

The Voter Trap

Juan Carlos Cisneros¹

April 14, 2026

Abstract

Why do cross-pressured voters persistently support parties that oppose their economic interests? We propose that asymmetric party brand histories constrain which cultural narratives are credible in electoral competition: the party with a deep cultural brand can activate identity-shifting frames that the economic party cannot replicate without destroying its own coherence. We formalize this mechanism through a brand-state model with three main results. First, the voter trap holds under sufficient conditions on evidence quality, cultural salience, and brand asymmetry. Second, the trap deepens dynamically: each election cycle of brand-consistent cultural campaigning lowers the signal threshold required to sustain the trap, and under long institutional memory the trap becomes absorbing. Third, platform separation is sustained endogenously by the brand mechanism, ruling out Downsian convergence. We apply the framework to the 1898 Wilmington Coup and the Jim Crow equilibrium, using evidence from Ottinger and Posch [2025] on elite-orchestrated propaganda in Southern newspapers.

1 Introduction

A recurring puzzle in democratic politics is the cross-pressured voter: an individual whose economic and cultural identities pull toward opposite parties. Working-class conservatives in the 1890s American South, dislocated industrial workers in the 2016 Rust Belt, and evangelical poor communities in twenty-first-century Brazil all share a common feature: their economic interests point clearly toward the redistributive left, yet they vote for parties whose platforms are anchored to cultural conservatism. This pattern, popularized by Frank

¹Universitat Pompeu Fabra. Email: juancarlos.cisneros@upf.edu. I acknowledge the use of Claude Sonnet 4.6 and Claude Opus 4.6 for discussions, revisions, and formatting of the paper. All remaining errors are my own.

[2004] and reframed as the “Kansas paradox” in the subsequent literature, has generated decades of multi-disciplinary debate. The voting behavior literature has shown that the naive “voting against interests” framing is partly misconceived: self-interest is consistently weak as a predictor of political attitudes [Sears and Funk, 1991], the apparent working-class shift rightward is substantially a Southern phenomenon driven by racial realignment [Bartels, 2006, Kuziemko and Washington, 2018], and voters act on social identities and partisan loyalties as much as on policy reasoning [Achen and Bartels, 2016]. A generation of formal models has established that multi-dimensional competition can rationally lead voters to sacrifice economic gains for identity affirmation [Roemer, 1998, Shayo, 2009, Bonomi et al., 2021]. The deeper question, then, is *what constrains which identity dimensions parties can credibly activate, and how do these constraints shape the persistence and depth of cross-pressured voting?*

This paper proposes an answer rooted in the history of party brands. The cross-pressured voter is not making an error. She is making the best available choice from a menu that has been asymmetrically constrained by each party’s accumulated record of campaigns and governing actions.

The key asymmetry is this. A party with a long history of cultural campaigning can credibly propose a model of current events that foregrounds cultural identity; cross-pressured voters, facing a genuine cultural signal in the public evidence, rationally adopt this model. The opposing party, anchored to an economic brand, cannot credibly replicate the cultural frame without destroying the coherence of its own established record. Because brand credibility is publicly known, the economic party cannot neutralize the cultural narrative, and the cross-pressured voter cannot condition on a counter-campaign that will not arrive. The result is a *voter trap*: a self-reinforcing configuration in which the voter rationally votes against her benchmark economic interests, common knowledge of the mechanism does not dissolve it, and the trap deepens over time as the cultural party’s brand strengthens.

The clearest historical illustration comes from the 1890s American South. The Populist Party of the 1890s was the first major American party to build a redistributionist cross-racial coalition: Black Republicans and poor white farmers joined in “fusion” governments across the South that delivered lower railroad rates, expanded public education, and multiracial elected governance. This coalition posed an existential threat to Democratic hegemony. Facing the prospect of a Populist-Republican majority, Southern Democratic elites orchestrated a systematic propaganda campaign through party-affiliated newspapers, foregrounding racial outrage to reframe elections as contests over racial cultural order rather than economic class. Ottinger and Posch [2025] document this campaign using data from thousands of historical newspapers: a ten-percentage-point increase in the local Populist-Republican fusion vote

share raised the probability of anti-Black content in Democratic-affiliated newspapers by 4.5 percentage points. This effect was supply-driven, concentrated in Democratic-affiliated papers and absent in politically independent outlets, consistent with elite orchestration rather than demand-driven audience preferences. The campaign culminated on November 10, 1898, when armed white supremacists overthrew the elected multiracial Republican government of Wilmington, North Carolina (the only successful coup d'état on American soil) and installed an administration that enacted the disenfranchisement laws locking in Jim Crow for the next seven decades [Wormser, 2004]. Section 9 applies the voter trap model formally to this episode.

Contemporary American politics exhibits patterns consistent with the same mechanism. The MAGA campaign of 2016 ran a cultural narrative centered on immigration, national identity, and cultural grievance that may have shifted the effective identity weights of working-class white voters whose economic circumstances aligned with the redistributive left. Autor et al. [2020] document that workers in trade-exposed communities, whose manufacturing employment collapsed following import competition, shifted toward Republican candidates; their economic grievance became a vehicle for cultural realignment rather than redistributive political action. Bursztyn et al. [2020] show that Trump's electoral success eroded the social stigma around publicly expressing culturally divisive views, making future cultural campaigns more credible. This amounts to a deepening of the cultural brand coherence prior that our model formalizes in Proposition 2. The MAGA live campaign has remained electorally viable across three election cycles, a persistence consistent with the stable and absorbing trap of Corollary 2.

The mechanism is not limited to the United States. In Brazil, poor and working-class voters who benefited most from the Partido dos Trabalhadores' *Bolsa Família* and minimum-wage programs voted in large numbers for Jair Bolsonaro in 2018, whose campaign foregrounded evangelical Christianity, anti-communism, and public security. This cultural narrative was coherent with the Brazilian right's long-standing brand but largely outside the PT's credible range. The voter trap framework suggests a reading of this case: the Brazilian right's cultural brand was credibly entrenched over decades, while the PT's brand anchored to economic redistribution may have precluded a credible cultural counter-campaign.

We develop a formal model of the voter trap with three components. The first is a *brand-state mechanism*: parties accumulate a publicly observable history of platforms, narratives, and governing actions that evolves according to an axiomatized system (properties G1–G4) and generates an endogenous credibility object, the coherence prior $\Pi_{p,t}$, that restricts which future campaigns voters will find believable. The second is a *posterior-odds model adoption rule*: voters evaluate competing interpretive models by comparing their evidential fit against

the voter’s prior model, weighted by brand coherence, and adopt the model that wins this joint test. The third is a *saliency mechanism*: the adopted model shifts the voter’s identity weight between economic and cultural dimensions, displacing her effective policy ideal and determining her vote.

The voter trap arises when this chain of adoption and saliency shift leaves a cross-pressured voter voting R (the culturally-entrenched party, formally defined in Section 2) even though party L (the economic party) remains objectively closer on the economic dimension. Critically, the trap does not require any bias in voter updating, any deception, or any failure of common knowledge. The voter knows she is being targeted by a cultural campaign, knows her opponent is brand-constrained, and knows the electoral rules. She votes R because the cultural model genuinely fits the current public evidence under a credible brand, her effective bliss point has shifted under that model, and L ’s inability to counter is common knowledge.

Three sets of results characterize the voter trap as a conditional configuration: the results identify sufficient conditions under which the trap arises, rather than proving that those conditions hold generically in equilibrium. The first result establishes *trap formation*: the voter trap holds when three conditions are jointly satisfied—the cultural party’s narrative is both evidentially supported and brand-credible, the resulting identity shift is strong enough to change the voter’s effective preferences, and the economic party cannot mount a credible counter-narrative due to its accumulated brand history. The second result establishes *dynamic deepening*: the trap is self-reinforcing. Each election cycle in which the cultural party runs its brand-consistent campaign strengthens the brand asymmetry, making future traps easier to trigger and harder to escape. We formalize this by showing that the minimum strength of cultural signal needed to sustain the trap declines monotonically over time. Under long institutional memory, this threshold eventually falls below any bounded signal support, and the trap becomes *absorbing*—it holds for all future realizations of the public evidence, not just favorable ones. The third result establishes *robustness*: the cultural party does not need to know the full distribution of voter types to implement the trap. The conditions that generate and sustain the trap depend on publicly observable quantities (the signal environment, the brand histories, and a lower bound on the share of cross-pressured voters recoverable from electoral returns), not on the party’s private information about voter preferences.

This paper contributes to three intersecting literatures. The first is the debate on cross-pressured voting and the political economy of identity. Frank [2004] popularized the observation; Bartels [2006] and Kuziemko and Washington [2018] showed the pattern is concentrated in the American South, driven by racial realignment; Achen and Bartels [2016] argued that

voters choose on social identities rather than policy reasoning. On the formal side, Roemer [1998] demonstrated that multi-dimensional competition can rationally depress redistribution; Shayo [2009] and Bonomi et al. [2021] formalize identity-driven voting; and Gennaioli and Tabellini [2025] provide the direct antecedent, in which parties choose propaganda intensity to shift voters’ identity weights. Our model departs from GT in three respects: the set of credible cultural narratives is constrained by accumulated brand history (properties G1–G4), making effectiveness path-dependent; campaign effectiveness is governed by an endogenous, asymmetric coherence prior $\Pi_{p,t}$ rather than a fixed symmetric parameter χ ; and brand accumulation generates dynamic deepening (Proposition 2) and absorbing traps (Corollary 2), consequences absent when history does not constrain the action space.

The second is the literature on narrative competition, persuasion, and propaganda. Kammenica and Gentzkow [2011] study optimal persuasion with a fixed prior; Eliaz and Spiegel [2020] model narrative competition as a contest over causal structure; we add the brand-coherence constraint that determines which narratives are live options at any date. On the empirical side, Ottinger and Posch [2025] document elite orchestration of racial propaganda in the 1890s South; Bursztyn et al. [2020] study the behavioral consequences of culturally divisive media; and Eliaz et al. [2025] show that false narratives can mobilize voters even when the falsehood is known. Our model provides a supply-side theory of *when* and *why* culturally loaded content is produced as a strategic political decision and characterizes the brand conditions under which true narratives trap voters without any false content.

The third is the party reputation and credibility literature. Banks and Sobel [1987], Alesina and Cukierman [1990], and Besley and Coate [1997] study how accumulated behavior constrains current credibility. Our brand-state mechanism departs by making credibility a fully public, path-dependent summary statistic ($B_{p,t}$, common knowledge) and focusing on the asymmetric constraint this imposes on narrative competition rather than on equilibrium selection among reputation-consistent paths.

The remainder of the paper proceeds as follows. Section 2 introduces the model environment. Sections 3–6 develop the brand-state mechanism, live campaigns, model adoption, and salience. Section 7 defines and characterizes the voter trap. Section 8 analyzes stability, platform separation, and testable implications. Section 9 applies the framework to the Wilmington Coup and the Jim Crow equilibrium. Section 10 concludes. Proofs are collected in Appendix A; robustness to electorate uncertainty in Appendix C.

2 Environment

The political game. Time is discrete and indexed by $t = 0, 1, 2, \dots$. The game begins at $t = 0$; each party enters with an initial brand state $(B_{L,0}, B_{R,0})$ that encodes its pre-game history of campaigns and governing actions, the formal evolution of which is given in Section 3. Two parties, $p \in \{L, R\}$, compete for office. The game has two logically distinct frequencies. First, *campaign competition* occurs at every t : parties choose campaigns (each consisting of a policy platform and an interpretive model), brand states evolve, and voters update their adopted models as the public realization x_t arrives. Second, *elections* are held at dates $\mathcal{T}_e = \{T, 2T, 3T, \dots\}$ for a fixed inter-election length $T \geq 1$. Between election dates, no vote is cast; at each $t \in \mathcal{T}_e$, voters vote based on their current effective bliss point, formalized in Section 6. The electoral rule is simple majority throughout. The franchise, the rule determining the winner, and the inter-election length T are common knowledge and remain fixed; institutional changes are beyond the scope of this paper.

Policy space. Policy is a pair

$$Y = (\tau, q) \in \mathcal{Y} := \mathbb{R}_+ \times \mathbb{R},$$

where $\tau \in \mathbb{R}_+$ is a proportional tax rate and $q \in \mathbb{R}$ is a social policy position. Following Gennaioli and Tabellini [2025], higher values of q correspond to more liberal social policy. The tax rate is restricted to be non-negative since it represents a proportional rate; the social policy index has no natural sign restriction, and values of q can range from maximally conservative to maximally progressive. Two parties $p \in \{L, R\}$ maintain a policy ordering throughout:

$$\tau_R < \tau_L, \quad q_R < q_L. \tag{1}$$

Party L is the economic-left and socially progressive party; party R is the economic-right and socially conservative party. Proposition 3 establishes that this ordering is sustained endogenously in equilibrium: the brand mechanism generates dimension-specific adjustment costs that make platform convergence strictly dominated for forward-looking parties. We state the ordering as a maintained assumption here for expositional clarity, since the formal result requires the brand dynamics developed in Section 3.

The electorate. Voters are indexed by $i \in [0, 1]$. Each voter belongs to one economic class and one cultural type:

$$(e_i, c_i) \in \{U, L\} \times \{C, P\},$$

where U and L denote upper and lower economic class, and C and P denote conservative and progressive cultural type. This yields four voter groups: UC , UP , LC , and LP . Voters of types UC and UP are not cross-pressured: UC voters prefer party R and UP voters prefer party L on both economic and cultural dimensions jointly; under the policy ordering (1), their equilibrium vote is weakly determined and they are not the focus of the analysis. Similarly, LP voters prefer party L on both dimensions and vote L in any equilibrium consistent with (1). The lower-class conservative voter, henceforth LC , is the pivotal type throughout the analysis: she is *cross-pressured*, meaning her economic and cultural identities pull toward opposite parties.

Preferences. Following Gennaioli and Tabellini [2025], voter (e_i, c_i) has utility over policy $Y = (\tau, q)$ given by

$$W^{ij}(\tau, q) = (1 + \epsilon_i)(1 - \tau) - \frac{\tau^2}{2} + (\nu + b\psi_j)\tau - \frac{\kappa}{2}(q - \psi_j)^2, \quad (2)$$

with normalization $\epsilon_U = \epsilon$, $\epsilon_L = -\epsilon$, $\psi_P = \psi$, $\psi_C = -\psi$, and parameters $\epsilon, \psi > 0$. The rational bliss point of voter ij is

$$\tau_{ij}^* = \nu + b\psi_j - (1 + \epsilon_i), \quad q_{ij}^* = \psi_j. \quad (3)$$

The parameter restriction $\epsilon > b\psi$ ensures that economic class dominates cultural type in determining the rational preferred tax rate, so that the cross-pressured voter's economic pull toward L is well-defined at the rational benchmark.

Reference identity ideals. We take as primitive for each voter two *reference identity ideals*:

$$Y_i^E = (\tau_i^E, q_i^E), \quad Y_i^C = (\tau_i^C, q_i^C).$$

These summarize where the voter would locate policy if the economic or cultural dimension fully determined political choice. They are naturally grounded in the rational bliss points (3): Y_i^E corresponds to the bliss point under economic prioritization, and Y_i^C to the bliss point under cultural prioritization.

Importantly, we do not require voters to choose a single identity. Instead, each voter places continuously varying weights on economic and cultural identity, and those weights are endogenous to the political environment at each date t , as formalized in Section 6.

Assumption 1 (Cross-pressured pivotal voter). *For the lower-class conservative voter LC ,*

the reference identity ideals satisfy

$$\tau_{LC}^E > \tau_{LC}^C, \quad q_{LC}^E > q_{LC}^C. \quad (4)$$

That is, LC is economically to the left and culturally to the right.

Together with (1), Assumption 1 implies that party L is closer to LC 's economic ideal and party R is closer to LC 's cultural ideal. The central question of this paper is when, and why, the cultural dimension governs LC 's vote even though her economic interests point toward L .

Public information. At each date t , all agents observe a public history

$$h_t = \left\{ x_s, c_{L,s}, c_{R,s}, r_s, y_s^{\text{obs}} \right\}_{s=0}^{t-1}$$

and a current high-dimensional public realization $x_t \in \mathcal{X} \subseteq \mathbb{R}^K$.¹ The history h_t records past public realizations, party campaigns $c_{p,s}$ (each consisting of a policy platform and an interpretive model, defined formally in Section 4), election outcomes $r_s \in \{L, R\}$ (the identity of the period- s winner), and observed policy outcomes $y_s^{\text{obs}} \in \mathcal{Y}$ produced by the incumbent during its time in office (the realized consequences of the winning party's platform under period- s conditions). Observed outcomes are payoff-relevant: they reflect the realized consequences of the incumbent's platform in each period it governs, and are part of the common information available to all voters and parties. Because K can be large, the history h_t quickly becomes high-dimensional, which is precisely why voters demand parsimonious interpretive models: no voter processes the full raw history directly, and which dimensions a model foregrounds will determine which political issues appear salient.

Beliefs over electorate composition. Neither party directly observes individual voter types. Let Θ denote the space of possible electorate compositions, and let

$$\rho_t \in \Delta(\Theta)$$

¹The event space \mathcal{X} is the full conditional support of realizations given the history x_{t-1} , which naturally includes events assigned near-zero probability under a voter's default economic model but positive probability under a culturally motivated causal model. Orchestrated or amplified cultural signals, such as those produced by a coordinated propaganda apparatus, are accommodated as realizations in the tail of this support: the party's campaign model $m_{p,t}$ assigns them high likelihood precisely because it foregrounds the cultural dimension. Manufactured narratives that accompany such signals enter the analysis through the same mechanism: a narrative whose causal structure assigns high probability to otherwise-improbable events can shift the posterior-odds comparison in the voter's model-adoption step even when the underlying event has low probability under the default model. For a formal treatment of false-narrative construction and its connection to posterior-odds manipulation, see Eliaz et al. [2025].

denote the prior over electorate composition at date t . The prior ρ_t is updated from observable electoral history—vote shares, turnout patterns, and realized outcomes recorded in h_t —and is common knowledge across both parties. Under simple majority, each party chooses its campaign to maximize expected vote share under ρ_t .

We assume throughout that ρ_t is common knowledge across both parties. Allowing heterogeneous party priors over Θ would introduce a distinct miscalculation mechanism—a party that overestimates the size of the cross-pressured segment invests too heavily in cultural campaigning, while one that underestimates it may cede the narrative space to its opponent at low cost. Although such misperceptions are empirically plausible and can reinforce the voter trap through a compounding information failure, we set them aside here for two reasons. First, common priors establish a conservative baseline: the voter trap result does not require any asymmetry in how parties read the electorate—it arises even when both parties have identical and correct beliefs about voter composition. Second, the distinct logic of miscalculation-driven traps warrants its own treatment; we set it aside in the present analysis.

Mechanism overview. The remainder of the model proceeds in stages, each building on the last. Each party’s observable history of platforms, narratives, and governing actions accumulates into a *brand state*, which determines which new campaigns voters will find credible (Section 3). At each date, parties propose *campaigns*—pairs of a policy platform and an *interpretive model*. An interpretive model is a parsimonious account of which forces in the high-dimensional public environment are most relevant for understanding the choice between the two parties; competing models foreground different dimensions of the same public events. The set of campaigns a party may credibly run is restricted to those that are coherent with its brand history, consistent with publicly salient evidence, and capable of building a majority coalition—we call these the party’s *live campaigns* (Section 4). Voters evaluate competing models by comparing their fit to the public evidence against the fit of the model they currently hold, weighted by the proposing party’s brand credibility; the model that clears this bar is adopted (Section 5). The adopted model determines which policy dimensions are *salient*—both materially important and surprising relative to existing beliefs—and thus how much weight the voter places on economic versus cultural identity (Section 6). Her effective policy ideal is a weighted average of her two reference identity ideals, and she votes for the party whose platform is closest to it. The *voter trap* arises when this chain of model adoption and salience shift leaves a cross-pressured voter voting against her benchmark economic interests, in a configuration that is locally stable and dynamically self-reinforcing (Section 7).

Party objectives. Both parties are *office-seeking*: each maximizes its discounted expected stream of electoral victories,

$$U_p = \mathbb{E} \left[\sum_{t \in \mathcal{T}_e} \beta^t \cdot \mathbf{1}(p \text{ wins at } t) \right], \quad \beta \in (0, 1), \quad (5)$$

taking the opponent’s strategy and the voter adoption rule as given. Winning at t means implementing platform $Y_{p,t}$ and holding office for T periods until the next election. Parties are forward-looking: today’s campaign choice affects the brand state $B_{p,t+1}$, which constrains the set of live campaigns available at future dates. This forward-looking incentive is what gives party R a motive to invest in cultural campaigning even when a single-period analysis might not justify it: the brand built today is the coherence advantage exploited tomorrow. The office-seeking payoff (5) also clarifies why electoral feasibility $\mathcal{F}_{p,t}^{\text{elec}}$ is a binding constraint: a campaign that does not achieve an expected majority has zero probability of delivering the office payoff, so no forward-looking party would run it.

3 Brand States and Bayesian Brand Dilution

Brand state. For each party $p \in \{L, R\}$ and date t , let

$$B_{p,t} \in \mathcal{B}$$

denote the party’s *brand state*. The brand state is a publicly observed summary of party p ’s history: past policy platforms, past interpretive models, and realized governing actions while in office. Because $B_{p,t}$ is derived entirely from the public history h_t , it is common knowledge among all voters and parties.

We require $(\mathcal{B}, d_{\mathcal{B}})$ to be a compact metric space. In the simplest case, $B_{p,t}$ is a scalar summarizing where party p sits on a left-right continuum, updated campaign by campaign. At the other extreme, it could be a high-dimensional latent vector encoding the full texture of the party’s messaging history: which issues it emphasized, which arguments it deployed, and with which social groups it aligned itself. What matters for our results are not the details of the representation but the structural primitives introduced next: a brand-image mapping, four properties on the update rule, and two properties on the coherence prior.

Brand image. Each campaign $c \in \mathcal{Y} \times \mathcal{M}$ has a *brand image*

$$\phi(c) \in \mathcal{B}, \quad (6)$$

representing the ideological fingerprint that the campaign leaves in the party’s public identity. A party that repeatedly runs cultural campaigns (foregrounding racial, religious, or nationalist themes) converges to a brand state near $\phi(c^C)$ encoding “this is a cultural party”; a party that runs economic campaigns converges near $\phi(c^E)$. What matters for future credibility is not the full richness of the campaign—its specific slogans, endorsements, or media strategy—but the ideological direction it signals, which is what ϕ extracts. The brand image defines the target toward which the brand state moves when a given campaign is run.

Historical support. The *historical support* of party p ’s brand at date t is

$$\text{supp}(B_{p,t}) := \left\{ \phi(c_{p,s}) : s < t \right\} \subseteq \mathcal{B},$$

the set of brand images of all campaigns party p has run up to date t . Since the full campaign history is part of the public record h_t , the historical support is common knowledge.

Brand-state dynamics. The brand state evolves according to

$$B_{p,t+1} = G_p(B_{p,t}, c_{p,t}, y_{p,t}^{\text{obs}}), \tag{7}$$

where $c_{p,t} = (Y_{p,t}, m_{p,t})$ is the campaign chosen by party p at date t and $y_{p,t}^{\text{obs}}$ records governing actions observed while party p holds office. We impose four maintained properties on G_p .

- G1. Reinforcement.** Running a campaign coherent with the current brand strengthens future coherence for that campaign type: if $c_{p,t}$ is in the historical support of $B_{p,t}$, the coherence prior assigned to campaigns of that type is weakly higher under $B_{p,t+1}$ than under $B_{p,t}$.
- G2. Persistence.** Brand states have long memory. G_p places positive weight on all lags of the party’s campaign history, with weights that decline in lag order. Recent campaigns receive more weight than distant ones, but no past campaign is fully forgotten.
- G3. Separation.** Party brands are independent: G_L depends only on L ’s own campaign history, and G_R depends only on R ’s. The brand state of one party does not enter the dynamics of the other. G3 rules out the possibility that party R ’s cultural campaign directly damages party L ’s economic brand. If such cross-party damage operated in the same direction as within-party brand deepening—weakening L ’s economic coherence without undermining R ’s cultural credibility—it could only deepen the asymmetry and strengthen the trap. The opposite case, in which R ’s cultural campaigning induces

voter skepticism of all cultural signaling including R 's own, would attenuate the trap. We treat this as empirically dominated in the episodes we study but acknowledge it as a potential limitation of G3.²

G4. Monotonicity. The coherence prior $\Pi_{p,t}(c \mid B_{p,t})$ is weakly decreasing in the distance metric $d_{\mathcal{B}}(\phi(c), \text{supp}(B_{p,t}))$: campaigns whose brand image is further from the party's historical support receive weakly lower coherence scores.

Properties G1–G4 formalize empirically supported observations about how brand credibility accumulates—that repeated campaign choices compound into coherence (G1–G2), that cross-party brand dynamics are negligible (G3), and that off-brand campaigns receive discounted credibility scores (G4).³ The paper's theoretical contribution is to identify the sufficient state conditions under which a voter trap arises *given* these axioms, not to derive the axioms from a deeper information-theoretic primitive.⁴

EMA parametrization. A natural functional form for G_p consistent with G1–G4 is the exponential moving average (EMA):

$$B_{p,t+1} = (1 - \delta) B_{p,t} + \delta \phi(c_{p,t}), \quad (8)$$

²The “narrative fatigue” channel operates through voter attention rather than party brand states: if prolonged cultural-signal saturation reduces the salience responsiveness η over time, the cultural identity weight $\Delta_{LC,C,t}$ becomes less responsive to campaign signals and the trap weakens. In our framework, this corresponds to a time-varying η_t rather than a violation of G3 (which governs cross-party brand contamination). The distinction matters: G3 concerns the production of brand capital, while narrative fatigue concerns the demand-side response to it.

³Each axiom finds empirical support in the party brand literature. G1 (reinforcement) is consistent with Petrocik 1996's issue-ownership evidence: parties that repeatedly emphasize an issue gain a measurable credibility advantage on it over opponents who do not. G2 (persistence) is consistent with the long-run brand stability documented in manifesto data by Budge 1994: party issue emphasis is strongly autocorrelated across election cycles. G3 (separation) is consistent with the same issue-ownership evidence: credibility advantages accumulate on separate issue domains with negligible cross-party transfer, so a party's campaign on its owned dimension does not alter the opposing party's established credibility on its own issues [Petrocik, 1996]. G4 (monotonicity) is consistent with the logic of long-run reputation [Kreps, 1990]: a party that has historically committed to a particular campaign type acquires credibility precisely because its track record makes that campaign predictable; campaigns outside that track record lack this backing and therefore receive discounted coherence scores.

⁴The additive-in-log-associations structure underlying our posterior-odds score (Section 5) is an instance of a context-dependent utility representation in the tradition of Apestegua and Salvanti [2026]. Our voter's adoption rule (16) is deterministic (argmax) rather than stochastic (logit), so the axiomatic characterizations of Apestegua and Salvanti [2026] do not apply directly; they do, however, provide a natural stochastic analogue in which $\eta \rightarrow \infty$ recovers our argmax. The voter trap's distinct contribution is to make the weights over dimensions (the analogue of the exogenous topic weights σ in Apestegua and Salvanti 2026) endogenous to salience, and to constrain the supply of credible contexts through brand-state dynamics.

where $\delta \in (0, 1)$ is the memory decay rate and ϕ is the brand-image mapping (6). The EMA requires \mathcal{B} to be a convex subset of a normed space (so that the convex combination is well-defined); in the general framework, \mathcal{B} need only be a compact metric space. The EMA is adopted where needed for three reasons: it satisfies G1–G4 with a single free parameter δ ; it yields closed-form iteration under repeated play of a fixed campaign c ($B_{p,t+T} = (1 - \delta)^T B_{p,t} + [1 - (1 - \delta)^T] \phi(c)$); and it is a contraction mapping with fixed point $\phi(c)$, ensuring that brands converge to a well-defined steady state.

Scope of the EMA parametrization. The main results divide according to whether they use the EMA. Proposition 1, Lemma 1, and Lemma 2 use only the abstract G1–G4 and P1–P2 primitives; any update rule satisfying these axioms delivers the same qualitative conclusions. Proposition 2 and Corollary 2 additionally exploit the EMA’s geometric-discounting formula $B_{p,t+T} = (1 - \delta)^T B_{p,t} + [1 - (1 - \delta)^T] \phi(c)$ to establish convergence rates and the absorbing-trap threshold; extending these results to other update rules requires verifying analogous contraction properties case by case. The EMA’s geometric structure is empirically grounded: voters weight recent party actions more heavily than distant ones [Fiorina, 1981, Healy and Lenz, 2014].

Forward-looking consistency of EMA dynamics. The EMA dynamics are consistent with forward-looking party optimization rather than myopia. A forward-looking party L might appear to have an incentive to run occasional off-brand cultural campaigns to maintain a “mixed brand” and preserve option value on the cultural dimension. Axiom G3 (brand separation) eliminates this incentive: a mixed-brand campaign by party L dilutes its economic coherence via Lemma 1 without generating cultural credibility, since $B_{L,t} \ll B_{R,t}$ ensures that the posterior-odds comparison on the cultural dimension is lost regardless of how frequently L invests in cultural campaigns. Under G3, the EMA fixed point of a mixed strategy is an intermediate brand state that weakens party L ’s competitive advantage on the economic dimension without building credibility on the cultural one. The option value of a mixed brand is therefore zero for a party whose only winning dimension is economic coherence with LC voters, and the EMA dynamics reflect the optimal forward-looking campaign policy given this structural constraint.

Brand state as a sufficient statistic. The brand state $B_{p,t}$ is a public sufficient statistic of the campaign history h_t with respect to the coherence prior $\Pi_{p,t}$: by G1–G4, only the summary $B_{p,t}$ is payoff-relevant for future credibility assessments—the full sequence of past campaigns affects future credibility only through its contribution to $B_{p,t}$. This maintained tractability assumption (G3 in particular) means voters need only track the current brand state rather than an infinite campaign history. The EMA is a natural aggregation rule for

this purpose: it discounts the past geometrically and depends on a single free parameter δ governing memory depth.⁵

Coherence prior. Given brand state $B_{p,t}$, voters assign a common *coherence prior*

$$\Pi_{p,t}(c \mid B_{p,t}) \in [0, 1] \quad (9)$$

to campaign $c \in \mathcal{Y} \times \mathcal{M}$. This is the prior probability that party p can and will sincerely run and implement campaign c , given what the party has publicly been. Since $B_{p,t}$ is common knowledge, the coherence prior is shared across all voters.⁶ We impose two maintained properties on $\Pi_{p,t}$, complementing G1–G4.

P1. Limit credibility. For any campaign c whose brand image $\phi(c)$ lies in $\text{supp}(B_{p,t})$, the coherence prior approaches certainty as the brand state concentrates on that image: $\Pi_{p,t}(c \mid B_{p,t}) \rightarrow 1$ as $d_{\mathcal{B}}(B_{p,t}, \phi(c)) \rightarrow 0$.

P2. Ordinal agreement. The coherence prior ranks campaigns consistently: $\Pi_{p,t}(c \mid B_{p,t}) > \Pi_{p,t}(c' \mid B_{p,t})$ whenever c is closer to the party’s historical brand support than c' . Under the common-prior specification, this is automatic. Under the heterogeneous extension (see footnote above), P2 requires that all voters agree on the ranking, though they may disagree on magnitudes.

Our main results require only that G4 (monotonicity) and P2 (ordinal agreement) jointly imply $\Pi_{p,t}(c_R \mid B_{R,t}) > \Pi_{p,t}(c_L \mid B_{L,t})$ whenever party R runs a cultural campaign and party L runs an economic one. Any coherence prior consistent with P1–P2 delivers the same qualitative conclusions.

Brand-feasible campaigns. The set of campaigns that are credible options for party p at date t is

$$\mathcal{F}_{p,t}^{\text{brand}} := \{c \in \mathcal{Y} \times \mathcal{M} : \Pi_{p,t}(c \mid B_{p,t}) > 0\}. \quad (10)$$

⁵The brand state $B_{p,t}$ plays the role of a “reputation state” in the tradition of Kreps [1990] and Mailath and Samuelson [2006]: it is the payoff-relevant summary of the party’s campaign history that governs future voter beliefs, analogous to how a firm’s reputation summarizes its past actions for customers. Unlike standard reputation models, however, $B_{p,t}$ is public (not the firm’s private type) and updates deterministically via the campaign history rather than through noisy performance signals.

⁶One could allow voter-level heterogeneity in coherence priors, indexed $\Pi_{p,t}$, with voters differing in how they weight brand history. Postulate P2 (ordinal agreement) would require all voters to agree on the ranking of campaigns by coherence. Since our results use only the ordinal ranking $\Pi_{p,t}(c_R \mid B_{R,t}) > \Pi_{p,t}(c_L \mid B_{L,t})$ for cultural vs. economic campaigns, heterogeneous magnitudes with ordinal agreement would not change qualitative conclusions. We adopt the common-prior notation for simplicity.

Section 4 imposes one further feasibility condition—*electoral feasibility*, which requires the campaign to command an expected majority under ρ_t —to define the full set of live campaigns $\mathcal{F}_{p,t} \subseteq \mathcal{F}_{p,t}^{\text{brand}}$. The brand-feasibility threshold (10) is the only explicit restriction on the campaign space. One might expect an additional “narrative feasibility” constraint requiring campaigns to address dimensions already prominent in the public evidence, so that a purely economic campaign cannot ignore a salient cultural shock. The model does not impose such a restriction explicitly, because the posterior-odds adoption rule in Section 5 generates this discipline endogenously: a campaign that omits dimensions already salient in voters’ prior models scores low in the PO comparison and is therefore not adopted—making explicit exclusion redundant.

Bayesian brand dilution. A campaign $\tilde{c}_{p,t}$ constitutes a *brand pivot* for party p at date t if it proposes a platform or model that departs from the historical support of $B_{p,t}$, so that $\Pi_{p,t}(\tilde{c}_{p,t} \mid B_{p,t})$ is low for a positive measure of voters.

Definition 1 (Brand-destroying pivot). *A brand pivot $\tilde{c}_{p,t}$ is brand-destroying if there exists $k \geq 1$ such that*

$$\mathbb{E} \left[\Pi_{p,t+k}(c^{\text{hist}} \mid B_{p,t+k}) \mid \text{pivot } \tilde{c}_{p,t} \text{ at } t \right] < \Pi_{p,t}(c^{\text{hist}} \mid B_{p,t})$$

for a positive measure of voters i , where c^{hist} denotes any campaign in the historical support of $B_{p,t}$.

The defining feature of a brand-destroying pivot is its forward cost: by entering $B_{p,t+1}$ through (7), the pivot permanently reduces future coherence priors on the party’s established campaign types, even if the party later attempts to return to its historical platform. We call this mechanism *Bayesian brand dilution*: voters rationally update that a party willing to deviate from its brand once may do so again, making all future campaigns less credible.

Lemma 1 (Off-brand pivots are brand-destroying). *Under G1–G4 and P2, any off-brand pivot $\tilde{c}_{p,t} \notin \mathcal{F}_{p,t}^{\text{brand}}$ is brand-destroying in the sense of Definition 1.*

Two consequences of brand dilution deserve emphasis. First, *incumbency advantage* emerges endogenously. During politically quiet inter-election periods, the public realization x_t generates small fit differentials between competing models, so the posterior-odds comparison is dominated by coherence priors. A challenger who copies the incumbent’s platform does not automatically tie the race: the incumbent’s continuation campaign is more coherent with its own brand state, and copying it earns only a lower coherence score. Incumbency advantage is thus a consequence of brand-state persistence under G2, not an assumption.

Second, brand dilution operates *asymmetrically* across parties whenever their brand states are anchored to different regions of the campaign space. A party with a long history of campaigns at one extreme faces particularly large brand-destruction costs from pivoting: the coherence prior on any such pivot is low not because voters rule it out in principle, but because the public record provides strong evidence against it. A party anchored to an economic brand cannot credibly pivot to cultural campaigning without destroying future coherence on its established platform, while its opponent, whose brand already includes cultural appeals, faces no such cost. This asymmetry has direct implications for the voter trap, characterized formally in Section 7.

4 Campaigns and Live Campaigns

Voters face an enormous public record, a high-dimensional history h_t of economic outcomes, political events, and governing actions, that no one processes in full. What organizes this record into something actionable is an *interpretive model*: a parsimonious account that foregrounds certain forces and dimensions while abstracting from others. Parties compete not only over policy platforms but over which model voters should use to read the world, since the model a voter adopts determines which issues appear salient and thus how she votes. The key insight, formalized below, is that model selection and issue selection are the same decision: there is no stage at which a party first chooses an issue and then designs a model—the model itself encodes the emphasis.

Campaigns. A campaign by party p at date t is a pair

$$c_{p,t} = (Y_{p,t}, m_{p,t}) \in \mathcal{Y} \times \mathcal{M},$$

consisting of a policy platform $Y_{p,t} = (\tau_{p,t}, q_{p,t})$ and an *interpretive model* $m_{p,t}$.

The model space. A model is an element $m \in \mathcal{M}$ that assigns a marginal likelihood to the observed public history and current realization (h_t, x_t) . Different models may foreground different dimensions of x_t , organize evidence through different latent forces, and impose different lag structures on the raw history. We do not impose a parametric form on \mathcal{M} ; the posterior-odds comparison in Section 5 naturally favors more parsimonious models once nuisance parameters and irrelevant variation are integrated out. The marginal likelihood $\mathcal{L}_{i,t}(m)$ integrates over all parameters of model m , penalizing models that spread probability mass over dimensions of x_t that fail to materialize—a Bayesian Occam’s razor that limits

the effective model space at any date to parsimonious models consistent with the realized evidence. This approach to modelling models connects to two strands of the persuasion literature that our framework bridges. Schwartzstein and Sunderam [2021] study an *ex-post* model persuasion problem: the sender observes the data realization and then proposes an interpretive model to fit it, exploiting the receiver’s willingness to update toward well-fitting frames. Aina [2025], by contrast, studies *ex-ante* model design: the sender commits to an interpretive framework before the realization, tailoring the model to the anticipated distribution of signals. In our setting, the brand state $B_{p,t}$ bridges these two perspectives. The historical record encoded in $B_{p,t}$ represents an ex-ante commitment: the party has, over many periods, effectively pre-committed to a class of models through its repeated campaign choices. Within that committed class, the current campaign selects the model that best fits the realized x_t —the ex-post component. Brand coherence thus functions as an endogenous constraint on the ex-post optimization, linking the two approaches in a single dynamic framework.

More broadly, our model belongs to the literature on narrative adoption initiated by Shiller [2017], who argues that narratives spread through economies and polities like epidemics, shaping beliefs and behavior beyond what fundamentals alone can explain. Our contribution is to provide a decision-theoretic foundation for *why* voters adopt narratives—via the posterior-odds comparison—and for *why some narratives are live options while others are not*—via the coherence prior and the live campaign set. Issue selection and model selection are the same decision: a party that announces an issue emphasis without a model that fits the corresponding components of x_t will lose the posterior-odds contest, so the fit comparison voters perform disciplines what can be credibly claimed.

Voter’s prior model. Before the current campaign, voter i holds a *prior model*

$$m_{i,t}^0 \in \mathcal{M},$$

which is the model she currently uses to organize public information. The prior model reflects accumulated public evidence and prior campaign exposures recorded in h_t . Parties do not observe individual prior models, but the distribution of $m_{i,t}^0$ across the electorate is informed by common knowledge: after party p wins office with a campaign built around model $m_{p,s}$, it becomes common knowledge that a substantial share of the electorate holds models tilted in that direction.

Fit. For any candidate model m , define its *fit* as the marginal likelihood of the public evidence:

$$\mathcal{L}_{i,t}(m) := p_m(h_t, x_t). \quad (11)$$

The voter’s prior model $m_{i,t}^0$ provides the benchmark fit $\mathcal{L}_{i,t}(m_{i,t}^0)$ against which any proposed campaign model is evaluated.

Electoral feasibility. A campaign $c = (Y, m)$ must satisfy one condition beyond brand feasibility to be a live option for party p at date t : it must be electorally viable. Brand feasibility was defined in Section 3; we introduce electoral feasibility here. The discipline that campaigns cannot ignore dimensions already salient in voters’ current models emerges from the posterior-odds competition in Section 5 rather than being imposed as a separate constraint. After party L wins on an economic narrative, the distribution of voters’ prior models $m_{i,t}^0$ is tilted toward the economic components of x_t . A model that assigns negligible weight to those dimensions will have a low Bayes factor relative to $m_{i,t}^0$ for economically minded voters and will lose the posterior-odds contest. Party R ’s optimal response is therefore a *mixed model*: one that acknowledges the economic components of the current realization while reweighting toward the cultural components where x_t carries a salient cultural signal. The strength of that cultural signal disciplines how far R can shift the narrative emphasis, through the fit comparison voters perform rather than by assumption.

Electoral feasibility. A campaign c is *electorally feasible* for party p at date t if it is expected to generate a majority under the common prior ρ_t :

$$\mathcal{F}_{p,t}^{\text{elec}} := \left\{ c \in \mathcal{Y} \times \mathcal{M} : \mathbb{E}_{\rho_t}[\text{VS}_p(c)] > \frac{1}{2} \right\}. \quad (12)$$

where $\text{VS}_p(c)$ denotes the vote share of party p under campaign c . A campaign that shifts the cross-pressured segment toward p but loses enough voters elsewhere to fall below majority is not a live option. Throughout, $\text{VS}_p(c)$ is computed holding the opponent’s campaign fixed at its equilibrium strategy σ_{-p}^* ; the electoral feasibility condition is therefore a best-response requirement that is well-defined within the MPE concept of Section 7.

Definition 2 (Live campaign set). *The live campaign set for party p at date t is*

$$\mathcal{F}_{p,t} := \mathcal{F}_{p,t}^{\text{brand}} \cap \mathcal{F}_{p,t}^{\text{elec}}. \quad (13)$$

A campaign $c_{p,t}$ is live for party p at date t if and only if $c_{p,t} \in \mathcal{F}_{p,t}$.

Brand feasibility constrains what the party can credibly commit to given who it has historically been. Electoral feasibility constrains which campaigns can build a majority coalition

under simple majority. Together they define the space of campaigns that are simultaneously credible and coalition-viable; the posterior-odds competition of Section 5 then determines which campaigns within $\mathcal{F}_{p,t}$ can actually shift voter beliefs.

The conditions that make a cultural campaign live for party R are precisely the conditions that make a cultural counter-campaign infeasible for party L . Brand feasibility permits the cultural campaign for R , whose historical record includes cultural appeals, while blocking it for L , whose brand is anchored to an economic platform. The same cultural signal in x_t that raises the Bayes factor of R 's mixed model is insufficient to rescue a cultural model proposed by L , whose prior voters already expect economic emphasis and whose adoption would fail the posterior-odds comparison for too large a share of the electorate. This asymmetry is the structural foundation of the voter trap result in Section 7.

5 Posterior-Odds Model Adoption

Posterior-odds score. Voter i evaluates party p 's proposed model $m_{p,t}$ by computing its posterior odds relative to the prior model $m_{i,t}^0$, weighted by the coherence of the associated campaign:

$$\text{PO}_{ip,t}(m_{p,t}) := \frac{\mathcal{L}_{i,t}(m_{p,t})}{\mathcal{L}_{i,t}(m_{i,t}^0)} \times \Pi_{p,t}(c_{p,t} \mid B_{p,t}). \quad (14)$$

The first factor is the Bayes factor of the proposed model against the prior model: it measures how much better $m_{p,t}$ fits the current public evidence relative to what the voter already believed. The second factor is the coherence prior from Section 3: it discounts the proposed model by the probability that party p can and will sincerely implement it.⁷ A model with high evidential fit but low brand coherence scores poorly; so does a brand-coherent model that fits the evidence no better than the voter's existing model.

For later reference, we define shorthand for the two components of the posterior-odds score evaluated for voter LC and party p 's campaign:

$$\Lambda_{p,t} := \frac{\mathcal{L}_{LC,t}(m_{p,t})}{\mathcal{L}_{LC,t}(m_{LC,t}^0)}, \quad \pi_{p,t} := \Pi_{p,t}(c_{p,t} \mid B_{p,t}). \quad (15)$$

⁷The score admits an intuitive decomposition into an evidence channel $\Lambda_{p,t}$ and a brand-coherence channel $\pi_{p,t}$. The logit (Luce) structure is standard; Apesteguia and Salvanti [2026] show (Theorem 1) that it is the unique structure, within a fixed context, in which the log-ratio of choice probabilities depends only on a pairwise score and not on what other alternatives are available. Under that Luce structure, the *additivity* of the log-score into $\ln \Lambda_{p,t} + \ln \pi_{p,t}$ —and hence the product form $\text{PO}_{ip,t} = \Lambda_{p,t} \cdot \pi_{p,t}$ —follows from the substantive assumption that evidence fit and brand coherence enter the voter's evaluation as *multiplicatively separable* channels. We specify model adoption as a deterministic argmax, which corresponds to the $\eta \rightarrow \infty$ limit of a Luce rule over $\{m_{i,t}^0, m_{L,t}, m_{R,t}\}$ with scores $\ln \text{PO}_{ip,t}$.

$\Lambda_{p,t}$ is the Bayes factor of party p 's proposed model against voter LC 's prior model; $\pi_{p,t}$ is the brand coherence of p 's campaign. The posterior-odds score for voter LC is then $\text{PO}_{LC,p,t}(m_{p,t}) = \Lambda_{p,t} \pi_{p,t}$.

Model adoption. Voter i adopts the campaign model that achieves the highest posterior-odds score, provided it beats the prior model:

$$m_{i,t}^{\text{eff}} \in \text{argmax} \left\{ 1, \text{PO}_{iL,t}(m_{L,t}), \text{PO}_{iR,t}(m_{R,t}) \right\}. \quad (16)$$

The outside option 1 corresponds to retaining the prior model $m_{i,t}^0$: a campaign model is adopted only if it strictly improves on the voter's current model of the world. If neither party's model clears the bar, the voter remains with $m_{i,t}^0$ and her effective bliss point is determined by that model in Section 6.

Three features of the adoption rule (16) are worth noting. First, equation (16) describes competition between fully specified interpretive models, not between signals about a fixed underlying state. This departs from standard Bayesian persuasion [Kamenica and Gentzkow, 2011, Bergemann and Morris, 2019], where a sender designs a signal structure to shift receiver beliefs about a common unknown, and from competing-narratives frameworks [Eliaz and Spiegler, 2020], which model narrative competition as a contest over the causal structure voters use to assign responsibility. Here, parties propose rival models that differ in which forces they posit as organizing the public evidence; voters update by comparing model fit weighted by brand credibility, and the winning model is adopted wholesale.

Second, the GT belief distortion [Gennaioli and Tabellini, 2025] is a limiting case in which \mathcal{M} is restricted to identity-defined models and the effectiveness of cultural appeals is governed by a fixed propaganda intensity χ that scales the identity weight directly. Our formulation departs from GT along two dimensions: the model space is unrestricted, allowing continuous mixtures of economic and cultural frames; and the effectiveness of cultural campaigns is endogenous through $\Pi_{p,t}$ (the brand-coherence prior that determines whether a model is adopted at all) rather than fixed exogenously by χ .

Third, the posterior-odds comparison in (14) is computed by each voter individually against her own prior model $m_{i,t}^0$. How other voters interpret the same public events does not affect whether a model is adopted, but it does affect the effective salience once a model is adopted—because the material stakes on each dimension in Section 6 depend on the parties' platforms, which are themselves shaped by beliefs about who is persuadable. This is why the voter trap can be self-reinforcing even when its mechanics are common knowledge: the common knowledge of brand constraints and voter model distributions is already factored

into the set of live campaigns, leaving no room for a deviation that would dissolve the trap.

6 Saliency, Identity Weights, and Effective Bliss Points

Issue-specific stakes. Let $\ell \in \{E, C\}$ index the economic and cultural dimensions. Define the *stake* of dimension ℓ for voter i under model m as the model-implied expected payoff difference between the two parties on that dimension, where the expectation is over *realized policy outcomes* given each party’s announced platform:

$$\Omega_{i\ell,t}(m) := \left| \mathbb{E}_m \left[u_{i\ell}(Y_{L,t}^*) - u_{i\ell}(Y_{R,t}^*) \mid h_t, x_t \right] \right|, \quad (17)$$

where $u_{i\ell}$ is the dimension- ℓ component of voter i ’s utility and $Y_{p,t}^*$ denotes the policy outcome realized when party p implements its announced platform in the environment characterized by model m and public realization x_t . The expectation is over the model-specific distribution over outcomes—not over the platforms themselves, which are publicly observed—and captures how voter i expects each platform to translate into consequences on dimension ℓ under different interpretive frameworks. Each interpretive model m thus serves a dual role: it organizes the public evidence x_t (determining fit in the PO comparison) and it maps announced platforms into expected policy consequences (determining stakes). A model that foregrounds cultural forces will both explain cultural components of x_t well and predict that cultural policy platforms have large consequences—linking the fit and saliency channels through a single interpretive object. Stakes are high on dimension ℓ when, under model m , the two parties are expected to produce materially different outcomes on that dimension. The baseline stakes are purely material; see Section 10 for a discussion of social stakes.

Saliency. For any model $m \in \mathcal{M}$ and dimension $\ell \in \{E, C\}$, let $p_{m,\ell}(x_{\ell,t} \mid h_t)$ denote the marginal likelihood of the ℓ -th component of x_t under model m , obtained by integrating out all dimensions $\ell' \neq \ell$. Under model m , define the *saliency* of dimension ℓ for voter i at date t as

$$S_{i\ell,t}(m) := \Omega_{i\ell,t}(m) \cdot \left| \log \frac{p_{m,\ell}(x_{\ell,t} \mid h_t)}{p_{m_{i,t}^0,\ell}(x_{\ell,t} \mid h_t)} \right|. \quad (18)$$

Saliency rises whenever the candidate model and prior model disagree about the public realization, regardless of whether the candidate finds it more or less likely. This design implies that an unusually strong economic realization—one the cultural model fails to explain—can raise economic saliency and erode the voter trap; the conditions under which such dissolution occurs are formalized in Definition 4 and Corollary 2. The second factor is the absolute log-

likelihood ratio of the proposed model against the voter’s prior model, restricted to the ℓ dimension. The absolute value ensures that a dimension becomes salient whenever the two models disagree sharply about $x_{\ell,t}$, regardless of direction: a dimension is salient both when the adopted model finds the realization much *more* likely than the prior (the new model explains the evidence better) and when it finds it much *less* likely (the dimension is contentious—the models disagree about what the evidence means, drawing voter attention to that dimension). Saliency is thus the product of stakes and model disagreement.⁸ A dimension is salient when it is both materially important *and* when the two models offer sharply different accounts of the current public realization on that dimension.

Continuous identity weights. Under effective model $m_{i,t}^{\text{eff}}$, voter i places weights

$$\Delta_{i,t} = (\Delta_{iE,t}, \Delta_{iC,t}) \in [0, 1]^2, \quad \Delta_{iE,t} + \Delta_{iC,t} = 1,$$

on the economic and cultural dimensions, determined by a softmax over saliency:

$$\Delta_{i\ell,t}(m_{i,t}^{\text{eff}}) = \frac{\exp\{\eta S_{i\ell,t}(m_{i,t}^{\text{eff}})\}}{\sum_{\ell' \in \{E,C\}} \exp\{\eta S_{i\ell',t}(m_{i,t}^{\text{eff}})\}}, \quad (19)$$

where $\eta > 0$ is the voter’s *saliency responsiveness*.⁹ A voter with high η concentrates identity weight sharply on the most salient dimension; a voter with low η mixes the two identities even when saliency is asymmetric. We treat η as a voter-level parameter identified by the curvature of identity switching: agents with higher identity flexibility (or lower discomfort from holding mixed identities) have lower η . In the present formulation, the softmax over saliency is the sole determinant of identity weights; social enforcement mechanisms (collective action costs, social punishment for identity deviants) would reinforce the trap by raising the community-level costs of identity switching, but are not modeled here.

⁸This definition shares the spirit of Bordalo et al. [2012, 2013]: a dimension is salient when it stands out relative to a reference point, here the voter’s prior model $m_{i,t}^0$ rather than a reference lottery. The key structural difference is that our saliency is model-conditional—computed separately for each candidate model m —and resolved once the voter selects her effective model via the PO adoption rule, so the saliency of a dimension can change discretely when a new model is adopted.

⁹The softmax form of (19) is a standard Luce aggregator over the two dimensions $\{E, C\}$: voter weight on dimension ℓ is proportional to $\exp(\eta S_{i\ell,t})$, where η governs how sharply voters concentrate on the most salient dimension. Apesteguia and Salvanti [2026] study a related contextual utility structure in which a similar softmax describes *choice* probabilities over parties given an *exogenous* vector of topic weights σ ; our setting differs in that (a) the voter’s choice rule at the model-adoption stage is deterministic (argmax), and (b) the weights over dimensions here are *endogenous* outputs of a saliency-driven softmax, rather than primitive inputs. Note that the saliency scores in (18) are evaluated under the *candidate* model m rather than under $m_{i,t}^{\text{eff}}$, so (19) is determined ex post once model adoption is resolved; there is no fixed-point in the adoption step.

Remark 1 (Relation to GT). As $\eta \rightarrow \infty$, voter i places almost all weight on the dimension with higher salience, recovering the hard identity choice on the voter side of Gennaioli and Tabellini [2025]. The full GT equilibrium is not nested, because GT’s propaganda intensity χ —which scales how effectively cultural appeals shift voter identity weights, and is exogenous and symmetric across parties in GT—has no analog in the brand-constrained framework; what is shared is the voter’s identity-switching mechanism. Finite η allows continuous mixing: a voter can be partially captured by the cultural frame without switching identity entirely—the voter need not fully switch identity to cross the voting cutoff $\bar{\Delta}_{LC}$.

Effective bliss point. The effective policy ideal of voter i at date t is the identity-weight average of her two reference ideals:

$$Y_{i,t}^\Delta := \Delta_{iE,t}(m_{i,t}^{\text{eff}}) Y_i^E + \Delta_{iC,t}(m_{i,t}^{\text{eff}}) Y_i^C. \quad (20)$$

Voting. Write $Y_{i,t}^\Delta = (\tau_{i,t}^\Delta, q_{i,t}^\Delta)$. Because the GT utility (2) is quadratic and separable in τ and q , it can be written as

$$W^{ij}(\tau, q) = -\frac{1}{2}(\tau - \tau_{ij}^*)^2 - \frac{\kappa}{2}(q - q_{ij}^*)^2 + \text{const},$$

so maximizing utility is equivalent to minimizing quadratic loss from the rational bliss point (τ_{ij}^*, q_{ij}^*) . Under model $m_{i,t}^{\text{eff}}$, identity weights shift the effective bliss point from (τ_{ij}^*, q_{ij}^*) to $Y_{i,t}^\Delta$; the voter then evaluates platforms from this new centre using the same loss function:

$$V_{ip,t} := -\frac{1}{2}(\tau_{p,t} - \tau_{i,t}^\Delta)^2 - \frac{\kappa}{2}(q_{p,t} - q_{i,t}^\Delta)^2, \quad (21)$$

where $\kappa > 0$ is the weight on social policy inherited directly from (2). Model adoption shifts *where* the voter stands; κ governs how she evaluates distance from wherever she stands. She votes for the party with the higher score at each $t \in \mathcal{T}_e$.

7 The Voter Trap

Equilibrium. We characterize *Markov Perfect Equilibria* (MPE) of the game between the two parties, with voters as passive (reactive) players. A strategy for party p is a mapping

$$\sigma_p : \mathcal{B} \times \mathcal{B} \times \mathcal{X} \times \Delta(\Theta) \longrightarrow \mathcal{Y} \times \mathcal{M}$$

from the payoff-relevant state $(B_{L,t}, B_{R,t}, x_t, \rho_t)$ into a campaign. An MPE is a profile (σ_L^*, σ_R^*) such that each party’s strategy maximizes (5) given the opponent’s strategy, the voter adoption rule (16), the identity weight formula (19), and the brand-state law of motion (7). Voters are *passive* in the strategic sense: they do not choose strategies in the game tree and take party campaigns as given. However, their model-adoption and voting decisions are individually rational best responses to the information environment—voters maximize posterior-odds scores and vote for the party closest to their effective bliss point. This distinction resolves the apparent tension between passivity and the rational adoption required by condition (iii) of the voter trap definition (Definition 3): passivity refers to strategic non-participation, not to irrationality. The passivity assumption makes the voter trap result *conservative*: if the trap holds when voters take no action beyond voting, introducing social enforcement mechanisms—costly collective action, social punishment for identity deviants—can only widen the trap by raising the community-level costs of identity switching. The model therefore provides a lower bound on trap durability.

Remark 2 (Existence of MPE). *Under the EMA parametrization (8) with a finite model space $|\mathcal{M}| < \infty$ and compact policy space \mathcal{Y} , the brand state $B_{p,t}$ is a finite-dimensional vector, the action space $\mathcal{Y} \times \mathcal{M}$ is compact, and the state space $(B_{L,t}, B_{R,t}, x_t, \rho_t)$ is a compact subset of $\mathbb{R}^d \times \mathcal{X} \times \Delta(\Theta)$. Existence of MPE in stationary strategies then follows from standard fixed-point results for discounted stochastic games with compact state and action spaces (e.g., Maskin and Tirole [2001]), provided the transition kernel induced by the EMA update and the signal distribution is continuous in the state variables. The voter adoption rule (16) introduces a potential discontinuity (the argmax over PO scores), but this is generically resolved when the PO scores are distinct, which holds with probability one under continuous signal distributions. When \mathcal{M} is unrestricted, compactness of the action space is not guaranteed. The general model should therefore be understood as an axiomatic framework that characterizes the mechanism, with the EMA-plus-finite- \mathcal{M} special case (Appendix B) serving as the existence and uniqueness benchmark.*

The MPE concept imposes appropriate discipline here for two reasons. First, strategies depend only on the payoff-relevant state, ruling out equilibria sustained by off-path threats that are not credible given a party’s current brand state—a party cannot credibly threaten a pivot it would never find optimal given its own history. Second, the forward-looking dimension of (5) is captured naturally: party R internalizes that running the cultural campaign today shifts $B_{R,t+1}$ in its favor, deepening the trap at future election dates.

Remark 3 (Scope of results). *The main results below are state-contingent characterizations: they identify sufficient conditions on the state $(x_t, B_{R,t}, B_{L,t})$ under which the voter trap holds,*

given that parties play live campaigns in $\mathcal{F}_{p,t}$. This is a deliberate partial-equilibrium framing. The contribution is to characterize when the trap holds, not to solve for the full equilibrium path that generates each state. Three features of the game-theoretic apparatus are nonetheless load-bearing. First, Remark 5 verifies that the cultural campaign is party R 's best response whenever the trap conditions are satisfiable, so the characterization does not rest on an off-equilibrium premise. Second, the dynamic-deepening result (Proposition 2) requires that R chooses the cultural campaign in equilibrium, because the deepening operates through the EMA update along the equilibrium path. Third, Appendix B develops a fully parametric special case (binary model space, scalar brand state) in which the unique MPE strategies are derived in closed form (Proposition 4) and verified to generate the state conditions of Proposition 1 endogenously. The general model provides the axiomatic framework; the parametric case provides the equilibrium verification.

The benchmark. The voter trap is defined relative to a benchmark vote: what voter LC would do absent any campaign-induced shift in her model adoption. Under the prior model $m_{LC,t}^0$, the economic dimension dominates and LC votes for party L .

Definition 3 (Voter trap). *Voter LC is in a voter trap at election date $t \in \mathcal{T}_e$ if the following four conditions hold simultaneously.*

- (i) **Benchmark vote:** Under the prior model $m_{LC,t}^0$, voter LC votes L —her economic interests dominate and party L is closer on the economic dimension.
- (ii) **Induced trap vote:** Under the equilibrium adopted model $m_{LC,t}^{\text{eff}}$, voter LC votes R —the cultural weight crosses the voting cutoff, $\Delta_{LC,C,t}(m_{LC,t}^{\text{eff}}) > \bar{\Delta}_{LC}$.
- (iii) **Rational adoption:** The adoption of $m_{LC,t}^{\text{eff}}$ is a rational best response by LC given $(B_{p,t}, h_t, x_t)$ and the common knowledge structure— LC is not making an error given what she perceives and knows.
- (iv) **No effective counter-campaign:** No campaign in $\mathcal{F}_{L,t}$ simultaneously achieves higher posterior odds than $m_{LC,t}^{\text{eff}}$ for voter LC and pushes $\Delta_{LC,C,t}$ below $\bar{\Delta}_{LC}$.

Condition (iii) is what distinguishes the voter trap from manipulation or deception. Voter LC knows that party R is running a cultural model to shift her vote; she knows that party L is constrained by brand incoherence; she knows the electoral rules. All of this is common knowledge—and yet she votes R , because the adopted model genuinely fits the current evidence better than her prior, her effective bliss point has genuinely shifted under that model, and L 's inability to counter is itself common knowledge so she cannot condition on

a counter-campaign that will not arrive. The trap is not in LC 's rationality. It is in the information environment, brand structure, and higher-order beliefs jointly, and it is self-reinforcing precisely because common knowledge of it does not dissolve it.

Remark 4 (Welfare criterion). *The voter trap is welfare-reducing in the following precise sense. Consider a symmetric-brand benchmark in which $B_{L,t} = B_{R,t}$, so that both parties can contest both dimensions with equal credibility. Under brand symmetry, neither party has a coherence advantage on cultural campaigns, condition (i) of Proposition 1 fails, and voter LC votes according to her benchmark economic bliss point Y_{LC}^E . The voter trap arises only under brand asymmetry, and voter LC 's expected material payoff (evaluated at the policy outcomes she would choose under undistorted identity weights) is strictly lower under the trap than under the symmetric benchmark. The normative content of the voter trap therefore rests on a counterfactual-information criterion: the voter does worse because informational constraints on party credibility distort the menu of interpretive models available to her, not because she fails to optimize given the menu she faces.*

Lemma 2 (Derived voting cutoff). *Under the policy ordering (1), the voting payoff (21), and Assumption 1, voter LC votes for party R at date $t \in \mathcal{T}_e$ if and only if her cultural identity weight satisfies $\Delta_{LC,C,t} > \bar{\Delta}_{LC}$, where*

$$\bar{\Delta}_{LC} := \frac{(\tau_L - \tau_R)(\tau_{LC}^E - \bar{\tau}_p) - \kappa(q_L - q_R)(q_{LC}^E - \bar{q}_p)}{(\tau_L - \tau_R)(\tau_{LC}^E - \tau_{LC}^C) - \kappa(q_L - q_R)(q_{LC}^E - q_{LC}^C)}, \quad (22)$$

with $\bar{\tau}_p := (\tau_L + \tau_R)/2$ and $\bar{q}_p := (q_L + q_R)/2$ denoting the policy midpoints. The cutoff $\bar{\Delta}_{LC} \in (0, 1)$ under the following joint restrictions on primitives: (a) voter LC prefers party L at full economic identity ($\Delta_{LC,C,t} = 0$, the benchmark vote condition); (b) voter LC prefers party R at full cultural identity ($\Delta_{LC,C,t} = 1$); and (c) the denominator in (22) is strictly positive, ensuring preference is strictly monotone in $\Delta_{LC,C,t}$. In the parametric benchmark calibration of Appendix B ($\tau_L = 1$, $\tau_R = 0$, $q_L = 1$, $q_R = 0$, $\kappa = 1$, voter LC ideals from Table 1), condition (c) holds and the cutoff evaluates to $\bar{\Delta}_{LC} \approx 0.556$, consistent with the simulation results in Figure 1. The closed form (22) relies on the quadratic-loss, separable structure of the utility function (21); alternative specifications (e.g., CES or non-separable preferences) would yield different cutoff expressions while preserving the qualitative prediction of an interior threshold.

Proposition 1 (Voter trap). *Suppose Assumption 1 and Lemma 2 hold, and fix live campaigns $c_{p,t} \in \mathcal{F}_{p,t}$ for $p \in \{L, R\}$. Using the notation (15), let \bar{S}_{LC} denote the salience threshold at which $\Delta_{LC,C,t}$ reaches $\bar{\Delta}_{LC}$ (from Lemma 2). The voter trap holds for voter LC at date t if the following three conditions on primitives are satisfied.*

(i) **Evidential and brand advantage of R:**

$$\Lambda_{R,t} \pi_{R,t} > \max \left\{ 1, \frac{\mathcal{L}_{LC,t}(m_{L,t})}{\mathcal{L}_{LC,t}(m_{LC,t}^0)} \cdot \Pi_{L,t}(c_{L,t} | B_{L,t}) \right\}.$$

Party R's cultural model clears both the evidence bar (Bayes factor) and the credibility bar (brand coherence), and does so better than L's best live model.

(ii) **Saliency amplification:** Under $m_{R,t}$, the cultural saliency for voter LC exceeds the threshold,

$$S_{LC,C,t}(m_{R,t}) > \bar{S}_{LC}.$$

The public realization x_t carries a cultural signal strong enough that adopting $m_{R,t}$ shifts $\Delta_{LC,C,t}$ above $\bar{\Delta}_{LC}$.

(iii) **Brand constraint on L:** No live campaign $c \in \mathcal{F}_{L,t}$ simultaneously achieves

$$\frac{\mathcal{L}_{LC,t}(m)}{\mathcal{L}_{LC,t}(m_{LC,t}^0)} \cdot \Pi_{L,t}(c | B_{L,t}) > \Lambda_{R,t} \pi_{R,t}$$

and

$$S_{LC,C,t}(m) \geq \bar{S}_{LC}.$$

Party L's brand history prevents any campaign within $\mathcal{F}_{L,t}$ from both out-scoring R's model in the PO comparison and sustaining a cultural saliency shift sufficient to dissolve the trap.

Under (i)–(iii), voter LC is in a voter trap in the sense of Definition 3: she votes R even though L remains closer on the economic dimension, and her adoption of $m_{R,t}$ is a rational best response given the common knowledge structure.

Remark 5 (Incentive compatibility of R's cultural campaign). Under the office-seeking objective (5), the cultural campaign is party R's best response whenever conditions (i)–(iii) are satisfiable from the current state. If voter LC is trappable, running the cultural campaign secures LC's vote; the alternative—an economic campaign within R's brand-feasible set—does not shift LC's effective bliss point past the voting cutoff and leaves LC voting L at the benchmark. The cultural campaign therefore dominates any brand-feasible alternative that does not trap LC, and party R strictly prefers it whenever the additional votes from trapping LC are pivotal for a majority. The voter trap is thus not only internally consistent but incentive-compatible for the trapping party: party R endogenously chooses to run the campaign that generates the trap.

One clarification bears on the precise role of brand feasibility in condition (iii). The brand-feasibility threshold (10) is deliberately permissive: a campaign is brand-feasible whenever the average coherence prior is strictly positive. The model does not require brand *infeasibility* to exclude party L 's cultural campaigns. The relevant barrier is the posterior-odds comparison (14): even if L 's cultural pivot clears the brand-feasibility threshold ($\Pi_{L,t} > 0$), its coherence score $\pi_{L,t}$ will be lower than R 's $\pi_{R,t}$ for the same campaign type, so the PO score $\Lambda_{L,t} \cdot \pi_{L,t}$ is dominated by $\Lambda_{R,t} \cdot \pi_{R,t}$. The brand structure endogenizes the set of campaigns L can *win*, not the set it can run.¹⁰

Assumption 2 (Regularity). *The live-campaign correspondence $\mathcal{F}_{L,t}(B_{L,t}, x_t)$ is upper hemicontinuous in $(B_{L,t}, x_t)$.*

Assumption 2 is a standard continuity condition on correspondence-valued maps. It holds, for example, whenever the posterior-odds scoring function is continuous in the state variables, which follows from the EMA Lipschitz property and the continuity of $\Pi_{p,t}$ in $B_{p,t}$.

Corollary 1 (Local stability of the trap). *Suppose the conditions of Proposition 1 hold with strict slack $\epsilon > 0$: specifically, $\Lambda_{R,t}\pi_{R,t} > \max\{1, \mathcal{L}_{LC,t}(m_L)/\mathcal{L}_{LC,t}(m^0) \cdot \pi_{L,t}\} + \epsilon$ and $S_{LC,C,t}(m_{R,t}) > \bar{S}_{LC} + \epsilon$. Define the PO slack function $F_1(c, B, h, x) := \Lambda_{R,t}\pi_{R,t} - \max\{1, \Lambda_{L,t}\pi_{L,t}\}$ and the salience-margin function $F_2(c, B, h, x) := S_{LC,C,t}(m_{R,t}) - \bar{S}_{LC}$, and let*

$$\mathcal{N}_\epsilon := \left\{ (c', B', h', x') : F_1(c', B', h', x') \geq \frac{\epsilon}{2} \text{ and } F_2(c', B', h', x') \geq \frac{\epsilon}{2} \right\}.$$

Whenever G_p is Lipschitz continuous (as under the EMA parametrization (8)) and $\Pi_{p,t}$ is continuous in $B_{p,t}$, the neighborhood \mathcal{N}_ϵ is open and non-empty, and all four voter trap conditions of Definition 3 hold throughout \mathcal{N}_ϵ . Condition (iv) (no live escape) holds throughout \mathcal{N}_ϵ by upper hemicontinuity of $\mathcal{F}_{L,t}$ in $(B_{L,t}, x_t)$ (a maintained regularity condition): any campaign in L 's live set at the perturbed state still faces a posterior-odds score that loses to R 's model by at least $\epsilon/2$. Support for party R by voter LC is therefore a local maximum over the reachable set of campaigns.

The three conditions identify the minimal configuration generating a voter trap. The product $\Lambda_{R,t}\pi_{R,t}$ in condition (i) is the key object: neither evidential fit alone (a culturally salient event interpreted by a party with no cultural brand) nor brand coherence alone (a culturally consistent campaign in a period with no salient signal) suffices. Condition (iii) is

¹⁰The binary brand structure (cultural vs. economic poles) is a simplification. Real party brands encode distinguishable emphases across immigration, trade, and nationalism as separate cultural frames, each with its own coherence history. The core logic does not depend on the binary structure: what the trap requires is that R 's coherence prior on the relevant campaign model exceeds L 's, whether the issue space has two dimensions or twenty.

what makes the outcome a *trap* rather than a *realignment*: L 's brand history prevents any live counter-campaign from simultaneously outperforming R 's model in the PO comparison and sustaining the salience shift. Corollary 1 establishes that this configuration is not a knife-edge: whenever the conditions hold with strict slack, there is an open neighborhood of states within which the trap persists, even against deviations by party L within its live set. This structural infeasibility, not any form of irrationality or manipulation, is the precise sense in which the voter is trapped.

8 Dynamics and Trap Characterization

The static analysis of Section 7 establishes that a voter trap can arise at a given election date. This section asks two further questions about the structure of the trap over time. First: once a voter trap forms, how durable is it? We characterize *stable* traps—configurations in which the conditions for the trap hold with increasing slack over successive election cycles, so that escape requires an increasingly extreme realization. Second: how much does party R need to know about the electorate to sustain the trap? We characterize *robust* traps—configurations in which the trap holds not merely at a single type space Θ but across all type spaces consistent with a finite set of moment conditions on voter composition, in the spirit of Ollar and Penta [2025].

8.1 Dynamic Deepening and Trap Stability

Dynamic deepening. The voter trap need not be a one-period phenomenon. If party R continues running the cultural campaign over multiple inter-election periods, properties G1–G4 imply that $B_{R,t}$ shifts toward the cultural campaign type, raising the coherence prior $\Pi_{R,t+k}(c^C \mid B_{R,t+k})$ for all $k \geq 1$. Simultaneously, Lemma 1 implies that any cultural counter-pivot by party L is brand-destroying, so $\mathcal{F}_{L,t+k}$ does not expand to include credible cultural campaigns. The brand-coherence advantage in Corollary 1's stability neighborhood therefore grows weakly over time: the minimum Bayes factor required for the trap to hold is non-increasing, expanding the set of realizations under which the trap persists.

Proposition 2 (Dynamic deepening). *Suppose the conditions of Corollary 1 hold at election date $t \in \mathcal{T}_e$ with slack $\epsilon > 0$. If party R runs the coalition-optimal cultural campaign at every narrative period between t and $t + T$, then under G1–G4:*

- (a) **Brand deepening:** $\pi_{R,t+T} \geq \pi_{R,t}$ and $\pi_{L,t+T} \leq \pi_{L,t}$, with strict inequality in $\pi_{R,t+T}$ whenever $x_{C,s}$ exceeds the stationary mean of x_C under the equilibrium cultural campaign for at least one $s \in [t, t + T)$.

(b) **Trap deepening:** The minimum Bayes factor $\underline{\Lambda}_t$ required for condition (i) of Proposition 1 to hold is non-increasing: $\underline{\Lambda}_{t+T} \leq \underline{\Lambda}_t$. The set of public realizations under which the voter trap holds therefore expands over successive election cycles.

As $t \rightarrow \infty$ with R running the cultural campaign each period, $B_{R,t} \rightarrow \phi(c_R^C)$ and $\pi_{R,t} \rightarrow 1$ by P1; the trap becomes absorbing in finite time whenever $x_{C,t}$ has bounded support and R 's cultural model is not completely refuted by any realization in the support (i.e., $\Lambda_{R,t} \geq \underline{\lambda} > 0$ for all $x_{C,t}$ in the support).

Stable traps. Proposition 2 characterizes how the trap deepens. We now ask when a trap is *stable* in the strong sense: when dissolving it requires a sequence of realizations that are increasingly implausible given the prevailing model. Define the *cultural-signal advantage* $z_t := x_{C,t} - x_{E,t}$, the difference between the cultural and economic signal components at date t . Higher z_t corresponds to a stronger cultural signal relative to the economic signal, and R 's cultural model is more persuasive when z_t is large. The trap holds when z_t is large enough, and dissolution requires an economic signal strong enough to push z_t below the dissolution threshold.

Definition 4 (Stable trap). *A voter trap at election date t is stable if there exists a threshold $\underline{z}(t) \in \mathbb{R}$ such that (i) the trap conditions hold for all cultural-signal advantages $z_t \geq \underline{z}(t)$, and (ii) the threshold is decreasing over time: $\underline{z}(t+T) \leq \underline{z}(t)$ whenever party R runs the coalition-optimal campaign between t and $t+T$. A trap is strongly stable if $\underline{z}(t) \rightarrow -\infty$ as $t \rightarrow \infty$.*

A stable trap is one in which escape requires progressively more extreme economic events. Under G1–G4, the dissolution threshold $\underline{z}(t)$ is linked to the slack ϵ in Corollary 1: a larger slack means that a larger shift in the cultural realization is needed to move the Bayes factor or salience outside \mathcal{N}_ϵ . Since Proposition 2(b) shows the minimum Bayes factor required for the trap is non-increasing, the dissolution threshold inherits the same monotonicity.

Corollary 2 (Trap stability). *Under the conditions of Proposition 2, every voter trap arising in equilibrium is stable. If, additionally, the cultural-signal advantage z_t has bounded support and the regularity condition of Proposition 2 holds, then the trap is strongly stable and becomes effectively absorbing in finite time.¹¹*

¹¹“Effectively absorbing” means that no realization within the equilibrium signal support $\text{supp}(z_t)$ can dissolve the trap after the finite date at which $\underline{z}(t)$ falls below the support's lower bound. Escape then requires exogenous structural shocks that produce realizations outside this support. Paradigmatically, the 1965 Voting Rights Act altered the feasible campaign set by institutional mandate rather than through any realization within the ordinary signal distribution. The term denotes institutional lock-in: dissolution requires a qualitative change in the political environment, not merely an extreme realization of the ongoing signal process.

The mechanism behind strong stability is the compounding brand asymmetry: as R 's cultural brand deepens via G1, the threshold $\Lambda_{R,t}\pi_{R,t}$ in Proposition 1(i) grows, raising the bar for any counter-campaign by L to outperform R 's model. With bounded support on z_t , there is a finite date after which no realization within the support can generate a counter-model for L that both clears the PO bar and pushes $\Delta_{LC,C,t}$ below $\bar{\Delta}_{LC}$.

Endogenous dissolution. The voter trap dissolves endogenously when a public realization x_t generates a Bayes factor for L 's economic model that exceeds $\Lambda_{R,t}\pi_{R,t}/\pi_{L,t}$ and simultaneously reduces cultural salience below $\bar{\Delta}_{LC}$. By Corollary 2, this requires a cultural-signal advantage z_t below the dissolution threshold $\underline{z}(t)$, which is non-increasing. In a weakly stable trap, such realizations remain possible and dissolution occurs when an especially strong economic signal arrives. In a strongly stable trap, $\underline{z}(t)$ eventually falls below the support of z_t : no realization within the ordinary range of variation can restore L 's credibility gap, and the trap becomes absorbing in finite time. Only an exogenous shock that alters the feasible campaign set—most directly, an institutional intervention that breaks the brand constraint—can then dissolve it. Paradigm-shifting economic crises, such as the Great Depression or the 2008 financial crisis, represent precisely such realizations: by generating public signals of exceptional magnitude in the economic dimension, they push z_t below the dissolution threshold $\underline{z}(t)$, making the economic frame overwhelmingly salient and creating conditions under which L 's economic model can outperform R 's cultural model in the PO comparison. These episodes are precisely the empirically salient events that the model predicts are *necessary* to break a strongly stable trap from within.

8.2 Endogenous Platform Separation

The preceding analysis conditions on the policy ordering (1). A natural question is whether forward-looking parties would converge toward each other's platforms, eliminating the trap through Downsian competition. We show that the brand dynamics generate endogenous, dimension-specific adjustment costs that prevent convergence. Since a party's strategy maps the payoff-relevant state into a *campaign* $c_{p,t} = (Y_{p,t}, m_{p,t})$, the equilibrium platform choice is determined jointly with the narrative model. The brand image $\phi(c_{p,t})$ depends on the full campaign: a platform shift toward the opponent's position moves ϕ away from the party's historical brand support and incurs a brand-dilution cost under G1–G2. Forward-looking parties internalize this cost through the discounted objective (5).

Proposition 3 (Endogenous platform separation). *In the parametric benchmark (Appendix B) with party objective (5) and EMA brand dynamics (8):*

(a) **Cultural party.** When the voter trap is active at state $(x_t, B_{R,t}, B_{L,t})$, party R 's equilibrium cultural campaign (Y_R, m^C) captures all cross-pressured voters: $VS_R = \mu_{UC} + \mu_{LC}$. Any campaign deviation—whether a platform shift, a model shift, or both—that moves $\phi(c_{R,t})$ strictly below $\phi(c^C) = 1$ dilutes $B_{R,t+1}$ without capturing additional voters, since all four voter types are at their respective corners under the policy ordering. The deviation is strictly dominated for all $\beta > 0$.

(b) **Economic party.** Any campaign deviation by party L falls into one of two cases.

(i) Cultural pivot (shifting narrative or platform toward R 's cultural position): electorally ineffective by Lemma 1, since $B_{L,t} < B_{R,t}$ implies $PO_{L,t}^{(C)} < PO_{R,t}$; and brand-destroying by G1–G2, since the pivot shifts $B_{L,t+1}$ toward 1 and dilutes future economic coherence.

(ii) Economic convergence (shifting τ_L toward τ_R while maintaining the economic model): reduces the economic policy gap $|\tau_L - \tau_R|$ without improving cultural competitiveness, weakening L 's differentiation on the only dimension where it holds a brand advantage.

The equilibrium economic campaign (Y_L, m^E) strictly dominates any deviation for all $\beta > 0$.

(c) **No Downsian convergence.** The policy ordering $\tau_R < \tau_L$, $q_R < q_L$ is sustained in equilibrium by the brand mechanism. Platform convergence requires off-brand campaign choices that are dominated under the forward-looking objective (5) for all discount factors $\beta \in (0, 1)$.

Part (b)(i) formalizes the fundamental dilemma facing the economic party: a cultural pivot is simultaneously the most tempting response to the trap (if the cultural frame is winning, why not compete on that dimension?) and the most destructive (it concedes L 's only viable brand asset). The historical application in Section 9.2 illustrates this directly. During Phase III of the North Carolina simulation, elements of the Fusion coalition capitulated to the Democrats' racial framing: Marion Butler's *Caucasian* ran racist cartoons and presented the Populists as "the true white man's party." In the model, this is a cultural campaign pivot by party L at $t = 12$ –13. The pivot does not win a single additional election (since $B_{Fus,t} < B_{D,t}$, the Fusion coalition can never win the PO comparison on the cultural dimension) but devastates the coalition's future economic coherence. Butler's capitulation is an instance of the strict dominance result in part (b)(i); the Populist Party did not survive as an independent political force after the 1898 campaign. The general lesson extends to any

economic party facing a brand-advantaged cultural opponent: the cultural dimension is a losing front precisely because the opponent’s accumulated brand capital makes it so.

Part (a) reveals a complementary result for the cultural party: when the trap is active, the trap mechanism itself makes platform convergence unnecessary. Party R already wins all cross-pressured voters through the cultural salience channel; no platform adjustment toward L ’s economic position can capture additional voters, because the non-cross-pressured types (UC, LP, UP) are locked into their respective parties by the policy ordering. The only effect of a platform shift is brand dilution, which weakens the coherence prior for future cultural campaigns and makes future trap maintenance harder. A forward-looking party R therefore maintains its equilibrium platform, not because convergence is exogenously ruled out, but because the brand-constrained game makes it strategically dominated.

8.3 Testable Implications

The voter trap model generates predictions that distinguish it from unconstrained identity-politics theories in which both parties can contest any dimension with equal credibility. We state two as formal corollaries of the preceding results.

Corollary 3 (Brand-depth prediction). *Fix the public realization x_t and the signal parameters. The minimum cultural signal strength required to sustain the trap at date t is strictly decreasing in the cultural party’s brand state $B_{R,t}$ and strictly increasing in the economic party’s brand state $B_{L,t}$. Equivalently, a party with a deeper cultural brand can activate the voter trap at weaker signal realizations than a party with a shallower brand, holding the information environment fixed.*

In the parametric benchmark, this follows directly from the closed-form PO threshold: the minimum cultural signal gap $\underline{z}_t = (\sigma^2/\alpha)[\log(1 - B_{\text{Fus},t}) - \log B_{D,t}]$ is decreasing in $B_{D,t}$ and increasing in $B_{\text{Fus},t}$. The brand-depth prediction therefore separates the voter trap from frameworks in which propaganda effectiveness is symmetric or exogenous (e.g., the χ parameter in Gennaioli and Tabellini 2025).

Empirical test. The cross-section of persuasion effectiveness across regions or election cycles should covary with measures of accumulated brand investment (e.g., historical cultural-newspaper density, prior cultural vote share), holding signal intensity fixed. Regions where the cultural party’s brand is deeper should exhibit larger vote swings toward R for the same level of cultural-signal exposure.

Corollary 4 (Absorbing-trap exclusion restriction). *In a strongly stable trap (Corollary 2), once the dissolution threshold $\underline{z}(t)$ falls below the support of z_t , economic shocks within the*

signal support have zero marginal effect on the vote share of the cultural party. The sensitivity of VS_R to economic realizations $x_{E,t}$ is zero in the absorbing region, positive outside it.

This is an exclusion restriction: in the absorbing region, economic conditions should be uninformative about vote share. Models without absorption dynamics predict continuous sensitivity of vote share to economic conditions at all brand-state levels. The prediction can be tested by interacting economic-shock measures with proxies for brand asymmetry and checking whether the interaction coefficient is negative and, in the limit, drives the economic-shock coefficient to zero.

Robustness to electorate uncertainty. The trap conditions of Proposition 1 depend on publicly observable quantities (the signal environment and brand histories), not on the party’s private information about voter preferences. Appendix C formalizes this observation: the evidential-credibility and salience conditions hold identically for all type spaces, and the cross-pressured share μ_{LC} is point-identified from the observed electoral swing between a pre-trap and trap election. A single moment condition on the swing suffices to verify that the trap is robust across all electorates consistent with that bound. The voter trap is therefore a *low-information mechanism*: party R does not need detailed knowledge of the electorate to implement it.¹²

9 Historical Illustration: The Wilmington Coup and Jim Crow Durability

This section illustrates the model’s architecture by mapping the formal objects onto the political economy of the 1890s American South, culminating in the 1898 Wilmington Coup and the Jim Crow equilibrium that followed. The exercise is a historical illustration, not a formal test: we do not estimate Bayes factors, coherence priors, or identity weights, but instead provide an interpretive mapping from the historical record to the model’s primitives. A natural temptation when illustrating a formal model with historical evidence is to selectively emphasize the features of an episode that the model captures while eliding those it does not—ironically oversimplifying the history to improve the model’s apparent fit. We resist this temptation: Appendix section D discusses what the model does *not* capture, including

¹²One might ask whether party L could escape the trap by activating a new campaign dimension (e.g., democratic integrity or anti-corruption) on which R has no established brand advantage. Establishing a credible brand on a new dimension requires the EMA to build from a near-zero initial state, which takes approximately $O(1/\delta)$ election cycles (roughly 7 with $\delta = 0.15$). Moreover, if R can absorb the new frame through its existing cultural brand, the effective dimensionality of competition remains unchanged.

coalition coordination asymmetry, social punishment, and extralegal violence. The episode is a compelling candidate for the voter trap mechanism: a cross-racial redistributionist coalition posed a genuine threat to an entrenched cultural party; that party responded with a systematic narrative campaign that exploited its brand advantage; and the cross-pressured voter adopted the cultural model in a pattern consistent with the trap conditions.

9.1 Historical Background

The Populist threat and Fusion victories. The political landscape of the 1890s American South was defined by the rise of the People’s Party, the first major American party to openly advocate redistribution regardless of race. Populist leaders allied with Black Republican voters, recognizing that a cross-racial coalition of the economically distressed was their only path to electoral viability against the entrenched Democratic Party. The 1892 North Carolina gubernatorial election revealed the arithmetic of this threat: combined Republican-Populist support exceeded Democratic totals by over 7,000 ballots. The Populist-Republican “fusion” ticket swept the 1894 and 1896 state elections, winning the governorship and the legislature. Wilmington, then North Carolina’s largest city and majority-Black, elected a Fusion government that included Black aldermen, a Black chief of police, and a Black coroner. By 1896, approximately 1,000 Black citizens held elected or appointed office statewide, and Democratic representation had fallen to 26 of 120 state House seats. The Fusion experiment had proven that a cross-racial economic coalition could win and govern in the post-Reconstruction South.

The White Supremacy Campaign. The Democratic Party recognized the Fusion coalition as an existential threat. Rather than challenging the Fusion record on economic grounds, a terrain where Democrats were vulnerable, Democratic campaign chairman Furnifold Simmons deployed a systematic cultural campaign: a statewide propaganda apparatus designed to reframe the election as a referendum on racial cultural identity rather than economic policy. Simmons coordinated a network of newspapers, including Josephus Daniels’ *Raleigh News & Observer* (built from a bankrupt sheet to the state’s dominant paper), the *Charlotte Observer*, the *Wilmington Morning Star and Messenger*, and dozens of allied weeklies, that reproduced the same racial frame daily. Artist Norman Jennett produced approximately 75 front-page cartoons depicting Black officeholders as buffoons and Fusionists as race traitors; Daniels later admitted that stories of a “reign of terror” by Black men were fabricated. The campaign’s most explosive episode centered on the Manly editorial: Alexander Manly, editor of Wilmington’s *Daily Record*, challenged the Black rapist myth in print, and Democrats reprinted his editorial across the state for weeks as evidence of racial threat. Ottinger and

Posch [2025] quantify this campaign using full-text newspaper data across the South: a ten-percentage-point increase in local Fusion vote share raised the probability of anti-Black content in Democratic papers by 4.5 percentage points, an effect concentrated in partisan outlets and absent in independent papers, ruling out a demand-side explanation.

The coup and Jim Crow. On November 8, 1898, Red Shirt paramilitaries stalked polling places; Democrats recaptured the legislature with 53% of the statewide vote. Two days later, in Wilmington, where the Fusion coalition still controlled city government, former Congressman Alfred Waddell led over 1,500 armed white men to the offices of Manly’s *Daily Record*, destroyed the press, and burned the building. Elected Fusion officials were forced to resign at gunpoint; an estimated 14 to 60 or more Black residents were killed; at least 1,400 fled the city permanently. It remains the only successful coup d’état against an elected government on American soil [Zucchini, 2020]. The new government swiftly enacted literacy tests and poll taxes that collapsed registered Black voters from 126,000 to 6,100 by 1902. Similar disenfranchisement laws swept the South between 1890 and 1908. Jim Crow had locked in.

9.2 Parametric Benchmark: Simulating the 1890s Voter Trap

The Democratic Party plays the role of party R (the culturally-entrenched party) and the Populist-Republican Fusion plays the role of party L (the economic party). We write $B_{D,t}$ and $B_{Fus,t}$ for party-specific brand states throughout. The poor white farmer of the 1890s South is the model’s LC voter: economically, his interests aligned with Populist redistribution (railroad reform, cheap credit, public education), placing his economic ideal closer to the Fusion platform; culturally, he occupied the conservative end of the racial hierarchy, placing his cultural ideal closer to the Democrats. Assumption 1 is satisfied. This illustration is not a causal identification of the mechanism: violence, organizational fracture, and disenfranchisement all played independent roles, and the unitary-actor assumption is most plausible for the Democratic side (Simmons’ centralized apparatus) but is an approximation for the organizationally fragmented Fusion coalition.

Calibration. We calibrate the parametric special case of Appendix B to the 1890s North Carolina episode. All primitives are given explicit functional forms; full technical details appear in Appendix B. Table 1 reports the structural parameters and their historical motivation.

Table 1: Parametric benchmark: structural parameters and historical motivation

Parameter	Value	Historical motivation
$Y_L = (\tau_L, q_L)$	(1, 1)	Fusion platform: high redistribution, progressive racial policy
$Y_R = (\tau_R, q_R)$	(0, 0)	Democratic platform: low taxes, racial cultural conservatism
$Y_{LC}^E = (\tau^E, q^E)$	(0.8, 0.7)	Farmer's economic ideal: closer to Fusion on redistribution and social policy
$Y_{LC}^C = (\tau^C, q^C)$	(0.4, 0.2)	Farmer's cultural ideal: closer to Democratic racial conservatism
κ	1.0	Equal weight on economic and cultural policy dimensions
b_R (initial Democrat brand)	0.75	Deep cultural brand from Reconstruction-era Redeemer campaigns
b_L (initial Fusion brand)	0.20	New coalition with strong economic identity but no cultural history
δ (memory decay)	0.15	Roughly 7 election cycles to brand convergence
η (salience responsiveness)	2.0	Moderate identity-weight sensitivity to salience gaps
$\bar{\Delta}_{LC}$ (voting cutoff)	$5/9 \approx 0.556$	Derived from primitives (see Appendix B)

The initial Democrat brand $b_R = 0.75$ reflects decades of Reconstruction-era racial campaigns; the Fusion brand $b_L = 0.20$ reflects the coalition's novelty (strong economic coherence but no cultural history), and with $\delta = 0.15$ brand states converge over roughly 7 cycles. The voting cutoff $\bar{\Delta}_{LC} \approx 0.556$ means that the poor white farmer switches when his cultural identity weight exceeds roughly 56%, a non-trivial threshold reflecting the genuine economic advantage of the Fusion platform.

The North Carolina simulation. We model four historical phases by varying the signal environment: **Phase I** (Reconstruction, $t = 0-4$): cultural signals dominate; **Phase II** (Agricultural Crisis, $t = 5-9$): strong economic signals; **Phase III** (White Supremacy Campaign, $t = 10-14$): the strongest cultural signals in the calibration; **Phase IV** (Jim Crow, $t = 15-19$): moderate but persistent cultural signals. Figure 1 shows the simulation across all four phases.

The North Carolina Story: A Voter Trap Simulation

Democrat wins: 15/20 | Fusion wins: 5/20

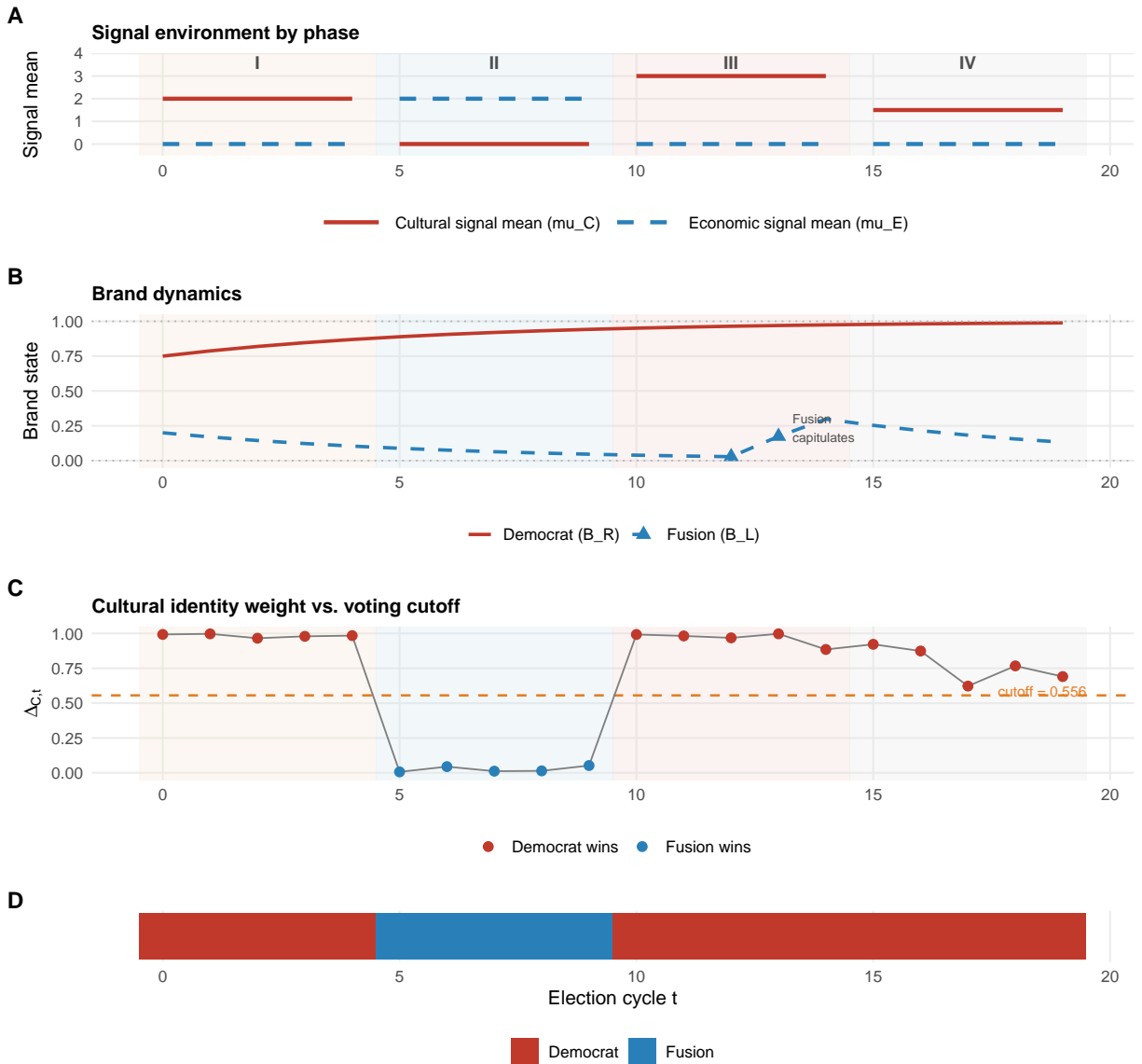


Figure 1: **The North Carolina voter trap simulation.** Panel (A): phase-specific signal means μ_E (economic) and μ_C (cultural) across 20 election cycles. Panel (B): brand dynamics — the Democratic brand $B_{D,t}$ (red) deepens toward its absorbing limit even during Fusion victories; the Fusion brand $B_{Fus,t}$ (blue) declines toward zero, with a capitulation-induced spike at $t = 12-13$. Panel (C): cultural identity weight $\Delta_{C,t}$ relative to the voting cutoff $\bar{\Delta}_{LC} = 5/9 \approx 0.556$; red dots denote Democratic victories (trap active), blue dots Fusion victories. Panel (D): election outcomes — Democrats win 15 of 20 elections, losing only during the agricultural crisis (Phase II). An interactive version of this simulation is available at https://jccisneros.com/uploads/nc_story_app.

Three findings stand out. First, despite the Fusion electoral victories in Phase II (blue dots), the Democratic brand $B_{D,t}$ continues deepening throughout. The Democrats campaign culturally even in opposition, while the Fusion brand $B_{\text{Fus},t}$ declines. The Fusion victories are structurally fragile: they depend on the persistence of strong economic signals, not on any accumulated brand capital. Second, when the propaganda campaign launches in Phase III, $B_{D,t}$ is already near 0.95: the PO condition is overwhelmingly satisfied and the trap holds in every election of the phase. Third, by Phase IV, the trap region has grown so large that it holds for essentially any signal realization. The trap has become absorbing, consistent with Corollary 2.

A key insight from the simulation concerns the Fusion coalition’s capitulation. During Phase III, Marion Butler’s *Caucasian* ran racist cartoons and presented the Populists as “the true white man’s party,” captured in the model as party L deviating to the cultural campaign at $t = 12$ –13. Proposition 3(b)(i) formalizes why this pivot was strictly dominated: it could not win the PO comparison (since $B_{\text{Fus},t} < B_{D,t}$) yet destroyed the economic coherence that was the coalition’s only viable asset.

The simulation findings connect directly to the supply-side evidence in Ottinger and Posch [2025]. In the model, cultural content is strategically produced by party R (the Democrats) because racial framing is brand-coherent for them and brand-destroying for party L ; politically independent outlets face no corresponding strategic incentive to produce the same content. The empirical finding that anti-Black content rose sharply in Democratic-affiliated newspapers but not in independent outlets is therefore the supply-side signature of the mechanism: it rules out a demand-driven account in which voter preferences alone would have generated the cultural shift regardless of party strategy. The Phase III propaganda campaign in the simulation is calibrated to this asymmetry—only partisan supply can generate the brand-deepening dynamic that eventually absorbs the trap.

Jim Crow durability. By 1898, the dissolution threshold $\underline{z}(t)$ had fallen to the point where no reachable realization could dissolve the trap (Corollary 2). The subsequent disenfranchisement laws removed Black voters from the rolls entirely, eliminating the Fusion coalition at its demographic foundation and rendering the trap strongly stable. Jim Crow’s durability, seven decades from the Wilmington Coup to the Voting Rights Act of 1965, is consistent with a trap that has become absorbing through the compounding of brand asymmetry and institutional constraint. The Voting Rights Act itself represents an exogenous institutional shock: federal intervention that restored the demographic foundation for an economic-left coalition and reopened $\mathcal{F}_{\text{Fus},t}$ to campaigns that could contest cultural

salience.¹³

Scope and limitations. The model captures the electoral persuasion mechanism but abstracts from several features of the historical episode: coalition coordination asymmetry (Simmons’ centralized apparatus vs. the fragmented Fusion opposition), social punishment and ostracism (the “scalawag” stigma that raised the cost of defection from the cultural frame), and the violence and extralegal coercion that culminated in the Wilmington Coup itself. Additionally, Proposition 3 establishes endogenous platform separation for episodes where brand dynamics discipline campaign choices, but the framework abstracts from cases where institutional forces override the brand-maintenance incentive, as in the post-1965 American partisan realignment [Kuziemko and Washington, 2018, Bernini et al., 2025]. A detailed discussion of these limitations and the natural extensions they suggest is provided in Appendix D.

10 Concluding Remarks

This paper develops a formal model of the voter trap: a configuration in which a cross-pressured voter rationally votes against her benchmark economic interests because the structure of party brands asymmetrically constrains the menu of interpretive models available to her. The trap exists under three primitive conditions on evidence quality, cultural salience, and brand asymmetry; it deepens dynamically as the cultural party’s brand consolidates; and platform separation is sustained endogenously by the brand mechanism. Applied as a historical illustration to the 1890s American South, the model provides an interpretive framework for the electoral dynamics behind the Wilmington Coup of 1898 and the Jim Crow equilibrium that followed: a cross-racial redistributionist coalition dismantled by a systematic narrative campaign consistent with the model’s brand-asymmetry mechanism.

The voter trap is not a historical curiosity; the mechanism’s observable implications are visible in several contemporary settings, though we emphasize that these are suggestive illustrations rather than formal tests of the model. In the United States, the MAGA coalition exhibits features consistent with the stable trap characterized in Proposition 2: successive election cycles of culturally coherent campaigning have plausibly raised the Republican cul-

¹³Consistent with the model’s prediction that institutional shocks restore feasibility without instantaneously erasing accumulated brand asymmetry, Bernini et al. [2025] document that the VRA immediately triggered white countermobilization—a spike in white voter registration and support for racially conservative candidates concentrated in counties where Black political representation posed a credible electoral threat. Brand states and cultural salience dynamics thus persisted well into the post-VRA period, consistent with the trap requiring not only a change in institutional rules but sustained economic-signal realizations to unwind.

tural brand’s coherence prior, and the persistence of this coalition across three election cycles, despite significant economic underperformance in trade-exposed communities [Autor et al., 2020], is consistent with the model’s prediction of trap deepening. The normalization of previously stigmatized cultural expression documented by Bursztyn et al. [2020] is suggestive of the social-norm analog of brand deepening.

Across Latin America, the post-left electoral wave of the 2010s and 2020s may reflect similar dynamics. In Brazil, the evangelical-conservative cultural brand of the Brazilian right may have constrained the PT’s ability to credibly contest the cultural dimension, a pattern consistent with the brand asymmetry at the core of the voter trap mechanism.

Two directions are natural complements to the present framework. The first is micro-founding identity weight allocation through community-level enforcement: a richer action space in which voters impose social costs on those who deviate from the dominant cultural frame would reinforce the brand asymmetry by making the trap self-enforcing at the social as well as the electoral level (see Appendix D for further discussion). The second concerns structural estimation: brand states can be proxied from party platforms and newspaper content [following Ottinger and Posch, 2025], public realizations from economic indicators, and the cross-pressured share from electoral returns and survey data. The testable implications formalized in Section 8.3 provide the regression specifications needed to distinguish the voter trap from unconstrained identity-politics theories.

The voter trap is, at its core, a story about the limits of democratic representation. When brand asymmetry is large and persistent, the formal machinery of democracy (free elections, informed voters, competing parties) can produce systematic and durable misrepresentation of economic interests. Cross-pressured voters are not failing democracy; they are choosing optimally from a menu that has been shaped by historical accidents of brand accumulation. Understanding the structural conditions that create and sustain voter traps, and the exogenous shocks that can dissolve them (as the federal-level Voting Rights Act did), is a prerequisite for any theory of democratic reform that takes the informational and credibility constraints on political competition seriously.

References

- Christopher H. Achen and Larry M. Bartels. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press, Princeton, NJ, 2016.
- Chiara Aina. Tailored stories, 2025. Working paper.
- Alberto Alesina and Alex Cukierman. The politics of ambiguity. *Quarterly Journal of Economics*, 105(4):829–850, 1990.
- Jose Apesteguia and Leandro Salvanti. Choice by contextual associations, 2026. Working paper.
- David Autor, David Dorn, Gordon Hanson, and Kaveh Majlesi. Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*, 110(10):3139–3183, 2020.
- Jeffrey S. Banks and Joel Sobel. Equilibrium selection in signaling games. *Econometrica*, 55(3):647–661, 1987.
- Larry M. Bartels. What’s the matter with *What’s the Matter with Kansas?* *Quarterly Journal of Political Science*, 1(2):201–226, 2006.
- Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.
- Andrea Bernini, Giovanni Facchini, Marco Tabellini, and Cecilia Testa. Black empowerment and white mobilization: The effects of the voting rights act. *Journal of Political Economy*, 133(10):3078–3139, 2025.
- Timothy Besley and Stephen Coate. An economic model of representative democracy. *Quarterly Journal of Economics*, 112(1):85–114, 1997.
- Giampaolo Bonomi, Nicola Gennaioli, and Guido Tabellini. Identity, beliefs, and political conflict. *Quarterly Journal of Economics*, 136(4):2371–2415, 2021.
- Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience theory of choice under risk. *Quarterly Journal of Economics*, 127(3):1243–1285, 2012.
- Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013.

- Ian Budge. A new spatial theory of party competition: Uncertainty, ideology and policy equilibria viewed comparatively and temporally. *British Journal of Political Science*, 24 (4):443–467, 1994.
- Leonardo Bursztyn, Georgy Egorov, and Stefano Fiorin. From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–3548, 2020.
- Kfir Eliaz and Ran Spiegler. A model of competing narratives. *American Economic Review*, 110(12):3786–3816, 2020.
- Kfir Eliaz, Simone Galperti, and Ran Spiegler. False narratives and political mobilization. *Journal of the European Economic Association*, 23(3):983–1027, 2025.
- Morris P. Fiorina. *Retrospective Voting in American National Elections*. Yale University Press, New Haven, CT, 1981.
- Thomas Frank. *What’s the Matter with Kansas? How Conservatives Won the Heart of America*. Metropolitan Books, New York, 2004.
- Nicola Gennaioli and Guido Tabellini. Presidential address: Identity politics. *Econometrica*, 93(6):1937–1967, 2025.
- Andrew Healy and Gabriel S. Lenz. Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*, 58(1): 31–47, 2014.
- Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.
- David M. Kreps. Corporate culture and economic theory. In James E. Alt and Kenneth A. Shepsle, editors, *Perspectives on Positive Political Economy*, pages 90–143. Cambridge University Press, Cambridge, 1990.
- Ilyana Kuziemko and Ebonya Washington. Why did the democrats lose the south? bringing new data to an old debate. *American Economic Review*, 108(10):2830–2867, 2018.
- George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, New York, 2006.
- Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219, 2001.

- Mariann Ollar and Antonio Penta. Incentive compatibility and belief restrictions, 2025. Working paper.
- Sebastian Ottinger and Max Posch. The political economy of propaganda: Evidence from u.s. newspapers, 2025. Working paper.
- John R. Petrocik. Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, 40(3):825–850, 1996.
- John E. Roemer. Why the poor do not expropriate the rich: An old argument in new garb. *Journal of Public Economics*, 70(3):399–424, 1998.
- Joshua Schwartzstein and Adi Sunderam. Using models to persuade. *American Economic Review*, 111(1):276–323, 2021.
- David O. Sears and Carolyn L. Funk. The role of self-interest in social and political attitudes. *Advances in Experimental Social Psychology*, 24:1–91, 1991.
- Moses Shayo. A model of social identity with an application to political economy of nations. *Journal of Political Economy*, 117(2):248–290, 2009.
- Robert J. Shiller. Narrative economics. *American Economic Review*, 107(4):967–1004, 2017.
- Richard Wormser. *The rise and fall of Jim Crow*. Macmillan, 2004.
- David Zucchino. *Wilmington's Lie: The Murderous Coup of 1898 and the Rise of White Supremacy*. Atlantic Monthly Press, New York, 2020.

A Proofs

Proof of Lemma 1. By G2, G_p places positive weight on all lags of campaign history, so the pivot $\tilde{c}_{p,t}$ enters $B_{p,t+k}$ with strictly positive weight for all $k \geq 1$. Since $\tilde{c}_{p,t} \notin \mathcal{F}_{p,t}^{\text{brand}}$, it lies outside the historical support of $B_{p,t}$, and by G2 its positive weight shifts $B_{p,t+k}$ away from that support for each subsequent k . By G4, $\Pi_{p,t+k}(c^{\text{hist}} \mid B_{p,t+k})$ is weakly decreasing in the distance between c^{hist} and the support of $B_{p,t+k}$; because $B_{p,t+k}$ moves away from c^{hist} for all $k \geq 1$ (and strictly so for k small relative to δ by G1's reinforcement condition), the coherence prior on historical campaigns falls weakly, and strictly for a positive measure of voters by G4 and P2. \square

Proof of Lemma 2. Voter LC votes R iff $V_{LC,R,t} > V_{LC,L,t}$. Expanding (21) and rearranging:

$$V_{LC,R,t} > V_{LC,L,t} \iff (\tau_L - \tau_R)(\bar{\tau}_p - \tau_{LC,t}^\Delta) > \kappa(q_L - q_R)(\bar{q}_p - q_{LC,t}^\Delta).$$

Substituting $\tau_{LC,t}^\Delta = (1 - \Delta)\tau_{LC}^E + \Delta\tau_{LC}^C$ and $q_{LC,t}^\Delta = (1 - \Delta)q_{LC}^E + \Delta q_{LC}^C$ (where $\Delta \equiv \Delta_{LC,C,t}$) yields a condition that is affine and strictly monotone in Δ whenever condition (c) holds. Setting both sides equal and solving for Δ yields (22). Conditions (a) and (b) guarantee the solution lies in the interior $(0, 1)$: condition (a) says the inequality is violated at $\Delta = 0$ (voter prefers L under full economic identity), condition (b) says it holds at $\Delta = 1$ (voter prefers R under full cultural identity), and strict monotonicity under (c) implies a unique interior crossing. \square

Proof of Proposition 1. We verify each of Definition 3's four conditions.

Condition (i) (Benchmark vote). Under the prior model $m_{LC,t}^0$, no campaign-induced salience shift has occurred, so $\Delta_{LC,C,t} = \Delta_{LC,C}^0$ at the pre-campaign level. By the parameter restriction $\epsilon > b\psi$ and Assumption 1, voter LC prefers party L on the economic dimension at this identity weight: $V_{LC,L,t} > V_{LC,R,t}$ under $m_{LC,t}^0$.

Condition (ii) (Trap vote). By proposition condition (i), $\text{PO}_{LC,R,t}(m_{R,t}) = \Lambda_{R,t}\pi_{R,t} > \max\{1, \text{PO}_{LC,L,t}(m_{L,t})\}$. The adoption rule (16) therefore assigns $m_{LC,t}^{\text{eff}} = m_{R,t}$. By proposition condition (ii), $S_{LC,C,t}(m_{R,t}) > \bar{S}_{LC}$. By the softmax formula (19), this salience threshold implies $\Delta_{LC,C,t}(m_{R,t}) > \bar{\Delta}_{LC}$. By Lemma 2, voter LC votes R .

Condition (iii) (Rational adoption). Adoption of $m_{R,t}$ follows from the adoption rule (16), which is voter LC 's best response to the information environment given common knowledge of $(B_{p,t}, h_t, x_t)$ and the campaign structure. No error or bias is required: the PO score of $m_{R,t}$ genuinely exceeds that of every alternative model available to voter LC .

Condition (iv) (No live escape). This is exactly proposition condition (iii): no $c \in \mathcal{F}_{L,t}$ simultaneously achieves $\text{PO}_{LC,L,t}(m) > \Lambda_{R,t}\pi_{R,t}$ and $S_{LC,C,t}(m) \geq \bar{S}_{LC}$. Party L 's brand

history restricts $\mathcal{F}_{L,t}$ so that any campaign achieving the PO threshold fails the salience condition, and vice versa. \square

Proof of Corollary 1. Under the EMA parametrization (8), G_p is Lipschitz continuous in $(B_{p,t}, c_{p,t})$, so the coherence prior $\Pi_{p,t}(c | B_{p,t})$ is continuous in $B_{p,t}$. The PO slack function $F_1(c, B, h, x) := \Lambda_{R,t}(x, h)\pi_{R,t}(B_{R,t}) - \max\{1, \Lambda_{L,t}(x, h)\pi_{L,t}(B_{L,t})\}$ and the salience-margin function $F_2(c, B, h, x) := S_{LC,C,t}(m_{R,t}) - \bar{S}_{LC}$ are both continuous in the state (c, B, h, x) . The neighborhood

$$\mathcal{N}_\epsilon = \{F_1 \geq \frac{\epsilon}{2}\} \cap \{F_2 \geq \frac{\epsilon}{2}\}$$

is the intersection of two super-level sets of continuous functions; since $F_1 > \epsilon$ and $F_2 > \epsilon$ at the original state, \mathcal{N}_ϵ is open and non-empty. Conditions (i)–(iii) of Definition 3 hold throughout \mathcal{N}_ϵ by the argument of Proposition 1, since $F_1 \geq \epsilon/2 > 0$ ensures $m_{R,t}$ wins the PO comparison and $F_2 \geq \epsilon/2 > 0$ ensures $\Delta_{LC,C,t} > \bar{\Delta}_{LC}$. Condition (iv) holds throughout \mathcal{N}_ϵ by continuity of $\mathcal{F}_{L,t}$ in $(B_{L,t}, x_t)$: any campaign $c \in \mathcal{F}_{L,t}$ at the perturbed state still faces a PO score that is below R 's by at least $\epsilon/2$ (from $F_1 \geq \epsilon/2$), so no live L campaign can both dominate R 's model and restore $\Delta_{LC,C,t}$ below $\bar{\Delta}_{LC}$. \square

Lemma 3 (Brand-feasible set is non-expanding). *Under G1–G4, if party L runs only campaigns in $\mathcal{F}_{L,t}^{\text{brand}}$ between t and $t + T$, then $\mathcal{F}_{L,t+T}^{\text{brand}} \subseteq \mathcal{F}_{L,t}^{\text{brand}}$.*

Proof. Suppose $c \notin \mathcal{F}_{L,t}^{\text{brand}}$, so $\Pi_{L,t}(c | B_{L,t}) = 0$. By G4 (monotonicity), this requires $d_{\mathcal{B}}(\phi(c), \text{supp}(B_{L,t}))$ to be large. Between t and $t + T$, party L runs only brand-feasible campaigns (by hypothesis), so $\phi(c_{L,s}) \in \text{supp}(B_{L,t})$ for all $s \in [t, t + T]$. By G2 (persistence), $B_{L,t+T}$ places positive weight on all past brand images; the historical support $\text{supp}(B_{L,t+T}) = \text{supp}(B_{L,t}) \cup \{\phi(c_{L,s}) : s \in [t, t + T]\} \subseteq \text{supp}(B_{L,t})$, so the support does not expand beyond its pre- t set. By G4, $\Pi_{L,t+T}(c | B_{L,t+T}) = 0$, so $c \notin \mathcal{F}_{L,t+T}^{\text{brand}}$. Since c was arbitrary, $\mathcal{F}_{L,t+T}^{\text{brand}} \subseteq \mathcal{F}_{L,t}^{\text{brand}}$. \square

Proof of Proposition 2. Under the EMA law of motion (8), T applications of the coalition-optimal cultural campaign c_R^C yield

$$B_{R,t+T} = (1 - \delta)^T B_{R,t} + [1 - (1 - \delta)^T] \phi(c_R^C).$$

Since $(1 - \delta)^T \in (0, 1)$ and $\phi(c_R^C)$ lies in the direction of the cultural campaign type, $B_{R,t+T}$ is weakly closer to $\phi(c_R^C)$ than $B_{R,t}$. By G1 (reinforcement), the coherence prior on cultural campaigns is weakly higher under $B_{R,t+T}$: $\pi_{R,t+T} := \Pi_{LC,R,t+T}(c_R^C | B_{R,t+T}) \geq \Pi_{LC,R,t}(c_R^C | B_{R,t}) =: \pi_{R,t}$.

By Lemma 1, any cultural pivot by party L is brand-destroying, so $\mathcal{F}_{L,t+T}$ acquires no credible cultural campaigns not in $\mathcal{F}_{L,t}$, and $\pi_{L,t+T} \leq \pi_{L,t}$. This establishes part (a).

Part (b). Condition (i) of Proposition 1 requires $\Lambda_{R,t}\pi_{R,t} > \max\{1, \Lambda_{L,t}\pi_{L,t}\}$. At date t , this holds with slack ϵ , so the minimum Bayes factor sustaining the trap is $\underline{\Lambda}_t = \max\{1, \Lambda_{L,t}\pi_{L,t}\} / \pi_{R,t}$. Since $\pi_{R,t+T} \geq \pi_{R,t}$ and $\pi_{L,t+T} \leq \pi_{L,t}$,

$$\underline{\Lambda}_{t+T} = \frac{\max\{1, \Lambda_{L,t+T}\pi_{L,t+T}\}}{\pi_{R,t+T}} \leq \frac{\max\{1, \Lambda_{L,t}\pi_{L,t}\}}{\pi_{R,t}} = \underline{\Lambda}_t,$$

where the inequality uses $\pi_{R,t+T} \geq \pi_{R,t}$ (denominator increases) and $\pi_{L,t+T} \leq \pi_{L,t}$ (numerator's coherence component decreases); the Bayes factor $\Lambda_{L,t+T}$ in the numerator is weakly bounded above by $\Lambda_{L,t}$ for any fixed realization, since $\mathcal{F}_{L,t+T}^{\text{brand}} \subseteq \mathcal{F}_{L,t}^{\text{brand}}$ (party L 's brand-feasible set does not expand) and Λ_L is the maximum over that set. The set of realizations sustaining the trap therefore expands.

Strict inequality in (a). If $x_{C,s} > \mathbb{E}[x_C]$ for some $s \in [t, t+T)$, the cultural campaign in period s generates a brand update $\phi(c_{R,s}^C)$ that lies strictly further in the cultural direction, making $B_{R,t+T}$ strictly closer to the cultural pole. By the strict form of G1, $\pi_{R,t+T} > \pi_{R,t}$, so $\underline{\Lambda}_{t+T} < \underline{\Lambda}_t$.

Absorbing limit. For any fixed $\delta > 0$, over n election cycles $(1 - \delta)^{nT} \rightarrow 0$ as $n \rightarrow \infty$, so $B_{R,nT} \rightarrow \phi(c_R^C)$. By P1 (limit credibility), $\pi_{R,nT} \rightarrow 1$ as $B_{R,nT} \rightarrow \phi(c_R^C)$, and the minimum Bayes factor $\underline{\Lambda}_{nT} \rightarrow \max\{1, \Lambda_L\pi_L\}$, which is bounded. With bounded support on $x_{C,t}$ and $\Lambda_{R,t} \geq \underline{\lambda} > 0$ for all realizations in the support (the regularity condition in the proposition statement), the Bayes factor $\Lambda_{R,nT}$ of R 's cultural model exceeds $\underline{\Lambda}_{nT}$ for all realizations in the support after finitely many cycles. The trap is then absorbing. \square

Proof of Corollary 2. By Proposition 2(b), the minimum Bayes factor $\underline{\Lambda}_t$ required to sustain the trap is non-increasing under equilibrium play: $\underline{\Lambda}_{t+T} \leq \underline{\Lambda}_t$. The dissolution threshold $z(t)$ is the infimum of cultural-signal advantages at which the Bayes factor of R 's model falls below $\underline{\Lambda}_t$; since $\underline{\Lambda}_t$ is non-increasing, $z(t)$ is weakly decreasing. Definition 4(ii) holds. Definition 4(i) holds because the trap conditions hold throughout \mathcal{N}_ϵ by Corollary 1, and all realizations $z_t \geq z(t)$ generate Bayes factors above $\underline{\Lambda}_t$.

Strong stability. If z_t has bounded support $[\underline{z}, \bar{z}]$, then by the absorbing-state limit of Proposition 2, $\pi_{R,t} \rightarrow 1$ and $\underline{\Lambda}_t \rightarrow \max\{1, \Lambda_L\pi_L\}$ as $t \rightarrow \infty$. Since $\underline{\Lambda}_t$ converges to a finite limit while $\Lambda_{R,t}$ is bounded below on the support of z_t , there exists a finite date \bar{t} such that $\Lambda_{R,t} > \underline{\Lambda}_t$ for all realizations in the support and all $t \geq \bar{t}$. The trap is then effectively absorbing. \square

Proof of Proposition 3. We work in the parametric benchmark of Appendix B with binary

model space $\mathcal{M} = \{E, C\}$, scalar brand state $B_{p,t} \in [0, 1]$, and brand images $\phi(c^C) = 1$, $\phi(c^E) = 0$. Since a campaign is a pair $c_{p,t} = (Y_{p,t}, m_{p,t})$, a deviation from the equilibrium campaign involves a change in the platform $Y_{p,t}$, the model $m_{p,t}$, or both. The brand image $\phi(c_{p,t})$ depends on the full campaign: any campaign that includes a cultural model or moves the platform toward the cultural end yields $\phi > 0$; campaigns that maintain the economic model and platform yield $\phi = 0$.

Part (a): Cultural party R. Under the policy ordering (1) and Proposition 4, party R 's equilibrium campaign is (Y_R, m^C) with $\phi = 1$. When the trap is active, the vote shares are: type- UC voters prefer R on both dimensions (always vote R); type- LC voters adopt R 's cultural model and vote R (by Proposition 1); type- LP and type- UP voters prefer L on both dimensions (always vote L). Hence $VS_R = \mu_{UC} + \mu_{LC}$.

Now consider any deviation $c'_{R,t} = (Y'_R, m'_R)$ with $\phi(c'_{R,t}) < 1$. The deviation cannot increase R 's vote share above $\mu_{UC} + \mu_{LC}$: voters LP and UP prefer L on both dimensions under any policy ordering satisfying (1), so they cannot be captured by any platform or model shift by R . The deviation therefore has zero current-period electoral benefit. However, $\phi(c'_{R,t}) < 1$ implies $B_{R,t+1} = (1 - \delta)B_{R,t} + \delta\phi(c'_{R,t}) < (1 - \delta)B_{R,t} + \delta = B_{R,t+1}^*$, reducing the coherence prior at $t + 1$ and weakening future trap conditions. The deviation is therefore strictly dominated for all $\beta > 0$, since it yields weakly lower current payoff and strictly lower continuation value.

Part (b)(i): Cultural pivot by L. Suppose L deviates to a campaign with cultural model or culturally shifted platform, yielding $\phi(c'_{L,t}) > 0$. By Lemma 1, since $B_{L,t} < B_{R,t}$, the cultural campaign's PO score satisfies $PO_{L,t}^{(C)} = B_{L,t}\Lambda_{D,t} < B_{R,t}\Lambda_{D,t} = PO_{R,t}$. The cultural pivot therefore fails to win the PO comparison and does not change voter LC 's model adoption. Simultaneously, $B_{L,t+1} = (1 - \delta)B_{L,t} + \delta\phi(c'_{L,t}) > B_{L,t+1}^*$, shifting L 's brand toward the cultural end and diluting future economic coherence. The deviation yields zero current-period benefit and strictly negative continuation value; it is strictly dominated for all $\beta > 0$.

Part (b)(ii): Economic convergence by L. Suppose L deviates to platform $Y'_L = (\tau'_L, q_L)$ with $\tau'_L < \tau_L$ (moving toward R 's economic position) while maintaining the economic model. The deviation reduces the economic policy gap $|\tau'_L - \tau_R| < |\tau_L - \tau_R|$. Examining the voting cutoff (22): the numerator, which captures LC 's preference for L under full economic identity, decreases as τ_L moves toward τ_R (since the midpoint $\bar{\tau}_p$ shifts and the economic stakes shrink). The denominator, which captures the sensitivity of LC 's preference to the identity weight, also changes. Critically, reducing $|\tau_L - \tau_R|$ shrinks the economic stakes that are L 's only source of differentiation: the welfare gain from electing L over R falls, weakening the benchmark economic case that anchors condition (i) of the trap definition. Meanwhile,

cultural competitiveness is unchanged (the economic model still loses the PO comparison on the cultural dimension). The economic party's optimal strategy is therefore to maintain the platform that maximizes economic differentiation.

Part (c) follows from (a) and (b): both parties optimally maintain their equilibrium platforms under the forward-looking objective (5), so the policy ordering (1) is sustained in equilibrium. □

B Parametric Benchmark

This appendix develops the fully parametric special case used in Section 9.2. All objects introduced in the main text receive explicit functional forms; equilibrium campaign choices are derived in closed form; and the three main results—trap formation, dynamic deepening, and the absorbing limit—are verified directly.

B.1 Setup

Environment. Time is discrete with elections at every period ($T = 1$). The policy space and voter types follow the main model. For the historical calibration, party platforms are normalized to $Y_L = (1, 1)$ (Fusion) and $Y_R = (0, 0)$ (Democrats). Two voter types are present: UC (upper-class conservative, always votes R) and LC (lower-class conservative, the pivotal cross-pressured voter).

Scalar brand space. $\mathcal{B} = [0, 1]$, where 0 denotes a purely economic brand and 1 a purely cultural brand. Brand images: $\phi(c^E) = 0$, $\phi(c^C) = 1$, where $c^E = (Y_L, m^E)$ and $c^C = (Y_R, m^C)$ denote the equilibrium economic and cultural campaigns respectively. EMA dynamics: $B_{p,t+1} = (1 - \delta)B_{p,t} + \delta \phi(c_{p,t})$.

Coherence prior.

$$\pi_p(m \mid B_{p,t}) := \begin{cases} B_{p,t} & m = C, \\ 1 - B_{p,t} & m = E. \end{cases} \quad (23)$$

This satisfies G1–G4, P1–P2: coherence on cultural campaigns is increasing in $B_{p,t}$; coherence on economic campaigns is decreasing.

Bayes factor. Public signals $x_\ell \sim N(\theta_{m,\ell}, \sigma^2)$ with model E predicting mean α on the economic dimension (and 0 on cultural) and model C predicting mean α on the cultural dimension (and 0 on economic). The Bayes factor of model C against the voter’s prior model E simplifies to

$$\Lambda_{D,t} = \exp\left\{\frac{\alpha}{\sigma^2}(x_{C,t} - x_{E,t})\right\}. \quad (24)$$

Posterior-odds scores.

$$\text{PO}_{D,t} = B_{D,t} \cdot \Lambda_{D,t}, \quad (25)$$

$$\text{PO}_{\text{Fus},t}^{(E)} = 1 - B_{\text{Fus},t} \quad (\text{Fusion proposes economic model}), \quad (26)$$

$$\text{PO}_{\text{Fus},t}^{(C)} = B_{\text{Fus},t} \cdot \Lambda_{D,t} < \text{PO}_{D,t} \quad (\text{since } B_{\text{Fus},t} < B_{D,t}). \quad (27)$$

Saliency and identity weights. Stakes: $\Omega_E = |\tau_L - \tau_R|$, $\Omega_C = \kappa|q_L - q_R|$. Saliency: $S_{\ell,t} = \Omega_\ell \cdot (\alpha/\sigma^2)|x_{\ell,t}|$. Cultural identity weight under model C :

$$\Delta_{C,t} = \frac{1}{1 + \exp\{-\eta(S_{C,t} - S_{E,t})\}}. \quad (28)$$

Voting cutoff. Substituting the parameters of Table 1 into Lemma 2:

$$\bar{\Delta}_{LC} = \frac{1 \cdot (0.8 - 0.5) + 1 \cdot (0.7 - 0.5)}{1 \cdot (0.8 - 0.4) + 1 \cdot (0.7 - 0.2)} = \frac{0.3 + 0.2}{0.4 + 0.5} = \frac{5}{9} \approx 0.556.$$

B.2 Equilibrium

Proposition 4 (Parametric equilibrium). *In the parametric benchmark, the unique MPE campaign strategies are: party R (Democrats) plays $c^C = (Y_R, m^C)$ in every period; party L (Fusion) plays $c^E = (Y_L, m^E)$ in every period. Platform separation is maintained endogenously (Proposition 3).*

Proof. Democrats. Since party R plays c^C at every period and $\phi(c^C) = 1 > B_{D,0} = b_R = 0.75$, the EMA rule (8) implies $B_{D,t} \geq b_R = 0.75 > 1/2$ for all $t \geq 0$ by induction (the sequence is non-decreasing whenever $B_{D,t} < 1$). An economic campaign yields $\text{PO}_D^{(E)} = 1 - B_{D,t} < 1$ (since $B_{D,t} > 1/2$), so voter LC retains her prior and votes Fusion. The cultural campaign yields $\text{PO}_{D,t} = B_{D,t}\Lambda_{D,t}$, which exceeds $1 - B_{\text{Fus},t}$ whenever the trap conditions hold. The cultural campaign weakly dominates.

Fusion. A cultural pivot yields $\text{PO}_{\text{Fus}}^{(C)} = B_{\text{Fus},t}\Lambda_{D,t} < \text{PO}_{D,t}$ (since $B_{\text{Fus},t} < B_{D,t}$), so the pivot never wins the PO comparison. It also shifts $B_{\text{Fus},t+1}$ toward 1 by (8), diluting future economic coherence. The economic campaign strictly dominates. \square

B.3 Trap Formation and Dynamic Deepening

Under equilibrium play, brand states evolve as:

$$B_{D,t} = 1 - (1 - \delta)^t(1 - b_R), \quad (29)$$

$$B_{\text{Fus},t} = (1 - \delta)^t b_L. \quad (30)$$

Both converge monotonically: $B_{D,t} \rightarrow 1$ and $B_{\text{Fus},t} \rightarrow 0$. The minimum cultural signal $\underline{z}_t = x_{C,t} - x_{E,t}$ required for PO dominance is

$$\underline{z}_t = \frac{\sigma^2}{\alpha} \left[\log(1 - B_{\text{Fus},t}) - \log B_{D,t} \right], \quad (31)$$

which is non-increasing (approaches 0 from below as $t \rightarrow \infty$). This verifies Proposition 2: each election cycle lowers the signal threshold required to sustain the trap.

When $x_{C,t} - x_{E,t}$ has bounded support $[\underline{z}, \bar{z}]$, there exists a finite date \bar{t} after which $\underline{z}_t < \underline{z}$ —the trap holds for all signal realizations in the support, and the absorbing-trap prediction of Corollary 2 is verified directly.

B.4 Brand Dilution Under Capitulation

If Fusion deviates and plays C at dates t_1, t_2 , the brand update (8) shifts $B_{\text{Fus},t}$ toward 1 instead of toward 0. Over two periods of capitulation: $B_{\text{Fus},t+2} \approx (1 - \delta)^2 B_{\text{Fus},t} + [1 - (1 - \delta)^2] \cdot 1$. With $B_{\text{Fus},t} \approx 0.03$ and $\delta = 0.15$, this yields $B_{\text{Fus},t+2} \approx 0.29$ —a sixfold increase that persists for many subsequent periods. The deviation is simultaneously ineffective (the cultural pivot is PO-dominated since $B_{\text{Fus}} < B_D$) and dynamically costly (it destroys the economic coherence that is Fusion’s only viable asset).

Fusion's Dilemma: Brand Dilution vs. Staying On-Brand

Capitulating to racial framing is PO-dominated and brand-destroying

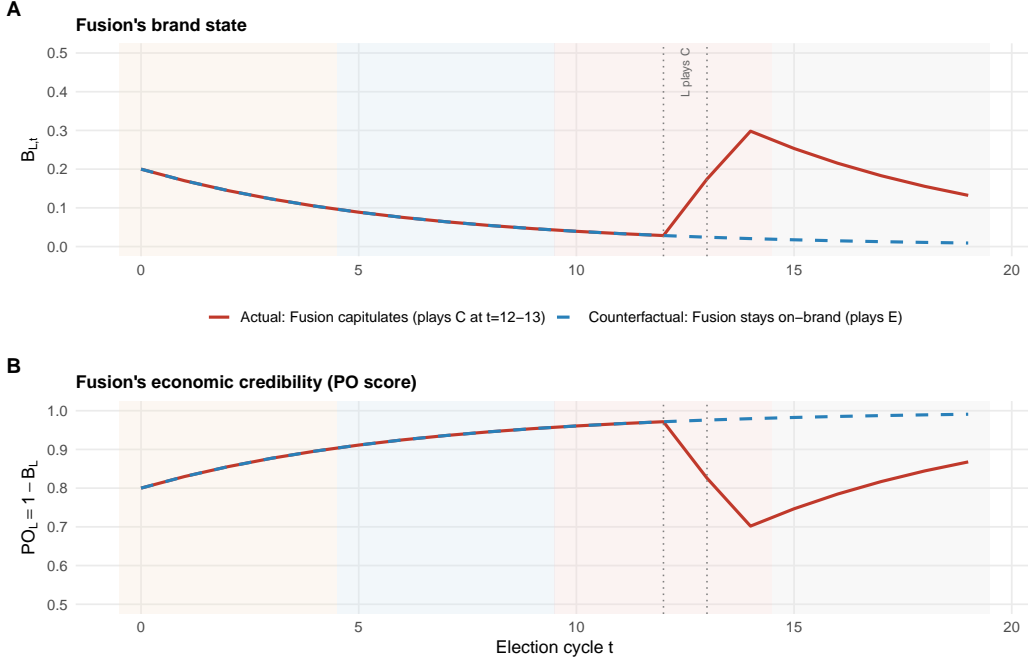


Figure 2: **Fusion’s dilemma: brand dilution under capitulation.** The figure traces the trajectory of Fusion’s economic brand state $B_{FUS,t}$ under two scenarios: the equilibrium path (Fusion maintains its economic platform) and the capitulation deviation (Fusion adopts cultural campaigns at dates t_1 and t_2). Under equilibrium play, $B_{FUS,t}$ converges toward zero, deepening the credibility gap with the Democratic Party. Under capitulation, $B_{FUS,t}$ shifts sharply upward, destroying economic coherence and rendering subsequent economic campaigns ineffective. The parametric benchmark uses $\delta = 0.15$, $B_{FUS,0} \approx 0.03$.

C Moment-Robust Traps and Implementation with Moment Conditions

The conditions of Proposition 1 are stated for a specific type space Θ and a common prior $\rho_t \in \Delta(\Theta)$. In practice, neither party directly observes the joint distribution of voter types; what is observable—through electoral returns, polling, and realized vote shares—are low-dimensional moment conditions on Θ . A voter trap that requires precise knowledge of the full type distribution to implement is fragile; one that holds across all type spaces consistent with a few observable moments is *moment-robust*.

Following the moment-condition approach of Ollar and Penta [2025], we say that a set of conditions is implementable under moment conditions $\mathcal{C}(\Theta)$ if it holds for every type space Θ' consistent with $\mathcal{C}(\Theta)$. Let $\mathcal{C}(\Theta) = \{E_{\Theta}[\varphi_k] = \mu_k : k = 1, \dots, K\}$ denote a finite collection of moment restrictions on the type distribution, where each φ_k is a bounded measurable function of voter types.

Definition 5 (Moment-robust trap). *A voter trap at election date t is robust to moment conditions \mathcal{C} if: for every type space Θ' consistent with $\mathcal{C}(\Theta)$, there exists a campaign $c'_{R,t} \in \mathcal{F}_{R,t}(\Theta')$ such that the voter trap conditions of Definition 3 hold for voter LC under Θ' .*

Sufficient conditions for moment-robustness. The key observation is that Proposition 1’s conditions depend on the type space Θ only through two channels: the electoral feasibility condition (which requires $\text{VS}_R(c_{R,t}) > 1/2$ under ρ_t) and the brand-constraint condition on $\mathcal{F}_{L,t}$ (which determines what L can credibly counter). The evidential-credibility condition (i) and salience condition (ii) are properties of the public realization x_t and the brand state $(B_{R,t}, B_{L,t})$, which are common knowledge and do not depend on the type distribution.

This separation implies that robustness of conditions (i) and (ii) is automatic—they hold for all type spaces. Full robustness of the trap additionally requires that L ’s live campaign set $\mathcal{F}_{L,t}$ does not expand under alternative type distributions in a way that introduces an effective counter-campaign. The moment conditions below are chosen to ensure both R ’s electoral feasibility and the invariance of L ’s brand-feasible set (which depends on brand states, not the electorate).

Electoral swing as the identifying statistic. A direct moment condition on the share of cross-pressured voters μ_{LC} would require that type to be observable—but cross-pressured voters have no incentive to reveal themselves, and their share is not recoverable from any single election’s returns alone. What is observable, however, is the *electoral swing*: the change in party R ’s vote share between consecutive elections. Under the model’s voting structure, UC voters vote R regardless of model adoption (they prefer R on both dimensions; see the proof of Proposition 1 in Appendix A), and UP and LP voters vote L regardless.¹⁴ Only type- LC voters switch across elections in response to whether the trap conditions hold. It follows that the swing in R ’s vote share between a reference election t_{ref} at which trap conditions did not hold and a candidate trap election t_0 at which they do is

$$\sigma_R(t_0) := \text{VS}_R(t_0) - \text{VS}_R(t_{\text{ref}}) = \mu_{LC}, \quad (32)$$

exactly— μ_{LC} is point-identified from the observed electoral swing between a pre-trap and trap election, both quantities recoverable from published electoral returns without any

¹⁴This maintained property requires that UC voters satisfy $\tau_{UC}^* < \tau_R$ and LP voters satisfy $\tau_{LP}^* > \tau_L$ across all x_t realizations in the equilibrium signal support—so that no realization within the equilibrium distribution generates a vote switch for these corner types. These are maintained parameter bounds, consistent with the policy-ordering assumption (1): the strict policy ranking ensures that corner types’ ideal points lie outside the range of electoral competition for all realizations the model generates in equilibrium.

polling or type-revelation mechanism.

Proposition 5 (Moment-robust voter trap). *Suppose the evidential-credibility condition (i) and salience condition (ii) of Proposition 1 hold with strict slack $\epsilon > 0$, and that the parameter restriction $\tau_{UC}^* < \tau_R$ holds. Let $t_{\text{ref}} \in \mathcal{T}_e$ be the most recent election at which trap conditions did not hold, and define the observed electoral swing $\sigma_R(t) := \text{VS}_R(t) - \text{VS}_R(t_{\text{ref}})$. Let $\mathcal{C}(\Theta)$ consist of the single moment condition $\sigma_R(t) \geq \underline{\sigma}$, where $\underline{\sigma} > 0$ is chosen large enough to guarantee $\text{VS}_R(c_{R,t}) > 1/2$ across all type spaces Θ' consistent with $\mathcal{C}(\Theta)$. Then:*

- (a) **Automatic robustness.** *Conditions (i) and (ii) hold identically for all type spaces Θ' consistent with $\mathcal{C}(\Theta)$, since they depend only on the public realization x_t and brand states $(B_{R,t}, B_{L,t})$.*
- (b) **Conditional robustness.** *Provided that L 's brand-feasible set $\mathcal{F}_{L,t}^{\text{brand}}$ contains no campaign satisfying condition (iii) of Definition 3, the full voter trap is moment-robust to $\mathcal{C}(\Theta)$: it holds for all Θ' consistent with $\mathcal{C}(\Theta)$.*

Under part (b), party R needs only observe the electoral swing between the current election and the most recent pre-trap baseline to verify the moment condition.

Part (a) is unconditional: evidence quality and salience amplification are properties of the public signal environment and brand histories, not of the electorate. Part (b) adds a maintained assumption: L 's brand-feasible set must not contain an effective counter-campaign. This assumption is verifiable from brand-state data. In the parametric benchmark, it holds whenever $B_{L,t}$ is sufficiently below $B_{R,t}$ that no live campaign in $\mathcal{F}_{L,t}$ achieves the PO threshold (see Appendix B for the closed-form brand-state condition). Electoral feasibility requires that R 's campaign commands a majority, which depends on μ_{LC} through vote-share arithmetic. Since the electoral swing (32) equals μ_{LC} under the model, a lower bound on the observed swing is a lower bound on the cross-pressured share, observable from any pair of pre-trap and in-trap electoral returns without requiring voters to report or reveal their types.

Proof of Proposition 5. Let Θ' be any type space consistent with $\mathcal{C}(\Theta)$.

Step 1 (Conditions (i) and (ii) are type-space-independent). Proposition 1's conditions (i) and (ii) are properties of $(x_t, B_{R,t}, B_{L,t})$, which are common knowledge and do not depend on the distribution of voter types. Since they hold with strict slack $\epsilon > 0$ at Θ , they hold identically at Θ' .

Step 2 (Brand constraint is type-space-independent for the PO component). The PO comparison in condition (iii) is performed by voter LC using her adoption rule (16). The result depends on $(x_t, B_{R,t}, B_{L,t})$ and not on the type distribution. For any $c \in \mathcal{F}_{L,t}^{\text{brand}}$, the strict slack in condition (i) ensures $\text{PO}_{LC,R,t}(m_{R,t}) > \text{PO}_{LC,L,t}(m) + \epsilon$, regardless of Θ' .

Step 3 (Electoral feasibility of R's campaign across Θ'). Under conditions (i) and (ii), all type- LC voters adopt $m_{R,t}$ and vote R by Lemma 2. Type- UC voters (upper class, conservative: $\epsilon_i = \epsilon > 0$, $\psi_j = -\psi$) have rational bliss point $\tau_{UC}^* = \nu - b\psi - 1 - \epsilon$, which lies below $\tau_R < \tau_L$, and cultural ideal $q_{UC}^* = -\psi$, which lies on the same side as $q_R < q_L$. Under the policy ordering (1), type- UC voters prefer party R on both economic and cultural dimensions for any identity weight, and hence vote R regardless of model adoption. Type- UP and type- LP voters prefer L on both dimensions and vote L regardless of model adoption (by an analogous argument). Therefore $\text{VS}_R(c_{R,t}) = \mu_{UC} + \mu_{LC}$ under any Θ' , and only type- LC voters change their vote across elections. By (32), the observed electoral swing $\sigma_R(t_0) = \text{VS}_R(t_0) - \text{VS}_R(t_{\text{ref}})$ equals μ_{LC} exactly: μ_{UC} cancels in the difference (it is constant across elections since UC voters never switch), and μ_{LP} and μ_{UP} contribute nothing to the swing. The moment condition $\sigma_R(t) \geq \underline{\sigma}$ therefore implies $\mu_{LC} \geq \underline{\sigma}$, and $\underline{\sigma}$ can be chosen (from the observed pre-trap vote share $\text{VS}_R(t_{\text{ref}}) = \mu_{UC}$) such that $\mu_{UC} + \mu_{LC} > 1/2$. Hence $c_{R,t} \in \mathcal{F}_{R,t}^{\text{elec}}(\Theta')$.

Steps 1–3 verify Definition 5: for every Θ' consistent with $\mathcal{C}(\Theta)$, the campaign $c'_{R,t} = c_{R,t}$ satisfies all voter trap conditions. \square

D Scope and Limitations of the Historical Application

The parametric simulation in Figure 1 demonstrates that the voter trap mechanism reproduces the qualitative arc of the 1890s episode quantitatively: Fusion victories during economic crisis, deepening brand asymmetry throughout, an absorbing trap under the White Supremacy Campaign’s cultural saturation, and the capitulation dynamics that destroyed the Fusion coalition’s economic brand. The model’s three formal conditions—evidential and brand advantage, salience amplification, and brand constraint on the opposition—each find specific historical counterparts in the calibration. Several critical features of the historical episode, however, lie outside the model’s scope; we highlight four.

Coalition coordination asymmetry. The model treats each party as choosing a single campaign from its feasible set; it does not distinguish between a centrally coordinated propaganda operation and a fragmented collection of independent outlets. In practice, the asymmetry in organizational capacity was decisive. Simmons ran a centralized operation with daily message coordination between Daniels, Jennett’s cartoon desk, a speakers’ bureau of prominent lawyers and politicians, and county-level “White Government Unions.” The Fusion opposition had no equivalent: Butler’s *Caucasian* operated independently of the Black press, which operated independently of white Republican papers, with no unified message, no shared editorial calendar, and no pooled resources. The Democratic campaign was financed through Simmons’ network of railroad, banking, and industrial interests who expected regulatory relief in return; the opposition survived on stock sales and local advertising. A richer model would endogenize campaign coordination as a function of organizational capacity, allowing the feasible set $\mathcal{F}_{p,t}$ to vary not only with brand history but with the party’s ability to solve its own collective-action problem.

Social punishment and ostracism. The model restricts voter behavior to voting: identity salience is shaped by party campaigns and public realizations, but voters impose no social costs on one another. The historical record is replete with such costs. Opponents of the Democratic line risked being labeled “scalawag” (a white Southerner who cooperated with Black citizens or Republicans) or “carpetbagger” (an outsider imposing unwelcome change)—terms that carried real social and economic penalties. As Red Shirt intimidation intensified, advertisers withdrew from opposition newspapers, readers feared being seen with them, and the economic base for independent journalism eroded. The social punishment channel reinforces the voter trap by raising the cost of acting on one’s economic interests when doing so requires visibly defecting from the dominant cultural frame. Microfounding identity weight allocation through community-level enforcement is a natural extension that

would make the trap self-enforcing at the social as well as the electoral level.

Violence and extralegal coercion. The voter trap is a model of electoral persuasion: it predicts how voters update beliefs and allocate identity weight, not how parties deploy physical force. The Wilmington Coup, the Red Shirt intimidation of voters at polling stations, and the systematic threat of lynching are outside the model’s formal scope. What the model *can* explain is the electoral conditions—Democratic majorities locked in by the voter trap—that gave the party the political cover and institutional capacity to act. The armed overthrow of Wilmington’s Fusion government was a historical contingency enabled by those electoral conditions, not a prediction of the formal mechanism. The disenfranchisement laws that followed were similarly an institutional action taken by officeholders whose electoral position was secured, in part, through the trap’s operation. A complete account of the 1890s episode requires supplementing the voter trap mechanism with models of collective violence [e.g., Wormser, 2004] and institutional consolidation.

Platform separation and the post-1960s realignment. Proposition 3 establishes that the policy ordering is sustained endogenously when the brand mechanism disciplines campaign choices. This result applies naturally to the 1890s episode, where the Fusion and Democratic platforms were stable across the simulation window and the operative strategic variable was cultural salience amplification. The framework does abstract, however, from episodes in which a party’s genuine policy commitment fundamentally reassigns its brand through external institutional forces that override the brand-maintenance incentive. The post-1965 party realignment in the American South is the canonical case: the Civil Rights Act and the Voting Rights Act shifted the national Democratic Party’s association from economic redistribution toward racial integration at the platform level, not merely through narrative framing. This platform-level brand shift created a new cohort of cross-pressured voters, Southern white working-class voters whose economic interests remained aligned with the Democratic platform but whose cultural ideal had migrated toward the Republican position, and the Republican Southern Strategy exploited exactly the salience-amplification mechanism our model formalizes. Bernini et al. [2025] document that the VRA immediately triggered white countermobilization consistent with Phase III dynamics; Kuziemko and Washington [2018] show that the Civil Rights Act caused a lasting partisan realignment among racially conservative whites. A natural extension would allow institutional shocks or generational turnover to override the brand-maintenance incentive of Proposition 3, shifting the baseline cultural association q_p and initiating a new cycle of brand-building and salience competition. Such an extension would nest the 1890s and 1960s episodes in a unified frame-

work and speak to modern instances of cross-pressured electorates in which the party of economic redistribution has acquired a cultural brand that is orthogonal or hostile to the preferences of its historic base.