Epistocracy and the Commitment Problem

Abstract. The epistocracy debate has turned largely on the character of the electorate — what voters know, whether competence filters can be designed well, and what filtered selection produces. Another strand has focused on the legitimacy of epistocratic justifications. I ask instead about the character of the institution. Epistocracy shifts some political authority from the median voter toward citizens selected on competence grounds, and existing critiques press legitimational, epistemic, abuse-based, and demographic objections against such proposals. I develop a different line, focusing on one especially plausible form of epistocracy: Jason Brennan's *epistocratic veto*, a competence-selected council empowered to override legislation it judges incompetent. Drawing on Acemoglu and Robinson's account of democratization, I argue that the franchise solves a credible-commitment problem between elites and the poor under conditions of revolutionary threat: universal suffrage puts redistributive authority in majority hands in a way elites cannot easily reverse, because reversing it requires revoking the franchise itself. The institutionally realistic veto hands some of that authority back to a body whose preferences systematically diverge from the median voter's, reintroducing the commitment problem suffrage solved. I give empirical and theoretical reasons to worry that the magnitude of the problem could be significant. A formal appendix maps the veto to Acemoglu and Robinson's imperfect-democracy parameter and shows that the range of conditions under which democratic institutions avert revolution strictly shrinks.

1. Introduction

Jason Brennan has argued that while democracies perform reasonably well, they are imperfect, and that shifting some political authority to those who know more might improve them. He considers several forms this shift might take; I focus on what he highlights as enjoying various advantages over other epistocratic forms, and what is in any case the hardest case for my argument, the *epistocratic veto*: a council appointed by competence criteria, empowered to overturn legislation it judges unreasonable or incompetent (Brennan 2016, pp. 215–220). The debate Brennan started has mostly focused on what voters know, whether the experts we would pick really are experts, and whether the filtering can be gamed. This paper asks a different question.

What institutional function does the franchise perform — not as a mechanism through which any given voter expresses a preference, but in shaping the equilibrium behavior of elites and mass publics under conditions of sharp inequality? Daron Acemoglu and James Robinson (2006) give a structural answer. The franchise, on their view, solves a credible-commitment problem. Before universal suffrage, elite offers of redistribution to a restive poor were never credible: what could be offered in a moment of crisis could be rescinded once the crisis passed, and both sides knew it. Extending the franchise changed that. It put authority over redistributive policy in the hands of the poor themselves, in a way elites could not easily reverse. That transfer is what historically averted revolution in societies where inequality was steep enough for the poor to threaten one. Whatever else democracy is, it is also an institution that holds certain elite temptations in check.

If that is right, it matters for Brennan. The veto hands back to a competence-selected body some of the redistributive authority the franchise was supposed to keep in majority hands — not all of it, and not on every issue, but enough to matter on the redistributive questions where majority and elite preferences most visibly diverge. My claim is that this reintroduces, in attenuated but nonzero form, the same commitment problem universal suffrage solved. The institutionally realistic veto carries a structural cost: a loss of commitment value to the coalition whose credible threat underwrote the original democratic settlement. That cost has been systematically absent from the

ledger the epistocracy literature has kept.

I do not claim the cost is decisive. Brennan is a comparative instrumentalist. He asks whether a regime produces better policy on average than feasible alternatives. My argument grants the question — it concedes, for the sake of argument, that expert judgment is epistemically superior to majority judgment on redistributive matters — and adds a line to the ledger on democracy's side. Whether that line tips the balance depends on how much it weighs and how much the epistemic benefits weigh against it. What I do claim is that the line is real, nonzero in institutionally realistic veto regimes, and has to be weighed. Still, the significance of this overall problem depends on the magnitude. The final response of Section 4 analyzes this question in detail, arguing that the cost's size is proportional to the size of the preference divergence between an epistocratic council and the median voter, and exploring the conditions under which the overall magnitude of the problem would be significant.

This is not the only line of critique against epistocracy, of course. Four broad lines have organized the debate so far. The *legitimational* line (Estlund 2008) argues that the invidious epistemic comparisons epistocracy requires cannot meet the demands of public justification. The *epistemic* line (Landemore 2013; Gunn 2019; Méndez 2022) argues that restricted electorates are cognitively impoverished in ways Brennan's competence filters cannot repair. The *abuse* line (Bagg 2018; Klocksiem 2019; Vandamme 2020; Somin 2022) argues that epistocratic selection will be gamed in predictable ways. The *demographic* line (Christiano 2008; Kolodny 2014) argues that competence criteria track privilege along dimensions that implicate broader equality concerns. Three of these lines focus on the character of the *electorate* — what voters know, how they decide, or who counts. The legitimational line raises a different question, about whether epistocratic justifications are publicly acceptable. The commitment argument asks instead about the character of the *institution*.

A short preview. The argument takes the epistocrat's case in its strongest form: it assumes that epistocratic selection is clean, that competence criteria are unbiased, that council members are competent in both their expertise and their deliberations, and that none of the abuse or gaming worries the existing critical literature focuses on are operative. It is not an abuse argument. Its mechanism turns instead on a structural feature the Acemoglu–Robinson framework makes precise — that the allocation of decisive authority over redistributive policy shapes what elites can credibly commit to over time, and that shifting decisive authority on redistribution from the median voter to a competence-filtered council degrades that commitment. Section 3.4 defends the formal modeling choice at the heart of this: the reduction of the veto to Acemoglu and Robinson's reduced-form parameter χ. The argument is also narrow in policy scope. It bears on institutions that substitute expert judgment for democratic redistributive authority on substantive-competence grounds — Brennan's veto on its institutionally realistic reading, some technocratic fiscal-authority proposals, and historically the Lochner-era substantive-due-process review of social legislation. It does not bear on rights-based judicial review, on mandate-bounded regulatory agencies, on technical advising, or on unfiltered sortition.

Section 2 lays out the Acemoglu-Robinson argument. Section 3 develops the main challenge to epistocracy. Section 4 takes up four responses; the fourth, the magnitude of the problem, is the most consequential and receives the longest treatment. Section 5 locates the argument relative to existing critiques of epistocracy, with attention to Bagg (2018). Section 6 draws methodological and policy upshots. An appendix formalizes the core results.

2. The Acemoglu-Robinson Commitment Device

Acemoglu and Robinson's argument begins with the structure of pre-democratic autocracy and shows how extending the franchise changes its equilibrium. It starts from an asymmetry in how authority is held under autocracy. The rich hold *de jure* power over policy — the constitutional authority to set taxes, determine transfers, and enforce property rights. The poor hold *de facto* power only during "revolutionary moments," periods when coordination is cheap enough that collective action can threaten the regime. Both claims are idealizations of a specific class of transitions; §5 returns to their empirical range.

Here is the basic picture. The poor's preferred tax rate is $\tau^p > 0$; the rich prefer $\tau^r = 0$. Call θ the rich's share of total income and μ the cost of revolution — the fraction of output destroyed if the poor revolt. When $\theta > \mu$, the poor would rather revolt than accept the autocratic status quo.

Faced with this, the rich have every incentive to buy off the poor. They can offer the preferred redistributive bundle now in exchange for peace. But the offer is not credible. Once the revolutionary moment passes, *de jure* power reverts to the rich, who have every incentive to roll taxes back to zero. The poor, anticipating this, reject the offer and revolt anyway.

Extending the franchise breaks the impasse. Under universal suffrage, τ^p is no longer a concession. It is an equilibrium output — what the regime implements each period, not because elites are honorable but because the institutional structure does not give them the option (Meltzer and Richard 1981; Acemoglu and Robinson 2006, ch. 4). The promise to redistribute becomes credible precisely because reversing it requires revoking the franchise itself, and revoking the franchise is costly.

Notice what is doing the work here, because it is the pivot of what follows. The commitment function has nothing to do with whether voters are knowledgeable, or whether they deliberate well, or whether they vote "rationally." The properties of the electorate that the epistocracy debate fixes on — competence, information, rationality — are not the properties that drive the commitment mechanism. What drives the mechanism is *who holds authority over redistributive policy.* Two regimes with identical electorates, identical information, and identical deliberative capacities can differ in their commitment properties if they differ in who gets the decisive vote. An institution that changes who decides on redistribution thereby changes commitment, even if it leaves the electorate and its epistemic qualities exactly where it found them.

Two empirical qualifications matter for what follows. First, real democracies do not reliably produce τ^p. How much they actually redistribute depends on electoral systems, coalition structures, ideological framing, and mass mobilization (Iversen and Soskice 2006, 2019; Boix 2003; Scheve and Stasavage 2016; Kenworthy and Pontusson 2005). But the commitment argument does not require that democracy hit τ^p on the nose. It requires that democracy *structurally support* median-voter authority over redistribution — that the median voter's preferences are the equilibrium reference point even when politics noisily pushes outputs around. The commitment function rides on that structural role, not on policy-magnitude determinism.

Second, there is a live empirical debate about whether historical franchise extensions were actually driven by AR's mechanism or by something else — elite-competition dynamics between incumbent and rising economic groups, the organizational capacity of conservative parties to compete under wider franchises, elite-dominated transitions (Ansell and Samuels 2014; Ziblatt 2017; Albertus and Menaldo 2018). §5 returns to this. For present purposes: the normative argument does not require that AR was historically dominant. It requires only that the mechanism is a live possibility now.

For the argument that follows, one extension of the AR model will be especially important. Acemoglu and Robinson handle imperfect democracies — democracies that favor elite preferences to some degree, short of reverting fully to autocracy — by introducing a single reduced-form parameter. They call it χ (Ch. 4 §5, pp. 115–116), and let it run from 0 to 1. χ tracks how much weight elite preferences get when the polity sets policy. At $\chi = 0$ the regime delivers τ^p and we are at full democracy. At $\chi = 1$ it delivers $\tau^r = 0$ and elites effectively run the show. Between the endpoints, higher χ means less redistribution. The usefulness of χ for my purposes is that it provides a single dimension along which an imperfect democracy can be compared to a fully democratic benchmark — and, as §3 will argue, the veto regime can be mapped onto that dimension.

From this setup they derive a central result: a democracy that redistributes too little — χ too high — is no longer credible to the poor as a commitment. The set of environments in which democratization averts revolution shrinks as χ rises. This is the structural feature I will argue the epistocratic veto inherits.[1]

One last point about scope. Acemoglu and Robinson develop the mechanism for transitions in societies where $\theta > \mu$ binds: agrarian and early-industrial polities with extreme inequality and credible revolutionary threats. In contemporary advanced industrial democracies, the constraints do not bind in the literal 19th-century sense. §3.3 takes up how the argument bears on polities where the conditions hold only in attenuated form.

3. The Epistocratic Veto as Commitment Failure

3.1. Which veto — and at what scope

Brennan discusses four epistocratic regimes: restricted suffrage, plural voting, enfranchisement lotteries, and the epistocratic veto. I focus on the veto for three reasons. Brennan himself highlights various advantages it enjoys over other epistocratic forms (Brennan 2016, pp. 215–220). It is the hardest case for my argument, because it preserves universal suffrage at the ballot box — so arguments targeting franchise restriction do not apply directly. And it admits the cleanest reduction to AR's χ. The appendix sketches the other three regimes (§A.4).

A word on *which* veto is in view. At Brennan 2016, pp. 215–220, the council is described in narrow terms. It overturns legislation judged "malicious, incompetent, or unreasonable" (p. 216), and the justificatory analog he invokes is judicial-review-style rights-and-procedural review rather than substantive policy override. On this narrow reading the council rarely vetoes, and never on ordinary redistributive grounds. If that is the veto on offer, my argument has nothing to say about it.

But Brennan also *motivates* the veto by appeal to the empirical gradient between information and preferences. On his account, well-informed preferences diverge from median preferences on fiscal and regulatory matters, and the comparative-instrumentalist case for the veto rests on that divergence (Brennan 2016, ch. 7; the ordinal ranking of regimes on demographic representativeness is developed in Brennan 2018, p. 59). This motivation will not survive a veto that never touches fiscal or regulatory policy. On a strictly rights-and-procedural veto, the information-preference gradient does no work.

[1] AR's is not the only public-choice argument linking narrow franchises to predictable pathologies; see also Lizzeri and Persico (2004) on clientelism and public-goods under-provision, and Bueno de Mesquita et al. (2003) on small-coalition rent extraction. I rely on AR because their argument is the most explicit about the specific mechanism — intertemporal credibility of redistribution — that franchise extension supplies and that competence-based restriction would take away.

So Brennan's text points two ways at once. His institutional specification is narrow. His argumentative case requires a scope wide enough for the gradient to bite. I will call the institutional specification the *narrow reading* and the scope required by the argumentative case the *broad reading*. The paper's target is the institutionally realistic veto, which sits closer to the broad reading, for two reasons.

First, the comparative-instrumentalist case for the veto requires that the council's judgment differ from the legislature's precisely on matters where expert competence matters. On Brennan's own account, redistributive policy is one of those matters. A veto whose scope excluded everything that made the epistocratic case would be a veto without a case.

Second, the institutional-drift literature gives good reasons to think counter-majoritarian bodies with substantive-competence authority expand their effective scope over time (Hirschl 2004 on juristocracy; Tushnet 2003). The mechanisms are familiar: precedent ratcheting, selection feedback on council composition, the political economy of doctrinal expansion. Constitutional design can slow these drifts. No comparative record shows that it stops them, and Brennan offers no drift-prevention proposal.

So the target is a veto that is not quite the formal minimum Brennan specifies at pp. 215–220 but is what his own argumentative case requires and what the comparative literature suggests such a body becomes. Readers who take the narrow reading to be Brennan's considered view should read the paper as arguing against the institutionally realistic form rather than the formal minimum. Readers who take the broad reading to follow from the comparative-instrumentalist motivation should read it as arguing against Brennan's considered view.

3.2. The veto as an imperfect democracy

Under the veto, a universally-enfranchised legislature passes redistributive legislation and a council reviews it. Let v be the per-period probability that the council strikes a given piece of legislation within its domain, and let τ_C be the council's preferred rate. The realized tax in a given period is then a simple lottery: τ^p with probability $1 - v$, and τ_C with probability v. The expected tax over many periods is a weighted average of the two.

This expected tax has a natural interpretation in the Acemoglu–Robinson framework. It is as if the polity were operating at an "effective χ" — call it χ_{E_v} — given by

$$\chi_{E_v} = v\,(\tau^p - \tau_C)/\tau^p.$$

Viewed through this lens, the veto regime is an imperfect democracy whose imperfection comes from a specific institutional source: expert review. At $\chi_{E_v} = 0$ we are at full democracy; at $\chi_{E_v} > 0$ we are somewhere between full democracy and elite rule.

One difference between the veto regime and AR's reduced form is worth flagging on the surface. Because the veto delivers a lottery rather than a deterministic tax, and because the poor's value function is strictly concave in the tax rate (it peaks at their preferred rate and falls off on either side, with the curvature coming from deadweight loss), the veto regime is *strictly worse* for the poor than an AR limited democracy at the same mean-matched χ — by Jensen's inequality. The mean-matching is therefore conservative: the real commitment cost is larger than χ_{E_v} alone would suggest, so comparing the veto to AR's reduced form at $\chi = \chi_{E_v}$ underestimates the bite of the

mechanism.[2]

What makes the reduction substantive is two empirical claims: $\tau_C < \tau^p$, and $v > 0$. The second is easy. A council that never vetoes is a ceremonial body, not the institution Brennan defends. The first is a claim about preferences. Council membership is selected on competence, and competence correlates empirically with income and information; Brennan himself flags the demographic gradient (Brennan 2018, pp. 55–60).

Response 4 of §4 explores in detail what the empirical literature can tell us about the information–preference gradient, and how much commitment cost it supports. For present purposes, two facts orient the analysis. The information–preference literature robustly documents a divergence between better-informed and median preferences on the size and scope of government (Althaus 2003; Caplan 2007), with qualifications from the responsiveness literature (Enns and Wlezien 2011; Soroka and Wlezien 2010; Branham, Soroka, and Wlezien 2017). And the premise the analysis actually needs — $\tau_C < \tau^p$ on net redistribution — functions in this paper as the epistocrat's own empirical premise, granted for the argument (Brennan 2016, ch. 7; §3.1).

The argument does not depend on conflating epistocratic elites with the rich. Two versions of the worry deserve explicit treatment. First, a reader might think that an income-representative council escapes the preference gap. But the gap rests on the information–preference gradient (Althaus 2003), which holds within demographic groups as well as across them: better-informed preferences are systematically less interventionist on economic policy even after controlling for income, race, and gender. To the extent the gradient extends to redistribution proper — as the epistocrat's case asserts — an income-representative council selected on competence can still have preferences less redistributive than the median voter's. Brennan's own ranking by demographic representativeness (Brennan 2018, p. 59) places the veto sixth of eight in any case, between the filtered enfranchisement lottery and restricted suffrage. Response 4 of §4 develops the magnitude question from this. Second, a reader might think that an income-representative council closes the practical-influence gap that the rich currently exploit in democratic politics, and so produces more redistribution than current legislatures do. But the AR mechanism turns on a different fact: whether the median voter has decisive structural authority over redistribution. Universal suffrage anchors that structural authority — the median voter can discipline a legislature that strays through electoral pressure, even when actual legislation reflects elite influence. A veto regime removes the anchor: council membership is selected on competence rather than by the median voter, so the median voter cannot discipline a council that strays. Per-period practical influence and over-time structural authority are different institutional facts.

3.3. The main argument, with scope

The core result is this. Universal suffrage is a credible commitment to the poor: the equilibrium tax is τ^p every period, and reversal requires costly revocation of the franchise itself. The veto regime with $\chi_{E_v} > 0$ shrinks the set of environments in which that commitment averts revolution strictly (Proposition A.1). In the shrunken region, a regime that would have been stable under democracy is destabilized under the veto.

A minimal example fixes what is at stake. Suppose inequality is high enough to make revolution feasible against the autocratic status quo — say $\theta = 0.7$ — and suppose that under full democracy

[2]The mapping matches the veto regime to AR's reduced form on the expected tax but does not equate the two distributions; AR's limited democracy delivers a deterministic $\tau(\chi)$ each period, while the veto regime delivers a lottery. Appendix §A.2 works out the Jensen wedge precisely.

the poor prefer the democratic equilibrium to revolution, but only by a narrow margin. Introduce a veto with $\chi_{E_v} = 0.06$ (the midpoint of the range Response 4 of §4 calibrates): the expected redistribution under the veto is about six percent below the democratic benchmark, a modest but positive shortfall. That shortfall lowers the poor's expected payoff under the regime; if the margin under democracy was thinner than the loss the veto introduces, the poor now prefer revolution. A polity that would have been stable under democracy becomes unstable under the veto. Proposition A.1 shows there is a non-trivial region of the parameter space where the flip occurs.

The scope of this result reflects the parameter structure of AR's model. The mechanism operates strictly where the revolutionary constraint $\theta > \mu$ binds — where inequality is sharp enough that revolution is feasible and its prospect is credible to both sides. This is the terrain of the 19th- and 20th-century transitions AR analyze, of some contemporary middle-income democracies, and of any polity where mass-mobilization threats are a genuine factor in elite decision-making. In wealthy, stable, nuclear-armed democracies with thick civil society, where $\theta < \mu$ binds strictly, literal revolution is off the equilibrium path and the canonical AR mechanism does not apply. Whether the argument has any bite on those polities depends on whether *cognate* structural mechanisms — mass protest, democratic backsliding, populist collapse, electoral realignment — carry the AR credible-threat role in attenuated form.

The recent literatures on democratic backsliding and populism (Rodrik 2018; Funke, Schularick, and Trebesch 2023; Guriev and Papaioannou 2022; Levitsky and Ziblatt 2018) document that democracies under distributional stress exhibit predictable failure modes — polarization, norm erosion, gatekeeping failure by mainstream parties. These failure modes share a family resemblance with AR's story in that both link distributional distress to regime instability, but they are not the AR mechanism running at lower intensity: they are different mechanisms. This paper imports AR, not the backsliding channels, and does not establish that those channels produce the same commitment-failure dynamics Proposition A.1 proves. What the literatures jointly suggest is that the polities where the AR mechanism is most clearly live — where one expects the commitment argument to bite hardest — are the same polities where cognate backsliding channels also run. For prior work pressing instability concerns against epistocratic and politically-meritocratic arrangements, see Fumagalli (2022), Talisse (2022), Landa and Pevnick (2025), and Wong (2025). These contributions raise instability worries from angles other than the AR commitment mechanism; the present paper complements rather than competes with them.

A rough *liveness criterion* identifies where AR's mechanism is live. It has three parts: *revolutionary-threat capacity* (does the polity exhibit mass-mobilization capacity around distributional grievance? — strike intensity, demonstration frequency, willingness to participate in collective action on redistributive issues), *elite expropriation-risk perceptions* (do elites take redistributive shocks to be a real political risk? — asset-holding patterns, capital-flight sensitivity, political-risk premia in sovereign debt), and *distributional-grievance salience* (do redistributive questions animate public opinion? — measures of inequality concern, perceived elite capture, trust in democratic institutions on redistributive questions). Regimes high on all three have the AR mechanism live. The contemporary United States scores high on distributional-grievance salience and moderate on revolutionary-threat capacity; European social democracies score lower on both; many Latin American middle-income democracies score high on all three. The argument's bite on each polity is proportional to its score.

A further empirical qualification is worth making. The AR mechanism requires the poor to perceive realized redistribution accurately enough to update their revolutionary calculus on it. If the regime is institutionally opaque — if the veto's role in shaping outcomes is hard to attribute, for example — the bite of the mechanism is attenuated correspondingly. A more general formulation could include

a perception parameter scaling with the regime's transparency. I leave the formal extension to future work.[3]

Brennan himself frames the evaluation as what he calls, after Michael Munger, the Big Pretty Pig Contest (Brennan 2016, pp. 204–207): realistic regimes are compared on aggregate substantive performance, each with its real-world flaws, rather than by matching an idealized version of one against a messy version of the other. On that standard, the commitment cost belongs on the ledger alongside whatever epistemic benefits the epistocrat claims. Whether the ledger tilts for or against the veto is a further question. My argument contributes a consideration, not a verdict.

A final note on what the argument does and does not require. It does not require that epistocrats are corrupt, biased, or incompetent. It requires two things only: the institutional role the veto plays, and the empirical preference gradient that makes $\tau_C < \tau^p$ in practice. The second premise is empirical rather than conceptual, so the argument is conditional on it, rather than constitutive in a theorem-strength sense. Responses 2 and 4 of §4 press on this premise.

3.4. Is the reduction doing real work?

One worry about the argument runs through the reduction itself: if any counter-majoritarian institution mapped to $\chi > 0$, the result would follow from AR's framework and have little to do with the veto in particular. It does not. What generates $\chi > 0$ in AR's sense is an institutional arrangement that substitutes some non-median preference for the median voter's on redistribution. Rights review does not do this; central banking does not; mandate-bounded agencies do not; the institutionally realistic veto does. Response 1 works out the diagnostic case by case. Two further technical concerns — the variance introduced by the stochastic veto, and the possibility that v is chosen strategically by the council rather than fixed — are addressed in the appendix remarks (§A.4). The reduction also inherits AR's scope conditions — a poor capable of collective action, a binding revolutionary constraint $\theta > \mu$, and no alternative commitment technology doing the franchise's work — which the liveness criterion of §3.3 makes operational.

4. Four Responses

Four responses to the argument are worth taking up in turn. The first questions its scope: does it prove too much, condemning ordinary counter-majoritarian institutions alongside the veto? The second questions the empirical premise that $\tau_C < \tau^p$: does competence-based selection really produce a council with less redistributive preference than the median voter? The third questions the uniqueness of the franchise as a commitment technology: might some other institution supply what the franchise supplies, so that shifting authority away from the franchise costs nothing in commitment terms? The fourth — the most consequential — questions the magnitude of the problem: even granting the mechanism, is the cost it identifies large enough to matter? I take it up last and at length.

Response 1: The veto is like judicial review — the argument proves too much

Modern liberal democracies contain many institutions that constrain majoritarian redistribution: constitutional courts, central banks, independent regulatory agencies, supermajority rules. If the commitment argument condemns the veto, shouldn't it also condemn these?

[3]Thanks to an anonymous referee for raising this point.

The diagnostic question is clear enough. Does the institution substitute its own substantive policy preferences for the median voter's on redistributive matters on the ground of competence? Institutions differ.

Constitutional courts review legislation on procedural and rights-based grounds. The boundary between rights review and competence review is not always crisp — Waldron (2006) has pressed a related point, arguing that rights review itself often collapses into substantive policy disagreement. But the two kinds of review take characteristically different justificatory forms. Rights review asks whether legislation violates a protected interest; competence review asks whether legislation is wise policy. The practical difference shows up in what legislation actually gets struck down. Central banks operate within mandates set by elected legislatures: they choose instruments, not targets. Independent agencies implement policy whose basic shape is set by elected principals. Epistocratic councils, by description, overturn legislation on substantive-competence grounds. On the institutionally realistic reading of the veto, redistributive legislation falls within those grounds (§3.1).

Institutions that support median-voter authority on redistribution have χ close to zero on the dimension that matters. The institutionally realistic veto does not. The argument is not a general anti-counter-majoritarian claim.

Edge cases confirm the diagnostic. The Lochner-era U.S. Supreme Court overturned redistributive legislation on substantive-competence grounds, and the argument predicts that it generated commitment costs. Constitutional theorists mostly treat Lochner as a failure mode rather than a model, and the juristocracy literature (Hirschl 2004) tracks the same structural property. Some contemporary regulatory bodies that have drifted into substantive redistributive territory exhibit high-χ features; the argument predicts commitment costs for them, too. This is a testable implication — one readers worried about juristocratic tendencies in contemporary constitutionalism can check against their own judgments.

Response 2: Clean selection and orthogonal filters

The second response denies the empirical premise that $\tau_C < \tau^p$. A competence-selected council, the argument runs, may not inherit the information-preference gradient documented in survey populations. Its members do not run for re-election, do not owe favors to party machines, and are not tied to donor networks. What gets them onto the body is what they know, not whose side they are on; the council's preferences could be *orthogonal* to its members' informational profile in a way the wider population's preferences are not. Brennan's own response to abuse worries in Chapter 8 of *Against Democracy* turns on a related thought — that whether an epistocratic regime would be abused worse than democracy is a comparative question to be settled empirically, not ruled out in principle (Brennan 2016, pp. 223–224). Extending this comparative strategy to the preference-gradient worry does require a further step Brennan does not take explicitly, namely that competence-based selection for a standing council differs in kind from partisan or coalitional selection. But it is a natural extension, and it is the response with the most purchase on the commitment argument.

Two distinct moves could deliver that result, and they are worth separating.

The orthogonal-filter move. One way to make this concrete is to use Brennan's own typology. His Hobbits (ignorant), Hooligans (biased partisans), and Vulcans (cool reasoners) invite a charitable reading of the veto — one Brennan does not develop explicitly, but which fits his broader picture —

on which the council corrects *hooligan bias* (motivated reasoning, factual error, availability-driven misjudgment) rather than installs *Vulcan preferences*. Debiasing, not substitution. On this reading, feasible competence criteria for a veto council might in principle be orthogonal to redistributive preference: basic civic literacy, factual accuracy on policy-relevant empirical questions, absence of gross misinformation. If the filter is orthogonal in that sense, filtered selection produces $\chi \approx 0$ and the commitment argument is silent (Appendix §A.4(c) works the case out for the filtered lottery; the same reasoning extends to the veto council).

The difficulty is empirical. The criteria Brennan actually discusses — political knowledge, factual accuracy about policy consequences, capacity for reasoned deliberation — correlate with economic-policy preference in surveyed populations (Althaus 2003; Caplan 2007), and the epistocrat's case extends the correlation to redistributive preference (§3.2). The correlation may reflect substantive disagreement, demographic composition, framing effects, or genuine debiasing; the literature does not decompose cleanly. What matters here is that it is real, and that it is precisely what generates Brennan's epistemic case in the first place. Debiasing and substitution produce similar shifts in τ_C whenever the debiasing correlates with the well-informed preferences the information literature tracks — which, on Althaus's own simulations, it does. An orthogonal filter, one correlated with factual accuracy but not with redistributive preference, is a filter Brennan has not specified and the empirical literature has not identified. The orthogonal-filter escape remains possible but not yet realized.

The structural-selection move. A second response grants the correlation between competence and redistributive preference in the wider population but denies that it carries over into the council. Courts and senior civil services sometimes behave in ways that their institutional setting — insulation from electoral pressure, professional norms, small-group deliberative culture — makes intelligible, and that depart from what members' partisan affiliations alone would predict (Shapiro 1981; Wilson 1989; Gibbons 2022 develops a version of the argument in the epistocratic setting). The hypothesis, applied to a veto council, is that the same forces suppress the information-preference gradient inside the council even while it persists outside.

Two difficulties face this move. The first is that it asks more of the structural-selection effect than the literature on judicial and civil-service behavior has shown. Insulation reshapes behavior on some dimensions, but erasing a population-level preference gradient is a stronger claim and requires dedicated evidence that has not been produced for the veto-council case. The second is that it runs against the endogenous-drift dynamic flagged in Response 4: as the council vetoes redistributive legislation, its composition shifts over time toward members whose preferences would justify past vetoes, amplifying rather than suppressing the gradient.

The argument does not rule out epistocracies with demonstrably orthogonal filters and strong structural debiasing. It bites on epistocracies whose competence filters correlate with the redistributive-preference gradient the empirical literature documents, and whose composition is subject to ordinary selection and drift dynamics — which is where institutionally realistic proposals sit.

Response 3: The franchise is not the only commitment technology

The third response concedes $\chi_{E_v} > 0$ but denies that it generates a commitment failure, on the ground that the franchise is not the only commitment technology available. Modern political economies are saturated with alternatives — constitutional entrenchment, independent fiscal institutions, reputation effects, Tiebout exit, international integration — and if any of these supplies what the franchise supplies, shifting redistributive authority to a council costs nothing in commitment terms.

This is the response on which the appendix's result is most clearly conditional: the strict-inclusion theorem holds *other commitment technologies fixed*.

My reply is that the canonical alternatives each fall into one of three patterns. They presuppose the franchise upstream: constitutional redistributive guarantees like Germany's *Sozialstaatsprinzip* or Brazilian social-rights clauses bind because legislatures and courts treat them as speaking in the people's name, and an entrenchment adopted *under* the veto regime would itself be subject to council review on competence grounds; reputation on the repeated-game treatment (Dixit 2003; Greif 2006) requires off-equilibrium punishments that are electoral, constitutional, or revolutionary — that is, franchise-based or franchise-backed. Or they protect only already-powerful classes: Tiebout exit shields mobile capital, not low-skill labor (Molloy, Smith, and Wozniak 2011), exactly the wrong class for the commitment problem. Or they collapse into advisory roles: fiscal councils produce forecasts and advice rather than binding redistributive authority, and international integration binds only when it aggregates democratic authorization that the unilaterally-acting veto regime does not supply.[4]

One escape from the argument deserves explicit treatment. López-Guerra (2014) proposes an *enfranchisement lottery*: a randomly selected fraction of the citizenry votes each election. Without competence filtering, the expected decisive voter equals the population median, expected $\chi = 0$, and the commitment argument has nothing to say. But this does not help the epistocrat. An unfiltered lottery is not an epistocracy; it is sortition. Brennan does not defend unfiltered sortition; Landemore (2020) does. An epistocrat who escapes the argument by adopting an unfiltered lottery has stepped outside the epistocratic camp. Together with the orthogonal-filter escape of Response 2, this sharpens the argument's target: *the commitment cost bites on competence-weighted regimes with substantive authority over redistributive legislation*, with the gradient-installing filter as the load-bearing element.

The list of alternatives is not exhaustive. Some technology I have not canvassed might do the work. What the canvass shows is that the canonical ones are not free-standing substitutes for the franchise's commitment function — they presuppose it, or cover only already-powerful classes, or reduce to advising. I have not ruled out a successful counterexample. I have not found one.

Response 4: The magnitude of the problem

One worry is that although *in principle* epistocracy faces a commitment problem, the actual magnitude of the problem may be quite small. For example, it may be no larger than the ordinary gap between what voters expect and what policy delivers — a gap democracies absorb every period without regime-level consequences. In that case, the present critique of epistocracy would not be very interesting: it would identify a possible cost of the veto without showing the cost to be an important one. In response, I make three main points. First, I consider what the empirical literature can tell us about the size of the policy divergence between council and median voter, and calibrate an illustrative range. Second, I argue that the epistocratic veto's effects may differ in a few ways from ordinary democratic gaps between voter expectations and policy: regarding whether they

[4]Non-democratic historical cases sometimes offered as counterexamples — Bismarckian social insurance, Meiji welfare programs, Lee Kuan Yew's Singapore — reward inspection and turn out not to help the epistocrat. Bismarck's welfare state was built to pre-empt socialist mobilization and was visibly inadequate to its commitment function by the Weimar collapse. Meiji reforms bought off rural elites, not a revolutionary poor. Singapore operates on a distinctive settlement (PAP dominance plus high-end technocracy plus Tiebout-adjacent exit for mobile capital) that does not generalize. None is a case of durable commitment to an otherwise-revolutionary poor coalition under sustained non-democratic governance.

are systematic, correctable, and attributable. Third, I present suggestive evidence that changes in taxation and spending of roughly the calibrated size do in fact produce political unrest. I then revisit the reply that a prudent council, or rising prosperity, would shrink the realized cost, and close with a summary.

(a) How large is the divergence?

The commitment cost is governed by two quantities (§3.2): how far the council's preferred level of redistribution falls below what the median voter wants ($\tau^p - \tau_C$), and how often the council actually blocks redistributive legislation (v). The first of these quantities could be measured empirically, at least in principle.

The relevant empirical studies compare the policy preferences of citizens who know more about politics with those of citizens who know less, holding income, race, gender, and other demographics fixed — in effect estimating what the public would prefer if everyone were well informed. Their best-documented finding concerns the size and scope of government: better-informed citizens favor less government intervention in the economy, more market-oriented policy, and freer trade (Althaus 2003; Caplan 2007). But the scope of government is not the same thing as redistribution to the poor. Some spending and regulation redistributes upward rather than downward, and some policies that look redistributive fail to redistribute — rent control is a classic example (Diamond, McQuade, and Qian 2019), and on conventional economic criticisms the minimum wage may be another (Neumark and Wascher 2008; contested by Cengiz, Dube, Lindner, and Zipperer 2019). On taxation the same studies mostly point the other way: better-informed citizens were more opposed to the regressive 2001 tax cuts, and telling people how few estates pay the estate tax raises support for it (Bartels 2008, chs. 6–7; Kuziemko, Norton, Saez, and Stantcheva 2015). And redistribution depends on how revenue is spent, not only on tax rates. So the literature does not directly measure the quantity the model needs — the gap between filtered and median preferences over net transfers to the poor — and it is difficult to measure $\tau_C < \tau^p$ empirically.

Here, one general dialectical point is relevant. As the preferred tax rate under epistocracy converges with the rate under democracy, the relative value of epistocracy also falls: if epistocracy does not influence policy, what is the point of it? The epistocrat could reply that the veto mainly influences policy on matters other than redistribution. But this would be surprising, given how much of policy concerns redistribution.

Another way to estimate the magnitude of the effect is analytically, by considering how potential divergences between median and council tax rates would be predicted to influence revolution rates in the model. For example, suppose τ_C/τ^p falls somewhere in the range [0.5, 0.8] — that is, the council's preferred rate is between half and four-fifths of the median's. Next, we need to supply a schematic veto frequency. For illustration, suppose the council vetoed at a rate within [0.05, 0.25].

Multiplying the two ranges together, the veto lowers expected redistribution by somewhere between 1 and 12.5 percent of the median voter's preferred rate, with a midpoint around 6 percent. In concrete terms: if the median voter's preferred tax-and-transfer flow is around 30 percent of national income, the veto would shave roughly 2 percentage points of national income off realized redistribution — about the size of a major austerity package, except repeated every period rather than enacted once. In the model, the effect of such a shortfall on revolution risk depends on where a polity starts. The shortfall lowers what the poor receive from the regime every period. For polities where the poor's preference for democracy over revolt was already narrow, that loss can flip the comparison, destabilizing a polity that democracy alone would have kept stable — and Proposition A.1 shows that the region where this happens is not a knife-edge. For polities far from that margin, there is

no flip, and the cost is simply the lost redistribution itself.

(b) Does the epistocratic effect differ from ordinary democratic gaps?

Still, even if epistocracy has some kind of effect here, perhaps the expected effect is no greater than similar effects under democracy. After all, democratic processes are messy, and voters cannot expect that democratic policies will reliably produce the preferred tax rates of the median voter.

There are, however, a few ways that the epistocratic effect might differ from the democratic one. First, the epistocratic effect would involve a systematic departure from the median voter. Ordinary democratic deviation from the median voter's tax rate looks more like noise: sometimes the deviation goes to the right of the median voter, at other times to the left, and over time the deviations tend to average out. The epistocratic deviation runs in one direction — toward less redistribution — so it accumulates rather than washing out. Second, the epistocratic effect is not correctable by election. When democratic politicians depart from the median voter, voters can reliably expect correction at the next election. By contrast, the epistocratic council cannot be voted out. Third, the epistocratic effect is attributable. The AR mechanism requires the poor to observe realized redistribution (§3.3), and a council veto of passed legislation is a discrete, public, datable act with an identifiable institutional author — a focal event of the kind that lowers the coordination cost of mass mobilization. The diffuse drift of coalition politics offers nothing comparable to organize around. Again, though, these points are speculative; it is difficult to precisely estimate how significant the commitment problem is.

(c) Do tax rates affect political unrest?

Another worry about magnitude concerns the effect of taxation on revolution. Perhaps there is little actual effect, and so even if epistocracy changed the tax rate, this would not generate relevant commitment problems. Here, there is some suggestive empirical evidence that tax rates do indeed influence political unrest. Across Europe's twentieth century, expenditure cuts were followed by significantly elevated demonstrations, riots, and general strikes, with unrest increasing in the size of the cuts, and with expenditure cuts far more potent than tax changes (Ponticelli and Voth 2020). Two more recent studies point the same way. Districts more exposed to the United Kingdom's post-2010 welfare cuts shifted measurably toward UKIP — an effect large enough that, on the study's estimates, the Brexit referendum might plausibly have gone the other way without austerity (Fetzer 2019). And German districts subjected to deeper austerity between 1930 and 1932 gave measurably higher vote shares to the Nazi Party (Galofré-Vilà, Meissner, McKee, and Stuckler 2021). This evidence should be read with care: it concerns spending cuts enacted by governments rather than council vetoes, and effect sizes vary across settings. Still, it tells against the assumption that fiscal shortfalls of the size calibrated in (a) are politically inert.

(d) Self-correction and prosperity

The dynamic-mitigation reply concedes the mechanism and the gradient, but argues that the realized cost is smaller than the static calculation of (a) suggests. Two channels narrow the gap between the static cost and the realized one. The first is the council's own prudence: a competent council might notice the destabilization risk and exercise the veto less aggressively in response. The second is prosperity: epistocratic governance might deliver enough widely shared growth to soften the demand for redistribution that drives the commitment problem. The two channels are independent and worth taking separately.

Self-correction. A competent council might notice destabilization risk and exercise the veto less aggressively in response. There is a real channel here, but it has a structure worth bringing out. The

case for installing the veto rests on the council's preferences on redistribution differing systematically from the median voter's. Granting this empirical premise — granted throughout, and sized in (a) — the same preference gap that gives the council its raison d'être also generates the commitment cost: when the council vetoes, the realized tax rate falls below what the median voter wants. Self-correction in response to destabilization risk is therefore also self-correction against the gap the veto was designed to act on.

Two limit cases bring this out. At one limit, the council exercises full self-restraint — the veto is never used on redistributive legislation — and the realized outcome is just what the median voter would have produced anyway. At that limit, the veto plays no role, and the original case for installing it has gone with the self-correction. At the other limit, the council exercises no self-restraint and produces the full static commitment cost. The dynamic cost sits between the two limits, and the epistocrat has to accept either some commitment cost or some erosion of the case for the veto.

Two further considerations are worth noting. First, an endogenous-drift dynamic erodes self-correction over time: as the council vetoes redistributive legislation, its composition shifts toward members whose preferences justify those past vetoes, and the gap widens rather than shrinking. Second, a related version of the response says the council might produce more redistribution than the median voter, because it understands implementation better. Two things tell against this. The gradient the epistocrat's own case asserts goes the wrong way for it: the divergence motivating the veto runs toward less redistribution, not more (§3.2). And even granting the implementation advantage, the AR mechanism turns on the realized tax rate rather than on implementation skill — a more efficient council producing the same realized rate gives the poor the same flow value.

Prosperity. The second channel works through growth rather than through council restraint. The thought is that epistocratic governance might deliver enough widely shared prosperity that the poor are absolutely better off than they would be under democracy's higher redistribution rate. If so, growth substitutes for redistribution and the commitment problem dissolves. The cleanest response is a scope clarification rather than a refutation. The paper's argument is conditional on the AR scope conditions: mass mobilization on distributional grievance must be feasible, elites must perceive expropriation risk, and distributional grievance must be salient. The §3.3 liveness criterion identifies the polities where these conditions hold. In a polity where prosperity has dissolved them, the AR mechanism does not apply at all, and the paper is silent by design. Within polities where the conditions still hold, the gap between the council's preferred rate and the median voter's preferred rate still bites on the poor's flow value — even where overall incomes are rising.

Whether epistocracy in fact delivers widely shared prosperity is a further empirical question. The comparative growth literature is contested. Acemoglu, Naidu, Restrepo, and Robinson (2019) find that democracy itself causally raises long-run growth — which cuts against the technocratic argument that experts produce better policy. And distributional perception adjusts slowly: the backsliding literatures cited in §3.3 document distributional grievance persisting through periods of absolute growth, particularly when the growth is perceived as concentrated among elites. The prosperity response needs three things to hold together: actual growth, broad distribution of that growth, and perception by the poor that the growth is broadly distributed. Where any of these fails, the response does not get off the ground. Where all three hold together, the polity in question probably no longer meets the AR scope conditions in the first place — and the paper's argument is silent in such polities by design.

(e) Summary

In sum, the magnitude question is hard to answer definitively: the key preference divergence is

difficult to measure directly, and the comparisons with ordinary democratic politics are structural rather than precisely quantified. Still, there is a tentative case for a significant effect. Plausible calibrations put the veto's impact at a few percentage points of national income sustained indefinitely; the effect would be systematic, uncorrectable by election, and easy to attribute, unlike ordinary democratic disappointments; fiscal changes of comparable size have measurably increased political unrest; and the forces that might shrink the cost — council self-restraint, or prosperity — operate only by weakening the case for the veto or exiting the argument's scope. So while the exact size of the commitment cost remains an open empirical question, it cannot simply be priced at zero — and whether epistocracy's benefits outweigh it, alongside the other objections this paper brackets, is a further question the paper does not settle.

5. Situating the Critique

The commitment argument is an additional line in the critique of epistocracy. I want to locate it relative to the existing lines, and take up one empirical challenge directly.

Legitimacy (Estlund). Estlund's Qualified Acceptability Requirement is a legitimacy test. For a regime to be legitimately authoritative over its citizens, its justification has to be acceptable to all reasonable views — not in the sense that everyone has to endorse it, but in the sense that no reasonable view can be asked to accept a justification that singles it out as cognitively inferior. Regimes defended on invidious epistemic grounds — regimes whose justifications rest on marking some citizens as knowing better than others — fail this test in a way universal suffrage does not (Estlund 2008, pp. 33–37). My argument works in a different register. It does not say epistocracy fails the legitimacy test (Estlund says that); it says that even granting epistocracy whatever legitimacy the epistocrat thinks it has, the consequences are worse than Brennan takes them to be.

Epistemic (Landemore, Gunn, Méndez, Umbers). The epistemic critique targets input-quality: how electorate composition and deliberative procedure affect aggregate decision accuracy. The commitment argument targets institutional-incentive structure: how the allocation of *de jure* authority shapes elite behavior under credible-threat conditions. The two lines go through independently, and the commitment argument survives whether the epistocrat's defenders (Gibbons 2022) or critics (Umbers 2020) prevail on the epistemic question. Landemore (2020) proposes open democracy — a deliberative-representative model built around open mini-publics using sortition — as a constructive alternative to electoral democracy. Because open democracy does not impose competence filtering on the sampled body, it does not install the $\tau_C < \tau^p$ gradient the commitment argument turns on; it falls outside my target, alongside the unfiltered enfranchisement lottery discussed under Response 4.

Abuse (Bagg). Bagg (2018) is the closest prior work. Bagg argues that the franchise deters elite rigging of the epistocratic selection mechanism: without universal suffrage, elites face weaker accountability against designing competence tests and qualification criteria in their own favor. Gibbons (2025) notes that obstacles to capture remain even without electoral accountability — legal challenges to rigged designs, and opponents' capacity to adapt to them — though he grants the abuse worry is still serious. Bagg's argument and mine share the structural public-choice methodology. What distinguishes mine is that it isolates a failure mode that persists *under the cleanest possible selection.* Even if tests are pristine, certifications honest, and the council demographically representative on every dimension Bagg tracks, the commitment problem arises from the council's *institutional role* — the specific micro-foundation (AR's reduced form, tied to Proposition A.1) Bagg does not develop.

Democratic equality (Christiano, Kolodny, Viehoff, Guerrero; Pettit). A broader family of arguments focuses on what the franchise expresses about citizens' standing (Christiano 2008; Kolodny 2014;

Viehoff 2014), or on its role in securing non-domination (Pettit 2012). Guerrero's (2014) lottocracy and Landemore's open democracy belong to this register as constructive alternatives. My argument is orthogonal. Where standing and non-domination arguments target expressive and relational properties of the franchise, the commitment argument targets its structural function as a commitment device. In modern unequal societies the two lines converge — privilege and revolutionary-threat coalitions track each other closely — but they are analytically independent.

The Ansell–Samuels challenge. A distinct line of empirical work (Ansell and Samuels 2014; Ziblatt 2017; Albertus and Menaldo 2018) rejects or qualifies the Acemoglu–Robinson account of franchise extension on historical grounds. On these authors' readings, 19th- and 20th-century franchise extensions were driven by elite-competition dynamics between incumbent and rising economic groups, by the organizational capacity of conservative parties to compete under wider franchises, or by elite-dominated transition processes — not by redistributive bargains with a revolutionary poor. Haggard and Kaufman (2016) press a related critique of AR's inequality mechanism, though their work is more sympathetic to the mass-mobilization channel than the others in this group. Several prominent cases (the UK 1867 Second Reform Act; most post-1980 democratizations) do not fit the AR pattern.

The normative argument does not depend on the AR mechanism being historically dominant in 19th-century transitions. It depends on the mechanism being a live possibility now, in the polities where epistocracy is proposed. The liveness criterion of §3.3 supplies the diagnostic: where revolutionary-threat capacity, elite expropriation-risk perceptions, and distributional-grievance salience are measurably present, the AR mechanism is live and the commitment argument bites. Readers who accept the critics' historical revisions should read the paper as showing that the veto is risky where the AR mechanism is live, and as recommending caution proportional to its aliveness.

6. Upshots

The epistemic and commitment cases for democracy come apart. The epistocracy debate has run almost entirely in the epistemic register. Even if that debate were resolved in the epistocrat's favor, the commitment argument would weigh independently against the veto. The balance of considerations is therefore worse for the veto than the epistemic literature alone would suggest. And the argument does not depend on a particular resolution of the epistemic debate; it survives whether the epistocrat is right or wrong about the link between competence and policy quality.

The policy implication is narrow. The argument bears on competence-weighted regimes with substantive authority over redistributive legislation — institutions that substitute expert judgment for democratic redistributive authority on substantive-competence grounds. It does not bear on expert advising, technical regulation, rights-based judicial review, mandate-bounded independent agencies, unfiltered sortition, or competence-weighted bodies whose authority excludes redistribution. Contemporary proposals in view include Brennan's veto on the institutionally realistic reading; proposals for technocratic fiscal authorities of the sort occasionally floated in response to backsliding concerns; and, historically, the Lochner-era Supreme Court's substantive-due-process review of social legislation. The argument applies to each. It is silent on the rest of the administrative state.

The argument also extends beyond the veto to Brennan's other epistocratic proposals — restricted suffrage, plural voting, and the *filtered* enfranchisement lottery — each of which installs a competence-weighted redistributive authority that maps to the same AR reduction (Appendix §A.4). The one variant the argument does not reach is the *unfiltered* enfranchisement lottery, which is not an epistocracy in the relevant sense. On the institutionally realistic reading, then, the commitment

critique weighs against Brennan's epistocratic menu as a whole, not only the veto.

One extension is worth flagging. The Acemoglu–Robinson mechanism is set up around redistributive bargains, but the structural logic — that decisive authority shifts what elites can credibly commit to under mass-mobilization pressure — is not specific to redistribution. Recent right-wing populist mobilization in the U.S. and Europe has organized around ethnic, nationalist, and cultural grievances at least as much as economic ones (Mudde 2007; Norris and Inglehart 2019). An epistocratic council whose competence filter correlates with cosmopolitan and culturally-liberal preferences could install a similar wedge on these issues, generating commitment failures that mirror the redistributive case. The normative valence here is more complex than in the redistributive case — illiberal majoritarian preferences may be exactly what one wants epistocratic constraints to check — but the structural mechanism operates regardless of how that tension is resolved. The formal extension is beyond the present paper; the diagnostic extends to it.

The argument does not show that democracy is substantively optimal. It shows that democracy has a commitment function that the institutionally realistic veto cannot replicate. It does not show that there is no place for expert advice in democratic decision-making. It shows that expert advice and competence-based override are different institutional roles. And it does not show that real-world democracies are perfectly committing — commitment is continuous, real democracies have varying χ, and the argument's conclusion is directional. What it does is add a consideration to the ledger on democracy's side, and leave the final balance to the weighing.[5]

Appendix A: Formal Argument

The argument is an application of Acemoglu and Robinson (2006) plus one observation about Jensen's inequality. This appendix states the result precisely and shows how it follows.

A.1. The Acemoglu–Robinson setup

Following Acemoglu and Robinson (2006, Ch. 4 §§4.1–4.2 and Ch. 6 §§5–11). A population of mass 1 contains fraction $\delta < 1/2$ rich and fraction $1-\delta > 1/2$ poor, with mean income $\bar{y}$. The rich each earn $y_r = \theta\bar{y}/\delta$, the poor each earn $y_p = (1-\theta)\bar{y}/(1-\delta)$, and $\theta \in (\delta, 1)$ indexes inequality (higher θ means a larger share of total income accrues to the rich). A linear tax $\tau \in [0,1]$ funds a lump-sum transfer $T = (\tau - C(\tau))\bar{y}$, where C is the deadweight-loss function (the efficiency cost of taxation), strictly convex with $C(0) = C'(0) = 0$. Indirect utility is $V^i(\tau) = (1-\tau)y_i + (\tau - C(\tau))\bar{y}$ for $i \in \{p, r\}$. Strict convexity of C makes V^p strictly concave with interior maximum τ^p, the poor's preferred rate — a fact that does real work in §A.2. The rich prefer $\tau^r = 0$.

The state $\mu_t \in \{\mu_L, \mu_H\}$ tracks whether a revolutionary moment is available in period t, evolving as a Markov chain: in the non-revolutionary state μ_L the poor cannot credibly threaten mass action, and in the revolutionary state μ_H they can seize $(1-\mu_H)\bar{y}/(1-\delta)$ per period in perpetuity. Revolution strictly dominates the autocratic flow iff $\theta > \mu_H$. Discount factor $\beta \in (0,1)$ weights future periods. Let $V_p(R, \mu_H) = (1-\mu_H)\bar{y}/[(1-\delta)(1-\beta)]$ denote the poor's revolution value and $V_p(D) = V^p(\tau^p)/(1-\beta)$ their value under full democracy.

Acemoglu and Robinson summarize imperfect democracies by a single reduced-form parameter $\chi \in [0,1]$, with full democracy at $\chi = 0$ and elite rule at $\chi = 1$. In their model the limited-democracy tax $\tau(\chi)$ is defined implicitly by a first-order condition (Acemoglu and Robinson 2006, eq. 4.16) and

[5] This paper was drafted with extensive help from Claude Opus 4.7. This approach to research reflects the idea that our job as philosophers is to discover truths, using the best tools that we can.

is strictly decreasing in χ, with $\tau(0) = \tau^p$ and $\tau(1) = \tau^r = 0$. For tractability I adopt the linear normalization $\tau(\chi) = (1-\chi)\tau^p$, which agrees with AR's $\tau(\chi)$ at both endpoints and preserves its strict monotonicity; the results below use only endpoint agreement, monotonicity, and the concavity of V^p, not linearity. Write $V_p(\chi) = V^p((1-\chi)\tau^p)/(1-\beta)$ for the poor's value under the reduced form at χ. Since V^p is strictly concave with maximum at τ^p, $V_p(\chi)$ is strictly decreasing in χ with $V_p(0) = V_p(D)$. The revolution-averted region $\{(\theta, \mu_H) : V_p(\chi) \geq V_p(R, \mu_H)\}$ therefore strictly shrinks as χ rises. This is the result the veto regime inherits.

A.2. The veto regime

Definition. An *epistocratic veto regime* E_v consists of a universally-enfranchised legislature that proposes τ^p each period and an expert council that vetoes the legislated rate with probability $v \in (0,1)$ independently each period, substituting its preferred rate $\tau_C \in [0, \tau^p)$. The period-t realized tax is the lottery $\tilde{\tau}_t$, equal to τ^p with probability $1-v$ and τ_C with probability v, drawn independently of the revolutionary-moment process μ_t. (The boundary case $v = 1$, in which the council always vetoes, is a deterministic reduction to AR's reduced form at $\chi = (\tau^p - \tau_C)/\tau^p$ and is covered by AR directly.) The parameters (v, τ_C) are treated as committed primitives of the regime; endogenizing them is taken up in §A.4. Let $V_p(E_v) = \mathbb{E}[V^p(\tilde{\tau})]/(1-\beta)$. Define

$$\chi_{E_v} := v\,(\tau^p - \tau_C)/\tau^p \in (0,1).$$

The expected tax is $\mathbb{E}[\tilde{\tau}] = (1-\chi_{E_v})\tau^p$, matching AR's reduced form at $\chi = \chi_{E_v}$ on the mean dimension. (Under AR's own implicitly-defined $\tau(\chi)$, a mean-matched χ exists and is unique by strict monotonicity; the linear normalization of §A.1 makes it χ_{E_v} in closed form. Nothing in the argument turns on the choice.)

One wrinkle distinguishes the veto from AR's reduced form: the veto delivers a lottery, not a deterministic tax. Since V^p is strictly concave, Jensen's inequality gives

$$V_p(E_v) = \mathbb{E}[V^p(\tilde{\tau})]/(1-\beta) < V^p(\mathbb{E}\tilde{\tau})/(1-\beta) = V_p(\chi_{E_v}).$$

So the veto is strictly worse for the poor than AR's reduced form at the same mean-matched χ. The mapping is conservative in this specific sense: on the mean-comparison dimension, the veto regime is at least as bad as AR's reduced form at χ_{E_v}. Composed with the monotonicity of $V_p(\chi)$ derived in §A.1, this gives the chain $V_p(E_v) < V_p(\chi_{E_v}) < V_p(D)$ for any $\chi_{E_v} > 0$.

A.3. The commitment result

Let $R(D) = \{(\theta, \mu_H) : \theta > \mu_H \text{ and } V_p(D) \geq V_p(R, \mu_H)\}$ denote the parameter region in which full democracy averts revolution, and let $R(E_v)$ denote the analogous region for the veto regime. Under AR's maintained assumptions, $R(D)$ has nonempty interior on the AR-standard case in which democracy offers redistribution credible enough to deter revolution with slack — i.e., $V_p(D) > V_p(R, \mu_H)$ strictly on a non-empty open subset of $\{\theta > \mu_H\}$.

Proposition A.1. Assume AR's revolutionary-moment conditions: $\theta > \mu_H$ and the poor are the median voter under democracy. For any $\chi_{E_v} \in (0,1)$ and any $(\theta, \mu_H) \in R(D)$,

$$V_p(E_v) < V_p(\chi_{E_v}) < V_p(D).$$

Hence $R(E_v) \subsetneq R(D)$, and the difference set $R(D) \setminus R(E_v)$ has nonempty interior.

Proof. The first inequality is Jensen (§A.2); the second is the monotonicity derived in §A.1. Composing, $V_p(E_v) < V_p(D)$. On the interior of $R(D)$, the democratic margin $V_p(D) - V_p(R, \mu_H)$ is strictly positive; continuity of $V_p(D)$ and $V_p(E_v)$ in θ and of $V_p(R, \mu_H)$ in μ_H therefore yields an open subregion on which $V_p(D) - V_p(R, \mu_H) < V_p(D) - V_p(E_v)$, i.e., on which $V_p(E_v) < V_p(R, \mu_H)$. Every such point lies in $R(D) \setminus R(E_v)$. □

In plain English: wherever democracy averts revolution with positive margin narrower than the Jensen+χ wedge the veto introduces, the veto destabilizes a regime democracy would have stabilized. The destabilized region is not a knife-edge; it has interior. This is the whole formal argument. Everything in §A.4 is a remark.

A.4. Remarks

On endogenizing v. In the main model v is a primitive. If v is chosen endogenously, the natural question is whether the elite or the council can always avert revolution by picking v small enough — as $v \to 0$, $V_p(E_v) \to V_p(D)$ and the veto regime collapses back to democracy. The answer is yes in principle but degenerately: a veto exercised with vanishing frequency is a ceremonial body, not the expert-review institution the epistocratic case requires. Once v is bounded below by whatever minimum frequency the institution needs to do its epistemic work, Proposition A.1 applies and the destabilized region is non-empty. The commitment problem does not dissolve under endogenization; it re-emerges at any v the epistocrat would actually endorse.

On variance. Because the veto delivers a lottery, mean-preserving spreads in $\tilde{\tau}$ strictly reduce the poor's flow value. A high-variance per-period-independent veto regime is strictly worse than a low-variance persistent one at the same mean. The Proposition's conclusion uses the per-period-independent baseline, which is the friendliest reading of Brennan's proposal on this axis; realistic councils with slow compositional turnover sit between the independent and fully-persistent cases.

On repression costs. AR allow the elite to suppress revolution at cost $\kappa > 0$. Repression enters the elite's problem, not the poor's regime-comparison condition; the strict inequality $V_p(E_v) < V_p(D)$ that drives Proposition A.1 is unaffected. Repression can, however, enlarge the region in which the elite prefers suppression to redistribution; in that region the commitment problem is moot because the elite is not attempting commitment. The argument's bite is on polities where repression is not a cheap option, which is AR's own scope condition.

On other Brennan regimes. The same reduction applies to Brennan's three other epistocratic regimes.

(a) *Restricted suffrage* maps to a χ that rises as franchise restrictions narrow the electorate: the pivotal (decisive) voter shifts continuously toward higher-income voters as the franchise cut tightens, with χ close to zero at near-universal franchise and close to one at narrow elite franchise.

(b) *Plural voting.* Weighting a subset of the electorate at $k > 1$ shifts the pivotal-voter locus toward that subset's preferred rate; the resulting χ scales with the size and weighting of the privileged subset and with the preference gap between the subset and the median.

(c) *Enfranchisement lottery.* A lottery with filtering rate σ positively correlated with income gives an effective χ that scales with σ and with the preference gap $(\tau^p - \tau_{\text{survivors}})/\tau^p$ between filtered survivors and the population median. An orthogonal filter gives $\chi = 0$; an unfiltered lottery

gives expected $\chi = 0$, though realized χ fluctuates sample-to-sample and a Jensen-type wrinkle parallel to §A.2 applies to the resulting variance. Either way the commitment argument is silent on unfiltered sortition.

Enlightened preference voting. Brennan (2016, chs. 7–8) discusses a scheme on which actual votes are corrected for what voters would prefer if better informed — what Brennan and Landemore (2022) call Rule by Simulated Oracle. To the extent the counterfactual correction reproduces the information–preference gradient the epistocrat's case asserts (§3.2), the scheme installs the well-informed quartile's preferred rate in place of the median voter's, and the resulting gap scales with the magnitude of the correction. The commitment argument applies. An orthogonal correction — one whose informational criteria are uncorrelated with redistributive preference — would escape it, but the empirical literature has not identified such a correction, just as for the orthogonal-filter escape of §4 Response 2.

On scope. The analysis is stationary within each μ-regime; transition dynamics across regimes are not modeled. Within-poor heterogeneity is suppressed (AR's Ch. 8 multi-class extensions apply). The abuse-clean idealization — honest selection, competent council — strengthens the commitment critique by ruling out self-dealing as a confounder.

References

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy.* Cambridge: Cambridge University Press.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy* 127 (1): 47–100.

Albertus, Michael, and Victor Menaldo. 2018. *Authoritarianism and the Elite Origins of Democracy.* Cambridge: Cambridge University Press.

Althaus, Scott L. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People.* Cambridge: Cambridge University Press.

Ansell, Ben W., and David J. Samuels. 2014. *Inequality and Democratization: An Elite-Competition Approach.* Cambridge: Cambridge University Press.

Bagg, Samuel. 2018. "The Power of the Multitude: Answering Epistemic Challenges to Democracy." *American Political Science Review* 112 (4): 891–904.

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

Boix, Carles. 2003. *Democracy and Redistribution.* Cambridge: Cambridge University Press.

Branham, J. Alexander, Stuart N. Soroka, and Christopher Wlezien. 2017. "When Do the Rich Win?" *Political Science Quarterly* 132 (1): 43–62.

Brennan, Jason. 2016. *Against Democracy.* Princeton: Princeton University Press.

Brennan, Jason. 2018. "Does the Demographic Objection to Epistocracy Succeed?" *Res Publica* 24 (1): 53–71.

Brennan, Jason, and Hélène Landemore. 2022. *Debating Democracy: Do We Need More or Less?* New York: Oxford University Press.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson, and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge, MA: MIT Press.

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

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *Quarterly Journal of Economics* 134 (3): 1405–1454.

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

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco." *American Economic Review* 109 (9): 3365–3394.

Dixit, Avinash K. 2003. "Trade Expansion and Contract Enforcement." *Journal of Political Economy* 111 (6): 1293–1317.

Enns, Peter K., and Christopher Wlezien, eds. 2011. *Who Gets Represented?* New York: Russell Sage Foundation.

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

Fetzer, Thiemo. 2019. "Did Austerity Cause Brexit?" *American Economic Review* 109 (11): 3849–3886.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2023. "Populist Leaders and the Economy." *American Economic Review* 113 (12): 3249–3288.

Fumagalli, Corrado. 2022. "Stability in Liberal Epistocracies." *Social Epistemology* 37 (1): 97–109.

Galofré-Vilà, Gregori, Christopher M. Meissner, Martin McKee, and David Stuckler. 2021. "Austerity and the Rise of the Nazi Party." *Journal of Economic History* 81 (1): 81–113.

Gibbons, Adam F. 2022. "Is Epistocracy Irrational?" *Journal of Ethics and Social Philosophy* 21 (2): 251–281.

Gibbons, Adam F. 2025. "Epistocracy and the Problem of Political Capture." *Public Affairs Quarterly* 39 (1): 19–42.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.

Guerrero, Alexander A. 2014. "Against Elections: The Lottocratic Alternative." *Philosophy & Public Affairs* 42 (2): 135–178.

Gunn, Paul. 2019. "Against Epistocracy." *Critical Review* 31 (1): 26–82.

Guriev, Sergei, and Elias Papaioannou. 2022. "The Political Economy of Populism." *Journal of Economic Literature* 60 (3): 753–832.

Haggard, Stephan, and Robert R. Kaufman. 2016. *Dictators and Democrats: Masses, Elites, and Regime Change*. Princeton: Princeton University Press.

Hirschl, Ran. 2004. *Towards Juristocracy: The Origins and Consequences of the New Constitutionalism*. Cambridge, MA: Harvard University Press.

Iversen, Torben, and David Soskice. 2006. "Electoral Institutions and the Politics of Coalitions: Why Some Democracies Redistribute More than Others." *American Political Science Review* 100 (2): 165–181.

Iversen, Torben, and David Soskice. 2019. *Democracy and Prosperity: Reinventing Capitalism through a Turbulent Century*. Princeton: Princeton University Press.

Kenworthy, Lane, and Jonas Pontusson. 2005. "Rising Inequality and the Politics of Redistribution in Affluent Countries." *Perspectives on Politics* 3 (3): 449–471.

Klocksiem, Justin. 2019. "Epistocracy Is a Wolf in Wolf's Clothing." *Journal of Ethics* 23 (1): 19–36.

Kolodny, Niko. 2014. "Rule Over None II: Social Equality and the Justification of Democracy." *Philosophy & Public Affairs* 42 (4): 287–336.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105 (4): 1478–1508.

Landa, Dimitri, and Ryan Pevnick. 2025. "Political Meritocracy and Democracy." *Annual Review of Political Science* 28: 253–271.

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

Landemore, Hélène. 2020. *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*. Princeton: Princeton University Press.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die*. New York: Crown.

Lizzeri, Alessandro, and Nicola Persico. 2004. "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform.' " *Quarterly Journal of Economics* 119 (2): 707–765.

López-Guerra, Claudio. 2014. *Democracy and Disenfranchisement: The Morality of Electoral Exclusions*. Oxford: Oxford University Press.

Meltzer, Allan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89 (5): 914–927.

Méndez, María Pía. 2022. "An Epistemic Problem for Epistocracy." *Social Epistemology* 36 (2): 153–166.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25 (3): 173–196.

Mudde, Cas. 2007. *Populist Radical Right Parties in Europe*. Cambridge: Cambridge University Press.

Neumark, David, and William L. Wascher. 2008. *Minimum Wages*. Cambridge, MA: MIT Press.

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge: Cambridge University Press.

Pettit, Philip. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.

Ponticelli, Jacopo, and Hans-Joachim Voth. 2020. "Austerity and Anarchy: Budget Cuts and Social Unrest in Europe, 1919–2008." *Journal of Comparative Economics* 48 (1): 1–19.

Rodrik, Dani. 2018. "Populism and the Economics of Globalization." *Journal of International Business Policy* 1 (1–2): 12–33.

Scheve, Kenneth, and David Stasavage. 2016. *Taxing the Rich: A History of Fiscal Fairness in the United States and Europe*. Princeton: Princeton University Press.

Shapiro, Martin. 1981. *Courts: A Comparative and Political Analysis*. Chicago: University of Chicago Press.

Somin, Ilya. 2022. "The Promise and Peril of Epistocracy." *Inquiry* 65 (1): 27–34.

Soroka, Stuart N., and Christopher Wlezien. 2010. *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge: Cambridge University Press.

Talisse, Robert B. 2022. "The Trouble with Hooligans." *Inquiry* 65 (1): 15–26.

Tushnet, Mark. 2003. "Alternative Forms of Judicial Review." *Michigan Law Review* 101 (8): 2781–2802.

Umbers, Lachlan Montgomery. 2020. "Democratic Legitimacy and the Competence Objection." *Res Publica* 26 (2): 283–293.

Vandamme, Pierre-Étienne. 2020. "What's Wrong with an Epistocratic Council?" *Politics* 40 (1): 90–105.

Viehoff, Daniel. 2014. "Democratic Equality and Political Authority." *Philosophy & Public Affairs* 42 (4): 337–375.

Waldron, Jeremy. 2006. "The Core of the Case against Judicial Review." *Yale Law Journal* 115 (6): 1346–1406.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Wong, Baldwin. 2025. "Trapped Between Anger and Apathy: On the Problem of Instability in Confucian Meritocracy." *Journal of Politics* 87 (1): 205–215.

Ziblatt, Daniel. 2017. *Conservative Parties and the Birth of Democracy*. Cambridge: Cambridge University Press.