where they diverge, the verdict is a contest between the analyst's account of a group's interest and the group's own, and the most the instrument can defend is the size of the divergence, the distance between what a group votes for and what, on a stated theory of its interest, would serve it. That distance is real and computable. Whether it convicts anyone is a judgment the arithmetic was never going to make, and the discipline the whole construction buys is the refusal to let a number pretend otherwise.

References

Achen, C. H., and Bartels, L. M. (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press.

Bartels, L. M. (2008). *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton University Press.

Brennan, G., and Lomasky, L. (1993). *Democracy and Decision: The Pure Theory of Electoral Preference*. Cambridge University Press.

Campbell, A., Converse, P. E., Miller, W. E., and Stokes, D. E. (1960). *The American Voter*. Wiley.

Caplan, B. (2007). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton University Press.

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.

Council on American-Islamic Relations. (2024). *CAIR Exit Poll of Muslim Voters: 2024 General Election*. CAIR.

Cramer, K. J. (2016). *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker*. University of Chicago Press.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241–1299.

Dawson, M. C. (1994). *Behind the Mule: Race and Class in African-American Politics*. Princeton University Press.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper & Row.

Elster, J. (1985). *Making Sense of Marx*. Cambridge University Press.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. Yale University Press.

Frank, T. (2004). *What's the Matter with Kansas? How Conservatives Won the Heart of America*. Metropolitan Books.

Gelman, A. (2008). *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press.

Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046.

Green, D. P., Palmquist, B., and Schickler, E. (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Hajnal, Z. L., and Lee, T. (2011). *Why Americans Don't Join the Party: Race, Immigration, and the Failure of Political Parties to Engage the Electorate*. Princeton University Press.

Hochschild, A. R. (2016). *Strangers in Their Own Land: Anger and Mourning on the American Right*. New Press.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129–146.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3), 405–431.

Jost, J. T., and Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, 33(1), 1–27.

Jost, J. T., Banaji, M. R., and Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 881–919.

Key, V. O. (1966). *The Responsible Electorate: Rationality in Presidential Voting, 1936–1960*. Harvard University Press.

Kinder, D. R., and Kam, C. D. (2009). *Us Against Them: Ethnocentric Foundations of American Opinion*. University of Chicago Press.

Kinder, D. R., and Sanders, L. M. (1996). *Divided by Color: Racial Politics and Democratic Ideals*. University of Chicago Press.

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., and Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *Journal of Politics*, 62(3), 790–816.

Lazarsfeld, P. F., Berelson, B., and Gaudet, H. (1944). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press.

Lupia, A., and McCubbins, M. D. (1998). *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge University Press.

Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

Miller, A. H., Gurin, P., Gurin, G., and Malanchuk, O. (1981). Group consciousness and political participation. *American Journal of Political Science*, 25(3), 494–511.

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences*, 115(19), E4330–E4339.

Pew Research Center. (2025). *Behind Trump's 2024 Victory, a More Racially and Ethnically Diverse Voter Coalition*. Pew Research Center.

Riker, W. H., and Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review*, 62(1), 25–42.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley.

Sears, D. O., and Funk, C. L. (1991). The role of self-interest in social and political attitudes. *Advances in Experimental Social Psychology*, 24, 1–91.

Sen, A. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 6(4), 317–344.

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the Theory of Games, Volume II* (pp. 307–317). Princeton University Press.

Sidanius, J., and Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press.

Stenner, K. (2005). *The Authoritarian Dynamic*. Cambridge University Press.

Tajfel, H., and Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin and S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.

Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.