Referenda and Partial Commitment in Policy-Making [*]

Carl Heese [†]

Referenda are among the most prominent instruments of collective decision-making. Many do not directly select a policy but rather reshape the space of policies available. We develop a theory of such referenda and related decision processes. We model them as two-stage processes in which a large number of privately informed agents each answer a binary question, and their collective response constrains the options from which a principal must then choose. Our analysis singles out referenda in which the collective acts as a gatekeeper that authorizes or blocks an extreme policy while otherwise delegating choice to the principal. These referenda maximize the principal's payoff guarantee (the worst-case payoff across equilibria and information structures) within the broad class of processes considered; when the policy space is fine, they robustly yield near-full-information outcomes. Our results provide a rationale for referenda used in practice and highlight commitment as an important dimension of political institutions.

Many referenda and other collective decision processes, such as polls, serve not to specify a single policy to be implemented, but rather to reshape the space of policies available—opening the door to certain actions, blocking others. The final decision is left to a principal, who is free to select any policy within the constraints imposed by the agents.

For example, shareholders in a company may vote in a *cap referendum* to set a ceiling on executive pay, while leaving the design of the final pay package to the board. Similarly, in local public finance, citizens can vote on a cap override to determine whether officials may exceed a statutory spending limit, but leave the details of the budget to the officials. In other contexts, *gateway referenda* may trigger a departure from the status quo without committing to a specific alternative, as in Italy's abrogative referenda, which repeal an existing law (or part of it) without specifying the replacement, or as in constitutional referenda authorizing revision of a country's constitution.[1]

---

[*] May 19, 2026. I thank Matthias Fahn, Stephan Lauermann, Wolfgang Pesendorfer, Wing Suen, Balázs Szentes, Marciano Siniscalchi, and audiences at the University of Vienna, the University of Hong Kong, NYU Shanghai, the Chinese University of Hong Kong, Hong Kong University of Science and Technology, the East Asia Theory Conference, and the 2025 CityU Microeconomics Conference for their valuable comments and suggestions. This work was supported by a grant from the University of Hong Kong (Projects 102010268 and 109001744) and from Hong Kong's Research Grant Council (ECS 27505725).

[†] The University of Hong Kong, Department of Economics, heese@hku.hk

This paper develops a theory of such referenda and of related processes that share this partial-commitment structure. In a *referendum*, each of a group of agents takes a binary action (e.g., casts a yes/no vote); the principal observes the collective action (the average m of the individual actions); and m determines a set $R(m)$ of feasible policies from which the principal then chooses. For instance, in the shareholder example, $R(m)$ excludes all pay packages above the cap unless m meets a certain cutoff.

The central feature of the theory is *partial commitment*: $R(m)$ always contains more than one policy, so that the collective action constrains but does not specify the outcome.[2] We allow the mapping R to be otherwise arbitrary. This shifts attention away from the familiar polar cases of full-commitment voting rules and non-binding cheap talk and toward the general question of how commitment affects policy outcomes.[3]

Our analysis provides a rationale for the use of minimal-commitment processes—those in which the collective acts as a gatekeeper for a single policy, being able to block or authorize only this policy. In the environments we study, minimal commitment plays a dual role: It sustains robust aggregation of the agents' private information while minimizing the principal's exposure to their coordination problems. As a result, the principal's worst-case outcomes are optimized. Gateway referenda and certain cap referenda emerge as the leading examples of minimal-commitment processes.

The baseline model considers a linear policy environment with common preferences under full information. A principal must choose a policy level x from a finite, ordered policy set. The policy has a common linear cost $c \cdot x$ and a common linear benefit $\omega \cdot x$ that depends on an unknown binary state ($\omega \in \{0, 1\}$).[4] The principal would like to condition her policy choice on the state (choosing high x when the state is 1 and low x when the state is 0), but she does not observe the state. She therefore consults a population of N agents (where N is large) who hold private information in the form of noisy (conditionally independent) signals and heterogeneous prior beliefs. The agents participate in a referendum which results in a set of feasible policies for the principal to choose from.

Our main results compare referenda in terms of their robustness. Our yardstick of comparison is the principal's *payoff guarantee*: her worst-case payoff across all equilibria

---

[1] The same logic characterizes city-level votes to initiate planning processes and neighborhood referenda to authorize redevelopment, for example, in the United Kingdom.

[2] We also study *generalized referenda*, which allow for commitment to either the status quo or a non-singleton set of policies. These decision processes formalize gateway referenda.

[3] Classic work on the Condorcet jury theorem studies voting rules that commit to a fully specified outcome as a function of votes, whereas cheap talk models assume no commitment. See, e.g., Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1997), Bhattacharya (2018), and Krishna and Morgan (2012) for seminal contributions to the Condorcet jury theorem literature, and Wolinsky (2002), Morgan and Stocken (2008), Battaglini (2017), and Levit and Malenko (2011) for the cheap talk literature.

[4] In extensions, we cover non-linear preferences and allow the principal and the individuals to have conflicting preferences under full information. Linearity is a standard assumption; see, for example, the classic public-goods formulation with linear cost and benefit in Ledyard and Palfrey (1999). It also has a motivation in terms of stochastic commitment; see Section 5.5.

and all admissible distributions of priors and signals.

Across our broad class of referenda, we obtain a sharp performance ranking, led by two simple classes of procedures: *collective vetoes of the maximum* (which are a type of cap referendum) and *gateway referenda*. Their common feature is minimal commitment, with the collective acting as a gatekeeper for one polar option—either the maximal policy or the status quo—while deliberately leaving the choice between the remaining alternatives to the principal.

Under a collective veto of the maximum, the maximal policy is excluded if m falls short of a given cutoff; otherwise, the principal has full discretion. We show that collective vetoes of the maximum maximize the principal's payoff guarantee. When the policy space is fine, they even yield near-full-information outcomes in every equilibrium (Theorem 1).

A gateway referendum enforces the status quo if the collective action falls short of a certain cutoff; otherwise, it excludes the status quo but leaves the principal full discretion in choosing the alternative. To encompass these processes, we generalize our definition of referenda to allow singleton commitment to the status quo ($R(m) = \{x_1\}$) while maintaining non-singleton policy sets $R(m)$ otherwise. Gateway referenda satisfy a theorem analogous to Theorem 1: The gateway referendum with the simple-majority cutoff $m_1 = \frac{1}{2}$ maximizes the principal's payoff guarantee among all generalized referenda, and it robustly yields near-full-information outcomes when the policy space is fine (Theorem 2).

The main results turn on the way commitment shapes both information transmission and the principal's exposure to coordination problems of the agents.

First, coordination becomes difficult under partial commitment since each agent faces a strategic environment in which he has two incentives. His action transmits information and thereby affects the principal's posterior beliefs; but it can also influence the set of feasible policies. The interaction of these two incentives gives rise to pervasive *coordination failures*, that is, to inefficient equilibria in which the agents miscoordinate on how to use their collective influence.

These coordination failures are not knife-edge phenomena. They arise for *all* referenda and across environments (Propositions 1 and 3). They arise when the principal and the agents have broadly aligned beliefs, but also when the principal is perceived as biased: that is, when the agents believe that the principal's prior is systematically too conservative, or too aggressive, relative to the beliefs held by a large part of the population. Consider, for instance, the aligned case. Here, one might expect that informative behavior leads to efficient outcomes. We show that this need not be the case: Near-truthful behavior often arises in equilibrium, with almost all agents matching actions to signals; however, such equilibria can be inefficient, because the collective action may rule out the ex-post optimal policy.

The logic is easiest to see in a simple example. Suppose that in state 1 the highest policy is ex-post optimal, and in state 0 the lowest policy is. Suppose also that the referendum restricts the principal to high policies if the collective action m exceeds a certain large cutoff—say, 85%—and to low policies otherwise. If the agents' private information is informative but not too precise, then even in state 1 only a moderate majority of agents (say, around 60%) will receive a signal favoring high policies, so m misses the cutoff with probability approaching one as $N \to \infty$. Thus, even though the collective action is informative enough for the principal to learn the state (as $N \to \infty$), the induced constraint prevents her from acting on what she learns.

Ex-post efficient equilibria often also exist (e.g., in the 85% example), but they rely on more delicate coordination than near-truthful behavior. The inefficient equilibria therefore reveal genuine failures to coordinate efficiently, not merely a limitation of the feasible policy correspondence. This is precisely the kind of coordination failure that motivates a worst-case analysis. The approach is also natural in light of the information-aggregation literature, where coordination failures are a recurring concern (see, e.g., Ekmekci and Lauermann, 2020, and Ali, Mihm and Siga, 2025).

Second, commitment shapes information transmission. With no commitment, actions are pure cheap talk, and babbling equilibria without any information transmission exist. In related settings, no-commitment processes often have *only* babbling equilibria (see, e.g., Battaglini, 2017). In contrast, we show that many partial-commitment referenda are instead robustly informative. The principal learns the state from the collective action along *every* equilibrium sequence, uniformly across a broad class of admissible information structures. We provide a sharp characterization of when this robust information aggregation obtains for an important class—monotone referenda with a single cutoff (Proposition 2 and Section 4.2). This result will imply that very limited commitment can be sufficient for learning, making the use of minimal-commitment procedures viable.

A collective veto of the maximum optimally balances the tension between information transmission and exposure to the agents' coordination failures via minimal commitment. The existence of *some* commitment makes the collective action consequential and robustly sustains learning (as Proposition 2 will show). On the other hand, the process leaves the principal almost full discretion, thereby minimizing exposure to coordination failures. At worst, coordination failures may rule out the maximum; when the policy grid is fine, this yields only a negligible loss.

The optimality of the gateway referendum follows a similar but more subtle logic, also based on the agents' coordination problems.

In the second part of the paper, we discuss several extensions of our baseline model and results. In Section 5.2 we provide conditions under which versions of our optimality results persist under a weaker robustness criterion that compares referenda in terms of the worst-case payoff only across equilibria for a fixed information structure. In Sections 5.3 and 5.4, we show that the core logic of our results survives when we relax the assumptions on the players' preferences. In Section 5.3, we consider a preference aggregation problem, in which groups of agents may have opposing preferences depending on the state.[5] This allows the principal's preferences to conflict with those of some agents even under full information, thereby aggravating the information transmission problem. In Section 5.4, we relax the assumption of linear costs and benefits, instead considering preferences that are single-plateau over the policy space. In Section 5.5, we show that our baseline model can also be interpreted in terms of stochastic commitment.

The paper makes two main contributions. First, it develops a theory of referenda and related collective decision processes as institutions of partial commitment. This theory yields a sharp lesson about the robustness of referenda: those referenda featuring minimal commitment may be particularly robust. Second, the logic of minimal commitment provides a rationale for two widely used procedures—cap referenda (including collective vetoes) and gateway referenda. In environments where information is dispersed across a large group but efficient collective action is difficult to coordinate, these referenda optimally navigate a central tension between information aggregation and exposure to coordination failures.[6]

Our model interpolates between the two polar cases of full commitment and no commitment, which have been used in the literature to study formal voting rules on the one hand (e.g. Feddersen and Pesendorfer, 1997) and informal political processes such as protests on the other (Battaglini, 2017). In particular, our analysis suggests that coordination problems may be more germane to policy-making than the full-commitment benchmark implies. In Section 6 we discuss in detail the literature and benchmark results on commitment, and in Section 7 we survey further related literature.

Although we study commitment in the context of referenda, the concept admits several broader interpretations. First, as in Battaglini (2017), commitment can be seen as a measure of how formal a political process is. From this perspective, our results imply that formality matters less for information aggregation than existing impossibility results might suggest (see Section 6). Second, commitment can be viewed as capturing how direct or indirect a democratic institution is. More commitment corresponds to less residual discretion for the principal and hence to a more direct institution; less commitment corresponds to more discretion and hence to a more indirect institution. Third, partial-commitment processes may be understood as modeling mandates: Collective support confers a degree of political authorization for subsequent action. Stronger support translates into a broader mandate—that is, a larger set of policies from which

---

[5] Models with such preferences have been used to study distributive politics, in which the state of the world determines which group will benefit from a policy. See Fernandez and Rodrik (1991), Ali *et al.* (2025), and Bhattacharya (2018).

[6] Our results are also relevant to current policy debates on democratic innovation (see, e.g., OECD, 2020) and to reforms aimed at expanding citizen participation through alternative institutions, such as the UK's Localism Act 2011, which includes neighborhood planning referenda resembling gateway referenda.

the principal may choose—while weaker support yields a narrower mandate.[7]

1 Model and Incentives

A policy x needs to be chosen from a finite set of options $\mathcal{P} = \{x_1, \dots, x_l\}$ with $x_1 < x_2 < \dots < x_l$. To simplify the algebra, we let $x_1 = 0$, $x_l = 1$, and $x_2 = 1 - x_{l-1} = \varepsilon > 0$. The policy has a common and constant marginal cost of $c = \frac{1}{2}$, and a common and constant marginal benefit given by an uncertain state $\omega \in \{0, 1\}$. Thus, each player's payoff from policy x in state ω is

$$x\left(\omega - \frac{1}{2}\right).$$

There is a principal who has a commonly known prior[8]

$$\frac{1}{2} < \Pr(\omega = 1) \leq 1 - \varepsilon.$$

In addition, there is a set of agents $\{1, \dots, N\}$ who hold private information. These agents are either *strategic* or *partisans*, as specified below:

- We call an agent a (non-strategic) *partisan* if he chooses a prescribed action regardless of the state. For each action $a \in \{0, 1\}$, the probability that a given agent is a partisan for action a is $0 < \rho_a < \frac{1}{2}$, independent of the state. The existence of partisans implies minimal noise and trembling-hand perfection (cf. the remark at the end of this section).[9] With the remaining probability $(1 - \rho_1 - \rho_0)$, each agent is strategic.

- Each strategic agent i receives a binary private signal $s_i \in \{0, 1\}$ drawn independently from a common distribution conditional on the state, and satisfying $0 < \Pr(s_i = 1 \mid \omega = 0) < \Pr(s_i = 1 \mid \omega = 1) < 1$. Thus signal 1 is an indication for state 1 and signal 0 an indication for state 0.

- Each strategic agent i holds a private prior belief $p_i \in [0, 1]$ about the likelihood of state 1; we refer to p_i simply as agent i's *type*. Types are drawn independently from a differentiable distribution F with support $[0, 1]$.

---

[7] The idea that officials have mandates to govern, and that these mandates are stronger for officials with greater support, has been explored before but without consideration of a strategic principal; see, e.g., Herrera, Llorente-Saguer and McMurray (2019) and Damiano, Li and Suen (2025).

[8] The right constraint ensures that the payoffs from an equilibrium without any information transmission cannot be arbitrarily close to the full-information payoffs. The left constraint is without loss. If instead $\varepsilon < \Pr(\omega = 1) < \frac{1}{2}$, our main results continue to hold with the appropriate modifications. For example, under that condition, a collective veto of the minimum maximizes the principal's payoff guarantee within the class of all referenda.

[9] This approach is standard in the existing work on information aggregation, which similarly considers agents with partisan preferences and restricts attention to equilibria in which they vote for their preferred policy (see, e.g., Feddersen and Pesendorfer, 1997, and Bhattacharya, 2013). In particular, it allows us to discuss existing benchmark results within our framework.

The distributions of the private signals and types (together with ρ_0 and ρ_1) constitute an *agents' information structure.* Given a *referendum* R (defined below), the timing is as follows:

1. Each agent i observes his private information and takes a binary action $a_i \in \{0,1\}$.
2. The principal observes the quantity $m = \frac{\sum_{i=1}^{N} a_i}{N}$, which we call the *collective action*, and chooses a policy $x \in \mathcal{P}$ subject to a constraint $R(m)$ determined by the referendum R.

A *referendum* is a left-continuous mapping from $[0,1]$ to the set of all non-singleton subsets of $\mathcal{P}$, with at most finitely many discontinuities, and none at 0. A referendum R thus maps a collective action m to a policy set $R(m)$, which we call the set of *feasible* policies given m.[10] Every referendum R takes the form of a step function; that is, there exist a finite number of cutoffs $0 < m_1 < \ldots < m_T < m_{T+1} = 1$ such that $R(m)$ is constant on $I_1 := [0, m_1]$ and on $I_j := (m_{j-1}, m_j]$ for $j = 2, \ldots, T+1$. We denote the constant value of R on the interval I_j by R_j.

Given a referendum R, a principal's pure strategy is a sequence $\Big(x(k)\Big)_{k=0,\ldots,N}$ mapping each possible realization $m = \frac{k}{N} \in [0,1]$ to a policy $x(k) \in \Big(R(m)\Big)$. We allow for randomization. A symmetric agents' strategy is a mapping $\sigma : [0,1] \times \{0,1\} \to [0,1]$, where $\sigma(p,s)$ represents the likelihood that a non-partisan agent with prior p and signal s chooses action 1. Strategy profiles are denoted by η. The analysis that follows focuses on sequences $(\eta_N)_{N\in\mathbb{N}}$ of weak perfect Bayesian equilibria in symmetric agents' strategies, letting the number N of the agents grow large. The presence of partisans implies trembling-hand perfection (Selten, 1988)—see the working paper version (Heese, 2026)—and that the mean action for each state is interior, i.e.,

$$q(\omega';\sigma) := \Big((1-\rho_1-\rho_0)\mathrm{E}\Big(\sigma(p,s) \mid \omega = \omega'\Big) + \rho_1\Big) \in (0,1) \text{ for } \omega' \in \{0,1\}.$$

1.1 Principal's Best Responses

The principal's best response can be described via a single cutoff, since it is driven by monotone Bayesian updating. Upon observing that k of the N agents have chosen action 1, the principal has the following posterior:[11]

$$\frac{\Pr(\omega = 1 \mid k;\sigma,N)}{\Pr(\omega = 0 \mid k;\sigma,N)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \frac{\binom{N}{k}}{\binom{N}{k}} \Big(\frac{q(1;\sigma)}{q(0;\sigma)}\Big)^{k} \Big(\frac{1-q(1;\sigma)}{1-q(0;\sigma)}\Big)^{N-k}.$$

---

[10] The assumption that R has at most finitely many discontinuities is without loss of generality for monotone referenda, i.e., those where $\min R(m)$ and $\max R(m)$ are weakly increasing in m, since the policy space $\mathcal{P}$ is finite.

[11] We typically indicate posteriors of an agent i with the subscript i, e.g., $\Pr_i(\omega = 1|p_i = p, s_i = s)$, but do not use a subscript for the principal's beliefs.

If $q(1;\sigma) \geq q(0;\sigma)$, the posterior $\Pr(\omega = 1 \mid k;\sigma,N)$ is increasing in k. Since the principal's expected payoff from x is

$$x\Big(\Pr(\omega = 1 \mid k;\sigma,N) - \frac{1}{2}\Big),$$

and since her prior exceeds $\frac{1}{2}$, she can be indifferent after at most one k.

This allows us to characterize the best response in terms of a single cutoff $\bar{k}$: Either $\frac{1}{2} < \Pr(\omega = 1|k;\sigma,N)$ for all $0 \leq k \leq N$—in which case we set $\bar{k} = N$—or there is a minimal $\bar{k}$ with $-1 \leq \bar{k} < N$ such that[12]

$$\Pr(\omega = 1|\bar{k};\sigma,N) < \frac{1}{2} \leq \Pr(\omega = 1|\bar{k}+1;\sigma,N). \tag{1}$$

If $q(1;\sigma) < q(0;\sigma)$, the posterior is decreasing in k and $\bar{k}$ is defined analogously.

The principal can only be indifferent at $k = \bar{k}+1$; hence any mixed best response can only involve mixing at $k = \bar{k}+1$. In the remainder, we consider only principal's strategies that are best responses.

1.2 Agents' Best Responses

Fix a referendum R with cutoffs $(m_1,\ldots,m_{T+1})$ and a strategy profile η. An agent i's best response is driven by two incentives. His action transmits information and thereby affects the principal's posterior beliefs and her preference (for 0 versus 1); also, it may influence which policies are feasible ex post. Importantly, agent i's action affects the policy outcome x only if a *pivotal event* occurs. His incentives are determined by the average effect of his action on the policy outcome across multiple pivotal events. We make these ideas precise below.

Agent i's pivotality depends on the realized number k_{-i} of other agents choosing action 1. For each $k \in \{0,\ldots,N-1\}$, let piv^k denote the event that $k_{-i} = k$. Then we have the following observations:

- Since the principal's preference for 0 versus 1 switches at the cutoff $\bar{k}$, the events $\text{piv}^{\bar{k}}$ and $\text{piv}^{\bar{k}+1}$ are the only ones in which agent i's choice could possibly change the policy outcome by changing the principal's preference.
- In any event piv^k with $k = \lfloor m_j \cdot N \rfloor$ for some $j \in \{1,\ldots,T\}$,[13] agent i's choice changes the feasible policy set, and may affect the policy outcome as a consequence.
- In any other event, agent i's choice does not affect the policy outcome.

---

[12] We abuse notation here and set $\Pr(\omega = 1|k = -1;\sigma,N) = 0$.

[13] Here, for any $z > 0$, $\lfloor z \rfloor$ denotes the largest non-negative integer that lies weakly below z.

The average effect of agent i's action on the policy outcome in $\omega' \in \{0,1\}$, across all pivotal events and realized strategies of the principal, is given by

$$U(\omega';\eta) := \sum_{k=0}^{N-1} \Pr(\text{piv}^k \mid \omega = \omega'; \eta, N) \cdot r(k) \quad \text{ for } r(k) := \mathrm{E}\Big(x(k+1) - x(k) | \eta\Big). \tag{2}$$

Hence, if agent i has signal $s_i = s$ and prior $p_i = p$, then his best response is action 1 if

$$\Pr_i(\omega = 1 \mid p_i = p, s_i = s)\; U(1;\eta) - \Pr_i(\omega = 0 \mid p_i = p, s_i = s)\; U(0;\eta) > 0. \tag{3}$$

That is, he compares the average effects in the two states, weighted by the appropriate posteriors.

We denote the pivotal events related to a shift of the principal's preference by $\text{piv}_0 = \text{piv}^{\bar{k}} \cup \text{piv}^{\bar{k}+1}$, and those related to a change in the feasible policy set by $\text{piv}_j = \text{piv}^{\lfloor m_j \cdot N \rfloor}$ for $j = 1, \ldots, T$.

2 Preliminary Results

We present two preliminary results to prepare for the characterization of robust referenda.

The first result is negative: Partial commitment creates a broad scope for *coordination failures*; that is, equilibrium sequences in which the agents' collective action binds the principal to inefficient policy sets (Proposition 1). The second result is positive: Many referenda nevertheless aggregate information robustly. We give simple sufficient conditions under which, for every admissible information structure and every equilibrium sequence, the principal learns the state from the agents' collective action. These conditions cover referenda with very weak forms of commitment (Proposition 2).

We also use these preliminary results to contrast partial commitment with the benchmark cases of full commitment and no commitment. The formal proofs are in the appendices. After the statement of the main results in the next section, we develop their logic with a detailed sketch.

2.1 Coordination Failure

Given *any* referendum R and *any* of its policy sets R_j, there is an equilibrium sequence in which the binding policy set is R_j with probability converging to 1 as $N \to \infty$. Thus, any policy set R_j can become binding in some equilibrium sequence. Whenever R_j excludes an ex-post optimal policy, i.e., $x = 0$ or $x = 1$, the principal is constrained to choose a suboptimal policy in at least one state. Thus the equilibrium sequence is inefficient.[14]

---

[14] On the other hand, if R always includes both $x = 0$ and $x = 1$, then there is always an (inefficient) equilibrium without any information transmission that we discuss below.

Proposition 1. *Consider any referendum R with cutoffs $0 < m_1 < \ldots < m_T < m_{T+1} = 1$. For any $j^* \in \{1, \ldots, T+1\}$, there exists an agents' information structure and a sequence of equilibrium strategies $(\sigma_N)_{N\in\mathbb{N}}$ for which*

$$\lim_{N\to\infty} \Pr(m \in I_{j^*}|\sigma_N, N) = 1;$$

hence the realized policy set is R_{j^*} with probability approaching one.

Many referenda also have efficient equilibrium sequences for all agents' information structures, as we show in the working paper version (Heese, 2026). So Proposition 1 shows that the agents face a coordination problem: They may miscoordinate to achieve an inefficient equilibrium instead of an efficient one. For comparison, the Condorcet jury theorem (as in Bhattacharya, 2013) implies that when the principal has unlimited commitment power in our setting, simple majority voting between $x = 0$ and $x = 1$ implies efficient outcomes in *all* equilibrium sequences as $N \to \infty$. That is, the possibility of miscoordination is a consequence of partial commitment.

2.2 Information Aggregation

Many referenda possess robust information aggregation properties. We say that an equilibrium sequence *aggregates information* if the principal learns the state from observing the agents' collective action (with probability approaching one as $N \to \infty$).

We study information aggregation for monotone referenda with a single cutoff. We identify two elementary properties of such referenda that jointly guarantee information aggregation uniformly across equilibria and admissible information structures (Proposition 2 below).

The first property rules out a "balanced" split of decision authority between the principal and the agent population. A single-cutoff referendum has *no balance* if

$$\max R(0) \neq \min R(1),$$

and it has *balance* otherwise. The second property is that the maximum feasible policy is non-constant,

$$\max R(0) < \max R(1).$$

Proposition 2. *Consider any monotone referendum R with a single cutoff $0 < m_1 < 1$ and any agents' information structure. If R has no balance and $\max R(0) < \max R(1)$, then information aggregates in all equilibrium sequences.*

We discuss the second property here and postpone the discussion of the first to later. The second property rules out pure cheap talk as a special case, that is, processes R

for which the policy set $R(m)$ does not vary with m. This includes the no-commitment process, where $R(m) = \mathcal{P}$ for all m.

Cheap talk—and, more generally, constant $\max R(m)$— implies a trivial (babbling) equilibrium with no information transmission. In this equilibrium, all non-partisans choose the same action, so their collective action conveys no information. The principal then chooses the same policy (namely, the maximum feasible one) after every observation; this is always optimal given her prior preference for high policies. Hence all agents are indifferent between all strategies, which sustains their uninformative behavior as a best response as well.

Thus cheap talk does not allow for information aggregation along all equilibrium sequences. Related cheap-talk settings deliver an even sharper negative conclusion: There are no equilibria other than babbling when the principal's and the agents' priors are sufficiently misaligned (Battaglini, 2017; Chen, 2025; Levit and Malenko, 2011; Morgan and Stocken, 2008).

Our result gives the opposite conclusion for partial-commitment referenda. Even allowing for large misalignments between the principal's prior and an arbitrarily large mass of the agents' priors, the referenda covered by Proposition 2 admit no babbling equilibrium sequences; indeed, all equilibrium sequences aggregate information. Perhaps surprisingly, this conclusion applies even when the commitment embedded in the referendum is minimal, for example when $R(1) = \mathcal{P}$ and $R(0) = \mathcal{P} \setminus \{1\}$. Section 6 discusses this contrast with the cheap-talk impossibility results in more detail.

3 Main Results: Robust Referenda

Our main analysis compares referenda in terms of their robustness. Our yardstick of comparison is the principal's *payoff guarantee*, the proportion of the full-information payoff that the principal obtains in the worst-case scenario, as $N \to \infty$. Formally, for a referendum R, the principal's payoff guarantee is defined as

$$G(R) := \inf_{(\eta_N)_{N\in\mathbb{N}},\pi} \Big(\liminf_{N\to\infty} \mathrm{E}(x \mid \omega = 1; \eta_N, N) - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \mathrm{E}(x \mid \omega = 0; \eta_N, N) \Big),$$

where we take the infimum over all equilibrium sequences $(\eta_N)_{N\in\mathbb{N}}$ and all agents' information structures π.[15]

---

[15] Here, $\liminf$ denotes the smallest accumulation point of a sequence. For any equilibrium sequence $(\eta_N)_{N\in\mathbb{N}}$, the smallest accumulation point of the principal's payoff is $\liminf_{N\to\infty} \frac{1}{2}\Big(\Pr(\omega = 1)\mathrm{E}(x \mid \omega = 1; \eta_N, N) - \Pr(\omega = 0)\mathrm{E}(x \mid \omega = 0; \eta_N, N)\Big)$. When the principal knows the state, she can achieve the full-information payoff $\frac{1}{2}\Pr(\omega = 1)$. Dividing the former quantity by the latter and taking the infimum over all π and $(\eta_N)_{N\in\mathbb{N}}$ yields $G(R)$. The online appendix provides a general equilibrium existence result, which clarifies that the payoff guarantee of a referendum is always well-defined: For *any* N, *any* referendum, and *any* agents' information structure, an equilibrium exists.

Our main results characterize referenda that maximize the principal's payoff guarantee; we call such referenda *robust-optimal.* The results are driven by a common logic, grounded in the observations in the preceding sections: On the one hand, some commitment is necessary to make the collective action consequential and thereby sustain information transmission (as discussed in Section 2.2). On the other hand, stronger commitment than necessary makes outcomes more vulnerable to the agents' coordination failures (cf. Proposition 1). The robust-optimal referenda therefore turn out to use minimal commitment: The collective acts as a gatekeeper for one polar option, while the principal otherwise retains discretion.

3.1 Collective Vetoes of the Maximum

We first consider the baseline class of all referenda. In this class, the preceding logic singles out referenda in which the collective can exclude the maximal policy while the principal otherwise retains discretion. We call these *collective vetoes of the maximum.* Real-world examples of such referenda include cap-referendum procedures such as shareholder votes on remuneration caps, in which the shareholder can veto an extraordinary pay package, and budget- or tax-cap referenda in local public finance.

Formally, a collective veto of the maximum is a referendum R of the form

$$R(m) = \begin{cases} \mathcal{P} \setminus \{1\} \text{ if } m \leq m_1, \\ \mathcal{P} \text{ if } m > m_1, \end{cases} \tag{4}$$

for some $m_1 \in (0,1)$.[16] Theorem 1 shows that collective vetoes of the maximum are robust-optimal.

Theorem 1. *Any collective veto of the maximum has a payoff guarantee of* $1-\varepsilon$ *and is robust-optimal among all referenda.*

The proof runs as follows. As we observed earlier, robust-optimality requires balancing two forces. On the one hand, some commitment is necessary for information transmission: If a referendum R *never* excludes the maximal policy $x = 1$, then there are uninformative equilibria in which the principal chooses $x = 1$ in both states, as argued in Section 2.2. This implies an upper bound for the payoff guarantee of

$$G(R) \leq 1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} < 1 - \varepsilon. \tag{5}$$

---

[16] More generally, in a *cap referendum*, the collective action decides which of two policy caps $x_c < x_C$ the principal has to obey. Low collective actions impose the more stringent cap x_c; high ones impose x_C. More precisely, for some cutoff $m_1 \in (0,1)$, the set of feasible policies is $R(m) = [0, \ldots, x_c]$ if $m \leq m_1$ and $R(m) = [0, \ldots, x_C]$ if $m > m_1$. Collective vetoes of the maximum are cap referenda with $x_c = 1 - \varepsilon$ and $x_C = 1$.

Here the last inequality holds because we assumed the principal's prior is not extreme, $\Pr(\omega = 1) \leq 1 - \varepsilon$; see Section 1.[17]

On the other hand, stronger commitment than necessary increases the principal's exposure to coordination failures. If a referendum R excludes $x = 1$ after *some* collective action, then Proposition 1 implies the existence of equilibrium sequences in which the corresponding constrained policy set binds with probability approaching one. Along such sequences, the principal cannot choose $x = 1$ in state 1, and therefore chooses at most $1 - \varepsilon$. This yields an upper bound for the payoff guarantee of

$$G(R) \leq 1 - \varepsilon. \tag{6}$$

If the referendum were to exclude $x = 1$ and *additional* policies after some collective action, the same argument would imply an even smaller upper bound for the payoff guarantee.

A collective veto of the maximum R^* balances these two forces by using minimal commitment: allowing the collective to exclude at most one policy. This ensures robust information aggregation—an immediate implication of Proposition 2—while minimizing the principal's exposure to coordination failures. Since the principal learns the state under R^* and can choose any policy other than $x = 1$ regardless of the agents' equilibrium play, in the worst case she chooses $x = 0$ in state 0 and $x = 1 - \varepsilon$ in state 1. Consequently, a lower bound for her payoff guarantee is

$$G(R^*) \geq 1 - \varepsilon.$$

By (5) and (6), $1 - \varepsilon$ is an upper bound for the payoff guarantee of *every* referendum. Hence collective vetoes of the maximum are robust-optimal, with a payoff guarantee of $1 - \varepsilon$.

Two remarks are in order. First, the theorem shows that collective vetoes of the maximum achieve near-full-information outcomes if the policy space is fine, i.e. if ε is small. Second, collective vetoes of the maximum also maximize the agents' welfare guarantee given the common ex-post preferences. We define the agents' welfare guarantee as the ratio of an agent's mean equilibrium payoff in the worst-case scenario to his mean full-information payoff. For a referendum R, it is[18]

$$\inf_{(\eta_N)_{N \in \mathbb{N}}, \pi} \left(\liminf_{N \to \infty} \mathrm{E}(x \mid \omega = 1; \eta_N, N) - \frac{\mathrm{E}(1 - p_i)}{\mathrm{E}(p_i)} \mathrm{E}(x \mid \omega = 0; \eta_N, N) \right).$$

This quantity is obtained by replacing the principal's prior with the agents' mean prior in the definition of $G(R)$. Bounds analogous to (5) and (6) for the agents' welfare guarantee

---

[17] The calculation is as follows: $1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)} \leq 1 - \frac{\varepsilon}{1-\varepsilon} < 1 - \varepsilon$.

[18] To evaluate the payoffs of partisan agents, we assume here and in the following that partisans hold extreme priors that conform to their prescribed choice $a \in \{0, 1\}$, i.e. $p_i = a$.

hold provided the agents' mean prior satisfies the same extremeness bound as the principal's prior, $\mathrm{E}(p_i) \leq 1-\varepsilon$. Thus, collective vetoes of the maximum are also agent-optimal except in extreme scenarios.

3.2 Gateway Referenda

We now enlarge the class of decision processes by allowing commitment to the status quo (interpreted as the policy $x = 0$). Such commitment is feasible in many applications. In fact, many real-world referenda decide whether to retain the status quo or to mandate some change, while delegating the details of that change to a subsequent institutional procedure. A constitutional referendum, for example, may decide whether to keep the current version of a country's constitution or to allow it to be revised, without specifying the outcome of the revision. An abrogative referendum—the main instrument of direct democracy in Italy—allows voters to repeal an existing law, or part of it, without itself enacting a replacement.

We call a mapping R a *generalized referendum* if for every $m \in [0,1]$, either $R(m) = \{0\}$ or $R(m)$ is a non-singleton policy set. This relaxes the requirement in the definition of a referendum that $R(m)$ be non-singleton for all m.

We define a *gateway referendum* as a generalized referendum R of the form

$$R(m) = \begin{cases} \{0\} & \text{if } m \leq m_1, \\ \mathcal{P} \setminus \{0\} & \text{if } m > m_1, \end{cases} \tag{7}$$

for some cutoff $m_1 \in (0,1)$. Thus, if the cutoff is not met, the status quo is retained; if it is met, the status quo is excluded and the principal can choose any positive policy. Conditional on change, the referendum is therefore only minimally binding.

The next theorem shows that the simple-majority gateway referendum is robust-optimal within this larger class of decision processes.

Theorem 2. *The gateway referendum with the simple-majority cutoff* $m_1 = \frac{1}{2}$ *has a payoff guarantee of* $1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)}\,\varepsilon$ *and is robust-optimal among all generalized referenda.*

The proof is in the appendix. Here we highlight the key step that is novel relative to the proof of Theorem 1. What remains unchanged from the latter is the argument for information aggregation. The simple-majority gateway referendum is a single-cutoff process with no balance and with $\max R(0) < \max R(1)$, so the same argument as used for Proposition 2 shows that it provides sufficient incentives for information to aggregate in all equilibrium sequences.[19]

---

[19] Proposition 2 was stated for referenda R, but the relevant part of the proof uses only that R has a single cutoff, has no balance, and satisfies $\max R(0) < \max R(1)$. It does not use that every feasible set is non-singleton.

However, unlike in the case of collective vetoes, here information aggregation does not by itself imply outcomes close to the ex-post optimum. The new difficulty is that, while the gateway referendum is only minimally binding conditional on change, it provides full commitment to the status quo $x = 0$ otherwise. If this commitment were to bind in state 1, the outcome $x = 0$ would be far from optimal.

The reason commitment to $x = 0$ does not obstruct robust-optimality is that, under the simple-majority gateway referendum, it can bind asymptotically only in the state where $x = 0$ is ex-post optimal. To establish this, we show that in any equilibrium sequence the mean action in state 1 exceeds the simple-majority cutoff as $N \to \infty$. Hence the cutoff is met with probability approaching one in state 1, and the status-quo commitment cannot bind there.

Lemma 1. *For any agents' information structure and any sequence of equilibrium strategies $(\sigma_N)_{N\in\mathbb{N}}$ for the gateway referendum with $m_1 = \frac{1}{2}$, it holds that*

$$\lim_{N\to\infty} q(1;\sigma_N) > \frac{1}{2}.$$

The logic behind Lemma 1 is closely related to that of the classical Condorcet jury theorem (Bhattacharya, 2013). The Condorcet jury theorem establishes that simple majority voting between $x = 0$ and $x = 1$ leads to ex-post optimal policies, asymptotically as $N \to \infty$. In the standard majority-voting environment, one proves $\lim_{N\to\infty} q(1;\sigma_N) > \frac{1}{2}$ by contradiction. One assumes $\lim_{N\to\infty} q(1;\sigma_N) \leq \frac{1}{2}$ and combines this with two ingredients: first, that the equilibrium mean actions satisfy $\lim_{N\to\infty} q(0;\sigma_N) < \lim_{N\to\infty} q(1;\sigma_N)$; and second, that agents' incentives are only driven by the pivotal event piv_1 at the majority cutoff $\frac{1}{2}$. Under these two conditions, the mean action is closer to the majority cutoff in state 1: $\lim_{N\to\infty} |q(0;\sigma_N) - \frac{1}{2}| > \lim_{N\to\infty} |q(1;\sigma_N) - \frac{1}{2}|$. The standard argument then implies that pivotality at the majority cutoff becomes an overwhelmingly strong indication for state 1, so that almost all non-partisans strictly prefer action 1. So $\lim_{N\to\infty} q(1;\sigma_N) = 1 - \rho_0 > \frac{1}{2}$, a contradiction.

In our setting, neither of these ingredients is immediate. The ordering of the mean actions may be reversed in some equilibria (we will construct examples of such equilibria in Section 2.2), and the agents' incentives may also depend on the additional pivotal event piv_0. In the appendix we show, however, that under the simple-majority gateway referendum both ingredients are recovered asymptotically: The reverse ordering cannot arise in equilibrium, and if the conclusion of Lemma 1 fails, then piv_0 becomes asymptotically irrelevant for the agents' incentives. The standard Condorcet-jury-theorem contradiction therefore applies and yields Lemma 1.

Lemma 1 implies that the simple-majority gateway referendum minimizes the principal's exposure to coordination failures and so is robust-optimal: Since information aggregates, and since the status-quo commitment does not bind in state 1, the principal chooses the ex-post optimal policy in state 1. Hence the only potential loss arises if, in state 0, the principal is forced to choose some $x > 0$; in this worst case, she then chooses the smallest feasible policy, $x = \varepsilon$. Consequently,

$$G(R^*) \;\geq\; 1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)}\varepsilon \;>\; 1 - \varepsilon.$$

To rule out the possibility of strictly higher payoff guarantees in other generalized referenda, the formal proof constructs inefficient equilibria similar to those of the preliminary results (Propositions 1 and 2). This establishes that the simple-majority gateway referendum is robust-optimal. For example, one might consider the generalized referendum defined by $R(m) = \{0\}$ for $m \leq \frac{1}{2}$ and $R(m) = \mathcal{P}$ for $m > \frac{1}{2}$. However, this design has balance $\Big(\max R(0) = \min R(1)\Big)$ and therefore admits an equilibrium sequence that does not aggregate information and implies a lower payoff guarantee.

The next section develops the logic underlying the preliminary results, Proposition 1 and Proposition 2, and provides detailed sketches of their proofs.

4 Coordination Failure and Information Aggregation Explained

4.1 The Logic of Coordination Failure

We explain why, for any referendum R, each of its policy sets R_j becomes binding along an equilibrium sequence of some informational environment (Proposition 1). Along such equilibrium sequences, the agents miscoordinate on how to use their collective influence: whenever R_j excludes an ex-post optimal policy, i.e., $x = 0$ or $x = 1$, the principal is constrained to choose a suboptimal policy in at least one state. Thus, the equilibrium sequence is inefficient.

Although such coordination failures arise across environments (cf. Section 5.1), this conclusion can already be obtained when restricting to those in which the agents' priors are relatively close to the principal's prior:

$$\Pr_F\Big(p_i \in [\underline{p}_1, \Pr(\omega = 1)]\Big) > 1 - \frac{\delta}{4}, \tag{8}$$

for a certain bound $\underline{p}_1$, and when restricting to particularly intuitive equilibria. The proof of Proposition 1 constructs sequences of equilibria in "approximately truthful" strategies. Formally, for $\delta > 0$, an agents' strategy σ is *(δ-)approximately truthful* if, for any given realized signal, the share of non-partisan types matching their action to their signal is at least $1 - \delta$. Equilibria in approximately truthful strategies (which we call approximately truthful equilibria) exist for small enough δ and some agents' information structure satisfying one additional property besides (8): The agents' signals are relatively uninformative:

$$\Pr(s_i = 1|\omega = 1) - \Pr(s_i = 1|\omega = 0) = \gamma \tag{9}$$

for small enough $\gamma > 0$.

The relevance of the two properties is best illustrated by connecting them to the scenario where the agents' actions are cheap talk and the prior is common. This is a pure common-value game, and, as such, it has an equilibrium in which *all* agents truthfully match their actions to their signals.

In the following we sketch how, for small enough parameters $\gamma > 0$ and $\delta > 0$ and some information structure with the properties (8) and (9), the agents' incentives given any approximately truthful strategy approximate those in the common-value cheap-talk game. As the proof shows, this will imply the existence of an approximately truthful equilibrium.

The agents' incentives are driven by the pivotal events piv_j corresponding to the principal's cutoff and the cutoffs defining the referendum. In the cheap-talk game, only the principal's cutoff $\bar{k}$ and its limit $m_0 := \lim_{N\to\infty} \frac{\bar{k}}{N}$ matter for incentives. By contrast, in a referendum, multiple cutoffs may affect the incentives. However, under certain conditions only m_0 matters, as we explain via example below.

Intuitively, a cutoff m_j is more relevant in shaping the agents' incentives when it is closer to the mean action, because in that case, the pivotal event in which $m = m_j$ is more likely. It is therefore key to compare the distance of m_0 to the mean action in each state with that of the other cutoffs. We will do this via a bound $M > 0$.

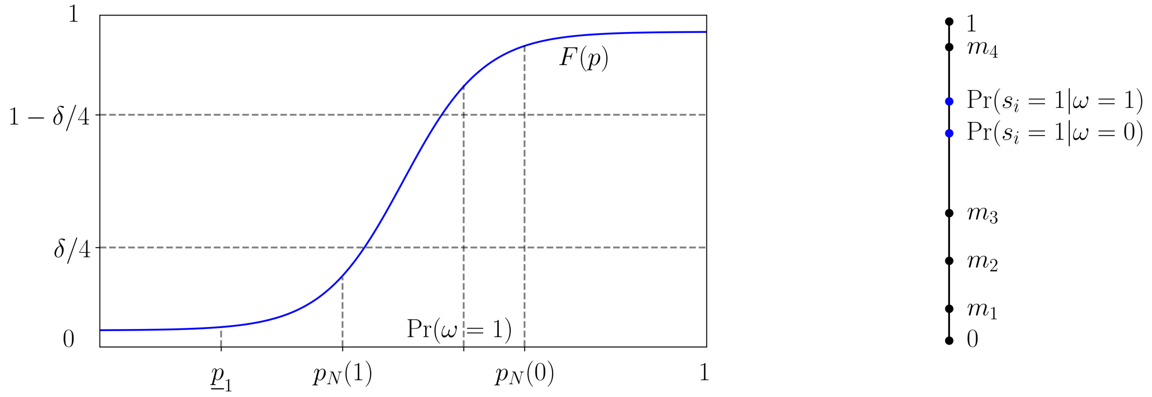

Figure 1: The agents' information structure, given by the prior distribution F (left) and the signal probabilities $\Pr(s_i = 1|\omega = \omega')$ (right). The approximately truthful equilibrium is given by the types $p_N(0)$ and $p_N(1)$ that are indifferent after signals 0 and 1, respectively: An agent i with signal $s \in \{0, 1\}$ chooses $a_i = 1$ if and only if $p_i \geq p_N(s)$.

Figure 1 depicts an example referendum and an agents' information structure satisfying (8) and (9) for small parameters $\gamma > 0$ and $\delta > 0$. The left panel shows the distribution of priors, nearly all of whose mass lies between the principal's prior and a bound $\underline{p}_1 < \Pr(\omega = 1)$ close to it. The right panel shows the signal probabilities $\Pr(s_i = 1|\omega = 1)$ and $\Pr(s_i = 1|\omega = 0)$.

The signal probabilities are chosen together with the parameter $\gamma > 0$ so that they lie between m_3 and m_4, and so that their distance from these cutoffs is relatively large, compared to γ.

The parameter $\delta > 0$ is chosen small enough so that any approximately truthful strategy σ_N has mean actions $q(\omega'; \sigma_N)$ so close to the signal probabilities that

$$q(1; \sigma_N) - q(0; \sigma_N) > \frac{\gamma}{2} \text{ and } \Big|m_j - q(\omega'; \sigma_N)\Big| > M\gamma \text{ for } j \in \{3, 4\},\ \omega' \in \{0, 1\}, \tag{10}$$

with $M > 0$ bounding the distance to the cutoffs m_3 and m_4. The first inequality implies that the principal's posterior is strictly increasing in the number of observed actions 1. As $N \to \infty$, by an application of the law of large numbers, her posterior becomes extreme—close to either 0 or 1—for observations close to either $q(0; \sigma_N)$ or $q(1; \sigma_N)$, respectively. Hence her best-response cutoff (at which her posterior is not far from $\frac{1}{2}$) must lie in between the mean actions, ensuring that

$$\Big|m_0 - \lim_{N\to\infty} q(\omega'; \sigma_N)\Big| \le \gamma \text{ for } \omega' \in \{0, 1\}.$$

Comparing this to (10), one sees that M bounds the relative distance of the principal's cutoff to the mean actions, compared to the distance of the referendum cutoffs m_j to the mean actions. As the proof of Proposition 1 shows, when M is relatively large (as illustrated by the example in the figure), the likelihood ratio of the pivotal events becomes extreme as N grows large. In particular, if agent i is pivotal, then it becomes almost certain that he is influencing the principal's preferences, rather than the feasible policy set:

$$\lim_{N\to\infty} \Pr_i(\text{piv}_0|\text{piv}; \sigma_N, N) = 1. \tag{11}$$

In other words, the agents' incentives are almost the same as if their actions were cheap talk. For small $\delta > 0$, the players' prior distribution is close to a common prior, so that altogether the incentives are close to those in the common-value cheap-talk game.

The formal proof then uses a fixed-point argument to construct an approximately truthful equilibrium, mirroring the truthful equilibrium of the common-value cheap-talk game.

The left panel of Figure 1 shows the equilibrium strategy in terms of its cutoff types, namely, the agent types $p_N(1)$ and $p_N(0)$ that are indifferent after signal 1 and after signal 0, respectively, with types above the cutoff choosing action 1 and types below action 0. Note that $p_N(1)$ lies below the δ-quantile and $p_N(0)$ above the $(1 - \delta)$-quantile. Thus, the equilibrium is δ-approximately truthful.

Finally, the mean actions of the equilibrium lie strictly between m_3 and m_4; see (10). So, by the law of large numbers, the realized collective action is in $I_4 = (m_3, m_4]$ and the realized policy set is R_4 with probability approaching one as $N \to \infty$. The proof of Proposition 1 extends the ideas of this example to show that *any* policy set R_j of *any* referendum can become binding with probability approaching one.

4.2 The Conditions of Information Aggregation

We explain the two properties that drive robust information aggregation for monotone referenda with a single cutoff: no balance and $\max R(0) < \max R(1)$.

We first explain that both conditions are necessary for robust information aggregation. The second property is necessary since constant $\max R(0) = \max R(1)$ entails a trivial (babbling) equilibrium with no information transmission, as discussed in Section 2.2. No balance is also necessary since balance makes possible the existence of equilibrium sequences that do not aggregate information. For example, consider

$$R(m) = \begin{cases} \{0, \dots, \frac{1}{2}\} & \text{for } m \leq \frac{1}{2}, \\ \{\frac{1}{2}, \dots, 1\} & \text{for } m > \frac{1}{2}. \end{cases}$$

Because the policy $\frac{1}{2}$ is feasible on both sides of the cutoff $m = \frac{1}{2}$, there exist *deadlock* equilibria in which the principal chooses $x = \frac{1}{2}$ after every observation, even though she updates her beliefs based on the collective action. When the principal plays such a constant strategy, all agents are indifferent between all strategies (their actions have no effect on the policy outcome). What matters is that for large N, there is an agents' strategy that induces the principal's constant strategy $x = \frac{1}{2}$ as a best response. This combination of strategies then constitutes an equilibrium.

The agents' equilibrium strategy has mean actions with $0 < q(1; \sigma_N) < q(0; \sigma_N) < \frac{1}{2}$ calibrated so that the principal is indifferent when the majority cutoff ($m = \frac{1}{2}$) is just met, i.e.[20]

$$\Pr\Big(\omega = 1 | k = \lfloor \frac{N}{2} \rfloor + 1; \sigma_N, N\Big) = \frac{1}{2}.$$

Since $q(1; \sigma_N) < q(0; \sigma_N)$, the principal's posterior is decreasing in k. Thus, if she observes $k < \lfloor \frac{N}{2} \rfloor + 1$, she prefers high policies but can choose at most $x = \frac{1}{2}$. If she observes $k \geq \lfloor \frac{N}{2} \rfloor + 1$, she prefers low policies but must choose at least $x = \frac{1}{2}$. Thus, it is optimal for her to choose $x = \frac{1}{2}$ in both cases.

Information does not aggregate: As $q(0; \sigma_N) < \frac{1}{2}$ for all N, in state 0 the realized number k is smaller than $\lfloor \frac{N}{2} \rfloor + 1$ and the principal's posterior is greater than $\frac{1}{2}$ with non-vanishing probability. So the principal does not learn the state when it is 0.

---

[20] The online appendix contains a formal proof that such a strategy exists when N is large enough.

Next, we sketch why the two conditions are jointly also sufficient for robust information aggregation, as claimed by Proposition 2. First, the condition $\max R(0) < \max R(1)$ rules out the possibility of "uninformative" equilibria σ_N, which we define via the property $q(0;\sigma_N) = q(1;\sigma_N)$.

Suppose σ_N is uninformative. Then the principal learns nothing from the realized collective action, so her equilibrium choice depends only on feasibility: Above the cutoff m_1 she selects the largest element of $R(1)$, and below the cutoff she selects the largest element of $R(0)$ (given her prior preference for high policies). Because $\max R(0) < \max R(1)$, an agent is thus pivotal only at the cutoff. The uninformativeness of the equilibrium implies that this pivotal event has the same probability in both states, so any agent with a uniform prior is indifferent before observing his signal. His best response (and that of nearby types) is to match action to signal, implying on average higher actions in state 1; that is, $q(0;\sigma_N) < q(1;\sigma_N)$. This contradicts the initial assumption of uninformativeness.

Second, non-balance causes the logic of the deadlock equilibrium to fail. The deadlock equilibrium is supported by the fact that agents cannot affect the policy outcome. Non-balance eliminates this possibility: It implies that an agent's average effect $U(\omega';\eta_N)$ on the policy outcome, given by (2), is non-zero and has the same sign in both states.

Lemma 2. *Consider any monotone referendum R, with a single cutoff $0 < m_1 < 1$, that has no balance and satisfies $\max R(0) < \max R(1)$. Then, for any equilibrium η_N,*

either $U(\omega';\eta_N) > 0$ for all $\omega' \in \{0,1\}$,
or $U(\omega';\eta_N) < 0$ for all $\omega' \in \{0,1\}$.

For intuition, let us revisit the referendum with balance defined earlier, where $\max R(0) = \min R(1) = \frac{1}{2}$, and decrease $\max R(0)$. Then the principal's best response to the agents' deadlock equilibrium strategy ceases to be constant: She chooses $\max R(0)$ if $m \leq m_1$ and $\min R(1)$ if $m > m_1$, but these policies no longer coincide. Hence the agents' actions now affect the policy outcome; an action of 1 moves the policy upwards at the cutoff, so the average effect is positive, i.e. $U(\omega';\eta_N) > 0$ for all $\omega' \in \{0,1\}$.

Lemma 2 implies a simple cutoff form for the agents' equilibrium strategies. Each agent trades off between the average effects of his action in the two states, weighted by his belief about the states. From the characterization (3) of the agents' best response, for each signal s there is a *unique* type $0 < p_N(s) < 1$ that is indifferent after observing s:

$$\frac{U(0;\eta_N)}{U(1;\eta_N)} = \frac{\Pr_i(\omega = 1 \mid p_i = p_N(s), s_i = s)}{\Pr_i(\omega = 0 \mid p_i = p_N(s), s_i = s)}. \tag{12}$$

Information aggregation follows once we show that the indifferent types do not drift to the extremes. In particular, it is enough to establish[21]

$$0 < \lim_{N \to \infty} p_N(1) < \lim_{N \to \infty} p_N(0) < 1. \tag{13}$$

Then, as $N \to \infty$, the mean action differs across signals and thus across the two states. Since the realized collective action concentrates around the mean action in each state, the principal learns the state from observing it.

Finally, the key observation underlying (13) is that, under partial commitment, a hypothetical failure of (13) would generically imply that, conditional on being pivotal, an agent becomes certain (as $N \to \infty$) that he is pivotal in affecting the principal's policy preference (as opposed to the feasible policy set); that is,

$$\lim_{N \to \infty} \Pr_i(\mathrm{piv}_0 | \mathrm{piv}; \eta_N, N) = 1. \tag{14}$$

The intuition is as follows. If (13) failed, then all non-partisans would choose the same action, implying extreme limit mean actions

$$\lim_{N \to \infty} q(0; \sigma_N) = \lim_{N \to \infty} q(1; \sigma_N) \in \{\rho_1, 1 - \rho_0\}.$$

The proof shows that then the principal's limit cutoff $m_0 = \lim_{N \to \infty} \frac{\bar{k}}{N}$ must coincide with the limit mean actions. Therefore, if $m_1 \notin \{\rho_1, 1 - \rho_0\}$, then for all large N the principal's cutoff is closer to the mean actions than m_1 is. As already noted in the proof sketch for Proposition 1, this relative closeness implies (14).

Given (14), the crucial point is that the principal's updating from piv_0 is bounded: It shifts her prior to a belief close to $\frac{1}{2}$, the indifference point. The agents' updating from piv_0 is therefore bounded as well. This, in turn, yields bounds on the limit indifferent types and hence implies (13).[22] It follows that (13) holds generically.

The appendix provides the full proof of Proposition 2, including the proof of Lemma 2, a precise comparison of the pivotal probabilities to derive (14), and an analysis of the knife-edge cases in which (13) may fail.

5 Extensions and Other Results

5.1 Other Coordination Failures: Disciplining a Biased Principal

The logic of coordination failure presented previously focused on environments in which the principal and the agents have broadly aligned beliefs (Section 4.1).

---

[21] Throughout, whenever limits of sequences such as $(p_N(s))_{N \in \mathbb{N}}$ or $(q(\omega'; \sigma_N))_{N \in \mathbb{N}}$ are considered, we work along convergent subsequences; since these sequences lie in compact sets, this is without loss.

[22] To be precise, (14) implies $\frac{U(0;\eta_N)}{U(1;\eta_N)} \approx \frac{\Pr_i(\mathrm{piv}_0 | \omega=0; \eta_N, N)}{\Pr_i(\mathrm{piv}_0 | \omega=1; \eta_N, N)}$, and the bounded updating from piv_0, i.e., the fact that $\lim_{N \to \infty} \frac{\Pr_i(\mathrm{piv}_0 | \omega=0; \eta_N, N)}{\Pr_i(\mathrm{piv}_0 | \omega=1; \eta_N, N)} \in (0, \infty)$, then implies (13).

Differences between the principal's and the agents' priors can also give rise to inefficiencies. If the principal is more pessimistic about the possible benefits of the policy than a sufficient mass of agent types, there are equilibria in which the agents miscoordinate on "disciplining" the principal's perceived bias: They constrain the principal to the highest policy range available, given the announced referendum, with probability close to 1. To state this result formally, we define $\bar{p} > \Pr(\omega = 1)$ as the type whose posterior likelihood ratio conditional on signal 0 equals the principal's prior likelihood ratio multiplied by $\frac{\Pr(s_i=1|\omega=1)}{\Pr(s_i=1|\omega=0)} \frac{\Pr(s_i=0|\omega=0)}{\Pr(s_i=0|\omega=1)} > 1$, and we call a referendum R *monotone* if both $\min R(m)$ and $\max R(m)$ are weakly increasing.

Proposition 3. *Consider any non-constant, monotone referendum with cutoffs* $0 < m_1 < \ldots < m_T < m_{T+1} = 1$. *There is some* $\bar{q} \in (m_T, 1 - \rho_0)$ *such that, when* $\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) > \bar{q}$, *there is a sequence of equilibrium strategies* $(\sigma_N)_{N \in \mathbb{N}}$ *for which* $\lim_{N\to\infty} \Pr(m_T < m \mid \sigma_N, N) = 1$.

The proof of Proposition 3 is in the appendix. The idea of the proof is to construct a sequence of equilibria σ_N in which the mean actions in both states exceed the highest cutoff of the referendum:

$$m_T < \bar{q} \leq q(0; \sigma_N) < q(1; \sigma_N) - \delta \text{ for some } \delta > 0. \tag{15}$$

This implies that the highest cutoff becomes almost certainly binding, by the law of large numbers, as the proposition claims.

The proof uses a formal fixed-point argument to show that the best response is a self-map on the set of strategies satisfying (15). We sketch the argument here. Take any strategy σ_N satisfying (15) and a best response of the principal to it. By an argument analogous to that of Section 4.1, if $\bar{q}$ is sufficiently high, then the principal's cutoff is much closer to the mean actions $q(\omega'; \sigma_N)$ than are the cutoffs of the referendum. As before, we can then show that, conditional on being pivotal, an agent's action is almost certain to be influencing the principal's preferences rather than the feasible policy set; that is, (11) holds.

Note, however, that unlike in Section 4.1, here the principal's priors are not close to the agents': The condition $\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) > \bar{q}$ means that the principal is more pessimistic about the policy than a large mass of non-partisan types. Consequently, it is not optimal for many agents to act truthfully (i.e., to act in line with their signals). Instead, an agent's best response arises from the following strategic reasoning: Conditional on the agent's action changing the principal's preferences, the principal must be close to indifferent. Now, if the agent's prior exceeds the principal's by a sufficiently large margin, then, regardless of his signal, the agent is not indifferent but strictly prefers to influence the principal to choose higher policies. Indeed, the proof of the proposition shows that, for all non-partisan types with prior above $\bar{p}$, the best response is action 1.

This best response strategy satisfies (15), since $\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) > \bar{q}$.

5.2 A Weaker Robustness Criterion

The logic underlying the optimality results of Section 3 can be extended to the following weaker robustness criterion. Fix an agents' information structure, and consider the principal's payoff guarantee—that is, her payoff in the worst-case scenario as $N \to \infty$—across all equilibrium sequences under that information structure. (That is, the payoff guarantee is defined in the same way as $G(R)$, except that the infimum is taken only over equilibrium sequences under the chosen information structure.) We characterize the referenda that are optimal in the sense of maximizing this payoff guarantee within the class of all monotone referenda with the single cutoff $m_1 = \frac{1}{2}$.

Take any such referendum and suppose $\Pr(\omega = 1) > \frac{1}{2}$, so that the principal's ex-ante preferred policy is the maximal one. As before, the key tension lies in how commitment affects both information transmission and exposure to coordination failures.

First, some commitment is necessary to sustain information transmission: If $\max R(0) = \max R(1)$, then there exists an equilibrium in which the agents' behavior is uninformative and the principal chooses the same policy regardless of the collective action. Thus, avoiding such equilibria requires $\max R(0) < \max R(1)$.

Second, stronger commitment increases exposure to coordination failures: When the principal's prior is sufficiently high, an argument analogous to that proving Proposition 3 implies the existence of equilibrium sequences in which $R(0)$ binds with probability approaching one, as $N \to \infty$. Intuitively, as $\Pr(\omega = 1) \to 1$, the cutoff $\bar{p}$ of Proposition 3 also approaches 1, implying $\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) < \bar{q}$ for any $\rho_1 < \bar{q} < \frac{1}{2}$; this is a mirror analogue of the sufficient condition in the proposition. (The details are in the online appendix.)

Along sequences in which $R(0)$ binds asymptotically, the maximal feasible policy in state 1 is $\max R(0)$, implying that the payoff guarantee (across all equilibrium sequences) is bounded above by

$$\max R(0) \leq 1 - \varepsilon.$$

The collective veto of the maximum achieves a payoff guarantee of $1 - \varepsilon$ (Theorem 1). Hence it is optimal among monotone referenda with cutoff $m_1 = \frac{1}{2}$, as the next proposition concludes.

Proposition 4. *Fix an agents' information structure. If the principal's prior is sufficiently high, the collective veto of the maximum maximizes the principal's payoff guarantee across all equilibrium sequences among all monotone referenda with a single cutoff* $m_1 = \frac{1}{2}$.

Now suppose $\Pr(\omega = 1) < \frac{1}{2}$. Again, to rule out uninformative equilibria, $\min R(m)$ must be non-constant, i.e. $\min R(0) < \min R(1)$. At the same time, when $\Pr(\omega = 1)$ is sufficiently small, an argument analogous to that of Proposition 3 implies the existence of equilibrium sequences in which $R(1)$ binds with probability approaching one as $N \to \infty$. Along such sequences, the minimal feasible policy in state 0 is $\min R(1)$, so the payoff guarantee (across all equilibrium sequences) is bounded above by

$$1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \min R(1) \leq 1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)}\varepsilon.$$

The simple-majority gateway referendum attains a payoff guarantee of $1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)}\varepsilon$.[23] Hence it is optimal among monotone generalized referenda with cutoff $m_1 = \frac{1}{2}$.

Proposition 5. *Fix an agents' information structure. If the principal's prior is sufficiently low, the gateway referendum maximizes the principal's payoff guarantee across all equilibrium sequences among all monotone generalized referenda with the single cutoff* $m_1 = \frac{1}{2}$.

5.3 Heterogeneous Ex-Post Preferences

In the baseline model, all players agree on the best policy once the state is known. Such unanimity is often unrealistic. For instance, a policy with distributive consequences may benefit different groups depending on the realized state; see, e.g., Fernandez and Rodrik (1991) or Ali *et al.* (2025). Similarly, the principal's ex-post ranking of policies may conflict with that of a nontrivial share of agents.

In this section we extend the model to account for these possibilities by allowing for state-dependent preference heterogeneity. In addition to partisans and "aligned" types who prefer higher policies in state 1 and lower ones in state 0, we allow for "opposed" types whose policy ranking is the reverse: They prefer lower policies in state 1 and higher ones in state 0. The fact that their preferences are strictly opposite to those of the principal in each state suggests that they have strong incentives to misrepresent or withhold information.

Formally, let us denote a non-partisan's type by $(p_i, \boldsymbol{t}_i)$, where $p_i \in [0, 1]$ is his prior and $\boldsymbol{t}_i = (t_i(0), t_i(1)) \in [0, 1]^2$ captures his marginal benefit from the policy in each state. The marginal cost $0 < c < 1$ is common, as in the baseline model. The opposed types are those with $t_i(0) > c > t_i(1)$.

Despite the extended scope for misrepresentation of private information in this model, all of our main results—in particular, those on information aggregation (Proposition 2)

---

[23] Theorem 2 has established this payoff guarantee under the otherwise standing assumption that $\Pr(\omega = 1) > \frac{1}{2}$. The same proof, with minor and straightforward modifications, extends to the case $\Pr(\omega = 1) < \frac{1}{2}$.

and robust-optimality (Theorems 1 and 2)—carry over provided the agents' aggregate preferences satisfy a certain monotonicity condition. (The proofs are in the online appendix.) A central observation is that, even with heterogeneous ex-post preferences, the aggregate best response can be summarized by a map Φ, which varies only in a one-dimensional incentive index. This map Φ is exogenously given via the type distribution.

Let us explain this observation. Fix a strategy profile η, and let $U(0;\eta)$ and $U(1;\eta)$ denote the average effect of an additional action 1 on the final policy in states 0 and 1, respectively. For an individual agent of type $(p_i, \boldsymbol{t}_i)$ with signal s, action 1 is a best response if

$$U(0;\eta)(1-p_i)\Pr(s_i = s|\omega = 0)\Big(t_i(0) - c\Big) + U(1;\eta)p_i \Pr(s_i = s|\omega = 1)\Big(t_i(1) - c\Big) \geq 0.$$

The map Φ summarizing the aggregate best response is defined as follows. Fix s (or equivalently, fix the likelihood ratio $l := \frac{\Pr(s_i=s|\omega=0)}{\Pr(s_i=s|\omega=1)} \in (0,\infty)$). We let $\Phi\big(U(0;\eta), U(1;\eta), l\big)$ denote the probability that a randomly drawn type $(p_i, \boldsymbol{t}_i)$ with signal s satisfies the above best-response condition.

The aggregate best response (the value of Φ at $\big(U(0;\eta), U(1;\eta), l\big)$) turns out to depend only on the single incentive index

$$z_1 := \frac{U(0;\eta)}{U(1;\eta)} \cdot l,$$

whenever $U(1;\eta) \neq 0$. This is evident from rewriting the above best-response condition for individual types: If $U(1;\eta) > 0$, then an agent of type $(p_i, \boldsymbol{t}_i)$ with signal s weakly prefers action 1 if

$$z_1 \cdot (1-p_i)\Big(c - t_i(0)\Big) \;\leq\; p_i\Big(t_i(1) - c\Big), \tag{16}$$

with the inequality reversed when $U(1;\eta) < 0$.

We now sketch how preference monotonicity matters. We say that preferences are *monotone* if, for each fixed sign of $U(1;\eta)$, the function Φ is continuously differentiable in z_1 and $\partial\Phi/\partial z_1$ has the same non-zero sign for all $z_1 \in (0,\infty)$.[24] Intuitively, monotonicity rules out non-monotone response patterns in which strengthening the incentive index z_1 could first increase and then decrease the share of types that favor the action 1. In other words, monotonicity ensures that aggregate best responses move in a single direction as incentives change.

Given preference monotonicity, we can modify the argument in the proof sketch for Proposition 2 to show that monotone referenda with a single cutoff and $\max R(0) < \max R(1)$ only have informative equilibria.

Suppose an equilibrium is uninformative. Then the principal learns nothing from the

---

[24] This monotonicity condition is an adaptation of the notion of "strong preference monotonicity" in Bhattacharya (2013).

realized collective action, so her equilibrium choice depends only on feasibility: Above the cutoff she selects $\max R(1)$, and below the cutoff she selects $\max R(0)$ (given her prior preference for high policies). Because $\max R(0) < \max R(1)$, an agent is then pivotal only at the cutoff m_1. The uninformativeness of the equilibrium implies that this pivotal event has the same probability in both states; thus the average effect of an additional action 1 is the same in both states, i.e. $U(1;\eta_N) = U(0;\eta_N) > 0$. However, since $U(0;\eta_N)/U(1;\eta_N) \in (0,\infty)$, the mean actions under the agents' best response satisfy $q(0;\sigma_N) \neq q(1;\sigma_N)$. This is a direct implication of (16) and the monotone likelihood ratio property of the signals, given preference monotonicity. It contradicts the initial assumption of uninformativeness.

This argument establishing informativeness is key. It allows us to repeat the proofs in the baseline analysis, with minor modifications, to show that all of our main results continue to hold in the heterogeneous-preference environment.

We conclude with three remarks. First, the monotonicity condition is satisfied in the baseline model, since all non-partisans share the same preference type. Second, non-monotonicity can overturn information aggregation of referenda for the same reason it can overturn the Condorcet jury theorem's aggregation result (see Bhattacharya, 2013). Third, the robust-optimal referenda of the form (4) and (7) also maximize the agents' ex-ante welfare guarantee, provided two additional conditions hold. The first is that the mean marginal benefit crosses the marginal cost across states,

$$0 \leq \mathrm{E}\big(t_i(0)\big) < c < \mathrm{E}\big(t_i(1)\big) \leq 1,$$

so that the principal and the mean agent share the same full-information ranking.[25] This is satisfied, for example, when the principal is a utilitarian social planner. Such incentives may arise from political-economy forces under a broad set of conditions, as seen in the literature on political agency and electoral accountability.[26] The second condition is

$$\frac{1-\mathrm{E}(p_i)}{\mathrm{E}(p_i)} \cdot \frac{c-\mathrm{E}\big(t_i(0)\big)}{\mathrm{E}\big(t_i(1)\big)-c} \geq \varepsilon,$$

which ensures that the agents' mean payoff from the constant policy outcome $x = 1$ remains bounded away from the mean full-information payoff.

---

[25] To evaluate the payoffs of partisan agents, we assume here that partisans hold extreme priors that conform to their prescribed choice $a_i \in \{0,1\}$, i.e. $p_i = a_i$.

[26] For the political agency literature, see Barro (1973) and Ferejohn (1986), among others; for the literature on electoral accountability, see, e.g., the survey in Ashworth (2012). Battaglini (2017) also provides an excellent discussion, with several explicit examples.

5.4 Non-Linear and Single-Plateau Preferences

In this section we relax the baseline model's assumption of constant marginal costs and benefits and consider non-linear payoffs $u(x, \omega)$. This extension broadens the scope of possible applications by allowing for intermediate policies to be optimal. Because the techniques developed for the baseline model cannot be directly transferred to this context, the analysis is much more involved; we therefore relegate it to the working paper version, Heese (2026). Here we introduce the setup, state the robustness result established there, and briefly explain the additional difficulties that this model presents, relative to the baseline model.

A natural example captured by our non-linear-payoff model is the problem of public-good provision with diminishing returns to scale, where providing the good is desirable but full capacity is inefficient. Similar examples arise in settings where compromises between opposing extremes are desirable.

We consider common payoffs $u(x, \omega)$, continuously differentiable in x, such that

- in state 0, the status quo $x = 0$ is optimal, with $u'(x, 0) < 0$ for all $x \in [0, 1]$;
- in state 1, a unique non-zero policy is optimal.

Letting $c(x) = -u(x, 0)$ and $b(x) = u(x, 1) - u(x, 0)$, we can express the payoffs as before in terms of cost and benefits:

$$u(x, \omega) = -c(x) + b(x)\omega \qquad \text{for } \omega \in \{0, 1\}.$$

The expected utility given a fixed prior p has a positive derivative at x if $\frac{c'(x)}{b'(x)} \leq p$. We focus on the case of "diminishing returns to scale," where[27]

$$\frac{c'(x)}{b'(x)} \text{ is strictly increasing.}$$

Under this condition, each player's ex-ante expected utility is single-plateau, with the peaks increasing in p. As in the baseline model, we assume that, given the principal's prior, there is a unique, non-zero optimal policy.

In the working paper version, we show that in this model, for a generic parameter region, the simple-majority gateway referendum aggregates information and achieves a payoff guarantee above $1 - \varepsilon$. Moreover, for parameter regions in which the prior-optimal policy coincides with the ex-post optimal policy in state 1, the simple-majority gateway referendum is robust-optimal. Thus, in these regions, the conclusion of Theorem 2 continues to hold.

The proof is not straightforward. In the baseline model, the principal's best-response correspondence can be described via a single cutoff $\bar{k}$. Under single-plateau preferences,

---

[27] The baseline model assumed that $\frac{c'(x)}{b'(x)}$ is constant.

by contrast, there may be multiple cutoffs $\bar{k}_{j+1,j}$ at which the principal's preference between adjacent policies x_j and x_{j+1} switches. This complicates the principal's side of the analysis.

It also complicates the agents' best-response problem, which is to compute how, in each state, choosing action 1 instead of 0 will affect the principal's policy choice and thus the average payoffs

$$U(1;\eta) := \mathrm{E}\Big(u(x,1)|a_i=1;\eta,N\Big) - \mathrm{E}\Big(u(x,1)|a_i=0,\eta,N\Big), \tag{17}$$

$$U(0;\eta) := -\Big(\mathrm{E}\Big(u(x,0)|a_i=1;\eta,N\Big) - \mathrm{E}\Big(u(x,0)|a_i=0,\eta,N\Big)\Big); \tag{18}$$

cf. (3). In addition, because payoffs are now non-linear in x, a policy switch from x_j to x_{j+1} need not have proportional implications across states (as it does in the baseline model).

Nevertheless, as the working paper shows, the strict monotonicity of $\frac{c'(x)}{b'(x)}$ imposes enough structure on the principal's and agents' incentives to recover information aggregation and a payoff guarantee above $1-\varepsilon$, for the simple-majority gateway referendum.

5.5 Stochastic Commitment: An Interpretation of the Baseline Model

The baseline model can also be interpreted as a model of supermajority voting between the policies $x'=0$ and $x'=1$ under stochastic commitment. The stochastic model retains the common payoff specification

$$x'\left(\omega-\frac{1}{2}\right) \quad \text{for } x' \in \{0,1\} \text{ and } \omega \in \{0,1\},$$

as well as the specification of the agents' side, from the baseline model. The only change is in how outcomes are determined. With probability $1-\varepsilon'$ the principal is bound by a supermajority rule and must choose $x'=0$ if $m \leq m_1$ and $x'=1$ if $m > m_1$. With the remaining probability $\varepsilon' > 0$, however, she is unconstrained and can choose either $x'=0$ or $x'=1$ after observing the collective action $m=\frac{k}{N}$.

This implies that, after the realization of any vote share m, a pure uncommitted choice induces an outcome probability

$$x(k) = \Pr(x'=1|k) \in \begin{cases} \{0,\varepsilon'\} & \text{if } m \leq m_1, \\ \{1-\varepsilon',1\} & \text{if } m > m_1. \end{cases}$$

Hence the principal's set of pure strategies is the same as in the baseline model with referendum R given by

$$R(m) = \begin{cases} \{0, \varepsilon'\} & \text{if } m \leq m_1, \\ \{1 - \varepsilon', 1\} & \text{if } m > m_1. \end{cases}$$

Because the common payoffs are linear in the probability of implementing $x' = 1$, each of the principal's strategies in the stochastic model has the same payoff consequences as its counterpart in the baseline model. It follows that the two models are strategically equivalent and therefore have the same set of equilibria.

6 Literature Discussion: Commitment in Policy-Making

We relate our analysis of partial commitment to two categories of models appearing in the information-aggregation literature: (i) models with full commitment to voting rules that map vote profiles to single policies, and (ii) models in which collective input is cheap talk (no commitment), which have been used to study informal political processes such as protests. These two strands of the literature provide natural benchmarks for our analysis. With regard to the first, our results show that coordination failures are more germane to policy-making under dispersed information than the full-commitment benchmark suggests. With regard to the second, they show that formality (commitment) matters less for information aggregation than existing results suggest.

Under full commitment, standard results such as the Condorcet jury theorem and its extensions show that, as $N \to \infty$, majority voting leads to the full-information outcome with probability approaching one under broad conditions (see, e.g., Feddersen and Pesendorfer, 1997, and Bhattacharya, 2013). These results provide a benchmark in which strategic coordination problems among agents disappear when the population is large enough. Our analysis implies that this benchmark result is fragile to arbitrarily small departures from full commitment. Here and throughout, we distinguish information aggregation (the principal learns the state) from full-information equivalence of outcomes.

This fragility can be seen most starkly in the interpretation of our baseline model as a model of majority voting with stochastic commitment, discussed in Section 5.5. Recall that in the stochastic model, the principal is bound by a majority rule with probability $1 - \varepsilon'$ and is unconstrained with probability ε'.

Under this interpretation, our baseline analysis implies that, for any $\varepsilon' > 0$, any supermajority rule admits equilibrium sequences (for some information structures) in which the policy $x' = 1$ is chosen in all states with probability at least $1-\varepsilon'$ as $N \to \infty$ (Propositions 1 and 3). Hence even a minimal likelihood of discretion can enlarge the equilibrium set in a way that upsets full-information equivalence. This observation complements work that identifies coordination failures and inefficiencies in related information-aggregation environments (see, e.g., Ekmekci and Lauermann, 2020, Mandler, 2012, Ali *et al.*, 2025, and Barelli, Bhattacharya and Siga, 2022).

In the literature on cheap talk, models without commitment have been used to study informal political processes such as protests and compare their properties with those of formal voting rules. Impossibility results in that literature show that when a large group of agents strategically communicates with a principal, only babbling equilibria exist under broad conditions (see, e.g., Morgan and Stocken, 2008, Battaglini, 2017, Levit and Malenko, 2011, and Chen, 2025). In our setting, by contrast, even procedures with almost no commitment robustly aggregate information in all equilibrium sequences as $N \to \infty$ (Proposition 2). Thus, if commitment is interpreted as a measure of how formal a political process is, our results imply that formality matters less for information aggregation than the impossibility results suggest.

To understand the logic behind the cheap-talk impossibility results, consider the following example (couched in the notation of this paper). Let R denote a cheap-talk mechanism (for example, $R(m) = \mathcal{P}$ for all m, so the collective action never constrains the principal). Suppose the principal's prior odds are $\frac{\Pr(\omega=1)}{\Pr(\omega=0)} = 9$, each agent has a lower prior with odds $\frac{1}{2}$, and a favorable signal $s_i = 1$ has likelihood ratio 3, while an unfavorable signal $s_i = 0$ has likelihood ratio $\frac{1}{3}$. Suppose moreover that each player prefers high policies whenever her posterior likelihood ratio exceeds 1.

Now consider the point of view of some agent i. Write m_{-i} for the other agents' reports, and let $\bar{k}$ denote the principal's cutoff, so that she chooses the high policy 1 if and only if the total number of high actions is at least $\bar{k}+1$ and the low policy 0 otherwise. At any profile m_{-i} at which agent i is pivotal, the principal's posterior odds conditional on m_{-i} and the action 1 from agent i must be at least 1, whereas her posterior odds conditional on m_{-i} and the action 0 must be below 1. This means

$$9 \cdot \frac{\Pr(m_{-i} \mid \omega = 1)}{\Pr(m_{-i} \mid \omega = 0)} \cdot 3 \;\geq\; 1 \;>\; 9 \cdot \frac{\Pr(m_{-i} \mid \omega = 1)}{\Pr(m_{-i} \mid \omega = 0)} \cdot \frac{1}{3}.$$

Equivalently,

$$\frac{1}{27} \leq \frac{\Pr(m_{-i} \mid \omega = 1)}{\Pr(m_{-i} \mid \omega = 0)} < \frac{1}{3}.$$

Now consider an agent who received a signal $s_i = 1$. Since his prior odds are $1/2$, his posterior odds conditional on the pivotal profile and the signal satisfy

$$\frac{\Pr_i(\omega = 1 \mid m_{-i}, s_i = 1)}{\Pr_i(\omega = 0 \mid m_{-i}, s_i = 1)} = \frac{\Pr_i(\omega = 1)}{\Pr_i(\omega = 0)} \cdot \frac{\Pr(m_{-i} \mid \omega = 1)}{\Pr(m_{-i} \mid \omega = 0)} \cdot \frac{\Pr(s_i = 1 \mid \omega = 1)}{\Pr(s_i = 1 \mid \omega = 0)} < \frac{1}{2} \cdot \frac{1}{3} \cdot 3 = \frac{1}{2}.$$

So even an agent with a signal 1 strictly prefers the report 0 at any pivotal profile. All non-partisan agents therefore babble, precluding informativeness. This is the core pivotal-inference logic in the impossibility results.

Under partial commitment, by contrast, agents are pivotal in multiple pivotal events, not just at the event described above where the principal's preference changes. This allows informative equilibria to survive even when the usual impossibility conditions on the players' priors hold and even when commitment is minimal.[28]

7 Further Related Literature

This paper is not the first to study collective choice environments with multiple pivotal events. In particular, Razin (2003) studies an environment in which multiple pivotal events arise because voters have signaling incentives—i.e., their actions may signal information to second-stage decision-makers—as in our setting. In his setting, information aggregation fails when decision-makers and agents have disjoint sets of ex-post optimal policies. Our results differ in that we do not focus on such extreme ex-post preference conflicts; instead, the inefficiencies in our model stem from coordination failures that can prevent full-information equivalence *despite* information aggregation, and we characterize referenda that robustly avoid such failures.

Other works have analyzed different types of signaling incentives,[29] as well as settings in which multiple pivotal events arise for reasons other than signaling, e.g. because of simultaneous voting on multiple issues (Ahn and Oliveros, 2012) or the existence of multiple voting thresholds (Damiano *et al.*, 2025).

To deal with the complexity created by the presence of multiple pivotal events, most existing work relies primarily on statistical tools such as large-deviation theory. Our paper develops new technical tools for this problem, combining large-deviation arguments with novel polynomial-algebraic methods. Specifically, we characterize information aggregation by using Descartes' rule of signs to bound the number of roots of a certain associated polynomial; see for example the proofs of Lemma 2 and Proposition 2 in the appendix.

We conclude our discussion of the literature with two remarks. First, a longstanding question in political science is the extent to which citizens vote "sincerely" (see, e.g., Farquharson, 1969 and Palfrey, 2009). Previous theoretical work has shown that in elections with full commitment power and large electorates, sincere voting typically fails to be an equilibrium strategy when voters are privately informed (see, e.g., Austen-

---

[28] Our model also differs from the existing work in allowing for heterogeneous priors among the agents, drawn from a distribution with full support, rather than assuming a common prior type. However, even in a common-type version of our baseline model, the presence of multiple pivotal events implies the existence of informative equilibria for a broad class of referenda, including collective vetoes of the maximum.

[29] We allude here to studies of settings in which early-stage collective behavior serves a signaling role. For example, in multiple-round voting, a first-round vote can convey information to second-round voters (Piketty, 2000) or shape candidates' second-round proposals (Castanheira, 2003; Meirowitz and Shotts, 2009). Likewise, pre-election polls (Coughlan, 2000; Fey, 1997) or costly protest activity (Lohmann, 1994) may transmit information that strategically affects subsequent political choices.

Smith and Banks, 1996). In contrast, Proposition 1 and its proof show that, under partial commitment, strategy profiles in which approximately all agents act sincerely can constitute an equilibrium for a nontrivial range of information structures. Roughly speaking, the presence of multiple pivotal events weakens the incentives to deviate from sincere behavior that arise under full commitment. Related observations have been made in settings with participation costs (Krishna and Morgan, 2012) or aggregate uncertainty about the fraction of uninformed voters (Acharya and Meirowitz, 2017).

Finally, we note a conceptual connection to the delegation literature (Alonso and Matouschek, 2008; Holmström, 1984). A referendum can be viewed as a form of *collective delegation* to a principal: The collective action determines the feasible set the principal faces ex post. This perspective connects delegation-style reasoning to settings in which information is dispersed among many agents and to collective-choice applications such as shareholder voting and public finance.

8 Concluding Discussion

Referenda are among the most prominent instruments through which collective input enters policy-making. Many referenda do not directly select a policy; they instead reshape the space of policies from which a principal may choose. This paper studies such processes as institutions of partial commitment and asks how their structure shapes policy outcomes when policy-relevant information is dispersed among a large population of strategic agents.

Our analysis shows that there is a case for minimal commitment. Minimal commitment creates strong enough incentives to sustain robust information aggregation; at the same time, it limits the extent to which coordination failures among the agents distort outcomes. Stronger commitment creates greater exposure to such failures and therefore magnifies the distortions they generate.

The paper identifies two simple forms of referenda, both widely seen in practice, that implement this logic: collective vetoes and gateway referenda. These processes allow the collective to block or authorize one polar option while leaving the detailed policy specification to a principal. When the policy space is fine, they deliver outcomes arbitrarily close to the full-information benchmark. They also maximize the principal's worst-case payoff guarantee across information structures and equilibria, within a broad class of decision processes.

Our analysis isolates one rationale for limited commitment, but it is unlikely to be the only one relevant in practice. Limited commitment may also help to provide legitimacy, accountability, or political cover, or to preserve discretion for a principal with expertise or access to information that becomes relevant only after collective input has been observed (cf. Kartik, Van Weelden and Wolton, 2017). However, our results show that the trade-off described above (between information aggregation and coordination failures) is sufficient to rationalize the use of minimal commitment.

Our model of referenda is meant as a first step. Future work could incorporate additional structure that is present in particular applications. For example, in shareholder voting, what factors determine how policy is decided after collective input is observed? Answering this question may require the introduction of legislative bargaining, competing principals, or other strategic specification problems. Such extensions would further clarify how policy outcomes depend on the institutional environment in which final policy is chosen—that is, on the micro-foundations of partial commitment.

More generally, the concept of partial commitment developed here may prove useful beyond the setting studied in this paper. The interpretations discussed in the introduction—in terms of the formality of political processes, the distinction between direct and indirect democracy, and mandates for a strategic principal—point to several possible directions. So do other collective-choice settings in which commitment is partial but the central problem is less one of information aggregation than of preference aggregation.

Appendix

Throughout the appendices, the following disclaimers apply.

1. For the principal, we only consider mixed strategies that are best responses. Given the best response characterization (1), we can thus identify a principal's mixed strategy with a pair $(\bar{k}, \tilde{x})$ where $\bar{k}+1$ is the principal's cutoff at which she mixes and chooses a policy x randomly according to the distribution $\tilde{x}$.
2. We identify converging subsequences with the original sequence and omit the subsequence notation. Converging subsequences exist in each instance since the sequences lie in compact sets.

A Mathematical Preliminaries

We collect the tools that will be used repeatedly in the proofs below. In particular, we will approximate probabilities of pivotal events as $N \to \infty$.

A.1 Basics of Large Deviation Theory

Take a binomial distribution X_n with success probability $q \in (0,1)$ and sample size n. Given any $m \in (0,1)$ with $mn \in \mathbb{N}$, the probability of exactly mn successes out of n trials is well known to be[30]

$$\Pr\left(X_n = mn\right) = \exp\Big(-n\mathrm{KL}(m,q) + o(n)\Big), \tag{19}$$

---

[30] Recall that a function f is $o\left(n\right)$ if $\frac{|f(n)|}{n}$ converges to 0 as $n \to \infty$.

where KL denotes the Kullback–Leibler divergence,

$$\mathrm{KL}(m,q) = m \log\left(\frac{m}{q}\right) + (1-m)\log\left(\frac{1-m}{1-q}\right).$$

A convenient derivation, due to Cramér (1938), uses a change of measure. Consider the binomial distribution under which the event is not rare but rather typical, $Z_n \sim \mathcal{B}(n,m)$. Then (19) follows from observing that

$$\frac{\Pr\left(X_n = mn\right)}{\Pr(Z_n = mn)} = \exp\Big(-n\mathrm{KL}(m,q)\Big) \text{ and } \Pr(Z_n = mn) = \exp\Big(o(n)\Big). \tag{20}$$

For the equation on the left, note that

$$\begin{aligned}\frac{\Pr(Z_n = mn)}{\Pr\left(X_n = mn\right)} &= \left(\frac{m}{q}\right)^{nm}\left(\frac{1-m}{1-q}\right)^{n(1-m)} \\ &= \exp\left(\log\left(\left(\frac{m}{q}\right)^{nm}\left(\frac{1-m}{1-q}\right)^{n(1-m)}\right)\right) \\ &= \exp\left(n\left(m\log\left(\frac{m}{q}\right) + (1-m)\log\left(\frac{1-m}{1-q}\right)\right)\right).\end{aligned}$$

The equation on the right holds because the probability mass function (PMF) of the binomial distribution peaks at its mean, implying $\Pr(Z_n = mn) \in [\frac{1}{n}, 1]$. But for any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in [\frac{1}{n}, 1]$, it holds that $x_n = \exp\Big(\log(x_n)\Big) = \exp\Big(o(n)\Big)$.

A.2 Taylor Approximations of the Kullback–Leibler Divergence

Below we give two approximations of the Kullback–Leibler divergence

$$\mathrm{KL}(m,q) = m \log\left(\frac{m}{q}\right) + (1-m)\log\left(\frac{1-m}{1-q}\right).$$

The first is for $m \approx q$.[31] For $m = q + \varepsilon'$ with small ε', we expand the log terms using the Taylor expansion $\log(1+x) \approx x - \frac{x^2}{2}$ around $x = 0$ to obtain

$$\begin{aligned}\log\frac{m}{q} &= \log\left(1 + \frac{\varepsilon'}{q}\right) \approx \frac{\varepsilon'}{q} - \frac{(\varepsilon')^2}{2q^2}, \\ \log\frac{1-m}{1-q} &= \log\left(1 - \frac{\varepsilon'}{1-q}\right) \approx -\frac{\varepsilon'}{1-q} - \frac{(\varepsilon')^2}{2(1-q)^2},\end{aligned}$$

and substitute:

$$\begin{aligned}\mathrm{KL}(m,q) \approx (q+\varepsilon')&\left(\frac{\varepsilon'}{q} - \frac{(\varepsilon')^2}{2q^2}\right) \\ &+ (1-q-\varepsilon')\left(-\frac{\varepsilon'}{1-q} - \frac{(\varepsilon')^2}{2(1-q)^2}\right).\end{aligned}$$

---

[31] For two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we write $a_n \approx b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Note that we do not retain the subscript.

Expanding the products and discarding all third-order terms, we have

$$(q+\varepsilon')\left(\frac{\varepsilon'}{q}-\frac{(\varepsilon')^2}{2q^2}\right)\approx\varepsilon'-\frac{(\varepsilon')^2}{2q}+\frac{(\varepsilon')^2}{q},\ \text{ and}$$
$$(1-q-\varepsilon')\left(-\frac{\varepsilon'}{1-q}-\frac{(\varepsilon')^2}{2(1-q)^2}\right)\approx-\varepsilon'-\frac{(\varepsilon')^2}{2(1-q)}+\frac{(\varepsilon')^2}{1-q}.$$

Noticing that ε' cancels out, and simplifying the coefficients of $(\varepsilon')^2$, we have

$$\begin{aligned}\mathrm{KL}(m,q)&\approx\left(-\frac{(\varepsilon')^2}{2q}+\frac{(\varepsilon')^2}{q}\right)+\left(-\frac{(\varepsilon')^2}{2(1-q)}+\frac{(\varepsilon')^2}{1-q}\right)\\&=\frac{(\varepsilon')^2}{2q}+\frac{(\varepsilon')^2}{2(1-q)}\\&=\frac{(\varepsilon')^2}{2}\left(\frac{1}{q}+\frac{1}{1-q}\right).\end{aligned}$$

We thus obtain the quadratic approximation

$$\mathrm{KL}(m,q)\approx\frac{(m-q)^2}{2q(1-q)}\ \text{ for }\ m\approx q. \tag{21}$$

For the second approximation, consider $q_1\approx q_2$. Note that $\frac{\partial}{\partial q}\mathrm{KL}(m,q)=-\frac{m}{q}+\frac{1-m}{1-q}$. We use linear Taylor approximations of $\mathrm{KL}(m,q_1)$ and $\mathrm{KL}(m,q_2)$ around the midpoint $\bar{q}=\frac{q_1+q_2}{2}$,

$$\begin{aligned}\mathrm{KL}(m,q_1)&\approx\mathrm{KL}(m,\bar{q})+\frac{\partial}{\partial q}\mathrm{KL}(m,q)_{|q=\bar{q}}(q_1-\bar{q}),\\\mathrm{KL}(m,q_2)&\approx\mathrm{KL}(m,\bar{q})+\frac{\partial}{\partial q}\mathrm{KL}(m,q)_{|q=\bar{q}}(q_2-\bar{q}),\end{aligned}$$

to approximate the difference of these two quantities:

$$\begin{aligned}\mathrm{KL}(m,q_1)-\mathrm{KL}(m,q_2)&\approx\frac{\partial}{\partial q}\mathrm{KL}(m,q)_{|q=\bar{q}}(q_1-q_2)\\&=\left(\frac{1-m}{1-\bar{q}}-\frac{m}{\bar{q}}\right)(q_1-q_2).\end{aligned}\tag{22}$$

A.3 Pivotal Probabilities

The relevance of the mathematical preliminaries in Sections A.1-A.2 for our collective choice model derives from the fact that, for any symmetric strategy σ of the agents, the number of actions 1 taken by $N-1$ agents follows a binomial distribution with success probability $q(\omega';\sigma)=(1-\rho_1-\rho_0)\mathrm{E}\Big(\sigma(s,p)|\omega=\omega'\Big)+\rho_1$. The pivotal events in our model thus correspond to point events of a binomial distribution, and (19) provides a suitable approximation of their likelihood. To be precise, if we let $q=q(\omega';\sigma)$ and $m=\frac{\lfloor m_jN\rfloor}{N}$ for $j>0$, then (19) becomes

$$\Pr(\mathrm{piv}_j|\omega=\omega';\sigma,N)=\exp\Big(-(N-1)\mathrm{KL}\Big(\frac{\lfloor m_jN\rfloor}{N},q(\omega';\sigma)\Big)+o\left(N\right)\Big). \tag{23}$$

Similarly, if we let $q = q(\omega'; \sigma)$ and $m = \frac{k}{N}$ for $k \in \{\bar{k}, \bar{k}+1\}$, then (19) becomes

$$\Pr(\text{piv}^k | \omega = \omega'; \sigma, N) = \exp\Big(-(N-1)\text{KL}\Big(\frac{k}{N}, q(\omega'; \sigma)\Big) + o(N) \Big). \tag{24}$$

B Information Aggregation

B.1 Proof of Lemma 2

Recall that Proposition 2's proof sketch in the main text established that any equilibrium is informative.

To prove Lemma 2, we show that the average effect cannot equal zero in both states,

$$U(0; \eta_N) \neq 0 \text{ or } U(1; \eta_N) \neq 0. \tag{25}$$

Lemma 2 follows from (25) since $U(0; \eta_N)$ and $U(1; \eta_N)$ must have the same sign. If the two average effects had opposite signs, then the best-response inequality (3) would point *all* non-partisan types in the same direction regardless of their signal, producing an uninformative best response and contradicting the informativeness of equilibrium.

Before we establish (25), a preliminary observation: Any principal's equilibrium strategy is necessarily non-constant on average, i.e. $\text{E}(x(k))$ is non-constant in k. Suppose it were not. The principal is never indifferent except possibly when observing $k = \bar{k}+1$, hence chooses a boundary point of the feasible set otherwise. Constant $\text{E}\Big(x(k)\Big)$ would mean chosen boundaries of the lower and upper feasible set $R(0)$ and $R(1)$ that are equal. Since $\min R(0) < \max R(0) < \max R(1)$ and $\min R(1) \neq \max R(0)$, equality implies the choice is $\text{E}\Big(x(k)\Big) = \min R(0) = \min R(1)$. However, this implies a contradiction: By Bayes-consistency, the prior $\Pr(\omega = 1) > \frac{1}{2}$ implies k with $\Pr(\omega = 1|k) > \frac{1}{2}$ and thus $\text{E}\Big(x(k)\Big) \in \{\max R(0), \max R(1)\}$, a contradiction.

Now we establish (25): For each k, let $r_N(k) := \text{E}\Big(x(k+1)\Big) - \text{E}\Big(x(k)\Big)$ denote the change in the principal's expected policy when the observed number of actions 1 increases from k to $k+1$. For any given agent, the number k_{-i} of actions 1 among the other agents is Binomial, so we can write the average effect of an additional action 1 as a binomial-weighted sum:

$$U(\omega'; \eta_N) = \sum_{k \in \{\lfloor m_1 N \rfloor, \bar{k}_N, \bar{k}_N+1\}} \binom{N-1}{k} q(\omega'; \sigma_N)^k \big(1 - q(\omega'; \sigma_N)\big)^{N-1-k} r_N(k).$$

Introduce the odds ratio $t := \frac{q(\omega'; \sigma_N)}{1 - q(\omega'; \sigma_N)} \in (0, \infty)$. Since $q^k(1-q)^{N-1-k} = (1-q)^{N-1} t^k$, the condition $U(\omega'; \eta_N) = 0$ is equivalent to a single polynomial equation in t,

$$P_N(t) := \sum_{k \in \{\lfloor m_1 N \rfloor, \bar{k}_N, \bar{k}_N+1\}} \binom{N-1}{k} r_N(k) t^k = 0,$$

where the polynomial's coefficients are the (scaled) policy jumps $\binom{N-1}{k} r_N(k)$. Suppose the coefficients at $\bar{k}, \bar{k}+1$, and $\lfloor m_1 N \rfloor$ are all non-zero. Then, the two jumps at $\bar{k}_N$ and $\bar{k}_N + 1$ have the same sign (if $q(0; \sigma_N) \leq q(1; \sigma_N)$, then $\Pr(\omega = 1|k = \bar{k}; \sigma_N) <$

$\frac{1}{2} \leq \Pr(\omega = 1|k = \bar{k}+1; \sigma_N)$ and both are weakly positive; if $q(0;\sigma_N) > q(1;\sigma_N)$, both are weakly negative), so the coefficient sequence has *at most one sign change.* This is trivially true if there are at most two non-zero coefficients. Further, since the principal's strategy is non-constant, the polynomial has *some* non-zero coefficients. By Descartes' rule of signs, such a polynomial has at most one positive real root. Since any equilibrium is informative ($q(0;\sigma_N) \neq q(1;\sigma_N)$), the corresponding odds ratios t differ across states, hence $U(0;\eta_N)$ and $U(1;\eta_N)$ cannot both be zero.

B.2 Proof of Proposition 2

Take any equilibrium sequence $(\eta_N)_{N\in\mathbb{N}}$ of a referendum meeting the conditions of Proposition 2. Using Lemma 2, the main text's sketch of proof already established the existence of unique indifferent types $0 < p_N(s) < 1$ for any signal s and any N; cf. (12).

Suppose that the limit indifferent types are interior, i.e.,

$$0 < \lim_{N\to\infty} p_N(1) < \lim_{N\to\infty} p_N(0) < 1. \tag{26}$$

Then the limit of the mean action differs across signals and thus across the two states, i.e., $0 < \lim_{N\to\infty} q(0;\sigma_N) \neq \lim_{N\to\infty} q(1;\sigma_N) < 1$. By an application of the law of large numbers, the realized share of actions 1 concentrates around the mean action in each state, implying that the principal learns the state from observing it. Thus, information aggregates.

It remains to show that information aggregation is also implied by the negation of (26). Note that the negation of (26) implies

$$\lim_{N\to\infty} p_N(0) = \lim_{N\to\infty} p_N(1) \in \{0,1\}. \tag{27}$$

We lead (27) to a contradiction in a generic case (Case 1), and show that it implies information aggregation in the complementary non-generic case (Case 2).

The arguments are based on a detailed analysis of point events for the realized number k_{-i} of actions 1 of the other agents: For any sequence $(m_N)_{N\in\mathbb{N}}$ with $m_N N \in \mathbb{N}$ for all N, we apply (19) to obtain

$$\Pr\Big(k_{-i} = m_N N|\omega = \omega'; \sigma_N, N\Big) = \exp\Big(-(N-1)\mathrm{KL}(m_N, q(\omega';\sigma_N)) + o(N)\Big),$$

and the left equation in (20) to obtain

$$\begin{aligned}\frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} &= \frac{\Pr(\omega = 1)}{1-\Pr(\omega = 1)} \\ &\cdot \exp\Big((N-1)\Big(\mathrm{KL}\Big(m_N, q(0;\sigma_N)\Big) - \mathrm{KL}\Big(m_N, q(1;\sigma_N)\Big)\Big)\Big).\end{aligned} \tag{28}$$

Specifically, we will consider the sequences given by $m'_N = \frac{\lfloor m_1 N\rfloor}{N}$, $m''_N = \frac{\bar{k}_N}{N}$, and $m'''_N = \frac{\bar{k}_N+1}{N}$, with $\bar{k}_N$ being the unique number satisfying (1). These sequences correspond to the pivotal events piv_1, $\mathrm{piv}_{0,\bar{k}_N}$, and $\mathrm{piv}_{0,\bar{k}_N+1}$. We make three preliminary observations:

First, as long as $\bar{k}_N \neq N$ for all N large enough, we have

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in (\kappa, \frac{1}{\kappa}) \text{ for } m_N \in \{m''_N, m'''_N\}, \tag{29}$$

for some $\kappa > 0$, by the defining property (1) of $\bar{k}_N$ (whenever we apply (29), we will rule out the case $\bar{k}_N = N$). Second, (27) implies

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in \{0, \infty\} \text{ for } m_N = m'_N. \tag{30}$$

Otherwise the inference from piv_1 would be bounded as $N \to \infty$. Then, for any N large enough, we would have either $\bar{k}_N = N$, so that only piv_1 would be relevant for the agents' best response, or, by (29), the inference from piv_0 would also be uniformly bounded. In either case, the indifferent types would be bounded away from 0 and 1.[32] Third, (27) implies $\lim_{N\to\infty} q(0; \sigma_N) = \lim_{N\to\infty} q(1; \sigma_N)$. We set

$$q^* = \lim_{N\to\infty} q(\omega'; \sigma_N), \text{ and } \Delta_N = q(1; \sigma_N) - q(0; \sigma_N) \text{ for any } N.$$

Case 1: $m_1 \neq q^*$ [33]

In this case we derive a contradiction to (27). For $\gamma > 0$, let $m_N^+(\gamma) = q^* + \gamma$ and $m_N^-(\gamma) = q^* - \gamma$. For any sequence $(m_N)_{N\in\mathbb{N}}$, when $\lim_{N\to\infty} m_N \neq q^*$, the linear approximation (22) applies, so that

$$(N-1)\Big(\text{KL}\Big(m_N, q(0;\sigma_N)\Big) - \text{KL}\Big(m_N, q(1;\sigma_N)\Big)\Big) \approx -(N-1)\Big(\frac{1-m_N}{1-q^*} - \frac{m_N}{q^*}\Big)\Delta_N.$$

For $m_N = m'_N$, the unbounded inference (30) implies

$$\lim_{N\to\infty} (N-1)\Delta_N \in \{-\infty, \infty\}.$$

We show case by case that

$$\lim_{N\to\infty} m''_N = \lim_{N\to\infty} m'''_N = q^*. \tag{31}$$

First suppose $\lim_{N\to\infty}(N-1)\Delta_N = \infty$. By definition, this implies $q(0;\sigma_N) < q(1;\sigma_N)$ for large N. Using the above linear approximation for $m_N = m_N^{\pm}(\gamma)$, we see that for any $\gamma > 0$,

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = \lfloor m_N N\rfloor; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = \lfloor m_N N\rfloor; \sigma_N, N)} = 0 \text{ for } m_N = m_N^-(\gamma),$$

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = \lfloor m_N N\rfloor; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = \lfloor m_N N\rfloor; \sigma_N, N)} = \infty \text{ for } m_N = m_N^+(\gamma).$$

---

[32] The indifferent types are pinned down by (12). Given (29) and (30), boundedness of the indifferent type would follow from setting $a = \Pr(\text{piv}_0|\omega = 0; \eta_N, N)$, $b = \Pr(\text{piv}_0|\omega = 1; \eta_N, N)$, $c = \Pr(\text{piv}_1|\omega = 0; \eta_N, N)$, $d = \Pr(\text{piv}_1|\omega = 1; \eta_N, N)$, and using the fact that for any $u, v, a, b, c, d > 0$, we have $\min(\frac{a}{b}, \frac{c}{d}) \leq \frac{au+cv}{bu+dv} \leq \max(\frac{a}{b}, \frac{c}{d})$.

[33] In the main text we assert that this is a generic case. This is true because (27) implies $q^* \in \{\rho_1, 1-\rho_0\}$.

In particular, there exist both collective actions k for which the principal's posterior $\Pr(\omega = 1|k;\sigma_N, N)$ exceeds $\frac{1}{2}$, and others for which it does not. Hence, $\bar{k}_N \neq N$ for large N. The monotonicity of the posterior further implies $\lim_{N\to\infty} m''_N \in (m^-_N(\gamma), m^+_N(\gamma))$ for all $\gamma > 0$, from which the claim (31) follows. The case in which $\lim_{N\to\infty}(N-1)\Delta_N = -\infty$ holds is analogous. Since $m_1 \neq q^*$, the relevant divergences differ in the limit, i.e.,

$$0 = \lim_{N\to\infty} \mathrm{KL}\Big(m_N, q(\omega';\sigma_N)\Big) < \lim_{N\to\infty} \mathrm{KL}\Big(m_1, q(\omega';\sigma_N)\Big)$$

for $m_N = m''_N$ and $m_N = m'''_N$. Thus,

$$\lim_{N\to\infty} \frac{\Pr(k_{-i} = m_N N|\omega, \sigma_N, N)}{\Pr(k_{-i} = m'_N N|\omega;\sigma_N, N)} = \infty$$

for all ω, $m_N = m''_N$, and $m_N = m'''_N$. Since the inference from each of m''_N and m'''_N is bounded, by (29), this implies interior limit cutoffs (cf. (2) and (12)), contradicting the initial assumption (27).

Case 2: $m_1 = q^*$ This case can be decomposed into several analogous subcases; we present one. Consider an equilibrium sequence such that $\lim_{N\to\infty} \Pr(\omega = 1|\mathrm{piv}_1;\sigma_N, N) = 1$, and such that, for any N, a type chooses the action 1 if and only if $p_i \leq p_N(s)$. Then, since $p_N(1) < p_N(0) \in (0, 1)$, it holds that $\rho_1 < q(1;\sigma_N) < q(0;\sigma_N) < 1 - \rho_0$. Recall that $p_N(s) \to 0$ or $p_N(s) \to 1$ by the assumption (27). If $p_N(s) \to 1$, then $q^* = 1 - \rho_0$. However, then $\Pr(\mathrm{piv}_1|\omega = 1;\sigma_N, N) \leq \Pr(\mathrm{piv}_1|\omega = 0;\sigma_N, N)$ and $\lim_{N\to\infty} \Pr(\omega = 1|\mathrm{piv}_1;\sigma_N, N) = 1$ cannot hold. Thus, $p_N(s) \to 0$, which implies $q^* = \rho_1$.

We now carefully examine the mean actions in each state,

$$q(\omega';\sigma_N) - m_1 = (1 - \rho_1 - \rho_0)\Big(\sum_{s=0,1} \Pr(s_i = s|\omega = \omega')\Big(F\Big(p_N(s)\Big)\Big)\Big) + \rho_1 - \rho_1.$$

Using simple algebra,[34] we see that

$$\frac{\Pr(s_i = 0|\omega = 1)}{\Pr(s_i = 0|\omega = 0)} \leq \lim_{N\to\infty} \frac{q(1;\sigma_N) - m_1}{q(0;\sigma_N) - m_1} \leq \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)},$$

which implies

$$\lim_{N\to\infty} \frac{q(\omega';\sigma_N) - m_1}{-\Delta_N} \in (0, \infty) \tag{32}$$

for all ω'.Using the approximation $q(\omega';\sigma_N) \approx q^*$, we restate the quadratic approximation (21) for $m = m_N$ and $q = q(\omega';\sigma_N)$ as follows:

$$\mathrm{KL}\Big(m_N, q(\omega';\sigma_N)\Big) \approx \frac{\Big(m_N - q(\omega';\sigma_N)\Big)^2}{2q^*(1 - q^*)}.$$

This approximation yields the following difference in divergences:

---

[34] To be precise, we use the fact that for any $u, v, a, b, c, d > 0$, we have $\min\left(\frac{a}{b}, \frac{c}{d}\right) \leq \frac{au+cv}{bu+dv} \leq \max\left(\frac{a}{b}, \frac{c}{d}\right)$.

$$
\begin{aligned}
&(N-1)\Big(\mathrm{KL}\Big(m_N, q(0;\sigma_N)\Big) - \mathrm{KL}\Big(m_N, q(1;\sigma_N)\Big)\Big) \\
&\approx \frac{(N-1)}{2q^*(1-q^*)}\Big(2m_N\Delta_N + q(0;\sigma_N)^2 - q(1;\sigma_N)^2\Big) \\
&\approx \frac{(N-1)}{2q^*(1-q^*)}\Big(2\Big(m_N - q(0;\sigma_N)\Big)\Delta_N - \Delta_N^2\Big).
\end{aligned}
$$

For $m_N = m'_N$, the unbounded inference (30) together with (32) then implies that $\Delta_N^2 N \to \infty$. Applying the central limit theorem and denoting by $\frac{k_N}{N}$ the realized share of actions 1 among all N agents, we have

$$
\lim_{N\to\infty} \Pr\Big(\Big|\frac{k_N}{N} - q(\omega';\sigma_N)\Big| < \frac{1}{4}|\Delta_N|\Big|\omega = \omega';\sigma_N, N\Big) = 1.
$$

Letting $m_N = \frac{k_N}{N}$, we see that with probability approaching one, as $N \to \infty$

$$
\begin{aligned}
&2\Big(m_N - q(0;\sigma_N)\Big)\Delta_N > \frac{3}{2}\Delta_N^2 \ \text{ if } \omega' = 1, \\
&2\Big(m_N - q(0;\sigma_N)\Big)\Delta_N < \frac{1}{2}\Delta_N^2 \ \text{ if } \omega' = 0;
\end{aligned}
$$

hence,

$$
(N-1)\Big(\mathrm{KL}\Big(m_N, q(0;\sigma_N)\Big) - \mathrm{KL}\Big(m_N, q(1;\sigma_N)\Big)\Big) \to \begin{cases} \infty & \text{if } \omega = 1, \\ -\infty & \text{if } \omega = 0. \end{cases}
$$

Given (28), this means the principal learns the state with probability approaching one, as $N \to \infty$: $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N, N) = 1$ if the state is 1 and $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N, N) = 0$ if the state is 0.

C Proof of Theorem 2

Fix a generalized referendum R. The proof of Theorem 2 proceeds in four steps.

1. We first establish information aggregation and show that in any equilibrium sequence, the state-contingent mean actions $q(0;\sigma_N)$ and $q(1;\sigma_N)$ converge to distinct interior limits (Section C.1).

2. We then prove Lemma 1, which states that under the simple majority gateway referendum, the limit mean action in state 1 exceeds the majority cutoff (Section C.2).

3. This implies that the status quo constraint binds only in state 0; in this way, the principal's payoff is guaranteed to be arbitrarily close to the full-information benchmark.

4. Finally, we show that any other generalized referendum fails to achieve this payoff guarantee (Section C.3).

C.1 Information Aggregation

As mentioned in the main text, the same proof as that of Proposition 2 establishes information aggregation. The only difference is that the non-generic case (Case 2) can be ruled out immediately, using the properties of the gateway referendum with a simple majority cutoff:

Proposition 2's proof in the appendix starts with the assumption (27) of non-interior limit indifferent types. This assumption implies that either only the 1-partisans choose the action 1 as $N \to \infty$ or all except the 0-partisans. Thus $q^* = \lim_{N\to\infty} q(\omega'; \sigma_N) \in \{\rho_1, 1-\rho_0\}$. Since the simple majority cutoff is $m_1 = \frac{1}{2}$ and since $\rho_1 < \frac{1}{2} < 1 - \rho_0$ (cf. Section 1) the case $m_1 = q^*$ (Case 2) can be ruled out. The other case (Case 1) leads to a contradiction, as before. We conclude that the limit indifferent types must be interior, which implies that the limit mean actions differ across the states, i.e.

$$\lim_{N\to\infty} q(1; \sigma_N) \neq \lim_{N\to\infty} q(0; \sigma_N). \tag{33}$$

Since the realized collective action concentrates around the mean action in each state, the principal learns the state from observing it. That is, information aggregates.

C.2 Proof of Lemma 1

We prove Lemma 1 by contradiction. Suppose that

$$\lim_{N\to\infty} q(1; \sigma_N) \leq \frac{1}{2}.$$

We begin with a preliminary observation. Given (33), the limit mean actions are strictly ordered. Then the principal's cutoff must lie strictly between them:

$$\lim_{N\to\infty} q(0; \sigma_N) < \lim_{N\to\infty} \frac{\bar{k}}{N} < \lim_{N\to\infty} q(1; \sigma_N) \qquad \text{or} \qquad \lim_{N\to\infty} q(1; \sigma_N) < \lim_{N\to\infty} \frac{\bar{k}}{N} < \lim_{N\to\infty} q(0; \sigma_N), \tag{34}$$

depending on which ordering in (33) holds. To see this, note that if one of the inequalities failed, then

$$\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(0; \sigma_N)\Big) \neq \lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(1; \sigma_N)\Big),$$

so by (24) the principal's inference from observing the collective action $\frac{\bar{k}}{N}$ would be unbounded. However, this cannot be, given the definition of $\bar{k}$; see (1).

Now, we consider the two possibilities in (33) and start with the second, reverse ordering since ruling it out establishes the first ingredient discussed in the proof sketch in the main text.

Step 1: The reverse ordering cannot arise. Suppose that

$$\lim_{N\to\infty} q(1; \sigma_N) < \lim_{N\to\infty} q(0; \sigma_N).$$

We distinguish two subcases.

Step 1a: The case $\lim_{N\to\infty} q(1;\sigma_N) < \frac{1}{2}$. Under the reverse ordering, the principal's posterior

$$\Pr(\omega = 1 \mid k; \sigma_N, N)$$

is decreasing in the realized number k of actions 1. Hence her constrained best response has the following form: $x(k) = 0$ for $k \leq \lfloor N/2 \rfloor$, it jumps up to some $x \geq \varepsilon$ at $k = \lfloor N/2 \rfloor + 1$, and it is weakly decreasing for larger k, while remaining strictly positive.

Since $\lim_{N\to\infty} q(1;\sigma_N) < \frac{1}{2}$, in state 1 the number of other agents choosing action 1 is Binomial with mode weakly below $\lfloor N/2 \rfloor$. Among all pivotal events, the event

$$k_{-i} = \left\lfloor \frac{N}{2} \right\rfloor$$

therefore has the highest probability in state 1; all other pivotal events lie further above the mode and are thus strictly less likely. At this event, an additional action 1 induces a discrete upward jump from $x = 0$ to some $x \geq \varepsilon$. By contrast, at all other pivotal events it induces only weakly smaller downward adjustments. Hence the average effect of an additional action 1 in state 1 is strictly positive:

$$U(1;\eta_N) > 0.$$

We claim that this contradicts the reverse ordering. To see this, consider the agents' best-response characterization (12). If $U(0;\eta_N) \leq 0$, then all non-partisans strictly prefer action 1, so

$$q(0;\sigma_N) = q(1;\sigma_N) = 1 - \rho_0,$$

a contradiction. If $U(0;\eta_N) > 0$, there are cutoff types $0 < p_N(1) < p_N(0) < 1$ pinned down by (12), and an agent chooses action 1 if and only if $p_i \geq p_N(s)$. Since signal 1 is more likely in state 1, this implies

$$q(0;\sigma_N) \leq q(1;\sigma_N),$$

again contradicting the reverse ordering. We conclude that the case

$$\lim_{N\to\infty} q(1;\sigma_N) < \lim_{N\to\infty} q(0;\sigma_N) \qquad \text{and} \qquad \lim_{N\to\infty} q(1;\sigma_N) < \frac{1}{2}$$

cannot arise.

Step 1b: The case $\lim_{N\to\infty} q(1;\sigma_N) = \frac{1}{2}$. Suppose now that

$$\lim_{N\to\infty} q(1;\sigma_N) = \frac{1}{2} \qquad \text{and} \qquad \lim_{N\to\infty} q(1;\sigma_N) < \lim_{N\to\infty} q(0;\sigma_N).$$

Then (34) implies that the principal's cutoff lies strictly above the majority cutoff,

$$\frac{1}{2} = \lim_{N\to\infty} q(1;\sigma_N) < \lim_{N\to\infty} \frac{\bar{k}}{N} < \lim_{N\to\infty} q(0;\sigma_N).$$

The large-deviation calculus in (23) and (24) then implies that an agent becomes certain of state 1 conditional on being pivotal. Consequently, all non-partisans choose action 1 in both states, so that

$$\lim_{N\to\infty} q(1;\sigma_N) = 1 - \rho_0 > \frac{1}{2},$$

contradicting the case assumption that $\lim_{N\to\infty} q(1;\sigma_N) = \frac{1}{2}$.

We have therefore ruled out the reverse ordering altogether. Hence one must have

$$\lim_{N\to\infty} q(0;\sigma_N) < \lim_{N\to\infty} q(1;\sigma_N). \tag{35}$$

Step 2: Under the ordering (35), the event piv_0 does not affect incentives if $\lim_{N\to\infty} q(1;\sigma_N) \le \frac{1}{2}$. Maintain the contradiction hypothesis

$$\lim_{N\to\infty} q(1;\sigma_N) \le \frac{1}{2}.$$

Together with (35), this implies

$$\lim_{N\to\infty} q(0;\sigma_N) < \lim_{N\to\infty} q(1;\sigma_N) \le \frac{1}{2}.$$

By (34), we then have

$$\lim_{N\to\infty} \frac{\bar{k}}{N} < \frac{1}{2}. \tag{36}$$

The observation (36) implies that, conditional on piv_0, the principal is constrained to choose $x = 0$ regardless of the action of the fixed agent i. Hence the event piv_0 does not affect agent i's incentives. The only relevant pivotal event is therefore the referendum-pivotal event

$$k_{-i} = \left\lfloor \frac{N}{2} \right\rfloor.$$

Step 3: Standard pivotal updating now yields a contradiction. Since (35) and the contradiction hypothesis $\lim_{N\to\infty} q(1;\sigma_N) \le \frac{1}{2}$ imply

$$\lim_{N\to\infty} \mathrm{KL}\Big(\frac{1}{2}, q(0;\sigma_N)\Big) > \lim_{N\to\infty} \mathrm{KL}\Big(\frac{1}{2}, q(1;\sigma_N)\Big),$$

the standard large-deviation calculus in (23) and (24) yields

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1 \mid \mathrm{piv}; \eta_N, N)}{\Pr(\omega = 0 \mid \mathrm{piv}; \eta_N, N)} = \infty.$$

That is, conditional on being pivotal, an agent becomes certain that the state is 1. Consequently, all non-partisans choose action 1, so that

$$\lim_{N\to\infty} q(1;\sigma_N) = 1 - \rho_0 > \frac{1}{2},$$

contradicting the maintained assumption $\lim_{N\to\infty} q(1;\sigma_N) \le \frac{1}{2}$.

This contradiction proves Lemma 1.

C.3 Payoff Guarantees

Information aggregation together with Lemma 1 implies that the only potential loss arises in state 0 if the principal is forced to choose some $x > 0$. In this worst case, she then chooses the smallest feasible policy, $x = \varepsilon$. Consequently, the payoff guarantee of the gateway referendum R^* with the simple majority cutoff has the lower bound

$$B = 1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)}\varepsilon > 1 - \varepsilon$$

What remains to be shown to prove Theorem 2's claim that R^* has a maximal payoff guarantee is that no generalized referendum R has a payoff guarantee that exceeds the lower bound B.

In the following, we assume that $G(R) > 1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)}\varepsilon$ for some R and lead this to a contradiction.

We start with two preliminary observations: First, Theorem 1 implies $G(R) \leq 1 - \varepsilon$ whenever $R(m) \neq \{0\}$ for all m. So $R(m) = \{0\}$ for some m. Second, we argue that, for all $j = 1, \ldots, T + 1$, either $R(m_j) = \{0\}$ or $\{0, 1\} \subseteq R(m_j)$. If $R(m_j) \neq \{0\}$, one can construct approximately truthful equilibrium sequences where $R(m_j)$ becomes almost certainly binding, as $N \to \infty$, using the same argument as in Proposition 1's proof sketch. If then $1 \notin R(m_j)$ or $0 \notin R(m_j)$, the principal's payoff in these equilibrium sequences is at most B, as $N \to \infty$. Thus, $R(m_j) \neq \{0\}$ implies $\{0, 1\} \subseteq R(m_j)$, under the assumption that $G(R) > B$.

The rest of the proof distinguishes three cases.

Case 1. Suppose $R(0) = \{0\}$ and consider an information structure for which $\rho_1 < m_1 < 1 - \rho_0$. By the identical logic as in Section 2.2 (and the details in the appendix section F), for any large enough N, there is a "deadlock" equilibrium in which the agents use a strategy with $\rho_1 < q(1; \sigma_N) < q(0; \sigma_N) < m_1$ and so that $\Pr(\omega = 1 | k = \lfloor m_1 N \rfloor + 1; \sigma_N, N) = \frac{1}{2}$ and the principal's best response is 0 after each m. This implies an upper bound of 0 for $G(R)$, in contradiction to the assumption $G(R) > 1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)}\varepsilon$.

Case 2. Suppose $R(1) = \{0\}$ and consider an information structure for which $\rho_1 < m_T < 1 - \rho_0$. Then, likewise, for any large enough N, there is a "deadlock" equilibrium in which the agents use a strategy with $m_T < q(0; \sigma_N) < q(1; \sigma_N) < 1 - \rho_0$ and so that $\Pr(\omega = 1 | k = \lfloor m_T N \rfloor + 1; \sigma_N, N) = \frac{1}{2}$ and the principal's best reply is 0 after each m. Again, this implies a contradiction to $G(R) > 1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)}\varepsilon$.

Case 3. Suppose $R(m_j) = \{0\}$ for some $j \in \{2, \ldots, T\}$, $R(0) \neq \{0\}$ and $R(1) \neq \{0\}$. We argue that for large enough N, there would exist an information structure and an equilibrium in which the agents choose a strategy σ_N with $q = q(0; \sigma_N) = q(1; \sigma_N)$. Since given such σ, the agents' collective action transmits no information, the principal's best response is her prior-optimal choice 1 if $1 \in R(m)$ and 0 if $R(m) = \{0\}$. The principal's payoff from this equilibrium is weakly worse than choosing 1 for all m; so a sequence of such equilibria would imply an upper bound for her payoff guarantee of $1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)} < B$, in contradiction to the assumption that $G(R) > B$.

The following shows that such an equilibrium would indeed exist. Fixing the principal's best response described above (she chooses 1 if $1 \in R(m)$ and 0 otherwise), it is sufficient to show that there is a strategy σ and a q so that $q = q(0; \sigma_N) = q(1; \sigma_N)$ and $U(0; \eta) = U(1; \eta) = 0$. Then, any strategy of the agents is a best response, including σ.

To prepare the argument, let j' denote the minimal number in $\{2, \ldots, T\}$ for which $R(m_{j'}) = \{0\}$ and j'' the minimal number in $\{j', \ldots, T\}$ for which $R(m_{j''+1}) \neq \{0\}$. This guarantees that the principal's choice switches from 1 to 0 when k switches from $\lfloor m_{j'-1} N \rfloor$ to $\lfloor m_{j'-1} N \rfloor + 1$, and it switches from 0 to 1 when k switches from $\lfloor m_{j''} N \rfloor$ to $\lfloor m_{j''} N \rfloor + 1$. So, in the pivotal event $\text{piv}_{j'-1}$, an additional action 1 moves the policy downwards, and in the pivotal event $\text{piv}_{j''}$, an additional action 1 moves the policy upwards. Now, consider an information structure with $\rho_1 < m_{j'-1}$ and $m_{j''} < 1 - \rho_0$; this constraint on the partisans ensures that for any $q \in [m_{j'-1}, m_{j''}]$, there is a strategy σ with $q = q(0; \sigma) = q(1; \sigma)$.

Consider σ so that $q(\omega'; \sigma) = q = m_{j'-1}$. Then, (23) implies $\lim_{N \to \infty} \Pr(\text{piv}_{j'-1} | \text{piv}; \eta, N) = 1$. Since an additional action 1 moves the policy downwards conditional on $\text{piv}_{j'-1}$, the average effect is negative, i.e. $U(\omega'; \eta) < 0$ for $\omega' \in \{0, 1\}$. Conversely, consider σ is so that $q = m_{j''}$. Then, (23) implies $\lim_{N \to \infty} \Pr(\text{piv}_{j''} | \text{piv}; \eta, N) = 1$. Since an additional action 1 moves the policy upwards conditional on $\text{piv}_{j''}$, the average effect is positive, i.e. $U(\omega'; \eta) > 0$ for $\omega' \in \{0, 1\}$. Finally, an application of the intermediate value theorem implies there is a strategy σ with $q(\omega'; \sigma) = q \in (m_{j'-1}, m_{j''})$ and $U(0; \eta) = U(1; \eta) = 0$. As explained above, this finishes the proof by contradiction in this last case.

References

Acharya, A. and Meirowitz, A. (2017). Sincere voting in large elections. *Games and Economic Behavior*, **101**, 121–131.

Ahn, D. S. and Oliveros, S. (2012). Combinatorial voting. *Econometrica*, **80** (1), 89–141.

Ali, S. N., Mihm, M. and Siga, L. (2025). The political economy of zero-sum thinking. *Econometrica*, **93** (1), 41–70.

Alonso, R. and Matouschek, N. (2008). Optimal delegation. *The Review of Economic Studies*, **75** (1), 259–293.

Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science*, **15** (1), 183–201.

Austen-Smith, D. and Banks, J. S. (1996). Information aggregation, rationality, and the condorcet jury theorem. *American political science review*, **90** (1), 34–45.

Barelli, P., Bhattacharya, S. and Siga, L. (2022). Full information equivalence in large elections. *Econometrica*, **90** (5), 2161–2185.

Barro, R. J. (1973). The control of politicians: an economic model. *Public choice*, pp. 19–42.

Battaglini, M. (2017). Public protests and policy making. *The Quarterly Journal of Economics*, **132** (1), 485–549.

Bhattacharya, S. (2013). Preference monotonicity and information aggregation in elections. *Econometrica*, **81** (3), 1229–1247.

— (2018). *Condorcet Jury theorem in a spatial model of elections.* Working paper.

Castanheira, M. (2003). Why vote for losers? *Journal of the European Economic Association*, **1** (5), 1207–1238.

Chen, K. (2025). *Communication with Multiple Senders.* Working paper.

Coughlan, P. J. (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *American Political science review*, **94** (2), 375–393.

Cramér, H. (1938). Sur un nouveau théoreme-limite de la théorie des probabilités. *Colloqué consacré a la théorie des probabilités, Actualités scientifiques et industrielles 736.*

Damiano, E., Li, H. and Suen, W. (2025). *Informative Voting in Large Elections.* Working paper.

Ekmekci, M. and Lauermann, S. (2020). Manipulated electorates and information aggregation. *The Review of Economic Studies*, **87** (2), 997–1033.

Farquharson, R. (1969). *Theory of Voting.* Yale University Press.

Feddersen, T. and Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica: Journal of the Econometric Society*, pp. 1029–1058.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public choice*, pp. 5–25.

Fernandez, R. and Rodrik, D. (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American economic review*, pp. 1146–1155.

Fey, M. (1997). Stability and coordination in duverger's law: A formal model of preelection polls and strategic voting. *American Political science review*, **91** (1), 135–147.

Heese, C. (2026). Referenda and partial commitment in policy making. *Available at SSRN 6680639.*

Herrera, H., Llorente-Saguer, A. and McMurray, J. C. (2019). The marginal voter's curse. *The Economic Journal*, **129** (624), 3137–3153.

Holmström, B. (1984). On the theory of delegation. In M. Boyer and R. E. Kihlstrom (eds.), *Bayesian Models in Economic Theory*, New York: North-Holland, pp. 115–141.

Kartik, N., Van Weelden, R. and Wolton, S. (2017). Electoral ambiguity and political representation. *American Journal of Political Science*, **61** (4), 958–970.

Krishna, V. and Morgan, J. (2012). Voluntary voting: Costs and benefits. *Journal of Economic Theory*, **147** (6), 2083–2123.

Ledyard, J. O. and Palfrey, T. R. (1999). A characterization of interim efficiency with public goods. *Econometrica*, **67** (2), 435–448.

Levit, D. and Malenko, N. (2011). Nonbinding voting for shareholder proposals. *The Journal of Finance*, **66** (5), 1579–1614.

Lohmann, S. (1994). Information aggregation through costly political action. *The American Economic Review*, pp. 518–530.

Mandler, M. (2012). The fragility of information aggregation in large elections. *Games and Economic Behavior*, **74** (1), 257–268.

Meirowitz, A. and Shotts, K. W. (2009). Pivots versus signals in elections. *Journal of Economic Theory*, **144** (2), 744–771.

Morgan, J. and Stocken, P. C. (2008). Information aggregation in polls. *American Economic Review*, **98** (3), 864–896.

OECD (2020). *Innovative citizen participation and new democratic institutions*. OECD Publications Centre.

Palfrey, T. R. (2009). Laboratory experiments in political economy. *Annual Review of Political Science*, **12** (1), 379–388.

Piketty, T. (2000). Voting as communicating. *The Review of Economic Studies*, **67** (1), 169–191.

Razin, R. (2003). Signaling and election motivations in a voting model with common values and responsive candidates. *Econometrica*, **71** (4), 1083–1119.

Selten, R. (1988). *Reexamination of the perfectness concept for equilibrium points in extensive games*. Springer.

Wolinsky, A. (2002). Eliciting information from multiple experts. *Games and Economic Behavior*, **41** (1), 141–160.

Online Appendix

D Proof of Theorem 1: The Remaining Equilibrium Existence Result

Theorem 1's proof in the main text relegated a general equilibrium existence result to the appendix, which clarifies that the payoff guarantee of a referendum is always well-defined: For *any* N, *any* referendum, and *any* agents' information structure, an equilibrium exists.

To show this result, we lift the fixed-point problem to a finite-dimensional space. We claim that equilibria can be represented as fixed points in the space of vectors $\Big(q(0;\sigma), q(1;\sigma), v\Big)$ with $\sum_l v_l = 1$ and $q(\omega';\sigma) \in [\rho_1, 1-\rho_0]$ for $\omega' \in \{0,1\}$. First, a principal's pure strategy is a mapping from the observed number $k \in \{0,\ldots,N\}$ of actions 1 to policy choices $x(k) \in \mathcal{P}$. Thus, we can represent mixed strategies by vectors $v = (v_1, \ldots, v_{|\mathcal{P}^{N+1}|-1}) \in [0,1]^{|\mathcal{P}^{N+1}|}$ with $\sum_l v_l = 1$. Second, recall that $q(\omega';\sigma) \in [\rho_1, 1-\rho_0]$ for $\omega' \in \{0,1\}$ given the partisans. We then observe that $\Big(q(0;\sigma), q(1;\sigma), v\Big)$ is a sufficient statistic for the best-response correspondence since the principal's inference only depends on $q(0;\sigma)$ and $q(1;\sigma)$, given (1), and since the pivotal events and their likelihood only depend on $(\mathbf{q}, v)$; compare to (3), and (2).Together, the observations imply that any equilibrium induces a fixed point vector $(\mathbf{q}, v)$ and vice versa, hence our representation. This representation will allow us a direct application of Kakutani's fixed point theorem to prove the existence result. Kakutani can be applied since the domain of vectors $(\mathbf{q}, v)$ is non-empty, compact and convex. Further, the correspondence that maps each vector $(\mathbf{q}, v)$ to its best response vectors is non-empty, convex, and has a closed graph.

E Coordination Failure: Proof of Proposition 1

Our proof strategy is as follows. Take any referendum with cutoffs $m_1, \ldots, m_{T+1}$. Fix a target $j^* \in \{1, \ldots, T+1\}$.

1. First, we present candidate equilibrium strategies for the principal and the agents (Section E.1).
2. Second, we show an auxiliary result about the ordering of the principal's best response cutoff $\frac{k}{N}$ and the agents' mean actions in the two states (Section E.2).
3. Third, we choose an information structure such that, for every candidate strategy, (a) for any principal's best response to it, the agents' incentives approximate those in the cheap-talk game as $N \to \infty$ and (b) the realized feasible policy set is R_{j^*} with probability approaching one (Section E.3).
4. Fourth, we construct the equilibrium sequences with a fixed-point argument (Section E.4).

E.1 Candidate Equilibrium Strategies

For any $x \in (0,1)$, denote by p_x the x-quantile of the prior distribution, $p_x = F^{-1}(x)$. We consider the following candidate strategies. The principal follows a mixed strategy randomizing between the pure strategy where she chooses

$$\begin{aligned} &\min R(\frac{k}{N}) \text{ if } k \leq k^*, \\ &\max R(\frac{k}{N}) \text{ if } k > k^*, \end{aligned}$$

and the pure strategy where she chooses

$$\begin{aligned} &\min R(\frac{k}{N}) \text{ if } k \leq k^* + 1, \\ &\max R(\frac{k}{N}) \text{ if } k > k^* + 1, \end{aligned}$$

for some $k^* \in \{1, \ldots, N-1\}$. We identify these two pure strategies with the cutoffs k^* and $k^* + 1$, and a mixture of the two with a random cutoff $\tilde{k}$ or its associated probability $z = \Pr(\tilde{k} = k^*)$.

The non-partisan agents follow strategies $\sigma_{\mathbf{p}}$ under which, after observing signal $s_i \in \{0, 1\}$, agent i chooses $a_i = 1$ if and only if $p_i \geq p(s_i)$, for some $p(s_i) \in (0, 1)$. We slightly abuse notation and write $p_{s_i} = p(s_i)$ for readability. We identify $\sigma_{\mathbf{p}}$ with $\mathbf{p} := \Big(p_0, p_1\Big)$ and consider the set

$$D(\delta) = \{\mathbf{p} : p_{\frac{\delta}{4}} \leq p_1 \leq p_{\frac{\delta}{2}},\ \Pr(\omega = 1) \leq p_0 < 1\}.$$

For any $\mathbf{p} \in D(\delta)$, the strategy $\sigma_{\mathbf{p}}$ is δ-approximately truthful under any information structure satisfying (8) and $\rho_a < \frac{\delta}{4}$ for $a \in \{0, 1\}$ (constructed in Section E.3 below). This is because after signal 0, the likelihood of action 1 is below

$$\rho_1 + (1 - \rho_1 - \rho_0)\Big(1 - F(\Pr(\omega = 1))\Big) \leq \rho_1 + (1 - \rho_1 - \rho_0)\frac{\delta}{4} < \delta;$$

where we used that $\Pr(\omega = 1) > p_{1-\frac{\delta}{4}}$, by (8), and for the second inequality we used the bounds for ρ_a. After signal 1, the likelihood of action 1 exceeds

$$\rho_1 + (1 - \rho_1 - \rho_0)\Big(1 - F(p_{\frac{\delta}{2}})\Big) \geq (1 - \frac{\delta}{4})^2 \geq 1 - \delta,$$

where for the first inequality we used that $\rho_a < \frac{\delta}{4}$ for $a \in \{0, 1\}$.

We denote by

$$\hat{\mathbf{p}}(\mathbf{p}, z) = \Big(\hat{p}(0; \mathbf{p}, z), \hat{p}(1; \mathbf{p}, z)\Big) \in [0, 1]^2$$

the cutoffs of the best response given $\mathbf{p} \in D(\delta)$ and z. The best-response cutoffs are well-defined since the properties of $\mathbf{p} \in D(\delta)$ and z imply $U(\omega') > 0$ for $\omega' \in \{0, 1\}$. They are given by (12). Often we drop the arguments $(\mathbf{p}, z)$ from the notation.

E.2 Ordering of the Principal's Cutoff and the Mean Actions

The principal's best response cutoff $\frac{\bar{k}}{N}$ to any strategy $\sigma_{\mathbf{p}}$ with $\mathbf{p} \in D(\delta)$ lies in between the strategy's mean actions when N is large and δ is small.

Lemma 3 (Interior Principal's Cutoff)**.** *There is some $N_1 \in \mathbb{N}$ and some $\gamma_1 > 0$ such* *that*

$$\frac{\bar{k}}{N} \in \Big(q(0;\mathbf{p}) + \gamma_1, q(1;\mathbf{p}) - \gamma_1\Big) \tag{37}$$

for all $N \geq N_1$, $\mathbf{p} \in D(\delta)$, *and* $\delta \leq \frac{1}{4}$.

Proof. The restriction $\delta \leq \frac{1}{4}$ implies $p_{\frac{\delta}{2}} < p_{1-\frac{\delta}{2}}$; thus, there is a uniform lower bound $\eta > 0$ such that

$$q(1;\mathbf{p}) - q(0;\mathbf{p}) \geq \eta \text{ for all } \mathbf{p} \in D(\delta) \text{ with } \delta \leq \frac{1}{4}, \tag{38}$$

given the full support and absolute continuity of the distribution of the agents' priors and the different likelihood ratios of the two signals in the two states.

We start by applying the first equation of (20) to obtain

$$\frac{\Pr\Big(k|\omega=1;\mathbf{p}\Big)}{\Pr\Big(k|\omega=0;\mathbf{p}\Big)} = \exp\Big(-N\Big(\mathrm{KL}\Big(\frac{k}{N}, q(1;\mathbf{p})\Big) - \mathrm{KL}\Big(\frac{k}{N}, q(0;\mathbf{p})\Big)\Big)\Big) \tag{39}$$

for any $k \in \{0, \ldots, N\}$. Consider

$$\begin{aligned}
A_1 &:= \inf_{\mathbf{p} \in D(\delta), \delta \leq \frac{1}{4}} \Big(\mathrm{KL}\Big(q(0;\mathbf{p}) + \gamma_1, q(1;\mathbf{p})\Big) - \mathrm{KL}\Big(q(0;\mathbf{p}) + \gamma_1, q(0;\mathbf{p})\Big), \\
A_2 &:= \inf_{\mathbf{p} \in D(\delta), \delta \leq \frac{1}{4}} -\Big(\mathrm{KL}\Big(q(1;\mathbf{p}) - \gamma_1, q(1;\mathbf{p})\Big) - \mathrm{KL}\Big(q(1;\mathbf{p}) - \gamma_1, q(0;\mathbf{p})\Big)\Big),
\end{aligned}$$

which are strictly positive for $\gamma_1 > 0$ sufficiently small by (38) and since the Kullback–Leibler divergence $\mathrm{KL}(m, q)$ has uniformly bounded partial derivatives on the parameter set $\{(m, q) : m \in [q(\omega';\mathbf{p}) - \gamma_1, q(\omega';\mathbf{p}) + \gamma_1]$ for some $\omega' \in \{0, 1\}, q \in \{q(0;\mathbf{p}), q(1;\mathbf{p})\}\}$.

Since $q(1;\mathbf{p}) > q(0;\mathbf{p})$, the function $\exp\Big(-N\Big(\mathrm{KL}\Big(\frac{k}{N}, q(1;\mathbf{p})\Big) - \mathrm{KL}\Big(\frac{k}{N}, q(0;\mathbf{p})\Big)\Big)\Big)$ is strictly increasing in k. This monotonicity implies

$$\begin{aligned}
&\frac{\Pr\Big(k|\omega=1;\mathbf{p},N\Big)}{\Pr\Big(k|\omega=0;\mathbf{p},N\Big)} < \exp(-NA_1) \text{ for any } k \text{ with } \frac{k}{N} \leq q(0;\mathbf{p}) + \gamma_1, \text{ and} \\
&\frac{\Pr\Big(k|\omega=1;\mathbf{p},N\Big)}{\Pr\Big(k|\omega=0;\mathbf{p},N\Big)} > \exp(NA_2) \text{ for any } k \text{ with } \frac{k}{N} \geq q(1;\mathbf{p}) - \gamma_1.
\end{aligned}$$

Since A_1 and A_2 are strictly positive, for any $\kappa > 0$, there is some $N(\kappa) \in \mathbb{N}$ such that for all $N \geq N(\kappa)$,

$$\begin{aligned}
&\frac{\Pr(\omega=1)}{\Pr(\omega=0)} \cdot \exp(-NA_1) < \kappa \text{ and} \\
&\frac{\Pr(\omega=1)}{\Pr(\omega=0)} \cdot \exp(NA_2) > \frac{1}{\kappa};
\end{aligned}$$

and the same bounds apply to the posterior likelihood ratio,

$$\frac{\Pr(\omega = 1|k; \mathbf{p}, N)}{\Pr(\omega = 0|k; \mathbf{p}, N)} < \kappa, \quad \text{for any } \frac{k}{N} \leq q(0; \mathbf{p}) + \gamma_1, \tag{40}$$

$$\frac{\Pr(\omega = 1|k; \mathbf{p}, N)}{\Pr(\omega = 0|k; \mathbf{p}, N)} > \frac{1}{\kappa} \quad \text{for any } \frac{k}{N} \geq q(1; \mathbf{p}) - \gamma_1. \tag{41}$$

Finally, we argue that we can choose $\kappa > 0$ small enough so that (37) holds uniformly, that is, for all $N \geq N_1 := N(\kappa)$, $\mathbf{p} \in D(\delta)$, and $\delta \leq \frac{1}{4}$. Suppose, on the contrary, that either $\frac{\bar{k}}{N} \geq q(1; \mathbf{p}) - \gamma_1$ or $\frac{\bar{k}}{N} \leq q(0; \mathbf{p}) + \gamma_1$. First, (40) and (41) imply that the posterior likelihood ratio crosses 1, so $\bar{k}$ is not degenerate, $\bar{k} \neq N$. Second, since each private signal realization is boundedly informative, we can choose $\kappa > 0$ small enough so that for $N \geq N(\kappa)$, (40) and (41) imply

$$\frac{\Pr(\omega = 1|\bar{k} + 1; \mathbf{p}, N)}{\Pr(\omega = 0|\bar{k} + 1; \mathbf{p}, N)} < 1 \text{ if } \frac{\bar{k}}{N} \leq q(0; \mathbf{p}) + \gamma_1,$$

$$\frac{\Pr(\omega = 1|\bar{k}; \mathbf{p}, N)}{\Pr(\omega = 0|\bar{k}; \mathbf{p}, N)} > 1 \text{ if } \frac{\bar{k}}{N} \geq q(1; \mathbf{p}) - \gamma_1.$$

However, this would contradict the minimality of $\bar{k}$; compare to its definition (1). □

E.3 The Information Structures

Lemma 4 (Approximate Cheap-Talk). *There exists $\delta_1 > 0$ small enough and an agents' information structure so that for any sequence of strategies $(\sigma_{\mathbf{p}_N})_{N \in \mathbb{N}}$ with $\mathbf{p}_N \in D(\delta_1)$ and any sequence of principal's best responses to $(\sigma_{\mathbf{p}_N})_{N \in \mathbb{N}}$,*

$$\lim_{N \to \infty} \Pr(\text{piv}_0 | \text{piv}; \eta_N, N) = 1. \tag{42}$$

Proof. First, we define the agents' information structure. Fix the previously chosen $j^* \in \{1, \ldots, T + 1\}$ and consider an agents' information structure satisfying (9), (8) for the bound $\underline{p}_1$ given by

$$\frac{\underline{p}_1}{1 - \underline{p}_1} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \frac{\Pr(s_i = 0|\omega = 1)}{\Pr(s_i = 0|\omega = 0)} \frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)}.$$

Further, the partisan probabilities $\rho_0 > 0$ and $\rho_1 > 0$ are taken to satisfy

$$\rho_a < \frac{\delta}{4}.$$

These choices imply that any strategy $\sigma_{\mathbf{p}_N}$ with $\mathbf{p}_N \in D(\delta)$ is δ-approximately truthful, as discussed in the preceding section.

The parameter $\delta > 0$ is taken to be small enough (smaller than some $\delta_1 > 0$) and the signal probabilities to satisfy the ordering

$$m_{j^*-1} < \Pr(s_i = 1|\omega = 0) < \Pr(s_i = 1|\omega = 1) < m_{j^*}$$

and to be much closer to each other than to the cutoffs m_{j^*-1} and m_{j^*}, so that there is

$\nu > 0$ and given any strategy $\sigma_{\mathbf{p}_N}$ with $\mathbf{p}_N \in D(\delta)$,[35]

$$q(\omega'; \sigma_{\mathbf{p}_N}) \in I_{j^*} \text{ for all } \omega' \in \{0,1\}, \tag{43}$$

and

$$\begin{aligned} &\nu + \mathrm{KL}\Big(q(\omega'; \sigma_{\mathbf{p}_N}), q(\omega''; \sigma_{\mathbf{p}_N})\Big) \\ &\quad < \min_{\omega \in \{\omega', \omega''\}} \Big(\mathrm{KL}(m_j, q(\omega; \sigma_{\mathbf{p}_N}))\Big) \text{ for all } j > 0 \text{ and } \omega', \omega'' \in \{0,1\}, \end{aligned} \tag{44}$$

We now establish the claim of the lemma. Combining (43)-(44) and (37), we see that for $\delta \leq \delta_1$ and $N \geq N_1$ as in Lemma 3, and any strategy $\sigma_{\mathbf{p}_N}$ with $\mathbf{p}_N \in D(\delta)$, it holds

$$\begin{aligned} &\nu + \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(\omega''; \sigma_{\mathbf{p}_N})\Big) \\ &\quad < \min_{\omega \in \{\omega', \omega''\}} \Big(\mathrm{KL}(m_j, q(\omega; \sigma_{\mathbf{p}_N}))\Big) \text{ for all } j > 0 \text{ and } \omega', \omega'' \in \{0,1\}, \end{aligned} \tag{45}$$

holds. Given (45), (23) and (24) imply (42). $\square$

We close this section, observing that, given (43)-(44), an application of the law of large numbers implies

$$\lim_{N \to \infty} \Pr(m \in I_{j^*} | \eta_N, N) = 1. \tag{46}$$

Remark 1. *Unlike in the main text, in the appendix we work with the Kullback-Leibler distances instead of Euclidean distances, which is more convenient for a precise analysis. In the main text, we mentioned that it is key that the Euclidean distance of* $q(\omega'; \sigma_N)$ *to* m_j *and* m_{j+1} *is large enough, as measured by the bound* $M > 0$ *as defined in* (10)*. The relevance of a large enough* $M > 0$ *is that it implies* (44)*, given* (10) *and* $\delta > 0$ *small enough.*[36]

E.4 The Equilibrium Construction

We construct equilibria as follows. We show there are $\mathbf{p}^* = \Big(p_0^*, p_1^*\Big) \in D(\delta)$ and a principal's mixing probability $z \in [0,1]$ so that

$$\hat{\mathbf{p}}(\mathbf{p}^*, z) = \mathbf{p}^*, \tag{47}$$

---

[35] In the case where $j^* = 1$, we use the convention $m_{j^*-1} = 0$, slightly abusing the notation since m_0 is also the principal's limit cutoff.

[36] To see this, take $\delta \leq \gamma$; then $|q(1; \sigma_N) - q(0; \sigma_N)| \leq 3\gamma$ given any δ-approximately truthful strategy σ_N. Together with (10) this means that $|q(1; \sigma_N) - q(0; \sigma_N)| \leq \frac{3}{M}|q(\omega'; \sigma_N) - m_j|$ for $j > 0$ and $\omega' \in \{0,1\}$. So, arbitrarily large M makes the Euclidean distance between the mean actions arbitrarily small relative to their distance to the referendum's cutoffs. One can show that, for $M > 0$ large enough, the Kullback-Leibler distance between the mean actions bounded above by their Kullback-Leibler distance to the cutoffs.

and

$$\frac{\Pr(\omega=1)}{\Pr(\omega=0)}\frac{\Pr\Big(k=\bar{k}+1|\omega=1;\mathbf{p}^*,N\Big)}{\Pr\Big(k=\bar{k}+1|\omega=0;\mathbf{p}^*,N\Big)}=1; \tag{48}$$

This means the agents' cutoff strategy is a best response to itself, given z, and the principal is indifferent after observing $\bar{k}+1$ actions 1, where $\bar{k}$ is the principal's cutoff defined in (1). The principal's mixed strategy z is thus also a best response, and together the players' strategies constitute an equilibrium.

The construction reduces equilibrium existence to a fixed point of a one-dimensional map T_N in the cutoff p_0. It works in three steps.

1. The first step shows that there is a continuous map from $p_0 \in [\Pr(\omega=1),1)$ to $p_1^*(p_0)$ so that the principal is indifferent at $\bar{k}+1$ given $\mathbf{p}=\Big(p_0,p_1^*(p_0)\Big)$ (Lemma 5 in Section E.4.2).
2. The second step shows that there is a continuous function that maps each $p_0 \in [\Pr(\omega=1),1)$ to a number $z^*\Big(p_0\Big)\in[0,1]$ such that the first part of the fixed-point equation (47) holds for $\mathbf{p}=\Big(p_0,p_1^*(p_0)\Big)$ (Lemma 6 in Section E.4.3) i.e.

   $$\hat{p}(1;\mathbf{p},z^*)=p_1^*(p_0). \tag{49}$$

3. The third step considers the mapping from p_0 to

   $$T_N(p_0):=\max\Big(\Pr(\omega=1),\hat{p}\Big(0;\mathbf{p},z^*(p_0)\Big)\Big),$$

   and shows that, for any N large enough, there is $\gamma_5>0$ so that (a) it is a continuous self-map on the compact interval $[\Pr(\omega=1),1-\gamma_5]$ and (b) it (only) has interior fixed points $p_{0,N}^*>\Pr(\omega=1)$ (Lemma 7 in Section E.4.4).

In all steps, we will identify a uniform upper bound on δ and a lower bound on N that are required for the arguments to hold.

E.4.1 Proof of Proposition 1

We present the proof of Proposition 1 based on the construction of interior fixed points $p_{0,N}^*$ of T_N.

Consider a sequence of interior fixed points $p_{0,N}^*$ of T_N. By definition, for any N and any interior fixed point $p_{0,N}^*$ of T_N, the agents' strategy $\mathbf{p}_N^*=\Big(p_{0,N}^*,p_{1,N}^*(p_{0,N}^*)\Big)$ and the principal's mixed strategy given by $z^*\Big(p_{0,N}^*\Big)$ constitute an equilibrium. Since $\mathbf{p}_N^*\in D(\delta)$, the corresponding sequence of agents' strategies is a sequence of δ-approximately truthful strategies. Since the limits of the strategies' mean actions ($N\to\infty$) lie strictly in the interval I_{j^*} by (43) and (44), an application of the law of large numbers implies $\lim_{N\to\infty}\Pr(m\in I_{j^*}|\eta_N,N)=1$. This proves Proposition 1.

E.4.2 Lemma about the Principal's Indifference

The principal will be indifferent if she observes that $k = k^*(p_0) + 1$ out of N agents have chosen the action 1. For any fixed $p_0 \in [\Pr(\omega = 1), 1)$, we define $k^*(p_0)+1$ as the minimal observed number of actions 1 such that the principal weakly prefers $x = 1$ given $p_1 = p_{\frac{\delta}{2}}$, i.e.,

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1 | \omega = 1; p_1 = p_{\frac{\delta}{2}}, p_0, N\Big)}{\Pr\Big(k = k^*(p_0) + 1 | \omega = 0; p_1 = p_{\frac{\delta}{2}}, p_0, N\Big)} \geq 1. \tag{50}$$

Note that $k^*(p_0)$ equals the best-response cutoff $\bar{k}$ given (p_0, p_1), so that (37) implies it is not degenerate, i.e. $k^*(p_0) \neq N$.

Lemma 5 shows that, for any candidate cutoff p_0, we can find a cutoff $p_1^*(p_0)$ such that, given the agents' strategy $\mathbf{p} = \Big(p_0, p_1^*(p_0)\Big)$, the principal becomes indifferent at $k^*(p_0) + 1$, the cutoff defined in the preceding section.

Lemma 5 (Indifference of the Principal)**.** *There exists $\delta_2 > 0$ such that for all $\delta \leq \delta_2$ there exists $N_2(\delta) \in \mathbb{N}$ and for all $N_2(\delta) \leq N$, there is a continuous function that maps each $p_0 \in [\Pr(\omega = 1), 1)$ to a number $p_1^*(p_0) \in [p_{\frac{\delta}{4}}, p_{\frac{\delta}{2}}]$ satisfying*

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1 | \omega = 1; p_1 = p_1^*(p_0), p_0, N\Big)}{\Pr\Big(k = k^*(p_0) + 1 | \omega = 0; p_1 = p_1^*(p_0), p_0, N\Big)} = 1. \tag{51}$$

Proof. We show the uniform existence of a number $p_1^*(p_0) \in [p_{\frac{\delta}{4}}, p_{\frac{\delta}{2}}]$ that solves (51). For this we establish the existence of some $\delta_2 > 0$ such that for each $\delta \leq \delta_2$ there is some $N_2(\delta) \in \mathbb{N}$ and

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1 | \omega = 1; p_1 = p_{\frac{\delta}{4}}, p_0, N\Big)}{\Pr\Big(k = k^*(p_0) + 1 | \omega = 0; p_1 = p_{\frac{\delta}{4}}, p_0, N\Big)} < 1 \tag{52}$$
$$\text{for all } N \geq N_2(\delta), \ \delta \leq \delta_2, \text{ and } p_0 \in [\Pr(\omega = 1), 1).$$

(note that we fix $p_1 = p_{\frac{\delta}{4}}$ here). Combining (50) and (52) and applying the intermediate value theorem then yields a $p_1^*(p_0) \in (p_{\frac{\delta}{4}}, p_{\frac{\delta}{2}}]$ such that the principal is indifferent given $p_1^*(p_0)$—i.e., (51) holds. Now, δ_2 and $N_2(\delta)$ for $\delta \leq \delta_2$ exist by the following argument. First, the minimality of $k^*(p_0) + 1$ implies a uniform bound $\gamma_2 > 0$ such that

$$\Pr(\omega = 1 | k = k^*(p_0) + 1; \mathbf{p}, N) - \frac{1}{2} \leq \gamma_2 \tag{53}$$

for all N, $\delta \leq \frac{1}{4}$, and $p_0 \in [\Pr(\omega = 1), 1)$. Second, note that the definition of $k^*(p_0)$ equals that of $\bar{k}$ for any $\mathbf{p} \in D(\delta)$ with $p_1 = p_{\frac{\delta}{2}}$. So, Lemma 3 implies

$$\frac{k^*(p_0)}{N} \in \Big(q(0; \mathbf{p}) + \gamma_1, q(1; \mathbf{p}) - \gamma_1\Big) \tag{54}$$

for all $\mathbf{p} \in D(\delta)$ with $p_1 = p_{\frac{\delta}{2}}$, any $\delta \leq \frac{1}{4}$, and any $N \geq N_1$; cf. (37). Third, note that

$q(0;\mathbf{p})$ and $q(1;\mathbf{p})$ are both strictly decreasing in p_1. Given this, (54), and the properties of the prior and signal distribution, there is $\delta_2 > 0$ sufficiently small and $\gamma_3 > 0$ so that

$$\frac{k^*(p_0)+1}{N} \in \Big(q(0;\mathbf{p}), q(1;\mathbf{p})\Big)$$

and

$$\frac{\partial}{\partial p_1}\mathrm{KL}\Big(\frac{k^*(p_0)+1}{N}, q(0;\mathbf{p})\Big) - \mathrm{KL}\Big(\frac{k^*(p_0)+1}{N}, q(1;\mathbf{p})\Big) \geq \gamma_3$$

for *all* $\mathbf{p} \in D(\delta)$, $\delta \leq \delta_2$, and $N \geq N_1$. Jointly, these observations and (39) imply that the posterior

$$\frac{\Pr\Big(k = k^*(p_0)+1|\omega = 1; \mathbf{p}, N\Big)}{\Pr\Big(k = k^*(p_0)+1|\omega = 0; \mathbf{p}, N\Big)}$$

is strictly increasing in p_1. Further, its derivative is bounded from below by an arbitrarily large number if N is arbitrarily large. Recalling (53) and (50), we see that for any fixed $\delta \leq \delta_2$, there is $N_2(\delta) \in \mathbb{N}$ such that (52) holds uniformly. As argued, (52) implies the uniform existence of $p_1^*(p_0) \in [p_{\frac{\delta}{4}}, p_{\frac{\delta}{2}}]$ solving (51).

Finally, we note that $p_1^*(p_0)$ is unique and continuous. It is unique because the posterior is strictly increasing. The continuity of $p_1^*(p_0)$ in p_0 follows from an application of the implicit function theorem. □

E.4.3 The Agents' First Fixed Point Equation

Lemma 6 shows that, for any candidate cutoff p_0, we can find a principal's mixed strategy $z^*\Big(p_0\Big)$ such that

$$\hat{p}\Big(1; \mathbf{p}, z^*(p_0)\Big) = p_1^*(p_0). \tag{55}$$

Lemma 6 (Mixing Lemma)**.** *There exists $\delta_3 > 0$ such that for all $\delta \leq \delta_3$ there is $N_3(\delta) \in \mathbb{N}$ and for all $N \geq N_3(\delta)$, there is a continuous function that maps each $p_0 \in [\Pr(\omega = 1), 1)$ to a number $z^*\Big(p_0\Big) \in [0,1]$ such that* (55) *holds.*

Proof. We fix $p_1 = p_1^*\Big(p_0\Big)$ for the duration of the proof of Lemma 6. The proof leverages the principal's indifference between the pure strategies with cutoffs $k^*(p_0)$ and $k^*(p_0)+1$ (we drop p_0 from its notation in the following). We will derive approximations of the indifference cutoff $\hat{p}(1;\mathbf{p},z)$ for the agents' best response, given either of the two pure strategies. The key is to establish that for small δ and large N,

$$\hat{p}(1;\mathbf{p},z) < p_1^*(p_0) \;\text{ if }\; z = 0, \;\text{ and} \tag{56}$$

$$\hat{p}(1;\mathbf{p},z) > p_1^*(p_0) \;\text{ if }\; z = 1. \tag{57}$$

Since $\hat{p}(1;\mathbf{p},z)$ is continuous in the probability z—see (2) and (12)—an application of the intermediate value theorem then implies the existence of a principal's mixed strategy $z^* \in (0,1)$ such that $\hat{p}(1;\mathbf{p},z^*) = p_1^*(p_0)$.

In the following, we first establish that the inequalities (56) and (57) hold in the double-limit where we take $N \to \infty$ and then $\delta \to 0$, i.e., we establish that $\lim_{\delta\to 0}\lim_{N\to\infty}\hat{p}(1;\mathbf{p},z) < p_1^*(p_0)$ if $z = 0$ and $\lim_{\delta\to 0}\lim_{N\to\infty}\hat{p}(1;\mathbf{p},z) > p_1^*(p_0)$ if $z = 1$. After that, we argue the existence of the uniform bounds δ_2 and $N_2(\delta)$ for $\delta \leq \delta_2$.

We start with the pure strategy where $1 - z = \Pr(\tilde{k} = k^* + 1) = 1$. This means that piv_0 is the event where $k_{-i} = k^*(p_0) + 1$ and

$$\Pr(\text{piv}_0|\omega = \omega';\mathbf{p},N) = \Pr(k_{-i} = k^* + 1|\omega = \omega';\mathbf{p}) \text{ for } \omega' \in \{0,1\}, \tag{58}$$

where $\Pr(k_{-i} = k^* + 1|\omega = \omega';\mathbf{p})$ is the likelihood of $k^* + 1$ out of $N - 1$ agents choosing the action 1. The principal is indifferent if she observes that $k = k^* + 1$ out of N agents have chosen the action 1, i.e.

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} = \frac{\Pr\left(k = k^* + 1|\omega = 0;\mathbf{p},N\right)}{\Pr\left(k = k^* + 1|\omega = 1;\mathbf{p},N\right)}. \tag{59}$$

Since the strategies given by $\mathbf{p}$ are δ-approximately truthful, as $\delta \to 0$ we have

$$1 - q(\omega';\mathbf{p}) \to \Pr(s_i = 0|\omega = \omega'), \tag{60}$$

so that

$$\begin{aligned}\frac{\Pr\left(k_{-i} = k^* + 1|\omega = 0;\mathbf{p},N\right)}{\Pr\left(k_{-i} = k^* + 1|\omega = 1;\mathbf{p},N\right)} &= \left(\frac{q(0;\mathbf{p})}{q(1;\mathbf{p})}\right)^{k^*+1}\left(\frac{1-q(0;\mathbf{p})}{1-q(1;\mathbf{p})}\right)^{N-k^*-2}\\ &\to \frac{\Pr\left(k = k^* + 1|\omega = 0;\mathbf{p},N\right)}{\Pr\left(k = k^* + 1|\omega = 1;\mathbf{p},N\right)}\cdot\frac{\Pr(s_i = 0|\omega = 1)}{\Pr(s_i = 0|\omega = 0)}.\end{aligned} \tag{61}$$

Recall the property (44). It implies (11), i.e., that in the limit as $N \to \infty$, an agent cares only about piv_0. Hence, the condition (12) for the indifference cutoffs $p(s)$ of the best response imply

$$\lim_{N\to\infty}\frac{\hat{p}(s;\mathbf{p},z)}{1-\hat{p}(s;\mathbf{p},z)} = \frac{\Pr\left(\text{piv}_0|\omega = 0;\mathbf{p},N\right)}{\Pr\left(\text{piv}_0|\omega = 1;\mathbf{p},N\right)}\cdot\frac{\Pr(s_i = s|\omega = 0)}{\Pr(s_i = s|\omega = 1)}. \tag{62}$$

If we combine (58)–(61) and take $\delta \to 0$, this indifference condition becomes

$$\lim_{\delta\to 0}\lim_{N\to\infty}\hat{p}(1;\mathbf{p},z) = \underline{p}_1 \tag{63}$$

for $s = 1$, with $\underline{p}_1$ given by

$$\frac{\underline{p}_1}{1-\underline{p}_1} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)}\cdot\frac{\Pr(s_i = 0|\omega = 1)}{\Pr(s_i = 0|\omega = 0)}\cdot\frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)}. \tag{64}$$

Next, consider the pure strategy where $z = \Pr(\tilde{k} = k^*) = 1$. For this strategy,

$$\Pr(\text{piv}_0|\omega = \omega';\mathbf{p}) = \Pr(k_{-i} = k^*|\omega = \omega';\mathbf{p}) \text{ for } \omega' \in \{0,1\}. \tag{65}$$

As $\delta \to 0$, we have

$$\frac{\Pr\Big(k_{-i} = k^* | \omega = 0; \mathbf{p}, N\Big)}{\Pr\Big(k_{-i} = k^* | \omega = 1; \mathbf{p}, N\Big)} = \left(\frac{q(0;\mathbf{p})}{q(1;\mathbf{p})}\right)^{k^*} \cdot \left(\frac{1-q(0;\mathbf{p})}{1-q(1;\mathbf{p})}\right)^{N-k^*-1}$$
$$\to \frac{\Pr\Big(k = k^*+1 | \omega = 0; \mathbf{p}, N\Big)}{\Pr\Big(k = k^*+1 | \omega = 1; \mathbf{p}, N\Big)} \cdot \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)}. \tag{66}$$

If we combine (59), (65), and (66), and take $\delta \to 0$, the indifference condition (12) becomes

$$\lim_{\delta \to 0} \lim_{N \to \infty} \hat{p}(1; \mathbf{p}, z) = \bar{p}_1 \tag{67}$$

for $s = 1$, with $\bar{p}_1$ given by

$$\frac{\bar{p}_1}{1-\bar{p}_1} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)} \cdot \frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)}. \tag{68}$$

Now, we combine the approximations (63) and (67) with the requirement (8) on the prior distribution to argue that (56) and (57) hold “in the double-limit.” The requirement (8) implies

$$\underline{p}_1 < p_{\frac{\delta}{4}} \text{ and } p_{1-\frac{\delta}{4}} < \Pr(\omega = 1). \tag{69}$$

Combining this with $\Pr(\omega = 1) \leq \bar{p}_1$ and $p_{\frac{\delta}{4}} < p_1^*(p_0) \leq p_{\frac{\delta}{2}}$ shows that

$$\underline{p}_1 < p_1^*(p_0) < \bar{p}_1. \tag{70}$$

Hence, (63) and (67) imply (56) and (57) in the limit, i.e., as $N \to \infty$ and then $\delta \to 0$. Note that we used Lemma 5 here, which guarantees the existence of $p_1^*(p_0)$ for $\delta \leq \delta_2$ and $N \geq N_2(\delta)$ (this also explains the order of limits).

Next we show that there are uniform bounds $\delta_3 > 0$ and $N_3(\delta) \in \mathbb{N}$ for $\delta \leq \delta_3$ such that (56) and (57) hold not just in the limit, but for any $\delta \leq \delta_3$, $N_3(\delta) \leq N$, and $\mathbf{p} \in D(\delta)$ with $p_1 = p_1^*(p_0)$. The limit analysis for $\hat{p}(1; \mathbf{p}, z)$ used two approximations, (60) and (62), and we argue that for any $\gamma_4 > 0$ we can find uniform bounds that ensure that both approximations hold up to an error term of at most γ_4. For (60), this is obvious for some δ small enough. For (62), this follows from observing that the convergence here is exponential in the difference $\min_{\omega' \in \{0,1\}} \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(\omega'; \mathbf{p})\Big) - \min_{\omega' \in \{0,1\}} \min_{j>0} \mathrm{KL}\Big(m_j, q(\omega'; \mathbf{p})\Big)$, and this difference is uniformly bounded from below, given (44) and (37), for any $N \geq N_1$, and $\mathbf{p} \in D(\delta)$ for $\delta \leq \delta_1$ and (with N_1 and δ_1 as in Lemmas 3 and 4). Hence, for any $\gamma_4 > 0$, the likelihood ratio $\frac{\hat{p}(1;\mathbf{p},z)}{1-\hat{p}(1;\mathbf{p},z)}$ is γ_4-close to its limit when N is above a certain bound $N_3(\delta)$. This way, the inequalities (69) and (70) do not only hold for the limit terms, but for $N \geq N_3(\delta)$,

$$\hat{p}(1; \mathbf{p}, z) < p_{\frac{\delta}{4}} \text{ for } z = 0 \text{ and} \tag{71}$$
$$p_{1-\frac{\delta}{4}} < \hat{p}(1; \mathbf{p}, z) \text{ for } z = 1. \tag{72}$$

Thus, (56) and (57) hold, given that $p_{\frac{\delta}{4}} < p_1^*(p_0) \leq p_{\frac{\delta}{2}}$, as stated above.

Finally, since (56) and (57) hold for any $\delta < \delta_3$, $N \geq N_3(\delta)$, and $\mathbf{p} \in D(\delta)$ with $p_1 = p_1^*(p_0)$, for any $p_0 \geq \Pr(\omega = 1)$, an application of the intermediate value theorem yields a $z^*(p_0)$ such that (55) holds. Since $\hat{p}\Big(1; \mathbf{p}, z\Big)$ is strictly monotone in z, the mixed strategy z^* is unique; since $\hat{p}\Big(1; \mathbf{p}, z\Big)$ is continuous in p_0 (this is because p_0, p_1, and z affect the likelihood of the pivotal events in a continuous way), z^* is also continuous in p_0. ☐

E.4.4 The Fixed-Point Construction

We use a fixed point argument to establish the existence of p_0^* such that the agents' strategy given by $\Big(p_0^*, p_1^*(p_0^*)\Big)$ is a best response to itself and the principal's mixed strategy $z\Big(p_0^*\Big)$.

Fix any N and δ that satisfy the uniform bounds of Lemma 6. This ensures that the mapping from $p_0 \in [\Pr(\omega = 1), 1)$ to the projection

$$T_N(p_0) := \max\Big(\Pr(\omega = 1), \hat{p}\Big(0; \mathbf{p}, z^*(p_0)\Big)\Big), \text{ with } \mathbf{p} = \Big(p_0, p_1^*(p_0)\Big),$$

is well-defined.

Lemma 7 (Fixed-Point Lemma)**.** *For any N, there is $\gamma_5 > 0$ so that the function T_N is a continuous self-map on the compact interval* $[\Pr(\omega = 1), 1 - \gamma_5]$ *and has only interior fixed-points.*

Proof. The mapping is continuous in $p_0 \in [\Pr(\omega = 1), 1)$ because $\hat{p}\Big(0; \mathbf{p}, z^*(p_0)\Big)$ is continuous in p_0. (This is because $p_1^*(p_0)$ and $z^*(p_0)$ are continuous in p_0 and all three parameters p_0, $p_1^*(p_0)$ and $z^*(p_0)$ affect the likelihood of the pivotal events in a continuous way.) For any fixed N, δ, and an agents' information structure, the best response $\hat{p}(0; \mathbf{p}, z^*)$ is uniformly bounded, i.e., $\hat{p}(0; \mathbf{p}, z^*) \leq 1 - \gamma_5$ for some $\gamma_5 > 0$ and all $\mathbf{p} \in D(\delta)$. So, the projection T_N is a continuous self-map on the compact interval $[\Pr(\omega = 1), 1 - \gamma_5]$. An application of Brouwer's fixed point theorem then yields a fixed point $p_{0,N}^*$.

We argue that any fixed point $p_{0,N}^*$ is interior, i.e., it is strictly greater than $\Pr(\omega = 1)$. To show this, we use (64). The best-response cutoff $\hat{p}(0; \mathbf{p}, z^*)$ relates to $\hat{p}(1; \mathbf{p}, z^*) = p_1^*(p_0)$ via the following equation:

$$\frac{p_1^*(p_0)}{1 - p_1^*(p_0)} = \frac{\hat{p}(0; \mathbf{p}, z^*)}{1 - \hat{p}(0; \mathbf{p}, z^*)} \cdot \frac{\Pr(s_i = 0|\omega = 1)}{\Pr(s_i = 0|\omega = 0)} \cdot \frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)}.$$

Comparing this to (64), we see that $\underline{p}_1 < p_1^*(p_0)$ (which holds since N and δ satisfy the uniform bounds of Lemma 6; cf. (70) in the proof of this claim) implies $\Pr(\omega = 1) < \hat{p}(0; \mathbf{p}, z^*)$. But this means that the boundary point $p_0 = \Pr(\omega = 1)$ is not a fixed point. ☐

F Construction of the Agents' Deadlock Equilibrium Strategy

We prove the remaining claim from Section 2.2 that there are agents' information structures and agents' strategies σ_N with mean actions $0 < q(1;\sigma_N) < q(0;\sigma_N) < m_1$ for which

$$\Pr(\omega = 1 | k = \lfloor m_1 N \rfloor + 1; \sigma_N, N) = \frac{1}{2}.$$

Take any $m_1 \in (0,1)$, any pair of mean actions $\mathbf{q} = \Big(q_0, q_1\Big)$ with $0 < q_1 \leq q_0 < m_1$, and let $m' = \frac{\lfloor m_1 N \rfloor + 1}{N}$. Then, note that

$$\frac{\Pr(k = m'N | \omega = 1)}{\Pr(k = m'N | \omega = 0)} = \exp\Big(- N\Big(\mathrm{KL}(m', q_1) - \mathrm{KL}(m', q_0)\Big)\Big),$$

by an application of (19).

If $q_1 = q_0$, the principal learns nothing from her observations and

$$\Pr(\omega = 1 | k = m'N; \mathbf{q}, N) = \Pr(\omega = 1) > \frac{1}{2}.$$

If $q_1 < q_0$, then $\lim_{N\to\infty} \mathrm{KL}\Big(m', q_1\Big) - \mathrm{KL}\Big(m', q_0\Big) > 0$, so

$$\lim_{N\to\infty} \Pr(\omega = 1 | k = m'N; \mathbf{q}, N) = 0.$$

By an application of the intermediate value theorem, for any N large enough, there are mean actions $0 < q_1 < q_0 < m_1$ for which

$$\Pr(\omega = 1 | k = m'N; \mathbf{q}, N) = \frac{1}{2}.$$

For any such q_0 and q_1, we can always find a strategy σ that induces these mean actions whenever the share of partisans choosing action 1 is not too large, $\rho_1 < q_1$.

G Coordination Failure: Proof of Proposition 3

For any large enough N, we construct an equilibrium strategy σ_N with the mean action exceeding the highest cutoff in each state,

$$m_T + \gamma < q(0;\sigma_N) < q(1;\sigma_N), \tag{73}$$

for some $\gamma > 0$. An application of the law of large numbers then yields the claim of Proposition 3

The equilibrium strategy is found among a parametric family of cutoff strategies. For each $L > 0$, let σ_L denote the strategy under which, after observing signal $s \in \{0,1\}$, a non-partisan agent chooses action 1 if and only if

$$p_i \geq p_L(s),$$

where $p_L(s)$ is implicitly defined by

$$L := \frac{\Pr_i(\omega = 1 \mid p_i = p_L(s), s_i = s)}{\Pr_i(\omega = 0 \mid p_i = p_L(s), s_i = s)} = \frac{p_L(s)}{1 - p_L(s)} \frac{\Pr(s_i = s \mid \omega = 1)}{\Pr(s_i = s \mid \omega = 0)}. \tag{74}$$

Thus, higher values of L correspond to more demanding cutoff types. In particular, $p_L(s)$ is strictly increasing in L for each signal realization s, and since signal 1 is more favorable to state 1 than signal 0, we have

$$p_L(1) < p_L(0) \qquad \text{for all } L > 0.$$

Hence the induced mean actions satisfy

$$q(0; \sigma_L) < q(1; \sigma_L).$$

We will work on a compact set of parameters $L \in [\underline{L}, \bar{L}]$, where

$$\underline{L} := \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot I^{-1} < \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot I =: \bar{L}. \tag{75}$$

for $I = \frac{\Pr(s_i=1|\omega=1)}{\Pr(s_i=1|\omega=0)} \frac{\Pr(s_i=0|\omega=0)}{\Pr(s_i=0|\omega=1)}$. These bounds are chosen so that, for any principal best response $(\bar{k}, \tilde{x})$ to a strategy σ_L, the likelihood ratio conditional on the pivotal event piv_0 lies in the interval $[\underline{L}, \bar{L}]$:

$$\frac{\Pr\Big(\mathrm{piv}_0 | \omega = 0; \sigma_L, (\bar{k}, \tilde{x}), N\Big)}{\Pr\Big(\mathrm{piv}_0 | \omega = 1; \sigma_L, (\bar{k}, \tilde{x}), N\Big)} \in [\underline{L}, \bar{L}]. \tag{76}$$

To verify (76), recall that piv_0 is the event that the realized number k_{-i} of other agents choosing action 1 is either $\bar{k}$ or $\bar{k} + 1$. Since

$$\frac{\Pr\Big(\omega = 1 \mid k = \bar{k}; \sigma_L, N\Big)}{\Pr\Big(\omega = 0 \mid k = \bar{k}; \sigma_L, N\Big)} < 1 \leq \frac{\Pr\Big(\omega = 1 \mid k = \bar{k} + 1; \sigma_L, N\Big)}{\Pr\Big(\omega = 0 \mid k = \bar{k} + 1; \sigma_L, N\Big)},$$

the likelihood ratio conditional on piv_0 must indeed lie in $[\underline{L}, \bar{L}]$ (we will also make sure that $\bar{k}$ is not degenerate, meaning that $\bar{k} \neq N$, so that the posterior likelihood ratios are well-defined.)

The next lemma packages the fixed-point argument that we will run on the compact parameter set.

Lemma 8 (Self-map lemma)**.** *There exists $\bar{q} \in (m_T, 1 - \rho_0)$ such that, for any distribution of priors satisfying*

$$\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) > \bar{q},$$

there are $\bar{N} \in \mathbb{N}$ and $\bar{\delta} > 0$ with the following property: for every $N \geq \bar{N}$, every $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$, and every principal best response $(\bar{k}_N, \tilde{x}_N)$ to σ_L,

(i) *the induced mean actions satisfy*

$$\bar{q} < q(0;\sigma_L) < q(1;\sigma_L),$$

(ii) *the agents' best response to* σ_L *and* $(\bar{k}_N, \tilde{x}_N)$ *is again a cutoff strategy* $\sigma_{L'}$ *with*

$$L' \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}].$$

Proof. Recall the definition of $\bar{p}$ from the main text; in mathematical notation,

$$\frac{\Pr_i(\omega = 1 \mid p_i = \bar{p}, s_i = 0)}{\Pr_i(\omega = 0 \mid p_i = \bar{p}, s_i = 0)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} I. \tag{77}$$

Fix an information structure satisfying

$$\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(\bar{p})\Big) > \bar{q}$$

for some $1 - \rho_0 > \bar{q} > m_T$ to be chosen below.

We begin with part (i). Since $q(0;\sigma_L) < q(1;\sigma_L)$ for all L, it remains to obtain a uniform lower bound on $q(0;\sigma_L)$. By continuity of $p_L(s)$ in L and the identity $p_{\bar{L}}(0) = \bar{p}$ (which follows by comparing (74) at $s = 0$ with (77) and the definition (75) of $\bar{L}$), there exists $\bar{\delta} > 0$ such that

$$\rho_1 + (1 - \rho_0 - \rho_1)\Big(1 - F(p_{\bar{L}+\bar{\delta}}(0))\Big) > \bar{q}.$$

Now fix any $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$. Since $p_L(s)$ is increasing in L,

$$p_L(s) \leq p_{\bar{L}+\delta}(s) \qquad \text{for all } s \in \{0,1\}.$$

Hence, using $p_L(1) < p_L(0)$,

$$\begin{aligned} q(\omega;\sigma_L) &= \rho_1 + (1 - \rho_0 - \rho_1) \sum_{s=0,1} \Pr(s_i = s \mid \omega)\Big(1 - F(p_L(s))\Big) \\ &\geq \rho_1 + (1 - \rho_0 - \rho_1)\big(1 - F(p_L(0))\big) \\ &\geq \rho_1 + (1 - \rho_0 - \rho_1)\big(1 - F(p_{\bar{L}+\delta}(0))\big) > \bar{q} \end{aligned}$$

for both states $\omega \in \{0,1\}$. This proves part (i).

Next, fix $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$ and consider any sequence of principal's best responses $(\bar{k}_N, \tilde{x}_N)$ to σ_L. We claim that

$$q(0;\sigma_L) < \lim_{N\to\infty} \frac{\bar{k}_N}{N} < q(1;\sigma_L). \tag{78}$$

(which implies $\bar{k}_N \neq N$). Indeed, if one of the inequalities failed, then

$$\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(0;\sigma_L)\Big) \neq \lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(1;\sigma_L)\Big),$$

so by (19) the principal's posterior at $\bar{k}_N$ would converge to either 0 or 1. But, by definition, the principal's posterior at $\bar{k}_N$ lies below $\frac{1}{2}$ and her posterior at $\bar{k}_N + 1$ weakly above.

Since a single agent's action is only boundedly informative, this directly contradicts the convergence to 0 or 1.

Choose $\bar{q} \in (m_T, 1 - \rho_0)$ large enough so that whenever $q(0; \sigma_L) \geq \bar{q}$, then

$$\mathrm{KL}\Big(m_T, q(\omega'; \sigma_L)\Big) > \mathrm{KL}\Big(m_0, q(\omega; \sigma_L)\Big) \qquad \text{for all } \omega', \omega \in \{0, 1\}, \tag{79}$$

where

$$m_0 := \lim_{N \to \infty} \frac{\bar{k}_N}{N}.$$

Such a choice is possible because when $q(0; \sigma_L)$ is close to $1 - \rho_0$, (78) implies m_0 is close to both mean actions, so the right-hand side of (79) is close to 0, whereas the left-hand side remains strictly positive since $m_T < 1 - \rho_0$.

Combining part (i) with (79), and using (23)- (24), we obtain the key large-deviation implication:

$$\Pr(\mathrm{piv}_0 \mid \mathrm{piv}; \sigma_L, N) \to 1. \tag{80}$$

That is, conditional on being pivotal, an agent almost certainly affects the principal's preference over policies instead of the feasible policy set.

We now prove part (ii). Fix any $L \in [\underline{L} - \delta, \bar{L} + \delta]$ and any principal best response $(\bar{k}_N, \tilde{x}_N)$ to σ_L. By (80) and the best-response characterization (12), the agents' best response is again of cutoff form, say $\sigma_{L'_N}$, with

$$L'_N \to \frac{\Pr\Big(\mathrm{piv}_0 | \omega = 0; \sigma_L, (\bar{k}_N, \tilde{x}_N), N\Big)}{\Pr\Big(\mathrm{piv}_0 | \omega = 1; \sigma_L, (\bar{k}_N, \tilde{x}_N), N\Big)}$$

as $N \to \infty$. By (76), the limit lies in $[\underline{L}, \bar{L}]$. Therefore, we can find a bound $\bar{N}$, for which

$$L'_N \in [\underline{L} - \delta, \bar{L} + \delta] \qquad \text{for all } N \geq \bar{N}.$$

This proves part (ii). □

We can now complete the proof of Proposition 3. For every $N \geq \bar{N}$, let Γ_N denote the correspondence on the compact interval $[\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$ that maps a parameter L to the set of parameters L' generated by the following two-step procedure:

(a) choose a principal best response $(\bar{k}_N, \tilde{x}_N)$ to σ_L,

(b) choose an agents' best response $\sigma_{L'}$ to σ_L and $(\bar{k}_N, \tilde{x}_N)$.

By Lemma 8, Γ_N maps the interval into itself. Moreover, it has non-empty, compact, convex values and a closed graph. Hence Kakutani's fixed point theorem yields a fixed point L^*_N of Γ_N for every $N \geq \bar{N}$.

The corresponding strategy $\sigma_{L^*_N}$ is an equilibrium strategy. By Lemma 8(i),

$$\bar{q} < q(0; \sigma_{L^*_N}) < q(1; \sigma_{L^*_N}).$$

Since $m_T < \bar{q}$, choose any $\gamma > 0$ such that $m_T + \gamma < \bar{q}$. Then

$$m_T + \gamma < q(0; \sigma_{L_N^*}) < q(1; \sigma_{L_N^*});$$

so the equilibrium sequence satisfies (73), as claimed.

H Optimality for a Given Information Structure

In Section 5.2 of the main text, we stated Propositions 4 and 5. Proposition 4 concerns the case in which the principal's prior is sufficiently high, and its key step is that monotone referenda with a single cutoff $m_1 = \frac{1}{2}$ admit equilibrium sequences along which the lower policy region $R(0)$ binds with probability approaching one. We prove this formally here, with an argument analogous to that underlying Proposition 3. Proposition 5 concerns instead a sufficiently low principal's prior, and the corresponding argument is symmetric.

Proposition 6. *Fix the agents' information structure and a monotone referendum with single cutoff* $m_1 = \frac{1}{2}$. *If the principal's prior* $\Pr(\omega = 1)$ *is sufficiently high, then there exist equilibrium sequences in which the realized policy set is* $R(0)$ *with probability converging to one as* $N \to \infty$.

Sketch. For any large enough N, we construct an equilibrium strategy σ_N with the mean action below the cutoff in each state,

$$q(0; \sigma_N) < q(1; \sigma_N) < \frac{1}{2} - \gamma, \tag{81}$$

for some $\gamma > 0$. An application of the law of large numbers then yields the claim.

We use the same parametric family of cutoff strategies σ_L as in the proof of Proposition 3, defined in (74). As shown there,

$$q(0; \sigma_L) < q(1; \sigma_L) \qquad \text{for all } L > 0.$$

We also use the same compact parameter interval $[\underline{L}, \bar{L}]$ from (75). As in the proof of Proposition 3, the posterior likelihood ratio conditional on the pivotal event piv_0 lies in the interval $[\underline{L}, \bar{L}]$; cf. (76).

The critical lemma in the proof of Proposition 3 was Lemma 8, and its analogue is the following.

Lemma 9 (Self-map lemma)**.** *For any* $\rho_1 < \bar{q} < \frac{1}{2}$, *if* $\Pr(\omega = 1)$ *is sufficiently high, then there exist* $\bar{N} \in \mathbb{N}$ *and* $\bar{\delta} > 0$ *such that, for every* $N \geq \bar{N}$, *every* $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$, *and every principal best response* $(\bar{k}_N, \tilde{x}_N)$ *to* σ_L,

(i)

$$q(0; \sigma_L) < q(1; \sigma_L) < \bar{q},$$

(ii) the agents' best response to σ_L *and* $(\bar{k}_N, \tilde{x}_N)$ *is again a cutoff strategy* $\sigma_{L'}$ *with*

$$L' \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}].$$

The proof is analogous to that of Lemma 8, except that part (i) is now based on a uniform *upper* bound on $q(1;\sigma_L)$ rather than a lower bound on $q(0;\sigma_L)$. When $\Pr(\omega = 1)$ is sufficiently high, the interval $[\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$ lies far to the right, so the cutoff types $p_L(s)$ are uniformly close to 1 for any L from this interval. Hence, for every $\bar{q} \in (\rho_1, \frac{1}{2})$, all sufficiently high $\Pr(\omega = 1)$, and all sufficiently small $\bar{\delta} > 0$,

$$q(0;\sigma_L) < q(1;\sigma_L) < \bar{q} \qquad \text{for all } L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}].$$

Part (ii) then follows exactly as in Lemma 8: Fix $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$. First, we show that the principal's cutoff $\frac{\bar{k}_N}{N}$ lies strictly between the two mean actions when N is large, i.e. (78) holds. Second, conditional on being pivotal and as $N \to \infty$, an agent almost certainly affects the principal's preference over policies instead of the feasible policy set, i.e. (80) holds. Third, there is a uniform bound $\bar{N}$ so that for any $N \geq \bar{N}$, the agents' best response is again a cutoff strategy with parameter in the same interval.

Finally, we conclude exactly as in the proof of Proposition 3. For every $N \geq \bar{N}$, let Γ_N denote the correspondence on $[\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$ that maps a parameter L to the set of parameters L' generated by:

(a) choosing a principal best response $(\bar{k}_N, \tilde{x}_N)$ to σ_L,

(b) choosing an agents' best response $\sigma_{L'}$ to σ_L and $(\bar{k}_N, \tilde{x}_N)$.

By Lemma 9, Γ_N maps the interval into itself. Further, it has non-empty, compact, convex values and a closed graph. Kakutani's fixed point theorem therefore yields a fixed point L^*_N for every $N \geq \bar{N}$. The corresponding strategy $\sigma_{L^*_N}$ is an equilibrium strategy. By Lemma 9(i),

$$q(0;\sigma_{L^*_N}) < q(1;\sigma_{L^*_N}) < \bar{q}.$$

Choose $\gamma > 0$ such that $\bar{q} < \frac{1}{2} - \gamma$. Then

$$q(0;\sigma_{L^*_N}) < q(1;\sigma_{L^*_N}) < \frac{1}{2} - \gamma,$$

so the equilibrium sequence satisfies (81), as claimed. □

I Heterogeneous Ex-Post Preferences

We consider a variation of the baseline model from Section 1 in which the voters have a private preference type, and payoffs can depend on the state in a general way. Except for this additional type dimension, the baseline model is unchanged.

Formally, each agent i is a (non-strategic) partisan for $a \in \{0,1\}$ and chooses $a_i = a$, with probability $0 < \rho_a < \frac{1}{2}$, as before. A non-partisan agent has a private type, which is a prior $p_i \in [0,1]$ and a pair $\boldsymbol{t}_i = \Big(t_i(0), t_i(1)\Big) \in [0,1]^2$, describing the type's constant marginal benefit from the policy choice in the two states. Types $\boldsymbol{t}_i$ are drawn from a distribution G and independently of priors, signals, and the state, and independently across agents.

A type $\Big(p_i, t_i(0), t_i(1)\Big)$'s payoff from x in ω is

$$x\Big(t_i(\omega) - c\Big),$$

Given a strategy profile η, a type prefers the action 1 if and only if

$$\begin{aligned} & p_i\Big(t_i(1) - c\Big)\ \frac{\Pr(s_i = s|\omega = 1)}{\Pr_i(s_i = s)} U(1;\eta) \\ -(1-p_i)\Big(c - t_i(0)\Big) & \ \frac{\Pr(s_i = s|\omega = 0)}{\Pr_i(s_i = s)} U(0;\eta) \geq 0. \end{aligned} \tag{82}$$

We assume that the mass of types for which $p_i\Big(t_i(1)-c\Big) = 0$ *and* $(1-p_i)\Big(c-t_i(0)\Big) = 0$ is zero and ignore these types in the following, without loss of generality. Finally, we generalize the notion of an agents' information structure π to mean the pair of a signal and a type distribution.

I.1 The Aggregate Preference Function Φ

We show that the equilibrium set in this generalized setting depends on the type distribution only through the *aggregate preference function*

$$\begin{aligned} & \Phi\Big(U(0;\eta), U(1;\eta), l\Big) \\ = & \ \rho_1 + (1-\rho_1-\rho_0)\Pr\Big(\{(p_i, \boldsymbol{t}_i) : p_i\Big(t_i(1)-c\Big)U(1;\eta) \geq (1-p_i)\Big(c - t_i(0)\Big)\cdot l \cdot U(0;\eta)\}\Big) \end{aligned}$$

for $l := \frac{\Pr(s_i=s|\omega=0)}{\Pr(s_i=s|\omega=1)}$, via two observations:

First, equilibria are equivalently characterized by a principal's strategy $(\bar{k}, \tilde{x})$ and a mean action pair $\mathbf{q} = \Big(q_0, q_1\Big)$ so that $(\bar{k}, \tilde{x})$ and $\mathbf{q}$ are best replies to $(\bar{k}, \tilde{x})$ and $\mathbf{q}$. To make sense of this, note that, for any strategy profile $\eta = \Big(\sigma, (\bar{k}, \tilde{x})\Big)$, the mean action pair $\mathbf{q}(\sigma) = \Big(q(0;\sigma), q(1;\sigma)\Big)$ pins down the set of principal's best replies; $\mathbf{q}$ and $(\bar{k}, \tilde{x})$ together pin down the average effects $U(\omega';\eta)$;[37] and the average effects are a sufficient statistic for the agents' best reply, given (82). In conclusion, $\mathbf{q}(\sigma)$ and $(\bar{k}, \tilde{x})$ are a sufficient statistic for the best reply correspondence, which yields the claimed equilibrium characterization.

Second, multiplying (82) by $\frac{\Pr(s_i=s)}{\Pr(s_i=s|\omega=1)}$ shows the best reply correspondence's mean action pairs depend on the type distribution only through Φ. Consequently, the same is true for the equilibrium set, as claimed.

Before we proceed to deriving analogs of the main results for the heterogeneous-preference setting, we recall from the main text that the value of Φ at $\big(U(0;\eta), U(1;\eta), l\big)$ only depends on

$$z_1 := \frac{U(0;\eta)}{U(1;\eta)} \cdot l.$$

whenever $U(1;\eta) \neq 0$, and that an agents' information structure has *monotone preferences* if Φ is continuously differentiable in z_1 and $\partial\Phi/\partial z_1$ has the same non-zero sign for all

---

[37] Cf. (1) and (2).

$z_1 \in (0, \infty)$.

I.2 Analog of Proposition 2

An analog of Proposition 2's statement about information aggregation holds in the heterogeneous-preference setting.

Proposition 3′. *Consider any monotone referendum with a single cutoff and any agents' information structure with monotone preferences. Information aggregates in all equilibrium sequences if the referendum has no balance and* $\max R(0) < \max R(1)$.

The proof of Proposition 3' closely follows that of Proposition 2. We therefore do not repeat the full argument, but only describe the necessary modifications.

In the baseline proof, the first step establishes Lemma 2. This lemma continues to hold in the heterogeneous-preference setting, with an identical proof.

The baseline argument then shows that, for each signal $s \in \{0, 1\}$, Lemma 2 implies the existence of a unique indifferent type $p_N(s) \in (0, 1)$. The proof proceeds by distinguishing two complementary cases:

$$0 < \lim_{N\to\infty} p_N(1) < \lim_{N\to\infty} p_N(0) < 1, \tag{83}$$

or

$$\lim_{N\to\infty} p_N(1) = \lim_{N\to\infty} p_N(0) \in \{0, 1\}, \tag{84}$$

and shows that each case leads either to a contradiction or to information aggregation.

In the heterogeneous-preference setting, equilibrium strategies are no longer characterized by two indifferent types. Instead, the analogous dichotomy is expressed in terms of the ratio of average effects. We distinguish the cases

$$\lim_{N\to\infty} \frac{U(0; \eta_N)}{U(1; \eta_N)} \in (0, \infty), \tag{85}$$

and

$$\lim_{N\to\infty} \frac{U(0; \eta_N)}{U(1; \eta_N)} \in \{0, \infty\}. \tag{86}$$

In the first case, monotonicity of preferences implies that the mean actions differ across states, i.e. $q(0; \sigma) \neq q(1; \sigma)$ (cf. (16)). Since the realized collective action concentrates around the mean in each state, the principal learns the state from the observed outcome, and information aggregates.

In the second case, note that in the baseline model (86) is equivalent to (84) via (12). Hence, the baseline proof can be adapted almost verbatim: (86) either yields a contradiction (generic Case 1) or implies information aggregation (non-generic Case 2). The required modifications are as follows.

1. In the baseline proof, (30) follows from the fact that the indifferent types are bounded away from 0 and 1. In the present setting, the same conclusion follows directly from (86): if (30) failed, then the ratio $\frac{U(0;\eta_N)}{U(1;\eta_N)}$ would remain bounded away from 0 and ∞, contradicting (86).

2. In Case 2 of the baseline proof, subcases are distinguished according to:
   (a) whether $\lim_{N\to\infty} \Pr(\omega = 1 \mid \text{piv}_1; \sigma_N, N)$ equals 0 or 1, and
   (b) whether types choose action 1 for $p_i \leq p_N(s)$ or for $p_i \geq p_N(s)$.

   In the baseline model, (b) is equivalent to the sign of $U(1;\eta_N)$: if $U(1;\eta_N) < 0$, a type chooses 1 iff $p_i \leq p_N(s)$, and if $U(1;\eta_N) > 0$, a type chooses 1 iff $p_i \geq p_N(s)$. Since the object $U(1;\eta_N)$ is also well-defined in the heterogeneous-preference setting and given this equivalence, the same case distinction and proof can be reused.

3. The first paragraph of Case 2 is replaced by:

   "This case can be decomposed into several analogous subcases; we present one. Consider an equilibrium sequence such that $\lim_{N\to\infty} \Pr(\omega = 1 \mid \text{piv}_1; \sigma_N, N) = 1$ *and* $U(1;\eta_N) < 0$ *for all* N*. Monotonicity of preferences then implies* $\rho_1 < q(1;\sigma_N) < q(0;\sigma_N) < 1 - \rho_0$*. By* (86)*, either* $\frac{U(0;\eta_N)}{U(1;\eta_N)} \to \infty$ *or* $\to 0$*. If the ratio diverges to* ∞*, then* $q^* = 1 - \rho_0$*, which contradicts* $\lim_{N\to\infty} \Pr(\omega = 1 \mid \text{piv}_1; \sigma_N, N) = 1$*. Hence the ratio converges to* 0*, implying* $q^* = \rho_1$*."*

4. In the subsequent analysis, the expression $F\Big(p_N(s)\Big)$ is replaced by $\Phi\Big(z_1(s), z_2\Big)$, where

$$z_1(s) = \frac{U(0;\eta_N)\Pr(s_i = s \mid \omega = 0)}{U(1;\eta_N)\Pr(s_i = s \mid \omega = 1)}, \qquad z_2 = \text{sign}\Big(U(1;\eta_N)\Big).$$

I.3 Analogs of Theorem 1 and 2

The analogs of Theorem 1 and 2 hold, provided monotone preferences and the condition that a majority of types ranks policy 0 highest in state 0 and policy 1 highest in state 1. Formally,

$$\rho_1 + (1 - \rho_1 - \rho_0)\Pr\Big(\{\boldsymbol{t}_i : t_i(1) - c > 0\}\Big) > \frac{1}{2}, \text{ and} \tag{87}$$

$$\rho_0 + (1 - \rho_1 - \rho_0)\Pr\Big(\{\boldsymbol{t}_i : c - t_i(0) > 0\}\Big) > \frac{1}{2}. \tag{88}$$

Theorem 1′. *Any collective veto of the maximum has a payoff guarantee of* $1 - \varepsilon$ *across all agents' information structures with monotone preferences and satisfying* (87) *and all equilibrium sequences. This maximizes the payoff guarantee across all referenda.*

Theorem 2′. *The gateway referendum with the simple majority cutoff* $m_1 = \frac{1}{2}$ *has a payoff guarantee larger than* $1 - \varepsilon$ *across all agents' information structures with monotone preferences and satisfying* (87) *and all equilibrium sequences. This maximizes the payoff guarantee across all generalized referenda.*

The proofs of Theorem 1' and Theorem 2' are almost the same as those of Theorem 1 and Theorem 2, whose proofs relied on Proposition 1 and Proposition 2.

Proposition 1 showed the existence of an information structure and a corresponding inefficient equilibrium sequence. It remains true since we only expanded the set of feasible information structures by allowing for heterogeneous ex-post preferences. We already proved an appropriate analog of Proposition 2 in Section I.2.

With these, the same proofs as before can be mimicked. In the proof of Theorem 2', one has to take care to replace and invoke (87) instead of its baseline model version condition $\rho_1 < \frac{1}{2} < 1 - \rho_0$ whenever the latter one was needed in an argument.

I.4 Agent-Optimal referenda

We provide sufficient conditions for the referenda defined by (4) and also the generalized referenda defined by (7) to be also agent-optimal in their respective settings. That is, they maximize the agents' payoff guarantee. By this we mean the percentage of the full-information payoff achieved in the worst-case scenario and as $N \to \infty$, by a social planner who has full information about the state and maximizes the agent's ex-ante welfare

$$\int_{p_i=0}^{1} p_i \mathrm{E}_G\Big(t_i(1) - c\Big) + (1 - p_i)\mathrm{E}_G\Big(t_i(0) - c\Big) dF(p_i).$$

The first condition is

$$0 \leq \mathrm{E}_G\Big(t_i(0)\Big) < c, \text{ and } 1 \geq \mathrm{E}_G\Big(t_i(1)\Big) > c$$

and means that, when the state is known, policies are ranked in the same way whether considering the principal's or the agents' ex-ante welfare (lower policies are strictly preferred in state 0 and higher ones in state 1). This means the agents' full information payoff is

$$\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(1)\Big) - c\Big)$$

and the agents' payoff guarantee is

$$\hat{G}(R) := \inf_{(\eta_N)_{N\in\mathbb{N}},\pi} \left(\liminf_{N\to\infty} \mathrm{E}(x \mid \omega = 1; \eta_N) - \frac{\mathrm{E}_F(1 - p_i)\Big(c - \mathrm{E}_G\Big(t_i(0)\Big)\Big)}{\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(1)\Big) - c\Big)} \mathrm{E}(x \mid \omega = 0; \eta_N) \right)$$

It differs from the principal's payoff guarantee simply by replacing $\Pr(\omega = 0)$ with $\mathrm{E}_F(1-p_i)\Big(c - \mathrm{E}_G\Big(t_i(0)\Big)\Big)$ and $\Pr(\omega = 1)$ with $\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(1)\Big) - c\Big)$.

The second condition is

$$\frac{1 - \mathrm{E}_F(p_i)}{\mathrm{E}_F(p_i)} \cdot \frac{c - \mathrm{E}_G\Big(t_i(0)\Big)}{\mathrm{E}_G\Big(t_i(1)\Big) - c} \geq \varepsilon. \tag{89}$$

and the relevant implication is that choosing $x = 1$ in both states yields an agent's ex-ante payoff smaller than $1 - \varepsilon$ times the full-information payoff.[38]

Given the two conditions, mimicking Section 3.1's proof of Theorem 1 shows that the referenda (4) maximize the agents' payoff guarantee across all referenda. Similarly, mimicking the proof of Theorem 2 in Appendix C shows that the referenda (7) maximize

---

[38] The payoff from choosing $x = 1$ in both states divided by the full-information payoff is $1 - \frac{\mathrm{E}_F(1-p_i)\Big(c - \mathrm{E}_G(t_i(0))\Big)}{\mathrm{E}_F(p_i)\Big(\mathrm{E}_G(t_i(1)) - c\Big)}$, which is smaller than $1 - \varepsilon$ if (89) holds.

the agents' payoff guarantee across all generalized referenda.