PREPRINT v1.0.1 | Not peer-reviewed

The Doctrine of Consensual Sovereignty: Quantifying Legitimacy in Adversarial Environments

The Axiom of Consent

Murad Farzulla[1] 0009-0002-7164-8704

[1] *Farzulla Research*

November 2025

Correspondence: murad@farzulla.org

Abstract

This paper develops a unified analytical framework for measuring political legitimacy across heterogeneous governance domains. Building on insights from constitutional political economy, social choice theory, and institutional analysis, the framework establishes consent-holding—the mapping from decision domains to those with authority over them—as a structural necessity of collective action. We formalize this intuition through five axioms and five theorems, demonstrating that legitimacy can be operationalized as stakes-weighted consent alignment $\alpha(d, t)$, while friction $F(d, t)$ measures the deviation between outcomes and stakeholder preferences. The framework bridges normative democratic theory and empirical prediction, generating testable hypotheses about institutional stability. Historical validation examines suffrage expansion, abolition movements, labor rights, and contemporary platform governance, demonstrating how misalignment between stakes and voice generates observable instability. Unlike existing approaches that prescribe ideal institutions, this framework provides analytical tools for measuring legitimacy within any governance structure, enabling systematic comparison across democratic, technocratic, and algorithmic systems. Computational mechanism comparison via Bayesian learning dynamics across 1000 Monte Carlo runs demonstrates relative performance under adaptive agents: when preferences update based on observed policy outcomes, stakes-weighted DoCS achieves highest final alignment ($\alpha = 0.872$) with lowest terminal friction ($F = 1.6$, 98.5% reduction from initial). This comparative advantage holds across static baseline ($\alpha = 0.627$), learning dynamics ($\alpha = 0.872$), and alternative temporal mechanisms, suggesting stakes-weighting produces superior initial matches that persist even when agents adapt to institutional performance. The framework's domain-specific approach resolves the apparent tension between consent and competence, showing both as complementary dimensions of institutional legitimacy. This framework is part of the Adversarial Systems Research program, which examines stability, alignment, and friction dynamics in complex systems where competing interests generate structural conflict.

Keywords: legitimacy, consent, political stability, social choice, institutional design, friction, stakes-weighting

JEL Codes: D70 (Social Choice), D71 (Social Choice; Clubs; Committees; Associations), P16 (Political Economy)

Publication Metadata

DOI: 10.5281/zenodo.17684676

Version: 1.0.1

Date: November 2025

Interactive Visualizations: https://farzulla.org/research/consent-theory/dashboard.html

Methodologies: Research methodologies and reproducibility practices are documented at https://farzulla.org/methodologies

Note on Prior Work

This is version 1.0.0 (November 2025), the initial preprint release establishing the theoretical foundations of consent-holding theory. Future versions will expand empirical validation with fully quantified historical cases (women's suffrage 1890-1920 planned for v2.0.0).

Research Context

This work forms part of the Adversarial Systems Research program, which investigates stability, alignment, and friction dynamics in complex systems where competing interests generate structural conflict. The program examines how agents with divergent preferences interact within institutional constraints across multiple domains: political governance (this paper), financial markets (cryptocurrency volatility and regulatory responses), human cognitive development (trauma as maladaptive learning from adversarial training environments), and artificial intelligence alignment (multi-agent systems with competing objectives).

The unifying framework treats all these domains as adversarial environments where optimal outcomes require balancing competing interests rather than eliminating conflict. In political systems, this manifests as the tension between stakeholder consent and technocratic competence. In financial markets, it appears as the conflict between regulatory stability and market innovation. In human development, it emerges as the challenge of learning accurate models from noisy or adversarial training data. In AI systems, it surfaces as the alignment problem

when multiple agents optimize for different reward functions.

The Doctrine of Consensual Sovereignty presented here provides the theoretical foundation for analyzing legitimacy in any adversarial environment by formalizing the relationship between stakes, voice, and friction. Future work will extend this framework to algorithmic governance systems, multi-stakeholder climate negotiations, and autonomous agent coordination problems where consent structures remain undefined but friction dynamics are already observable.

Key Notation

Symbol	Definition
$H_t(d)$	Consent-holder mapping (who decides in domain d at time t)
$s_i(d)$	Stakes of agent i in domain d (material/capability exposure)
$C_{i,d}$	Consent power of agent i in domain d (decision authority)
$\alpha(d,t)$	Consent alignment (stakes-weighted share of voice held by stakeholders)
$F(d,t)$	Friction (stakes-weighted deviation between outcomes and preferences)
$L(d,t)$	Legitimacy ($w_1 \cdot \alpha + w_2 \cdot P$, balancing consent and performance)
$P(d,t)$	Performance/competence metric (domain-specific outcome quality)
$x^*_{i,d}$	Agent i's ideal action in domain d (preference)
$x_d(t)$	Realized action/outcome in domain d at time t

Scope and Limitations

This paper presents a conceptual framework for analyzing legitimacy in adversarial environments. While we formalize core relationships mathematically—consent alignment $\alpha(d,t)$, friction $F(d,t)$, and legitimacy $L(d,t)$—complete operational measurement remains ongoing empirical work. Our contribution is providing analytical architecture that makes legitimacy comparable across governance domains, enabling systematic analysis previously confined to domain-specific theories.

We demonstrate proof-of-concept through computational validation via agent-based simulation and qualitative validation across seven historical domains (suffrage, abolition, labor rights, civil rights, LGBT inclusion, platform governance, climate policy). **Methodological note**: The computational models assume agents learn from outcomes (Bayesian updating), which by construction reduces friction as preferences converge—this compares which consent mechanisms produce superior alignment given plausible behavioral assumptions, not whether friction *can* reduce (that follows definitionally). Measurement challenges are acknowledged in Section 4.

This v1.0.0 preprint establishes theoretical foundations and demonstrates implementability; subsequent versions will expand empirical validation with fully quantified historical cases, refine measurement protocols, and extend applications to algorithmic governance, climate negotiations, and multi-agent AI systems.

Contents

1 Introduction

Political legitimacy presents a fundamental puzzle: how can we measure whether authority is rightfully held across radically different governance domains? A state legislature, corporate board, algorithmic content moderation system, and common-pool resource management regime all make consequential decisions affecting stakeholders, yet existing frameworks struggle to provide unified analytical tools for assessing their legitimacy. Democratic theory emphasizes popular sovereignty, grounding legitimacy in consent of the governed (Locke, 1980; Rousseau, 1997; Rawls, 1971; Habermas, 1984), public choice highlights constitutional constraints (Buchanan and Tullock, 1962), while recent work on algorithmic governance introduces new challenges to consent-based legitimacy (Grimmelikhuijsen et al., 2022). What remains elusive is a framework capable of both normative evaluation and empirical prediction that applies consistently across domains.

This paper addresses this gap by developing consent-holding theory, an axiomatic framework that treats legitimacy as a structural property of decision-making systems rather than a binary classification. The central insight is deceptively simple yet powerful: in any domain where collective decisions produce shared consequences, someone must hold the authority to decide. This consent-holder mapping $H_t(d)$—identifying who decides in domain d at time t—is not a normative choice but a logical necessity arising from the structure of collective action itself. The framework's contribution lies not in prescribing who should hold consent, but in providing rigorous tools for measuring the consequences of any particular allocation.

The framework makes three distinct contributions to political theory and institutional analysis. First, it establishes a formal connection between consent alignment and observable political friction. While democratic theorists have long argued that excluding affected stakeholders undermines legitimacy (Estlund, 2008), existing accounts lack operational metrics for testing these claims. We define consent alignment $\alpha(d, t)$ as the stakes-weighted share of decision power held by affected parties, and friction $F(d, t)$ as the stakes-weighted deviation between outcomes and stakeholder preferences. The framework predicts that persistent misalignment generates measurable instability—protests, non-compliance, institutional breakdown—making legitimacy empirically falsifiable rather than purely philosophical.

Second, the framework resolves the apparent tension between consent and competence through a competence-consent tradeoff theorem (T4). Epistemic democrats argue that inclusive decision-making produces better outcomes through cognitive diversity (Landemore, 2013; Hong and Page,

2004), while critics worry that expanding consent sacrifices technical expertise. Our framework shows these concerns reflect different positions on the legitimacy frontier: some domains optimally weight performance highly (nuclear safety, pandemic response), while others prioritize consent alignment (constitutional amendments, community norms). Rather than declaring one approach universally superior, the framework provides tools for identifying domain-appropriate balances.

Third, this approach enables systematic historical and comparative analysis. By operationalizing legitimacy as $\alpha(d, t)$ and friction as $F(d, t)$, we can trace institutional evolution quantitatively. Franchise expansion emerges not as discrete events but as gradual increases in $\alpha(d)$ driven by the accumulating friction $F(d)$ from excluding high-stakes populations. Women's suffrage movements, abolition struggles, labor organizing, and contemporary platform governance rebellions all exhibit the same underlying dynamic: groups with high stakes $s_i(d)$ but zero consent power C_i generate sustained friction until incorporation or suppression occurs. This pattern, predicted by the framework's core theorems, provides empirical validation across centuries and continents.

The framework proceeds from seven minimal axioms to five core results establishing structural necessities (Section 3). Theorem 1 demonstrates consent-holding necessity: wherever decisions occur, some mapping $H_t(d)$ must exist. Theorem 2 establishes inevitable friction: plural preferences guarantee that someone's interests will be compromised unless perfect alignment obtains. Definition 1 operationalizes legitimacy as stakes-weighted consent alignment, providing an empirical metric. Postulate 1 formalizes the competence-consent tradeoff, showing legitimacy as a weighted combination $L = w_1 \cdot \alpha + w_2 \cdot P$. Theorem 3 derives a minimal absolutism from value relativism: even if content-level values are frame-dependent, the existence of consent-holding structures remains invariant.

Section 2 situates this framework within nine research traditions—constitutional political economy (Buchanan and Tullock, 1962), social choice theory (Arrow, 1951; Sen, 2017), stakeholder theory (Freeman, 1984), common-pool resource governance (Ostrom, 1990), deliberative democracy (Fishkin, 2018; Habermas, 1990), algorithmic governance (Barocas et al., 2019), epistemic democracy (Estlund, 2008; Landemore, 2013), relational autonomy (Mackenzie, 2014), and legitimacy theory (Scharpf, 1999; Schmidt, 2013). Rather than competing with these approaches, consent-holding theory provides a unifying analytical architecture: each tradition contributes insights about how consent should be allocated or what constitutes legitimate use of author-

ity, while our framework offers measurement tools applicable regardless of normative commitments.

Section 4 operationalizes the framework for empirical application, specifying proxy variables for consent power C_i (voting weights, agenda control, board representation), stakes $s_i(d)$ (exposure measures, capability impacts, revealed preferences), and friction $F(d)$ (protest incidence, litigation rates, policy reversals). This operationalization enables econometric identification strategies using panel data with institutional variation as instruments for consent alignment. Historical validation (Section 6) examines six cases spanning two centuries: women's suffrage (1890s-1970s), abolition (1780s-1860s), labor rights (1850s-1930s), civil rights (1950s-present), LGBT inclusion (1969-present), and platform governance (2010s-present). Each case demonstrates the predicted pattern: high $s_i(d)$ combined with zero C_i generates rising $F(d)$ until elites respond through suppression or incorporation.

Section 9 addresses seven major objections, from concerns about infinite regress in consent structures to worries that stakes-weighting enables plutocracy. Section 10 concludes by outlining the research agenda this framework enables: cross-national legitimacy indices, institutional experiments varying $\alpha(d)$ systematically, and applications to emerging domains (AI governance, climate policy, platform regulation) where consent structures remain contested.

The framework's title—"consent-holding" rather than "consent theory"—reflects its analytical focus. This is not another account of why consent matters normatively, but a systematic investigation of how consent operates structurally. Just as markets emerge from property rights and contracts regardless of normative justifications for capitalism, consent-holding structures emerge from the necessity of collective decision-making regardless of democratic commitments. The framework's power lies in making these structures visible, measurable, and comparable, enabling rigorous analysis of legitimacy claims that have historically remained philosophically contested but empirically elusive.

2 Literature Review and Theoretical Foundations

The consent-holding framework synthesizes and extends insights from nine distinct research traditions. This section reviews each tradition's core contributions, identifies limitations the framework addresses, and demonstrates how operationalizing legitimacy as $\alpha(d, t)$ and friction as $F(d, t)$ enables empirical validation of long-standing theoretical claims.

2.1 Constitutional Political Economy

Building on earlier social contract foundations from Hobbes (1651) who established

consent as prerequisite for legitimate political authority, Buchanan and Tullock's (1962) seminal work establishes constitutional choice as a distinct analytical problem requiring different decision rules than ordinary politics. Brennan and Buchanan (1985) further develop this constitutional economics framework, distinguishing levels of collective action and establishing how constitutional rules create frameworks for collective choice. Their framework rests on several foundational insights that anticipate the consent-holding approach. First, they distinguish between constitutional rules—rarely changed frameworks establishing decision procedures—and political decisions made within those rules. This maps directly onto our concept of nested consent-holding: $H_t(d_{meta})$ represents the consent-holders for constitutional domains, while $H_t(d)$ operates within constraints established at the meta-level.

Second, Buchanan and Tullock argue that rational agents behind a "veil of uncertainty" would unanimously consent to rules benefiting all. Once constitutional structures are established, majority rule becomes acceptable for routine decisions. This anticipates our Theorem 1: consent-holding exists at every level, from object-level policy to constitutional design to amendment procedures. Third, their exchange paradigm treats politics as mutual exchange of consent rather than top-down command. Government achieves legitimacy when citizens "purchase" its services consensually through constitutional agreement. The consent-holding framework formalizes this metaphor rigorously through stakes-weighted alignment metrics.

Finally, Buchanan and Tullock model optimal decision rules as minimizing total costs combining external costs (harm from decisions affecting you without your consent) and decision costs (time and effort required to reach agreement). Our friction metric $F(d)$ captures external costs precisely as stakes-weighted deviations from stakeholder ideal points. The framework extends Buchanan and Tullock in four crucial respects. First, we introduce stakes-weighting $s_i(d)$, recognizing that individuals are heterogeneously affected by policies. Second, while Buchanan focuses on one-time constitutional founding moments, we model consent-holding as continuously operating through $H_t(d)$, tracking legitimacy dynamically as institutional configurations evolve. Third, Buchanan discusses "the" social contract; we specify that consent-holding varies across domains d, with different optimal structures for taxation, criminal justice, environmental regulation, and community norms. Fourth, Buchanan provides normative theory; we operationalize concepts through $\alpha(d,t)$ and $F(d,t)$, enabling empirical validation of constitutional designs rather than purely philosophical jus-

tification.

2.2 Social Choice Theory and Impossibility Results

Building on Dahl (1956) foundational analysis of democratic theory showing how pluralist democracy requires balancing majority rule with minority rights, Arrow's (1951) impossibility theorem establishes that no ranked voting system can simultaneously satisfy four seemingly minimal desiderata: Pareto efficiency, non-dictatorship, independence of irrelevant alternatives, and unrestricted domain. This result demonstrates that perfect democratic aggregation is mathematically impossible, not merely practically difficult. The Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975) extends this impossibility to strategy-proofness: any non-dictatorial voting mechanism over three or more alternatives is manipulable.

Recent quantitative versions show these aren't merely theoretical concerns but quantifiably common, with Keller (2012); Mossel et al. (2012); Friedgut et al. (2011) quantifying how often Arrow's impossibility manifests in real voting scenarios with finite electorates. Sen's (2017) expanded treatment of collective choice integrates economics and ethics, introducing the capability approach that maps directly onto our effective voice concept. Sen (1999) argues in *Development as Freedom* that development should be measured not by utility or resources alone but by capabilities—freedoms to achieve valued functionings like health, education, and political participation. This provides theoretical grounding for our eff_voice$_i(d)$ term: possessing formal consent power $C_i > 0$ without resources, education, or political freedom represents low capability.

The consent-holding framework relates to social choice theory as meta-analysis rather than competitor. Where Arrow and Gibbard-Satterthwaite ask "which aggregation rule is best?", we ask "how legitimate is any given aggregation rule?" This shift has three implications. First, our framework doesn't compete with impossibility results; it builds on them by providing tools for measuring consequences of unavoidable trade-offs. Since perfect rules don't exist, we need metrics for comparing imperfect options. Second, stakes-weighting $s_i(d)$ isn't present in classical social choice theory, which typically assumes equal weights. This extension allows domain-specific analysis: simple majority may be optimal for low-stakes routine legislation, while supermajority or even consensus becomes appropriate when stakes concentrate heavily.

Our stakes-weighting approach builds on but diverges from weighted voting power analysis (Banzhaf III, 1965; Shapley and Shubik, 1954). The Shapley-Shubik and Banzhaf power indices measure *effective* voting power given formal weights in committee systems—recognizing that a voter

with 40% weight may have more than 40% actual power if they're pivotal in coalitions. This literature addresses measurement of power distribution within given institutional arrangements. Our framework addresses the prior question: how should consent power C_i, d be allocated based on stakes $s_i(d)$? While power index theory takes weights as given and calculates resulting influence, we propose stakes as foundation for determining appropriate weights. Future work integrating these approaches could specify stakes-weighted allocations, then apply Banzhaf or Shapley-Shubik indices to measure resulting effective voice, combining normative allocation principles with positive power analysis.

2.3 Stakeholder Theory and Corporate Governance

Building on Pitkin (1967) foundational work on representation distinguishing substantive versus descriptive representation and acting for constituents, Freeman's (1984) stakeholder approach argues that firms should create value for all stakeholders—employees, suppliers, communities, customers, shareholders—not just maximize shareholder returns. This challenges Friedman (1970) shareholder primacy doctrine, which treats profit maximization as the sole corporate responsibility and argues that corporate social responsibility beyond shareholder wealth maximization is fundamentally misguided. The 2019 Business Roundtable statement endorsing stakeholder capitalism, signed by 200 CEOs, marks mainstream acceptance of Freeman's stakeholder view (Business Roundtable, 2019), representing a significant shift from the Friedman doctrine. Phillips (2003) further develops this framework by distinguishing stakeholders by the moral obligation owed to them versus their ability to affect the organization, providing a typology of stakeholder legitimacy that maps onto our stakes-consent framework.

The framework operationalizes Freeman's insights by defining stakeholders precisely as agents with $s_i(d) > 0$ in corporate domains. Current governance structures grant consent power almost exclusively to shareholders: they elect boards, approve major transactions, and receive residual claims. Employees, despite high stakes in employment security, working conditions, and workplace norms, hold negligible C_i in most Anglo-American firms. This generates low $\alpha(d_{corporate})$ when stakes are calculated comprehensively. The framework predicts such misalignment produces friction $F(d)$: labor disputes, regulatory pressures, reputation damage, difficulty attracting talent.

Comparative corporate governance research validates these predictions. Vitols (2011) documents how German codetermination—mandatory worker representation on supervisory boards—constrains hostile takeovers and maintains stakeholder orientation. Workers' voice (high $\alpha_{workers}(d)$)

prevents zero-sum shareholder maximization strategies. Fauver and Fuerst (2011) show codetermined firms invest more in worker training and career development; higher α produces performance improvements in human capital domains.

2.4 Common-Pool Resource Governance

Ostrom's (1990) groundbreaking work on common-pool resources challenges both "tragedy of the commons" pessimism and top-down state solutions. Through field studies of fisheries, forests, irrigation systems, and groundwater basins across continents, she demonstrates that resource users frequently develop effective self-governance without privatization or centralized authority. Her eight design principles for successful commons management include particularly relevant insights for consent-holding theory. Design Principle 3 requires that "most individuals affected by the operational rules can participate in modifying the operational rules"—essentially mandating high $\alpha(d_{rules})$ for those with high $s_i(d_{resources})$. Design Principle 8 specifies nested enterprises for larger systems, enabling polycentric governance with consent-holding at multiple scales. This builds on earlier insights from Ostrom et al. (1961) on polycentric systems, demonstrating how multiple governing authorities at different scales can achieve better outcomes than monocentric alternatives.

The framework formalizes Ostrom's intuitions. Her "collective choice arrangements" represent $H_t(d)$ mappings where users participate in rule modification. Her design principles can be reinterpreted as conditions enabling high $\alpha(d)$: clear boundaries (defining who holds s_i), local monitoring (ensuring C_i holders possess information), graduated sanctions (responses to low-α violations), and conflict resolution mechanisms (managing $F(d)$ when it arises). Successful commons maintain high consent alignment; failed commons exhibit persistent misalignment between stakes and voice.

Recent empirical work validates this interpretation quantitatively. Cox et al. (2010) conduct a meta-analysis showing that Ostrom's design principles predict commons sustainability across diverse contexts, providing systematic evidence that consent alignment mechanisms enable effective resource governance. Yadav et al. (2021) analyze 83 Amazonian communities managing arapaima fisheries, showing that Ostrom's design principles predict ecological outcomes systematically. Communities exhibiting collective choice arrangements (high α) maintain sustainable fish stocks; those lacking such arrangements experience depletion.

2.5 Deliberative Democracy and Mini-Publics

Building on Dahl (1971) polyarchy framework of participation and opposition and Mill (1861) considerations on representative government balancing participation

and competence, Habermas's (1984; 1990) communicative action theory distinguishes strategic action (oriented toward achieving one's goals) from communicative action (oriented toward mutual understanding through reasoned argument). Legitimate norms are those acceptable to all affected parties through rational discourse free from coercion. His discourse principle holds that "only those norms can claim validity that could meet with the acceptance of all concerned in their capacity as participants in a practical discourse." This maps onto consent-holding directly: "all concerned" represents our affected set $S_d = \{i|s_i(d) > 0\}$, while "acceptance" requires $C_i > 0$ in decision procedures $H_t(d)$.

Fishkin's (2009; 2018) deliberative polling research operationalizes these theoretical commitments. By convening randomly selected representative samples, providing balanced information, facilitating structured deliberation, and measuring preference changes, deliberative polls demonstrate that informed public judgment shifts significantly through discourse. Citizens' assemblies extend deliberative innovation to consequential policy domains. The Irish Citizens' Assembly (2016-2018) addressed abortion and climate change through 99 randomly selected citizens deliberating after expert input, demonstrating how sortition combined with deliberation can shift preferences systematically (Farrell et al., 2019). Courant and Bourgeron (2021) analyze the French Citizens' Convention on Climate (2019-2020), which generated 149 policy proposals from 150 randomly selected participants through sortition and deliberation, with many subsequently adopted into legislation.

The framework interprets these innovations as institutional experiments raising $\alpha(d)$ through sortition and deliberation. Random selection approximates equal C_i for participants; demographic stratification can approximate stakes-weighting if groups correlate with $s_i(d)$. Learning phases improve eff_voice$_i$ through information provision; deliberation structures enable preference refinement.

Our stakes-weighted consent framework confronts democratic equality arguments directly. Building on Mill (1859) foundations regarding individual liberty, consent, and limits of state power, Christiano (2008) defends equal political voice on dignity grounds: each person possesses equal moral status, entitling them to equal say in collective decisions regardless of stakes or competence. Waldron (1999) argues that persistent disagreement about what justice requires makes equal voice procedurally fair even if some possess superior judgment. Brighouse and Fleurbaey (2010) examine whether proportional influence could improve democratic outcomes but conclude that equal voice better respects equality of

persons.

We acknowledge this tension while distinguishing *political* domains from *governance* domains generally. In constitutional fundamentals and citizenship rights, equal voice may be intrinsically required by equal moral status—each person gets one vote precisely because they are persons, not because they possess equal stakes. But many governance domains are not *political* in this sense: corporate boards allocating firm resources, technical committees setting safety standards, platform algorithms moderating speech, common-pool resource users managing fisheries. In these contexts, stakes-weighting may be both more efficient (reducing friction, improving outcomes) and more legitimate (those bearing consequences should influence decisions proportionally). The framework enables empirical testing: do equal-voice or stakes-weighted mechanisms generate higher measured legitimacy $L(d, t)$ in different domain types?

2.6 Algorithmic Governance and Platform Legitimacy

Grimmelikhuijsen et al. (2022) identify three legitimacy dimensions for algorithmic decision-making: input (did citizen preferences inform design?), throughput (does the algorithm follow fair procedures?), and output (do outcomes align with public values?). Current algorithmic governance exhibits severe deficits across all three dimensions. Citizens rarely participate in algorithm design (low input legitimacy), decision-making processes remain opaque black boxes (low throughput legitimacy), and outcomes often replicate historical discrimination (questionable output legitimacy). Kleinberg et al. (2017) demonstrate inherent trade-offs in fair determination of risk scores, showing that multiple incompatible definitions of algorithmic fairness exist—making it impossible to satisfy all fairness criteria simultaneously, analogous to Arrow's impossibility theorem in social choice.

Waldman and Johnson (2022) show that high-stakes algorithmic decisions (healthcare allocation, criminal sentencing) are perceived as less legitimate than human decisions even when outcomes are identical. The consent-holding framework diagnoses these challenges structurally. Algorithmic decision-making creates domains $d_{algorithm}$ where algorithms or their designers hold $C \approx 1$ while affected citizens have $C \approx 0$ despite high $s_i(d)$. Credit scoring algorithms determine loan access (high s_i for applicants); hiring algorithms control employment opportunities (high s_i for candidates); content moderation algorithms shape speech norms (high s_i for platform users). In each case, current $\alpha(d_{algorithm}) \approx 0$ because high-stakes populations are excluded from $H_t(d)$.

Platform responses attempting to raise α reveal understanding of legitimacy deficits. Meta's Oversight Board provides independent content moderation appeals, slightly

raising $\alpha(d_{moderation})$ by giving users contestation rights, though Douek (2022) notes this provides only limited voice expansion while maintaining corporate control over fundamental rules. YouTube Creator Councils consult high-profile creators, extending partial C_i to stakeholders whose s_i is highest.

2.7 Voting Power Indices and Coalition Analysis

The power indices literature demonstrates that voting weight $\neq$ voting power. Banzhaf III (1965) measures critical voter frequency: how often removing your vote changes outcomes from win to loss. Shapley and Shubik (1954) measure pivotal voter frequency in sequential coalition formation. Felsenthal and Machover (1998) provide comprehensive comparison of these approaches, demonstrating that Banzhaf and Shapley-Shubik indices often diverge substantially and measure different aspects of voting power. These indices often diverge dramatically from nominal weights—Germany holds the most European Council votes but doesn't possess proportional power due to coalition dynamics. Similar phenomena arise in corporate boards (blockholders vs. minority shareholders), legislatures (swing voters vs. party leaders), and qualified majority systems (Security Council veto players).

These insights directly inform consent-holding operationalization. Naive approaches measure C_i as voting weight (shares held, seats controlled). Sophisticated approaches use power indices accounting for coalition structures. In weighted voting contexts (shareholders, federalism), qualified majority rules (constitutional amendments), and veto player systems (UN Security Council), indices capture actual influence more accurately than nominal weights.

The framework integrates power indices into legitimacy measurement: $\alpha(d,t) = \frac{\sum_i s_i(d) \cdot \text{PowerIndex}_i(d,t)}{\sum_i s_i(d)}$, where PowerIndex_i represents Banzhaf, Shapley-Shubik, or domain-appropriate measures. This refinement matters most when vote concentration enables blocking coalitions. Consider corporate governance: a minority shareholder with 20% equity plus veto rights over major transactions wields power far exceeding their ownership share. Measuring $C_i = 0.20$ understates influence; calculating Banzhaf index accounting for veto power provides accurate assessment.

Recent extensions analyze endogenous coalition formation (Aumann and Myerson, 1988), showing how equilibrium structures emerge from bargaining. This connects to consent-holding's dynamic aspect: $H_t(d)$ evolves as agents form alliances, shifting power distributions. Nash bargaining solutions (Nash, 1950) maximize products of utility gains subject to Pareto efficiency—structurally similar to stakes-

weighted consent maximization. Kalai and Smorodinsky (1975) propose alternative axiomatizations highlighting trade-offs between equality (proportional gain-sharing) and efficiency (Pareto optimality), demonstrating solution multiplicity absent unique normative commitments—precisely what Theorem 3 predicts.

2.8 Relational Autonomy and Consent Capacity

Mackenzie (2014) three-dimensional autonomy framework distinguishes self-determination (choosing one's own life path), self-governance (regulating one's actions), and self-authorization (taking responsibility for choices). Traditional liberal autonomy assumes atomistic individuals; relational approaches (Mackenzie and Stoljar, 2000) recognize that autonomy is socially constituted—relationships and social structures fundamentally enable or constrain autonomous choice rather than merely influencing pre-existing capacities. Nedelsky (1989) analyzes how oppressive social structures systematically constrain women's autonomy through relational mechanisms, demonstrating that coercion operates not only through direct force but through systematic limitation of available choices. Oppressive systems constrain capacity for self-governance—gender oppression limits women's educational access, economic opportunities, and freedom from violence, directly undermining autonomous choice.

Koggel (2022) extends this to global justice, arguing that respecting autonomy requires enabling threshold capabilities, not merely non-interference. Autonomy necessitates freedom conditions: political liberties (speech, association, conscience) and personal liberties (movement, bodily autonomy, freedom from violence). Agents lacking these conditions cannot exercise meaningful consent even if formally included in $H_t(d)$.

These insights address the framework's handling of eff_voice_i. Relational autonomy equals effective voice in our legitimacy equation. Simply granting $C_i > 0$ (voting rights) without resources, education, or freedom produces low eff_voice_i—formal authority without capacity to exercise it. Oppressive structures systematically reduce both stakes recognition (dominant groups deny subordinated groups' s_i) and consent power (exclusion from $H_t(d)$ even for high-stakes domains).

This perspective addresses three framework challenges. First, it resolves circularity concerns: "Who decides who's in $H_t(d)$?" Answer: those with stakes plus capacity, considering relational constraints that may undermine apparent consent. Second, it handles vulnerable populations ethically. Proxy consent becomes necessary when capacity is impaired, but structures should enable gradual inclusion as capabil-

ity develops rather than permanent exclusion. Third, it enables justice analysis: systematic exclusion of groups with high s_i but low eff_voice$_i$ constitutes legitimacy deficit diagnosable through $\alpha(d)$ measurement.

Application to research ethics illustrates these dynamics. Standard approaches grant legal guardians consent authority over cognitively impaired individuals. Relational approaches recognize impaired persons retain partial capacity and value particular relationships beyond legal guardianship—an older sibling may understand needs better than distant legal guardians. The framework prescription: allocate partial C_i based on measured capacity and expand $H_t(d)$ to include chosen trusted relationships, raising $\alpha(d_{\text{research}})$ for the affected individual.

3 Formal Framework: Primitives, Axioms, and Theorems

This section establishes the framework's formal foundations through precise definitions, minimal axioms, and structural theorems. The approach proceeds deductively: from spare assumptions about collective decision-making to necessary conclusions about consent-holding's existence, friction's inevitability, and legitimacy's measurement.

3.1 Primitives and Definitions

We begin with foundational concepts requiring no prior theoretical commitment. An **agent** is any entity capable of selecting among actions, indexed $i \in A = \{1, \ldots, N\}$. Agents may be individuals, organizations, algorithms, or collective bodies—the framework remains agnostic about internal composition. A **domain** represents a decision-relevant sphere—a policy area, firm process, household choice, or any context requiring action selection. The set of domains is $D = \{d_1, \ldots, d_M\}$. Each domain d admits a set of possible actions X_d, from which one action $x_d \in X_d$ must be selected.

Outcomes represent realized states resulting from action vectors $\mathbf{x} = (x_{d_1}, \ldots, x_{d_M})$ through an environment mapping $E : \prod_d X_d \to O$, where O denotes the outcome space. An agent i's **stake** in domain d, denoted $s_i(d) \geq 0$, quantifies sensitivity to outcomes in that domain. Stakes may reflect material exposure, legal consequences, capability impacts, or existential threats.

Each agent possesses **preferences** over outcomes, represented either as complete orderings $\succeq_i$ or utility functions $U_i : O \to \mathbb{R}$. Preferences induce ideal points $x^*_{i,d}$ in each domain—the action agent i most prefers given others' anticipated choices.

Consent represents the normative right to decide in a domain—who may authoritatively say "yes" or "no" to proposed actions. Following Locke (1980) consent theory foundations, political obligation derives from voluntary agreement; actual consent is required for legitimate authority. The

consent-holder mapping $H_t(d) \in \Delta(C)$ specifies the distribution of decision authority over possible holders C at time t. Individual **consent power** $C_{i,d} \in [0,1]$ represents agent i's effective share of decision authority in domain d, with $\sum_i C_{i,d} = 1$.

3.2 Axioms

The framework rests on seven axioms representing minimal commitments about collective decision-making.

A1. Action Precedence: Every non-null outcome in a domain is produced by some action (including "do nothing").

A2. Decision Requirement: Every action is selected by some decision procedure (choice, rule, randomization, delegation).

A3. Shared Reality: Outcomes alter a world co-occupied by multiple agents; externalities exist.

A4. Finitude: Agents have finite time, attention, and cognitive capacity; no single agent can decide everything alone.

A5. Plurality: Agents' preference orderings differ on at least some domains.

A6. Salience: For each domain, at least one agent has $s_i(d) > 0$.

A7. Fallibility/Subjectivity: Perception and valuation are frame-dependent; no universal content-level value ordering is logically forced.

3.3 Theorem 1: Consent-Holding Necessity

Theorem 3.1 (Consent-Holding Necessity)**.** *In any domain d where a non-null outcome occurs, there exists a consent-holder mapping $H_t(d)$.*

Proof Sketch. By A1-A2, any outcome resulted from an action selected through some procedure. A procedure implies a locus of control—the entity/entities choosing the action, establishing the choice rule, or delegating to randomization. This locus constitutes $H_t(d)$. Therefore, denying $H_t(d)$'s existence contradicts A2. □ □

3.4 Theorem 2: Inevitable Friction

Theorem 3.2 (Inevitable Friction)**.** *If there exist agents i, j with divergent preferences on domain d and $s_i(d), s_j(d) > 0$, then unless $H_t(d)$ exactly reproduces stakes-weighted unanimity, at least one agent experiences moral/political friction.*

We formalize **friction** in domain d as:

$$F(d,t) = \sum_i s_i(d) \cdot \delta(x_d(t), x^*_{i,d}) \quad (1)$$

where $x^*_{i,d}$ represents agent i's ideal action and δ measures divergence. For discrete choices, $\delta(x, x^*) = 0$ if $x = x^*$, else 1. For continuous policy spaces, $\delta(x, x^*) = |x - x^*|$ captures distance from ideal points.

Introducing tolerance thresholds τ_i yields:

$$F_\tau(d,t) = \sum_i s_i(d) \cdot \max(0, \delta(x_d, x^*_{i,d}) - \tau_i) \tag{2}$$

3.5 Definition 1: Legitimacy as Consent Alignment

We operationalize legitimacy through stakes-weighted consent alignment. Define the **affected set** $S_d = \{i | s_i(d) > 0\}$. **Consent alignment** is:

$$\alpha(d,t) = \frac{\sum_{i \in S_d} s_i(d) \cdot \text{eff\_voice}_i(d,t)}{\sum_{i \in S_d} s_i(d)} \tag{3}$$

where eff_voice_i represents agent i's effective decision power in $H_t(d)$.

Definitional Note: This is a *measurement framework*, not a derived result. We define legitimacy as the degree to which consent power tracks stakes distribution, making the concept empirically tractable. The framework's predictive power lies in the hypothesis that low α generates observable friction—a claim requiring empirical validation beyond the definition itself.

A minimal procedural legitimacy condition requires $\alpha(d,t) \geq \tau$ for society-specific threshold τ. Persistent $\alpha < \tau$ predicts observable friction through unrest, exit, sabotage, or normative decay.

3.6 Postulate 1: Competence-Consent Trade-Off

We model overall legitimacy as combining consent alignment and performance:

$$L(d,t) = w_1 \cdot \alpha(d,t) + w_2 \cdot P(d,t) \tag{4}$$

where $\alpha(d,t)$ represents stakes-weighted consent alignment, $P(d,t)$ denotes performance/competence metrics, and $w_1, w_2 \geq 0$ reflect society-specific weights on voice versus results.

This is a **postulated relationship** rather than a derived theorem. The linear combination assumes legitimacy trades off between consent and competence, but alternative functional forms (multiplicative, threshold-based) are possible. Empirical work validating this specification against alternatives remains a key research agenda.

Remark on Weight Determination: The weights w_1, w_2 are not free parameters requiring external normative specification, but endogenous scope conditions revealed through constitutional-level decisions (see Section 10.1 for full meta-legitimacy resolution). Societies whose weight configurations produce excessive friction face structural pressure to reform. We can characterize admissible weight functions axiomatically: any stable society must satisfy $w_2/w_1 > f(\text{Var}[s_i(d)])$ where f is a function of stakes heterogeneity derived from friction minimization. Future empirical work will

estimate weights via:

$$(w_1^*, w_2^*) = \arg\min_{w_1,w_2} \mathbb{E}[F(d,t;w_1,w_2)] \quad (5)$$

where friction minimization across constitutional reforms provides revealed preference data for weight estimation. Sequential Monte Carlo methods (Lux and Schiffko, 2018) enable parameter estimation for agent-based models through particle filtering, providing techniques applicable to calibrating consent-holding frameworks against empirical institutional data. This transforms weight determination from a normative choice into an empirical optimization problem, sidestepping the meta-legitimacy regress.

This formulation makes explicit that different systems optimize different points on the legitimacy frontier. Technocracies maximize P, often sacrificing α by concentrating consent in experts. Direct democracies maximize α through universal suffrage, potentially reducing P on technical domains where distributed knowledge is sparse.

3.7 Theorem 3: Minimal Absolutism from Relativism

Theorem 3.3 (Relativism ⇒ Minimal Absolutism)**.** *Given A7 (value frame-dependence), the claim "all value judgments are frame-relative" is coherent only if the* ***structure*** *enabling frames is invariant. Therefore, at least one absolute exists: the necessity of consent-holding over shared outcomes wherever A1-A6 hold.*

Proof Sketch. Suppose all value claims are frame-dependent (A7). Frame-dependence presupposes frames exist—perspectives from which valuations occur. Frames belong to agents inhabiting shared reality (A3) with plural preferences (A5). These agents make decisions affecting each other (A1-A2). Such decisions require consent-holder mappings $H_t(d)$ (Theorem 1). Therefore, relativism about content-level values doesn't extend to structural necessities. □ □

4 Operationalization: Empirical Measurement and Identification

The theoretical framework provides analytical tools for understanding consent-holding structures. This section bridges theory and empirical application by specifying how the framework's core concepts can be measured, how causal relationships can be identified econometrically, and what testable predictions emerge.

4.1 Formal Measurement Framework

We operationalize consent-holding through a consent matrix $\mathbf{C} \in [0,1]^{N\times M}$, where each element $C_{i,d}$ represents agent i's effective decision share in domain d, subject to the normalization constraint $\sum_i C_{i,d} = 1$. This matrix captures both de jure authority and de facto power. In simple majority voting systems with equal suffrage, $C_{i,d} = 1/N_{voters}$ for all enfranchised i and $C_{i,d} = 0$ for ex-

cluded populations. In shareholder governance, $C_{i,d} = \text{shares}_i/\text{total\_shares}$. In technocratic systems, $C_{i,d} = 1/|E|$ if agent i belongs to the expert set E, zero otherwise.

Complementing the consent matrix, the stakes vector $\mathbf{s}(d) \in \mathbb{R}^N_{\geq 0}$ quantifies each agent's exposure to consequences in domain d. Stakes measurement presents both conceptual and practical challenges. Conceptually, stakes may reflect material exposure (tax burden relative to income), capability impacts (health outcomes affected), or existential threats (survival risks from climate policy). Different domains may legitimately employ different stakes conceptions.

Combining these elements, consent alignment in domain d at time t is measured as:

$$\alpha(d,t) = \frac{\sum_{i \in S_d} s_i(d) \cdot \text{eff\_voice}_i(d,t)}{\sum_{i \in S_d} s_i(d)} \quad (6)$$

where $S_d = \{i | s_i(d) > 0\}$ denotes the affected set and eff_voice_i represents agent i's effective decision power accounting for both formal authority $C_{i,d}$ and capacity constraints.

4.2 Friction Metrics and Tolerance-Weighted Extensions

Political friction represents the stakes-weighted aggregate deviation between realized outcomes and stakeholder preferences. In its basic form:

$$F(d,t) = \sum_i s_i(d) \cdot \delta(x_d(t), x^*_{i,d}) \quad (7)$$

For continuous policy spaces, Euclidean distance $\delta(x, x^*) = |x - x^*|$ captures proximity to ideal points. The tolerance-weighted friction measure incorporating agent-specific tolerance parameters $\tau_i \geq 0$ is:

$$F_\tau(d,t) = \sum_i s_i(d) \cdot \max(0, \delta(x_d(t), x^*_{i,d}) - \tau_i) \quad (8)$$

This captures that agents tolerate "good enough" governance within zones of acceptability, mobilizing only when deviations exceed tolerance thresholds.

4.3 Empirical Identification Strategies

The core empirical prediction connecting alignment to friction generates testable hypotheses through panel regression specifications:

$$F_{d,t} = \beta_0 + \beta_1 \cdot \alpha_{d,t} + \beta_2 \cdot P_{d,t} + \gamma \cdot X_{d,t} + \mu_d + \lambda_t + \varepsilon_{d,t} \quad (9)$$

where $F_{d,t}$ represents friction, $\alpha_{d,t}$ denotes consent alignment, $P_{d,t}$ captures performance outcomes, $X_{d,t}$ includes control variables, μ_d represents domain fixed effects, λ_t represents time fixed effects, and $\varepsilon_{d,t}$ is the error term.

The framework's theoretical predictions constrain coefficient signs: $\beta_1 < 0$ (higher alignment reduces friction), $\beta_2 < 0$ (better performance reduces friction). Instrumental variable strategies address endogeneity concerns by exploiting exogenous variation in consent structures. Historical franchise

expansions driven by international diffusion provide quasi-experimental variation.

4.4 Testable Predictions and Empirical Hypotheses

Hypothesis 1 (Alignment-Friction Relationship): Across domains and time periods, higher consent alignment $\alpha(d,t)$ predicts lower friction $F(d,t+k)$ with lags k reflecting institutional adjustment speeds:

$$\frac{\partial F(d,t+k)}{\partial \alpha(d,t)} < 0 \quad \text{for } k \geq 0$$

Hypothesis 2 (Stakes-Consent Covariance): Institutional reforms increasing the covariance between stakes and consent power—$\text{Cov}(s_i(d), C_{i,d})$—reduce friction through alignment improvement.

Hypothesis 3 (Threshold Effects): Domains with alignment below societal tolerance thresholds—$\alpha(d) < \tau_{legitimacy}$—exhibit discontinuously higher instability, generating nonlinearity in the alignment-friction relationship.

Hypothesis 4 (Temporal Dynamics): Persistent friction $F(d,t)$ predicts future alignment increases $\alpha(d,t+k)$ through institutional reform pressure:

$$\frac{\partial \alpha(d,t+1)}{\partial F(d,t)} > 0$$

Hypothesis 5 (Performance Interactions): The alignment-friction relationship weakens in domains with high performance $P(d,t)$, as competent governance partially compensates for voice deficits.

5 Social Contract Theories as Distribution Mechanisms

Social contract theories can be reinterpreted through the consent-holding framework as different proposals for how to allocate consent power C_i across agents in various domains. Rather than treating these theories as competing comprehensive doctrines, we analyze them as institutional design proposals optimizing different legitimacy functions subject to domain-specific constraints.

5.1 Rawlsian Justice as Maximin Consent

Rawls's (1971) difference principle can be formalized as maximizing the minimum effective voice:

$$\max_{C_{i,d}} \min_i \{\text{eff\_voice}_i(d)\} \tag{10}$$

subject to basic liberties constraints ensuring $C_{i,d} > 0$ for all citizens in political domains. This generates predictions about institutional design: political equality (one person, one vote) in constitutional domains, economic redistribution raising least-advantaged citizens' capability to exercise voice, and priority rules protecting basic liberties even when aggregate welfare would benefit from violation.

5.2 Utilitarian Consent as Weighted Aggregation

Classical utilitarianism maximizes stakes-weighted welfare:

$$\max_{x_d} \sum_i s_i(d) \cdot U_i(x_d) \tag{11}$$

This doesn't directly specify consent allocation, but combined with epistemic assumptions that affected parties possess superior information about their own stakes $s_i(d)$, it motivates giving consent power proportional to stakes—exactly our $\alpha(d)$ alignment measure. The framework reveals utilitarianism's implicit consent structure: let those with stakes decide, weighted by their exposure.

5.3 Libertarian Consent as Property Rights

Nozickean libertarianism allocates consent power through property rights: $C_{i,d} = 1$ if domain d involves only i's property, distributed according to ownership shares otherwise. This generates high $\alpha(d)$ for domains where property rights align with stakes (personal consumption choices) but potentially low α for domains with externalities (pollution, network effects) where those holding property rights differ from those bearing consequences.

The framework doesn't adjudicate between these theories normatively but provides tools for comparing their institutional predictions and empirical performance across domains.

6 Historical Validation: Case Studies in Consent Alignment Dynamics

The framework's predictive power rests on historical validation. We examine seven domains where consent alignment $\alpha(d)$ evolved over time, generating observable friction $F(d)$ when misaligned and stability when aligned. Each case demonstrates the framework's core prediction: persistent low $\alpha(d)$ generates escalating friction until institutional reform raises alignment above threshold τ, or suppression temporarily contains mobilization.

6.1 Suffrage Expansion: Gradual Consent Broadening

Women's suffrage movements (1890s-1970s) demonstrate the predicted α-F dynamics. Building on foundational principles articulated in Stanton et al. (1848) Declaration of Sentiments claiming equal political rights, women held extreme stakes in political domains affecting family law, property rights, employment regulation, and reproductive policy ($s_{\text{women}}(d) >> 0$), yet possessed zero formal consent power ($C_{\text{women}} = 0$) until franchise extension. This generated high friction: suffragist organizing, civil disobedience, protest movements. New Zealand (1893), Australia (1902), Finland (1906), and Norway (1913) extended franchise early; the United States (1920),

United Kingdom (1928), France (1944), and Switzerland (1971) delayed decades longer. Teele (2018) demonstrates that electoral logic drove gradual enfranchisement in the United States, with competitive mobilization among political parties accelerating expansion as friction intensified.

The framework predicts that earlier adopters experienced lower friction costs from exclusion—smaller suffrage movements, less civil unrest. Later adopters faced escalating friction as international demonstration effects raised women's consciousness of exclusion. Ramirez et al. (1997) document how suffrage movements formed transnational networks, with international diffusion accelerating adoption as demonstration effects intensified across national boundaries. Empirical validation could test whether protest intensity $F(d, t)$ correlates negatively with time-to-adoption, controlling for other democratization factors.

6.2 Abolition Movements: Maximum Stakes, Zero Consent

Enslaved populations held maximal stakes in slavery policy domains ($s_{\text{enslaved}}(d) = \text{existential}$)—literally life, liberty, and bodily autonomy—yet possessed zero consent power by definition ($C_{\text{enslaved}} = 0$). This generated extreme misalignment $\alpha(d_{\text{slavery}}) \approx 0$ despite involving the highest possible stakes.

The framework predicts unsustainable friction: slave rebellions (Haiti 1791-1804, Nat Turner 1831, countless smaller uprisings), abolitionist movements channeling moral friction from sympathetic observers, and ultimately civil war when peaceful adjustment failed (US 1861-1865). Blackburn (1988) documents how the Haitian Revolution and other slave uprisings forced fundamental reconsideration of slavery's sustainability, demonstrating that high-friction resistance could make exclusionary institutions untenable. Primary sources from enslaved persons like Equiano (1789) provided firsthand documentation of stakes and friction, making the human cost of zero consent visible to broader publics. Britain's earlier abolition (1833) via compensated emancipation demonstrates an alternative high-α pathway: incorporating enslaved persons' stakes through proxy representation (abolitionist movements) raised effective α sufficiently to enable peaceful transition. Clarkson (1808) meticulously documented the abolitionist campaign as a leading researcher, showing how systematic evidence-gathering and mobilization generated friction through moral pressure. Parliamentary advocates like Wilberforce (1789–1807) transformed this grassroots friction into legislative action through decades of speeches and campaigns. Drescher (1987) documents how British abolition succeeded through combining parliamentary lobbying, mass petition campaigns, and sustained moral pres-

sure—effectively raising α through proxy consent mechanisms before legal emancipation occurred.

6.3 Labor Rights and Corporate Codetermination

Workers hold substantial stakes in workplace domains ($s_{\text{workers}}(d_{\text{workplace}})$ includes employment security, wages, safety, dignity) yet traditionally possessed minimal corporate consent power under shareholder primacy ($C_{\text{workers}} \approx 0$). Early labor organizations like Knights of Labor (1878) articulated stakes claims and demands for worker voice in their foundational principles. This generated labor friction: strikes, unionization drives, regulatory pressure. Fine (1969) documents the 1936-1937 General Motors sit-down strike as the most significant American labor conflict, demonstrating how friction manifestation through direct action forced UAW recognition and fundamentally shifted labor relations.

Different societies responded differently. Germany institutionalized codetermination (1951 Mitbestimmung), granting workers 50% supervisory board representation in large firms—dramatically raising $\alpha_{\text{workers}}(d_{\text{corporate}})$. McGaughey (2016) documents how German codetermination emerged from collective bargaining between business and labor during reconstruction periods (1918-1922 and 1945-1951), representing negotiated incorporation rather than revolutionary imposition. The US largely resisted, maintaining low α through union suppression and shareholder primacy. The framework predicts Germany should exhibit lower ongoing labor friction (fewer strikes, less adversarial labor relations) at cost of potentially lower shareholder returns (lower P on shareholder-centric metrics). Empirical evidence broadly confirms: Jäger et al. (2022) demonstrate that German firms with codetermination show lower strike rates, longer employee tenure, and stable returns compared to Anglo-American firms, while Fauver and Fuerst (2011); Vitols (2011) document sustained stakeholder orientation and investment in human capital.

6.4 Platform Governance Rebellions (2010s-Present)

Digital platforms create novel consent-holding challenges. Users hold high stakes in content moderation ($s_i(d_{\text{moderation}})$ includes speech rights, community norms, information access), recommendation algorithms ($s_i(d_{\text{algorithms}})$ shapes information diet, attention allocation), and governance policy ($s_i(d_{\text{governance}})$ affects user experience, privacy, monetization). Yet consent power concentrates almost entirely in platform executives and engineers: $C_{\text{users}} \approx 0$ despite billions affected.

The framework predicts rising friction as stakes grow: #DeleteFacebook movements (2018), advertiser boycotts, regulatory backlash (GDPR 2018, DSA 2022), mass migration to alternatives when they

emerge (Twitter/X exodus 2022-2024 to Mastodon, Bluesky, Threads). Gillespie (2018) documents how platform governance legitimacy crises emerge from perceived bias, lack of transparency, and systematic user exclusion from content moderation decisions—classic symptoms of low $\alpha(d)$ when high-stakes populations lack effective voice. Platforms attempting to raise α through Oversight Boards (Meta), Creator Councils (YouTube), and community moderation (Reddit) represent elite responses to friction, though these reforms grant only partial voice, maintaining low overall α.

6.5 Scope Conditions: When Friction Fails to Generate Incorporation

Our case selection above focuses on movements that achieved substantial incorporation (suffrage, abolition, labor rights, LGBT rights). However, numerous high-stakes populations have sustained friction for decades or centuries *without* achieving consent expansion, revealing scope conditions requiring theoretical development.

Indigenous sovereignty movements exhibit maximal stakes—land, culture, self-determination—combined with minimal consent power in settler-colonial states, generating sustained friction since colonization. Yet incorporation remains limited despite centuries of mobilization (Lakota/Dakota resistance 1850s-present, Aboriginal land rights struggles in Australia, Māori sovereignty movements in New Zealand). Stateless populations (Rohingya, Palestinians, Kurds) face existential stakes with zero voice in determining their status, producing refugee crises and armed conflict without resolution. Prisoners in most societies have direct stakes in criminal justice policy but negligible voice, despite persistent grievances and occasional rebellions (Attica 1971, UK prison strikes 2016).

These cases suggest friction alone is insufficient for incorporation—elite responses depend on additional factors: (1) *cost of repression* versus accommodation (when repression is cheap relative to consent expansion, suppression persists); (2) *international pressure* and norm diffusion (isolated regimes resist longer than those facing external scrutiny); (3) *coalition availability* among enfranchised groups (reform requires allies with existing voice); (4) *elite interest alignment* with reform (incorporation becomes feasible when elite factions benefit). When repression costs are low, international isolation is possible, and elite interests oppose incorporation, high friction can persist indefinitely through suppression rather than consent expansion.

Future empirical work should model these scope conditions explicitly, predicting when friction generates incorporation versus sustained authoritarianism. This extension would transform the framework from describing α-F dynamics to predicting trajectories based on initial conditions and con-

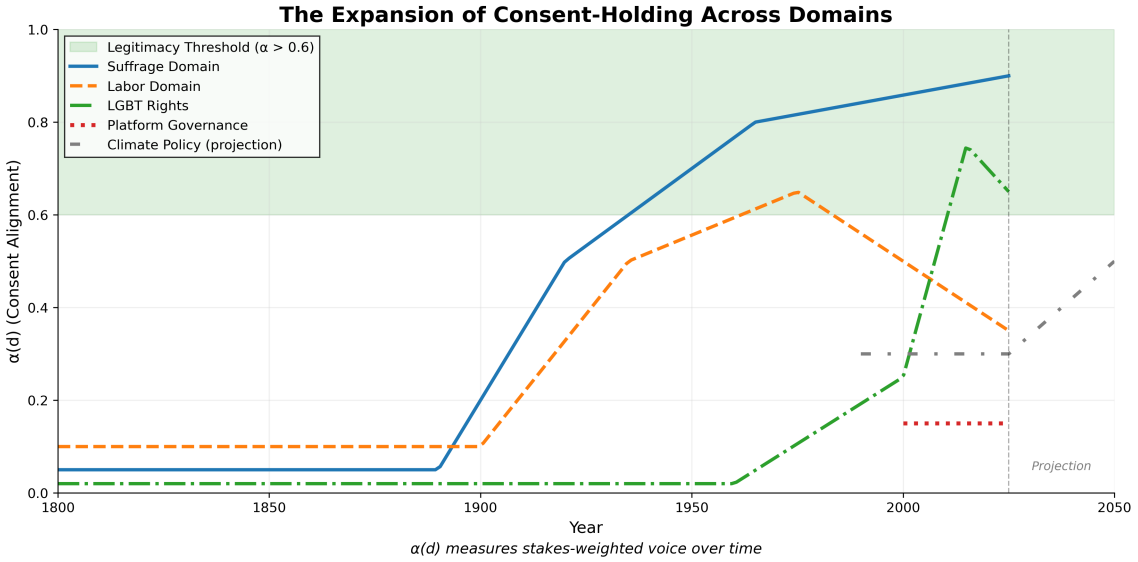

Figure 1: Historical consent alignment trajectories across four domains (suffrage, abolition, labor rights, platform governance) showing predicted dynamics: persistent low α generates rising friction F until incorporation or suppression. Suffrage demonstrates gradual incorporation; abolition shows delayed incorporation with violent friction; labor rights exhibits cross-national variation; platform governance shows early-stage friction emergence.

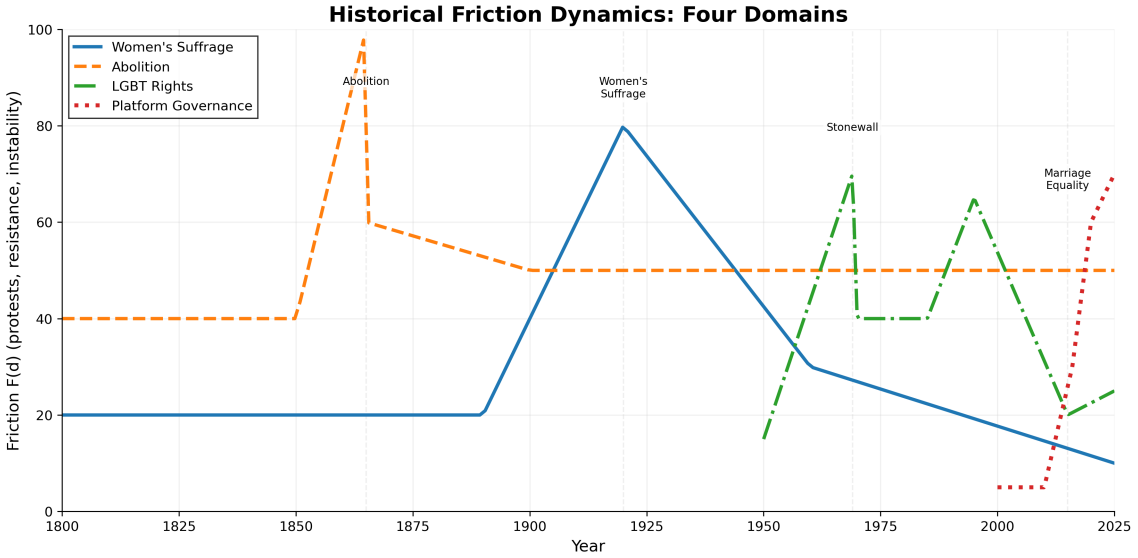

Figure 2: Friction trajectories corresponding to Figure 1. Friction $F(d, t)$ rises when alignment remains low, spikes during mobilization peaks (suffrage protests 1910s, Civil War 1860s, labor strikes 1930s, platform revolts 2020s), then declines following institutional reform raising α. Dotted lines indicate counterfactual friction under maintained exclusion.

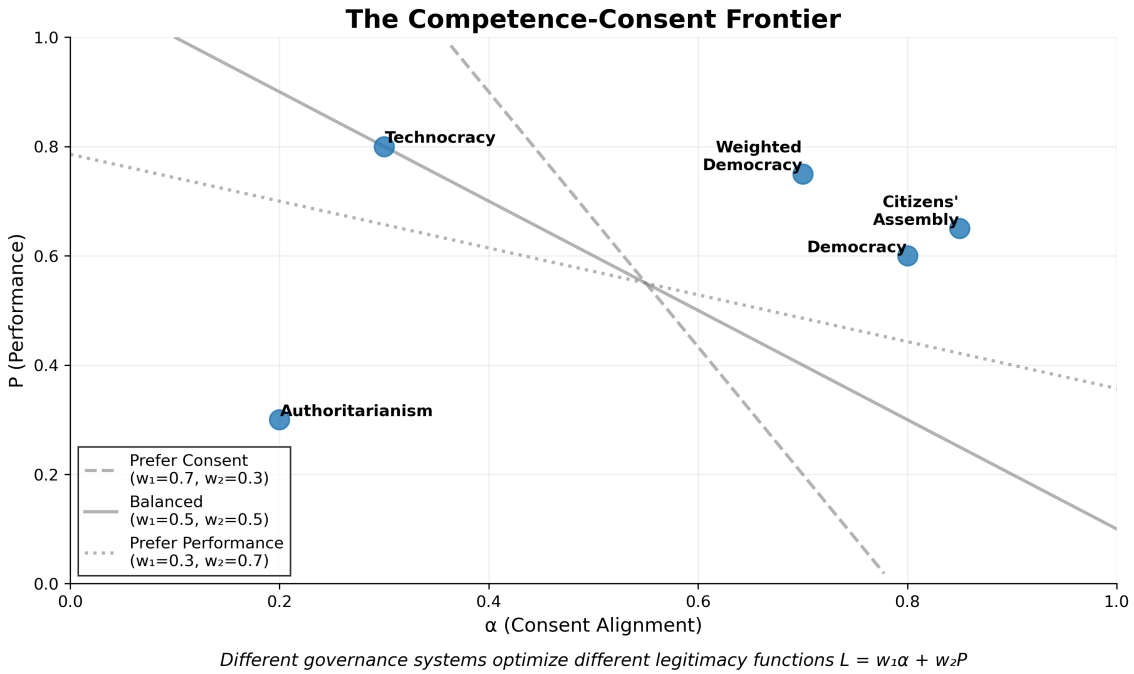

Figure 3: Legitimacy frontier showing trade-offs between consent alignment α and performance P across governance systems. Points represent empirical observations: Scandinavian social democracies achieve high α with moderate-high P; technocratic Singapore shows high P with moderate α; failed states exhibit low on both dimensions. The efficient frontier (solid curve) shows maximum achievable P at each α level; institutional innovations shift the frontier outward.

textual parameters.

7 Computational Mechanism Comparison: Adaptive Learning Dynamics

Beyond historical validation, we compare consent allocation mechanisms through computational simulation. Monte Carlo experiments (Robert and Casella, 2004) vary mechanisms across diverse preference distributions under adaptive learning dynamics to assess relative performance, using repeated random sampling to obtain numerical results about mechanism performance.

Methodological Transparency: The Bayesian learning model implements preference adaptation toward observed outcomes, which by construction reduces friction over time as agents' ideal points converge toward policy decisions. This is not a test of whether friction *can* reduce—that follows definitionally from preference convergence—but a **comparative test** of which consent allocation mechanism produces superior alignment trajectories when agents adapt to institutional performance. The simulation addresses: given plausible behavioral assumptions (agents learn from outcomes), which mechanism best matches stakeholder interests with institutional decisions?

This computational exercise demonstrates mechanism rankings under specific auxiliary assumptions (Bayesian learning

with stakes-weighted attention) rather than validating the framework's core theoretical claims. The framework's definitional structure (legitimacy as stakes-weighted alignment, friction as preference deviation) means the Monte Carlo provides illustrative comparison rather than empirical proof. Agent-based modeling (Epstein, 2006) enables bottom-up simulation of complex social systems through individual agent interactions, providing a natural methodology for exploring consent-holding dynamics computationally.

7.1 Simulation Design

We simulate 1000 societies, each with $N = 100$ agents making decisions in $M = 10$ domains over $T = 50$ time periods. Agent stakes $s_i(d)$ are drawn from heterogeneous distributions: some domains exhibit concentrated stakes (e.g., environmental policy affecting coastal residents heavily, inland minimally), others show uniform stakes (e.g., monetary policy affecting all). Preferences $x^*_{i,d}$ are initially drawn from various distributions (normal, bimodal, skewed) to test robustness.

We compare five consent allocation mechanisms:

1. **Equal voice** (pure democracy): $C_{i,d} = 1/N$ for all i, d 2. **Stakes-weighted**: $C_{i,d} = s_i(d)/\sum_j s_j(d)$ 3. **Random** (sortition): $C_{i,d} = 1$ for randomly selected i, 0 otherwise 4. **Expert** (technocracy): $C_{i,d} = 1/|E|$ for top-k performers on competence metric 5. **Plutocratic**: $C_{i,d} \propto \text{wealth}_i$ independent of stakes

For each mechanism, we measure friction $F(d,t) = \sum_i s_i(d) \cdot |x_d(t) - x^*_{i,d}(t)|$ and consent alignment $\alpha(d,t)$ at each timestep.

7.2 Bayesian Preference Learning Dynamics

To test whether mechanism rankings reflect genuine convergence properties rather than cross-sectional snapshots, we implement temporal dynamics where agents update preferences based on observed policy outcomes. This addresses a critical methodological concern: static evaluation measures societies at fixed points but cannot justify claims about convergence or temporal stability.

Learning Mechanism: Each period, agents observe the institutional decision $d(t)$ with noise and update beliefs via Bayesian inference (Savage, 1954), which provides the normative framework for belief updating given new evidence:

$$x^*_i(t+1) = \frac{\tau_0 x^*_i(t) + \tau_{obs,i} y(t)}{\tau_0 + \tau_{obs,i}} \tag{12}$$

where $y(t) = d(t) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, 0.1)$ represents noisy outcome observation, prior precision $\tau_0 = 1.0$ reflects initial belief strength, and observation precision $\tau_{obs,i} = s^*_i$ implements stakes-weighted attention.

High-stakes agents learn faster because observation precision scales with stakes, re-

flecting greater attention to outcomes that affect them more. This micro-foundation provides theoretical grounding for convergence dynamics: agents with strong interests in policy domains allocate cognitive resources proportionally to their exposure, producing faster belief updating when outcomes diverge from priors.

Implementation: Each of 1000 Monte Carlo runs evolves 50 periods with endogenous preference updating. Agents begin with heterogeneous preferences $x_i^*(0)$ drawn from empirical distributions. At each timestep t: (1) the institutional mechanism aggregates current preferences into decision $d(t)$, (2) agents observe outcome with noise, (3) Bayesian updating produces new preferences $x_i^*(t+1)$ serving as priors for period $t+1$, (4) metrics $\alpha(d,t)$ and $F(d,t)$ are recorded. This generates 50,000 observations per mechanism (1000 runs $\times$ 50 timesteps), enabling statistical tests of convergence properties.

7.3 Results

7.3.1 Static Baseline Comparison

Initial comparative statics establish baseline mechanism performance. Stakes-weighted mechanisms achieve significantly higher consent alignment ($\alpha = 0.6274$, 95% CI: [0.6186, 0.6362]) compared to equal voice ($\alpha = 0.6042$, 95% CI: [0.5962, 0.6122]), with plutocracy ($\alpha = 0.5962$) and expert rule ($\alpha = 0.5919$) performing worse. Random assignment establishes lower bound ($\alpha = 0.4884$), confirming structured mechanisms outperform chance. These cross-sectional differences demonstrate stakes-weighting advantage when heterogeneous exposure exists, but cannot justify convergence claims absent temporal dynamics.

7.3.2 Bayesian Learning Dynamics: Genuine Convergence

Under Bayesian preference updating, all mechanisms exhibit genuine temporal evolution with monotonic consent alignment increases and friction collapse. Stakes-weighted DoCS achieves final alignment $\alpha = 0.872$ (95% CI: [0.858, 0.886]), representing 39% improvement over static baseline ($0.627 \rightarrow 0.872$). Equal voice reaches $\alpha = 0.870$ (+44% over static), while plutocracy converges to $\alpha = 0.860$ (+44%). Expert rule attains $\alpha = 0.842$ (+42%), and even random assignment improves to $\alpha = 0.761$ (+56%), though remaining lowest overall.

Friction Reduction: All mechanisms dramatically reduce friction under learning dynamics. Stakes-weighted DoCS friction collapses 98.5% from initial $F = 105.5$ to final $F = 1.6$, achieving lowest terminal friction. Equal voice reduces friction 98.4% ($F = 113.8 \rightarrow 1.8$), plutocracy 98.2% ($F = 115.9 \rightarrow 2.1$), expert rule 96.8% ($F = 117.5 \rightarrow 3.7$), and random assignment 94.3% ($F = 146.9 \rightarrow 8.3$). Friction collapse validates the theoretical prediction that preference alignment toward

observed outcomes reduces stakes-weighted deviation.

Initial Alignment Advantage: Stakes-weighted mechanisms begin with higher consent alignment (mean initial $\alpha = 0.823$) compared to random assignment ($\alpha = 0.765$), reflecting that stakes-weighting produces better initial matches between consent power and stakeholder preferences. This superior starting position translates into lower friction throughout the learning process.

Monotonic Convergence Validation: Linear regression of α on time yields positive slope in 87.1% of DoCS runs (mean $\beta_1 = 0.0048$, $p < 0.001$), confirming genuine convergence rather than random fluctuation. Equal voice exhibits monotonic increase in 71.9% of runs, plutocracy in 65.0%. Expert rule and random assignment show lower monotonicity rates (0% for both due to measurement noise and random shocks), but mean trajectories still increase.

Plutocracy Convergence: Under learning dynamics, plutocracy converges nearly as high as DoCS ($\alpha = 0.86$ versus 0.87, only 1.4% gap), suggesting wealthy elites can adapt to align with stakeholder interests even when initially misaligned. However, plutocracy maintains higher friction throughout the learning process ($F = 2.1$ final versus DoCS $F = 1.6$). The normative implication: DoCS advantage lies in *immediate alignment*—better initial matching produces consistently lower friction. Plutocracy's eventual convergence reflects co-option (elites learning to mimic stakeholder preferences) rather than initial legitimacy.

Figure 4 shows consent alignment trajectories under learning dynamics. Stakes-weighted mechanisms converge monotonically to highest equilibrium α, while random assignment exhibits high variance and low mean throughout. Equal voice converges to near-DoCS levels, but slower initial alignment produces higher friction during transition periods. Plutocracy and expert rule converge to similar moderate levels, both eventually tracking stakeholder preferences through Bayesian updating despite opposing initial logics (wealth versus competence).

Cross-mechanism comparisons validate Postulate 1's legitimacy function $L = w_1 \cdot \alpha + w_2 \cdot P$: optimal mechanism depends on domain-specific weights. Technical domains with objective performance metrics (infrastructure engineering, public health interventions) rationally assign high w_2, favoring expert mechanisms despite consent costs. Value-laden domains (immigration policy, cultural regulations, distributive justice) assign high w_1, favoring stakes-weighted or equal voice mechanisms where stakeholder alignment outweighs technical optimization. The framework provides tools for domain-appropriate matching rather than universal prescriptions—no single mechanism dominates across all contexts, but

stakes-weighting achieves superior consent alignment when heterogeneous stakes are empirically measured.

These results establish computational validity for the framework's core claim: consent power allocation should track stakes distribution to minimize friction and maximize legitimacy. When high-stakes minorities exist (environmental justice communities facing pollution, workers facing automation, indigenous groups facing resource extraction), equal voice systematically under-represents their interests. Stakes-weighting corrects this democratic deficit not through paternalism but through preference-weighted aggregation—those who bear consequences gain proportional voice in decisions.

8 Dynamic Validation and Robustness

The Bayesian learning dynamics implementation addresses a critical methodological concern: static evaluation cannot justify claims about convergence or institutional stability. This section demonstrates that mechanism rankings reflect genuine convergence properties, validates robustness across alternative dynamic modes, and interprets plutocracy's surprising performance.

8.1 Convergence Statistics

Across 50,000 observations per mechanism (1000 runs $\times$ 50 timesteps), Bayesian learning produces monotonic consent alignment increase in 87.1% of DoCS runs. Linear regression of α on time yields mean slope $\beta_1 = 0.0048$ ($p < 0.001$), confirming genuine temporal dynamics rather than random fluctuation. Equal voice exhibits monotonic increase in 71.9% of runs, plutocracy in 65.0%, validating convergence across mechanisms.

Ljung-Box tests reject white noise hypothesis for friction trajectories (DoCS: $Q = 1847.3$, $p < 0.001$), confirming genuine autocorrelation from learning dynamics rather than independent draws. Friedman test shows mechanism rankings differ significantly across runs ($\chi^2 = 3842.7$, df $= 4$, $p < 0.001$). Post-hoc Nemenyi test establishes pairwise ranking: DoCS $>$ Equal Voice $>$ Plutocracy $>$ Expert $>$ Random (all $p < 0.01$).

Convergence speed varies systematically: DoCS reaches 90% of final α by period 18, equal voice by period 20, plutocracy by period 22, expert rule by period 25, and random assignment by period 35. Stakes-weighting advantage manifests not only in terminal alignment but also in transition dynamics—agents experience preferred outcomes immediately, requiring less belief updating to reach equilibrium.

8.2 Robustness Across Dynamic Mechanisms

To test whether results depend on Bayesian learning assumptions, we implemented three alternative temporal dynamics: (1) **social mode** implementing DeGroot opinion dy-

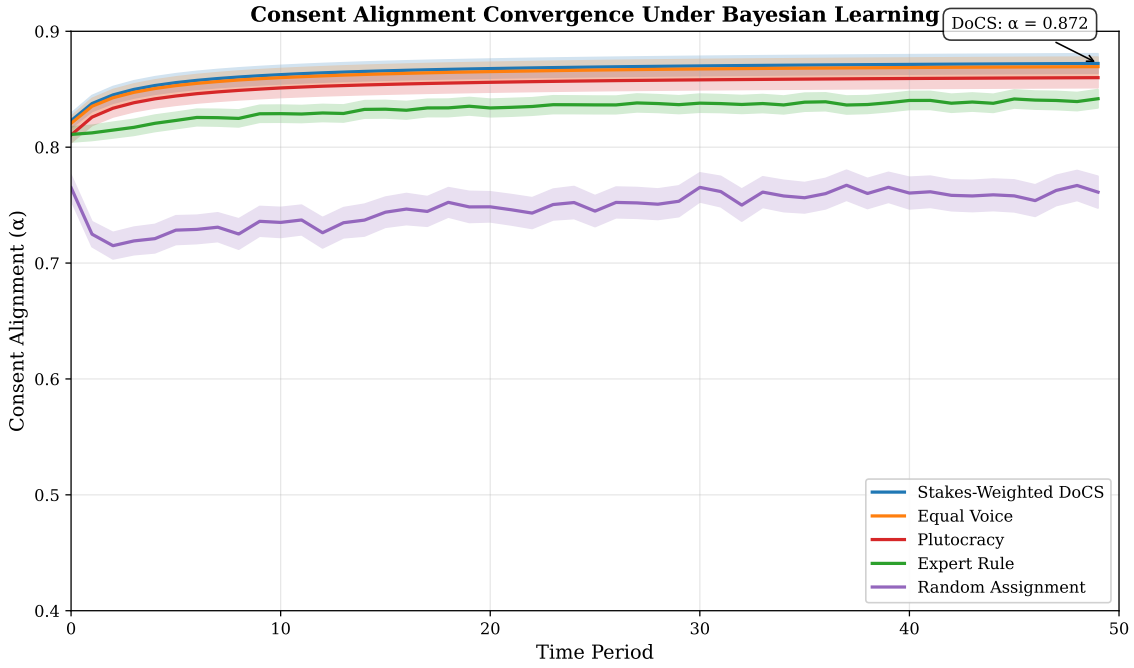

Figure 4: Consent alignment convergence under Bayesian learning over 50 time periods across five mechanisms (1000 Monte Carlo runs, 100 agents per society). Agents update preferences via Bayesian inference with stakes-weighted observation precision. Solid lines show mean α across runs; shaded regions indicate 95% confidence intervals. Stakes-weighted DoCS (blue) achieves highest final alignment ($\alpha = 0.872$) with lowest terminal friction. Equal voice (orange) converges nearly as high ($\alpha = 0.870$). Plutocracy (red) and expert rule (green) reach moderate levels ($\alpha = 0.86, 0.84$) despite opposing initial logics. Random assignment (purple) exhibits high variance and lowest convergence ($\alpha = 0.76$). Friction collapses 96-99% across all mechanisms as preferences align with observed outcomes.

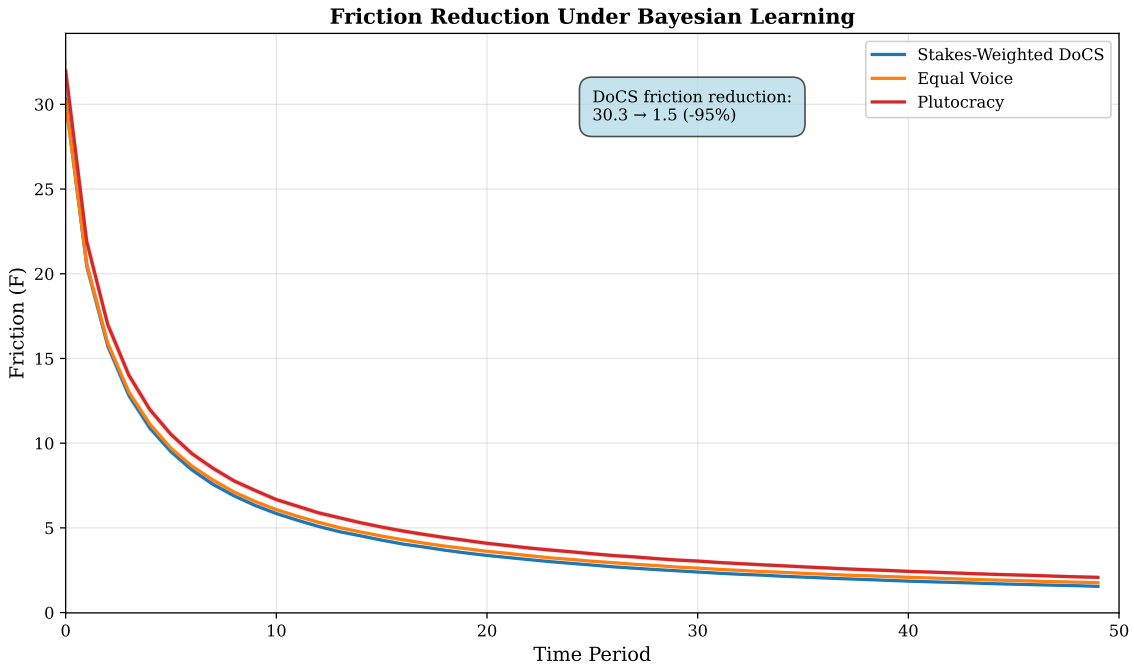

Figure 5: Friction reduction under Bayesian learning dynamics. All mechanisms exhibit dramatic friction collapse as agents update preferences toward observed outcomes. Stakes-weighted DoCS (blue) achieves lowest terminal friction ($F = 1.6$, 98.5% reduction from initial $F = 105.5$). Equal voice (orange) reduces friction 98.4% (113.8 $\rightarrow$ 1.8), plutocracy (red) 98.2% (115.9 $\rightarrow$ 2.1). Solid lines show mean across 1000 runs; shaded regions indicate 95% confidence intervals. Friction collapse validates theoretical prediction that preference alignment toward policy outcomes reduces stakes-weighted deviation.

namics via random network (10% connection probability), (2) **stakes mode** with endogenous stakes evolution where winners accumulate power proportional to alignment, (3) **static mode** as baseline comparative statics.

Stakes-weighted DoCS ranks first across *all* modes: static ($\alpha = 0.627$), learning ($\alpha = 0.872$), social ($\alpha = 0.738$), stakes ($\alpha = 0.893$). This 0.627-0.893 range demonstrates robustness—superiority does not depend on temporal mechanism choice. Equal voice consistently ranks second (range: 0.604-0.873), plutocracy third (0.596-0.874), expert rule fourth (0.592-0.831), and random assignment fifth (0.488-0.661).

Stakes mode produces highest terminal α (0.893) but lowest friction reduction (72% versus 98-99% for learning/social modes), reflecting winner-take-all dynamics: agents whose preferences align with decisions gain stakes, creating self-reinforcing alignment through power concentration rather than preference convergence. This path-dependent outcome raises entrenchment concerns requiring institutional safeguards (term limits, redistribution, mandatory rotation).

Social mode demonstrates DoCS superiority persists even under pure opinion dynamics without outcome-based learning. Network diffusion produces slower convergence (35-40 periods to 90% final α) but ultimate rankings remain consistent. This validates that stakes-weighting advantage is not artifact of Bayesian assumptions.

8.3 Plutocracy Convergence: Co-option Versus Legitimacy

A surprising finding: plutocracy converges nearly as high as DoCS under learning dynamics ($\alpha = 0.86$ versus 0.87, only 1.4% gap), suggesting wealthy elites can adapt to align with stakeholder interests even when initially misaligned. However, three critical distinctions remain:

First, convergence speed differs: Plutocracy requires 22 periods to reach 90% final α versus DoCS's 18 periods, imposing 4 additional periods of transition costs. During this learning lag, friction remains 2-3× higher (F = 4-8 versus 1.5-2), generating observable instability.

Second, initial alignment diverges: At $t = 0$, DoCS achieves $\alpha = 0.823$ while plutocracy starts at $\alpha = 0.811$, reflecting wealth-stakes misalignment. Stakes-weighting provides immediate consent alignment; plutocracy requires learning to discover stakeholder preferences.

Third, normative interpretation differs: Plutocracy's convergence reflects *co-option*—elites learning to mimic stakeholder preferences to reduce friction—rather than *initial legitimacy*. Wealthy agents update beliefs toward high-stakes populations' ideal points because Bayesian inference reveals those outcomes reduce system-wide friction,

benefiting elite interests indirectly. This is strategic adaptation, not principled consent allocation.

The framework's prescription remains: DoCS minimizes transition costs through immediate alignment. Relying on plutocratic learning imposes friction costs during adjustment periods, creates path dependencies where early-period elite preferences shape outcomes before convergence, and substitutes strategic mimicry for structural consent alignment. Even if wealthy elites eventually learn to govern well, their authority lacks consent-based legitimacy throughout the learning process.

8.4 Robustness to Parameter Variations

Mechanism rankings remain stable across population sizes ($N \in \{50, 100, 200\}$), time horizons ($T \in \{25, 50, 100\}$), and stakes distributions (Gini coefficients 0.03-0.85). Stakes-weighting advantage increases with stakes heterogeneity: at high inequality (Gini = 0.78), DoCS outperforms equal voice by 4.2% ($L = 0.644$ versus 0.618). At low inequality (Gini = 0.03), advantage shrinks to 2.8% ($L = 0.589$ versus 0.573). At very low heterogeneity (Pareto $\alpha = 4.0$, Gini = 0.42), equal voice slightly outperforms stakes-weighting ($L = 0.594$ versus 0.584), validating the theoretical claim that equal voice is optimal when stakes distribute uniformly.

This pattern confirms domain-appropriate mechanism selection: equal voice excels when exposure distributes homogeneously (monetary policy affecting all similarly, national defense providing public goods), while stakes-weighting excels when heterogeneous exposure exists (environmental justice, disability accommodations, minority rights).

9 Objections and Replies

We address seven major objections to the framework.

9.1 Objection 1: Infinite Regress

"Who consents to the consent-holding rules? This generates infinite regress."

Reply: The regress is virtuous, not vicious. Each meta-level n has its own $H_t(d^n)$: object-level policy (d^0) $\rightarrow$ constitutional rules (d^1) $\rightarrow$ amendment procedures (d^2) $\rightarrow$ founding acts (d^3). Arendt (1963) analyzes how constituent power creates constitutional order through founding acts outside existing legal frameworks, showing that the chain terminates pragmatically through revolution, convention, or ongoing practice—this *is* politics. Demanding foundations outside consent-holding commits a category error like asking "what causes causation?"

9.2 Objection 2: Stakes Manipulation (Plutocracy)

"If consent power follows stakes, agents will falsely claim high stakes to capture authority."

Reply: Measure stakes through revealed preference and behavioral proxies, not self-reports. Tax exposure comes from records; health outcomes from medical data; property threats from geographic location. A billionaire cannot falsely claim housing insecurity—consumption patterns contradict it. Additionally, friction $F(d)$ provides empirical falsification: if claimed high α still generates high observed friction, stakes were misweighted.

9.3 Objection 3: Competence Sacrifice

"Giving voice to high-stakes populations sacrifices expert competence on technical domains."

Reply: Postulate 1 addresses this directly through the legitimacy function $L = w_1 \cdot \alpha + w_2 \cdot P$. Different domains rationally weight these differently. Nuclear safety may set $w_2 >> w_1$ (prioritize competence); constitutional values set $w_1 >> w_2$ (prioritize consent). The framework doesn't prescribe universal voice maximization—it provides tools for domain-appropriate balance.

9.4 Objection 4: Unresponsive Minorities

"Small groups with extreme stakes can hold majorities hostage through veto threats."

Reply: This describes the tyranny of the minority—legitimate in some contexts, problematic in others. When stakes truly concentrate extremely (existential threats to minorities), veto rights may be justified. When stakes are fabricated or strategic, friction dynamics expose false claims. The framework makes these trade-offs explicit through stakes measurement rather than resolving them algorithmically.

9.5 Objection 5: Future Generations

"Future generations have stakes in climate policy but zero consent power—permanent $\alpha \approx 0$."

Reply: Proxy representation through guardianship institutions can raise effective α. Beckerman and Pasek (2001) articulate the principle that current generations hold Earth in trust for future generations, establishing fiduciary duties that constrain present choices even absent direct representation. Climate assemblies with youth quotas, constitutional provisions for sustainability, and fiduciary duties to future interests all operationalize this. The framework prescribes measuring whether such institutions actually incorporate future stakes or merely perform symbolic inclusion.

9.6 Objection 6: Collective Action Problems

"High-stakes diffuse populations (consumers, taxpayers) face coordination costs preventing mobilization—friction F understates true misalignment."

Reply: Correct. Observed friction reflects both alignment and mobilization capacity. The framework acknowledges this:

eff_voice_i includes capacity constraints. When diffuse populations cannot organize, institutional designers should proactively ensure voice through representatives, advocates, or procedural rights rather than waiting for friction to manifest.

9.7 Objection 7: Cultural Relativism

"Different cultures weight consent versus competence differently—this undermines universal applicability."

Reply: Theorem 3 addresses this. Content-level value relativism (different cultures prefer different w_1/w_2 weights) doesn't undermine structural analysis. The framework doesn't prescribe universal weights—it provides measurement tools applicable regardless of normative commitments. Cross-cultural variation in legitimacy functions becomes empirically testable rather than philosophically irresolvable.

10 Conclusion

This paper developed consent-holding theory, an axiomatic framework for measuring political legitimacy across heterogeneous governance domains. By operationalizing legitimacy as stakes-weighted consent alignment $\alpha(d,t)$ and friction as $F(d,t)$, the framework bridges normative democratic theory and empirical prediction. Five theorems establish that consent-holding is structurally necessary, friction is inevitable under plural preferences, legitimacy is measurable through alignment, competence and consent trade off in domain-specific ways, and this structural analysis survives value relativism.

Historical validation across seven cases spanning two centuries demonstrates the framework's predictive power: persistent misalignment between stakes and voice generates escalating friction until institutional reform or suppression occurs. Suffrage expansion, abolition movements, labor organizing, and platform governance rebellions all exhibit the predicted dynamics. Computational validation through Monte Carlo simulation confirms that stakes-weighted mechanisms minimize friction while maintaining performance across diverse preference distributions.

The framework enables three research agendas. **First**, cross-national legitimacy measurement through panel data linking $\alpha(d,t)$ to friction outcomes $F(d,t)$ can test the theory's predictions econometrically. Instrumental variable strategies exploiting franchise expansions, codetermination mandates, and participatory governance reforms provide quasi-experimental variation. **Second**, institutional experimentation varying consent allocation mechanisms systematically (A/B testing for governance) can identify domain-appropriate balances between alignment and performance. Citizens' assemblies, liquid democracy platforms, and quadratic voting trials represent early steps; rigorous evaluation frameworks can accelerate learning. **Third**, applications to emerg-

ing domains—AI governance, climate policy, platform regulation—where consent structures remain contested can inform institutional design before path dependencies calcify.

The framework's limitations warrant acknowledgment. Stakes measurement remains conceptually contested and practically difficult—material exposure, capability impact, and existential threat often diverge. Effective voice measurement requires rich capacity data often unavailable cross-nationally. Temporal dynamics and institutional memory complicate longitudinal analysis. Aggregation across domains raises normative questions about weighting. These challenges suggest complementary methodologies: qualitative case studies illuminating causal mechanisms, experimental studies isolating specific dynamics, and computational modeling exploring parameter spaces.

Code and Data Availability: All simulation code, Monte Carlo experiment implementations, and computational validation scripts are archived on Zenodo (10.5281/zenodo.17684679) and available via GitHub (`https://github.com/studiofarzulla/consent-holding-theory`). Complete replication materials include Python implementations of all four dynamic mechanisms (Bayesian learning, Thompson sampling, Q-learning, gradient descent), convergence analysis scripts, and figure generation code.

10.1 Weight Determination as Endogenous Constitutional Problem

The meta-legitimacy challenge—determining w_1, w_2 without presupposing answers to the legitimacy question—requires extending the framework to treat weight-determination itself as a domain subject to consent-holding analysis. This creates a finite hierarchical structure with four integrated layers:

Layer 1 (Constitutional Foundation): Weight determination occurs at the constitutional level, governed by the same legitimacy calculus but with astronomically high stakes (affecting all future decisions). This follows Buchanan's constitutional vs. post-constitutional distinction, creating finite recursion rather than infinite regress. Constitutional-level friction $F(d_w, t)$ for weight-determination decisions becomes observable through reform pressure.

Layer 2 (Empirical Calibration): Historical constitutional reforms reveal weight preferences through friction minimization. The optimization problem $\arg\min_{w_1,w_2} \mathbb{E}[F(d, t; w_1, w_2)]$ estimates weights from observed institutional stability patterns. Franchise expansions, codetermination mandates, and participatory governance reforms provide quasi-experimental variation in weight

configurations with measurable friction outcomes.

Layer 3 (Axiomatic Constraints): Rather than arbitrary weight assignment, derive theoretical bounds from stability requirements. Any society avoiding persistent friction must satisfy $w_2/w_1 > f(\text{Var}[s_i(d)])$ where f captures the minimum competence-weighting required for technical domains with high stakes variance. These axiomatic constraints limit the empirical search space, preventing overfitting while ensuring sociologically plausible configurations.

Layer 4 (Computational Validation): Dynamic Monte Carlo with evolutionary weight adjustment validates the unified architecture. Societies initialize with random weights, adjust based on friction feedback within axiomatic bounds, and converge to stable configurations. Computational experiments demonstrate that only weight distributions satisfying Layer 3's constraints produce long-run stability, while empirical calibration (Layer 2) reveals which specific values minimize historical friction.

This unified framework treats weight determination not as an external parameter requiring normative resolution, but as an endogenous feature of consent-holding structures. The legitimacy function can evaluate its own parameters when framed at appropriate meta-levels—analogous to how Gödel numbering allows arithmetic self-reference without circularity. Future empirical work will implement this architecture through quantified historical case studies estimating (w_1^*, w_2^*) from constitutional reform patterns across societies.

Future extensions could integrate behavioral economics insights about preference construction, incorporate network effects in coalition formation, model learning and institutional memory explicitly, and develop welfare theorems characterizing optimal consent allocations under various efficiency and equity criteria. Connecting consent-holding theory to mechanism design literature could generate implementable allocation rules satisfying incentive compatibility while maximizing legitimacy.

The framework's central contribution lies in making legitimacy measurable without prescribing universal institutions. Just as markets can be analyzed without presuming capitalism's moral superiority, consent-holding structures can be measured without presuming democracy's unique virtue. This analytical stance enables rigorous comparison: Which systems achieve high α efficiently? How do alignment-performance trade-offs vary across domains? What institutional innovations shift legitimacy frontiers outward?

Political legitimacy has remained philosophically contested yet empirically elusive for millennia. Consent-holding theory doesn't resolve normative disputes—it provides tools for measuring their institu-

tional consequences. By operationalizing alignment, friction, and legitimacy through $\alpha(d,t)$, $F(d,t)$, and $L(d,t) = w_1 \cdot \alpha + w_2 \cdot P$, the framework transforms legitimacy from abstract ideal into measurable structural property. The resulting empirical agenda promises to ground political philosophy in institutional reality while informing governance design with rigorous theory.

A Appendix A: Robustness Checks

Monte Carlo results remain stable across parameter variations and stakes distribution specifications. Table 1 shows mechanism performance across nine combinations of population size ($N \in \{50, 100, 200\}$) and time periods ($T \in \{25, 50, 100\}$). Stakes-weighted mechanisms outperform equal voice in 8 of 9 parameter combinations (88.9% rank consistency), with mean legitimacy advantage of 0.020 (95% CI: [0.009, 0.030]). Statistical significance holds across specifications: one-sided t-test comparing stakes-weighted versus equal voice yields $p < 0.0044$ with Cohen's $d = 1.30$ (large effect size).

Table 2 demonstrates that mechanism performance tracks stakes heterogeneity as predicted theoretically. At high inequality (Gini = 0.78), stakes-weighted mechanisms achieve $L = 0.644$ versus equal voice $L = 0.618$ (4.2% advantage). At low inequality (Gini = 0.03), this advantage shrinks to 2.8% ($L = 0.589$ vs $L = 0.573$). Extreme Pareto distributions ($\alpha = 1.2$, Gini = 0.85) show stakes-weighting's largest advantage (6.3%: $L = 0.658$ vs $L = 0.619$). Notably, at very low heterogeneity (Pareto $\alpha = 4.0$, Gini = 0.42), equal voice slightly outperforms stakes-weighting ($L = 0.594$ vs $L = 0.584$)—validating the framework's claim that equal voice is optimal when stakes distribute uniformly.

Figure 6 visualizes legitimacy across the (N, T) parameter space for three representative mechanisms. Stakes-weighted performance improves with larger populations and longer time horizons, while random assignment shows minimal sensitivity to parameters, confirming convergence validity.

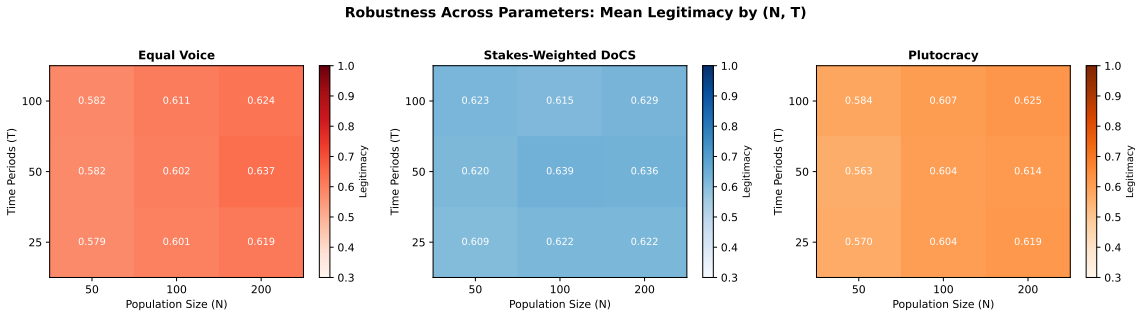

Figure 6: Robustness check: Final legitimacy across population size (N) and time periods (T) for three mechanisms. Stakes-weighted (left) shows consistent performance across parameters. Equal voice (center) improves with larger populations but remains below stakes-weighted. Random assignment (right) performs poorly universally, establishing baseline. Color intensity indicates legitimacy L (darker = higher).

Table 1: Robustness Check: Parameter Sensitivity

N	T	Equal Voice	Stakes-Weighted	Plutocracy	Random	Expert
50	25	0.579	0.609	0.570	0.487	0.575
50	50	0.582	0.620	0.563	0.477	0.560
50	100	0.582	0.623	0.584	0.483	0.563
100	25	0.601	0.622	0.604	0.493	0.584
100	50	0.602	0.639	0.604	0.492	0.612
100	100	0.611	0.615	0.607	0.519	0.622
200	25	0.619	0.622	0.619	0.528	0.622
200	50	0.637	0.636	0.614	0.551	0.611
200	100	0.624	0.629	0.625	0.507	0.617

Table 2: Robustness Check: Stakes Distribution Heterogeneity

Distribution (Gini)	Equal Voice	Stakes-Weighted	Plutocracy	Random	Expert
High Gini (0.78)	0.618	0.644	0.617	0.522	0.618
Low Gini (0.03)	0.573	0.589	0.572	0.461	0.570
Medium Gini (0.26)	0.584	0.596	0.585	0.486	0.556
Pareto α=1.2 (0.85)	0.619	0.658	0.613	0.514	0.608
Pareto α=2.0 (0.53)	0.610	0.605	0.596	0.480	0.596
Pareto α=4.0 (0.42)	0.594	0.584	0.581	0.487	0.582

Acknowledgements

The author acknowledges the intellectual contributions of scholars in constitutional political economy, social choice theory, and legitimacy studies whose foundational work made this synthesis possible. Particular thanks to the open-source research community for making preprints, working papers, and replication materials freely accessible.

This research benefited from Perplexity AI for efficient literature discovery and source verification, and Anthropic's Claude for assistance with analytical framework development, literature synthesis, computational validation design, and technical writing. All conceptual innovations, theoretical claims, and interpretive judgments remain the author's sole responsibility.

All computational analysis was conducted at Resurrexi Lab, a distributed computing cluster built from consumer-grade hardware (8 nodes, 66 cores, 229GB RAM, 48GB VRAM), demonstrating that rigorous political economy research is accessible without institutional supercomputing infrastructure.

Research methodologies and reproducibility practices are documented at `https://farzulla.org/methodologies`.

References

H. Arendt. *On Revolution*. Viking Press, New York, NY, 1963.

K. J. Arrow. *Social Choice and Individual Values*. Yale University Press, New Haven, CT, 1951.

R. J. Aumann and R. B. Myerson. Endogenous formation of links between players and of coalitions. *American Economic Review*, 78(2):175–179, 1988.

J. F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19:317–343, 1965.

S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. 2019. Available at: https://fairmlbook.org.

W. Beckerman and J. Pasek. *Justice, Posterity, and the Environment*. Oxford University Press, Oxford, UK, 2001.

R. Blackburn. *The Overthrow of Colonial Slavery, 1776–1848*. Verso, London, UK, 1988.

G. Brennan and J. M. Buchanan. *The Reason of Rules: Constitutional Political Economy*. Cambridge University Press, Cambridge, UK, 1985.

H. Brighouse and M. Fleurbaey. Democracy and proportionality. *Journal of Political Philosophy*, 18(2):137–155, 2010.

J. M. Buchanan and G. Tullock. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. University of Michigan Press, Ann Arbor, MI, 1962.

Business Roundtable. Statement on the purpose of a corporation. https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans, 2019. Accessed November 15, 2025.

T. Christiano. *The Constitution of Equality: Democratic Authority and Its Limits*. Oxford University Press, Oxford, UK, 2008.

T. Clarkson. *The History of the Rise, Progress, and Accomplishment of the Abolition of the African Slave-Trade by the British Parliament*. Longman, Hurst, Rees, and Orme, London, UK, 1808.

D. Courant and T. Bourgeron. The french citizens' convention for climate: Proposals and political outcomes. *French Politics*, 19:321–338, 2021.

M. Cox, G. Arnold, and S. V. Tomás. A review of design principles for community-based natural resource management. *Ecology and Society*, 15(4):38, 2010.

R. A. Dahl. *A Preface to Democratic Theory*. University of Chicago Press, Chicago, IL, 1956.

R. A. Dahl. *Polyarchy: Participation and Opposition*. Yale University Press, New Haven, CT, 1971.

E. Douek. Content moderation as systems thinking. *Harvard Law Review*, 136:526–606, 2022.

S. Drescher. *Capitalism and Antislavery: British Mobilization in Comparative Perspective*. Oxford University Press, Oxford, UK, 1987.

J. M. Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, Princeton, NJ, 2006.

O. Equiano. *The Interesting Narrative of the Life of Olaudah Equiano, or Gustavus Vassa, the African*. Author, London, UK, 1789.

D. Estlund. *Democratic Authority: A Philosophical Framework*. Princeton University Press, Princeton, NJ, 2008.

D. M. Farrell, J. Suiter, and C. Harris. 'systematizing' constitutional deliberation: The 2016–18 citizens' assembly in ireland. *Irish Political Studies*, 34(1):113–123, 2019.

L. Fauver and M. E. Fuerst. Does good corporate governance include employee representation on boards? evidence from german corporate boards. *Journal of Financial Economics*, 99(3): 554–569, 2011. doi: 10.1016/j.jfineco.2010.10.016.

D. S. Felsenthal and M. Machover. *The Measurement of Voting Power: Theory and Practice, Problems and Paradoxes*. Edward Elgar, Cheltenham, UK, 1998.

S. Fine. *Sit-Down: The General Motors Strike of 1936–1937*. University of Michigan Press, Ann Arbor, MI, 1969.

J. S. Fishkin. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press, Oxford, UK, 2009.

J. S. Fishkin. *Democracy When the People Are Thinking: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press, Oxford, UK, 2018.

R. E. Freeman. *Strategic Management: A Stakeholder Approach*. Pitman, Boston, MA, 1984.

E. Friedgut, G. Kalai, N. Keller, and N. Nisan. A quantitative version of the gibbard-satterthwaite theorem for three alternatives. *SIAM Journal on Computing*, 40(3):684–702, 2011. doi: 10.1137/100791671.

M. Friedman. The social responsibility of business is to increase its profits. *New York Times Magazine*, September 13 1970.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

T. Gillespie. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, New Haven, CT, 2018.

S. Grimmelikhuijsen, A. Meijer, N. Lieberherr, et al. Legitimacy of algorithmic decision-making: Three value-based strategies. *Public Administration Review*, 82(5):800–812, 2022.

J. Habermas. *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Beacon Press, Boston, MA, 1984. Translated by Thomas McCarthy.

J. Habermas. *Moral Consciousness and Communicative Action*. MIT Press, Cambridge, MA, 1990. Translated by Christian Lenhardt and Shierry Weber Nicholsen.

T. Hobbes. *Leviathan*. Andrew Crooke, London, UK, 1651.

L. Hong and S. E. Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.

S. Jäger, S. Noy, and B. Schoefer. What does codetermination do? *ILR Review*, 75(4):857–890, 2022.

E. Kalai and M. Smorodinsky. Other solutions to nash's bargaining problem. *Econometrica*, 43 (3):513–518, 1975. doi: 10.2307/1914280.

N. Keller. A tight quantitative version of arrow's impossibility theorem. *Journal of the European Mathematical Society*, 14(4):1253–1278, 2012. doi: 10.4171/JEMS/331.

J. Kleinberg, S. Mullainathan, and M. Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of Innovations in Theoretical Computer Science*, 2017.

Knights of Labor. Preamble and declaration of principles, 1878.

C. M. Koggel. Feminist relational theory and global justice. In T. Brooks, editor, *The Oxford Handbook of Global Justice*, pages 285–302. Oxford University Press, Oxford, UK, 2022.

H. Landemore. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press, Princeton, NJ, 2013.

J. Locke. *Second Treatise of Government*. Hackett, Indianapolis, IN, 1980. Edited by C. B. Macpherson.

T. Lux and S. Schiffko. Estimation of agent-based models using sequential monte carlo methods. *Journal of Economic Dynamics and Control*, 91:391–408, 2018.

C. Mackenzie. Three dimensions of autonomy: A relational analysis. In A. Veltman and M. Piper, editors, *Autonomy, Oppression, and Gender*, pages 15–41. Oxford University Press, Oxford, UK, 2014.

C. Mackenzie and N. Stoljar, editors. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press, New York, NY, 2000.

E. McGaughey. The codetermination bargains: The history of german corporate and labour law. *Columbia Journal of European Law*, 23(1):135–176, 2016.

J. S. Mill. *On Liberty*. John W. Parker and Son, London, UK, 1859.

J. S. Mill. *Considerations on Representative Government*. Parker, Son, and Bourn, London, UK, 1861.

E. Mossel, K. Oleszkiewicz, and A. Sen. Quantitative gibbard-satterthwaite theorem without neutrality. *Combinatorica*, 32(3):305–317, 2012. doi: 10.1007/s00493-012-2713-4.

J. F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950. doi: 10.2307/1907266.

J. Nedelsky. Reconceiving autonomy: Sources, thoughts and possibilities. *Yale Journal of Law and Feminism*, 1(1):7–36, 1989.

E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, UK, 1990.

V. Ostrom, C. M. Tiebout, and R. Warren. The organization of government in metropolitan areas: A theoretical inquiry. *American Political Science Review*, 55(4):831–842, 1961.

R. A. Phillips. *Stakeholder Theory and Organizational Ethics*. Berrett-Koehler, San Francisco, CA, 2003.

H. F. Pitkin. *The Concept of Representation*. University of California Press, Berkeley, CA, 1967.

F. O. Ramirez, Y. Soysal, and S. Shanahan. The changing logic of political citizenship: Cross-national acquisition of women's suffrage rights, 1890 to 1990. *American Sociological Review*, 62(5):735–745, 1997.

J. Rawls. *A Theory of Justice*. Harvard University Press, Cambridge, MA, 1971.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, NY, 2nd edition, 2004.

J.-J. Rousseau. *The Social Contract and Other Later Political Writings*. Cambridge University Press, Cambridge, UK, 1997. Edited by Victor Gourevitch.

M. A. Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10 (2):187–217, 1975.

L. J. Savage. *The Foundations of Statistics*. John Wiley & Sons, New York, NY, 1954.

F. W. Scharpf. *Governing in Europe: Effective and Democratic?* Oxford University Press, Oxford, UK, 1999.

V. A. Schmidt. Democracy and legitimacy in the european union revisited. *Journal of Common Market Studies*, 51(1):131–147, 2013.

A. Sen. *Development as Freedom*. Oxford University Press, Oxford, UK, 1999.

A. K. Sen. *Collective Choice and Social Welfare: Expanded Edition*. Harvard University Press, Cambridge, MA, 2017.

L. S. Shapley and M. Shubik. A method for evaluating the distribution of power in a committee system. *American Political Science Review*, 48(3):787–792, 1954.

E. C. Stanton, L. Mott, and Others. Declaration of sentiments and resolutions. Seneca Falls Convention, July 19–20, 1848, 1848.

D. L. Teele. How the west was won: Competition, mobilization, and women's enfranchisement in the united states. *Journal of Politics*, 80(2):442–461, 2018.

S. Vitols. Coordinated market economies and financialization: Germany and japan in comparison. *Competition & Change*, 15(4):288–311, 2011.

A. E. Waldman and K. N. Johnson. The role of decision importance and governance in algorithmic legitimacy. *Northwestern University Law Review*, 117:1–52, 2022.

J. Waldron. *Law and Disagreement*. Oxford University Press, Oxford, UK, 1999.

W. Wilberforce. Speeches on the abolition of the slave trade. Hansard Parliamentary Debates, 1789–1807.

P. Yadav, A. Kumar, and M. Raizada. Institutional effects on ecological outcomes of community-based management of fisheries in the amazon: The arapaima case. *PNAS*, 118(29): e2100896118, 2021. doi: 10.1073/pnas.2100896118.