Vague Politics *

Carl Heese †

A policy has to be selected. Opinions and information about it are dispersed among many agents, and a principal learns from observing their collective action. A partial-commitment mechanism maps the observed collective action to a menu of multiple policies, from which the principal must choose. Which mechanisms maximize the players' payoff guarantee across diverse information scenarios? It turns out certain *vague* mechanisms are optimal: they exclude only one policy from the full policy space, and do so only sometimes. Such mechanisms guarantee near full-information payoffs when the policy space is fine, and even outperform all full-commitment mechanisms in some scenarios.

The standard literature on information aggregation in collective choice problems largely considers mechanisms of unlimited commitment power. For example, models of majority voting typically map the agents' collective action to a single policy. In most applications, however, commitment to a precise policy is infeasible.

This paper considers mechanisms of *partial commitment*, which map the agents' collective action to a range of possible policies. The principal is committed to that range but may choose any policy from within it.

In addition to the practical infeasibility of precise commitments, there is another motivation for the idea of partial commitment: it directly addresses a central problem of democracy, namely, the distribution of authority between the populace and the officials of the state. Under a partial-commitment mechanism, the agents collectively determine the range of policies that may be implemented, then delegate the choice within that range to the principal. Historically, this question of delegation has been crucial in the design of democratic systems; it is the core of

*September 29, 2025. I thank Stephan Lauermann, Wolfgang Pesendorfer, Wing Suen, Balázs Szentes, and audiences at the University of Vienna, the University of Hong Kong, NYU Shanghai, 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

the dichotomy between direct and indirect democracy. For instance, the Founding Fathers of the United States debated how much autonomy should be granted to representatives of the people. They distinguished between *delegates*, who were expected to act as direct extensions of their constituents, and *trustees*, who were free to exercise their own judgment in making policy decisions (Burke, 1774; Madison, Hamilton and Jay, 1788). Delegation of decision-making is also an important issue in other applications, such as governance in large organizations.

We study a basic model of partial commitment in which a policy x must be selected from a finite, ordered policy space $\{0, x_2, \ldots, x_{l-1}, 1\}$. The key friction is that policy-relevant information is dispersed: it is uncertain whether the marginal benefit of the policy exceeds its marginal cost. Information about this binary state of the world is dispersed among N agents (where N is large). They hold private information about it in the form of conditionally independent, noisy binary signals, and heterogeneous prior beliefs: some are *partisan* (certain of one state), while others hold uncertain priors. The agents' information satisfies typical regularity conditions that have appeared in prior work on majority rules with unlimited commitment power (Bhattacharya, 2013).

As we are interested in simple, robust institutions, we consider simple mechanisms that have a small action space and respect the agents' anonymity. We call these mechanisms *processes of partial commitment*. Our main results characterize those processes of partial commitment that perform well across a wide range of information scenarios and all equilibria.

In any process, each agent observes his private information and chooses a binary action, 0 or 1. The agents' collective action, i.e., the average of their individual actions, is observed by a principal and determines a subset of the policy space, from which the principal selects the final policy. The principal's ex-ante optimal policy is the maximal one available to her.[1]

Our first result (Theorem 1) characterizes the processes of partial commitment that maximize the principal's ex-ante payoff guarantee (worst-case payoff) across all regular distributions of the agents' private information and all symmetric weak perfect Bayesian equilibria.[2] The theorem identifies a particular class of processes, namely *majority votes over exclusion of the maximal policy*, as being optimal, and it establishes that when the policy space is fine (i.e., when the parameter $\varepsilon :=$

[1]Our basic model assumes constant marginal cost and benefit. (This assumption is typical, for example, in the literature on public good referenda; see, e.g., Ledyard and Palfrey (2002).) Thus, given any prior, either the maximal or the minimal policy is optimal. In this paper we assume without loss of generality that the principal's prior is such that the maximal policy is optimal.

[2]We impose a simple tie-breaking refinement that implies trembling-hand perfection (Selten, 1988).

$1 - x_{l-1} = x_2$ is small), these processes achieve near full-information payoffs. The processes take the following form: If the share of agents choosing 1 exceeds a certain cutoff, the policy set remains unrestricted, i.e., the principal can choose any policy from $\{0, x_2, \ldots, 1\}$. Otherwise, the maximal policy is excluded; the principal must choose from $\{0, x_2 \ldots, x_{l-1}\}$. The fundamental feature of these processes is their vagueness: they exclude only one policy, and do so only sometimes.

For comparison, let us consider as a benchmark the case in which the principal has unlimited commitment power. Here, the Condorcet jury theorem (Bhattacharya, 2013; Feddersen and Pesendorfer, 1998) implies that full-information payoffs are achieved by a majority vote between the two extreme policies, $x = 0$ and $x = 1$—a process in which the principal never has more than one policy to choose from, in sharp contrast to the vagueness of the optimal processes identified in Theorem 1. The benchmark result suggests that under partial commitment, a simple majority vote between the most extreme policy sets possible, $\{0, x_2\}$ and $\{x_{l-1}, 1\}$, might be a natural candidate for an optimal process.

However, this candidate process is suboptimal.[3] The reason is that it allows for the agents to miscoordinate on inefficient equilibria. In fact, we give a broader result (Theorem 2) that says that processes of partial commitment *always* allow for such miscoordination: While some equilibria may yield full-information payoffs, other inefficient ones always exist. Specifically, there are natural inefficient equilibria in which nearly all agent types truthfully match their action to their signal.[4] Figure 1 illustrates such equilibria for the process described above (the majority vote between $\{0, x_2\}$ and $\{x_{l-1}, 1\}$). In the equilibrium on the left, the principal is constrained to choose a policy in $\{0, x_2\}$, even in the state where the ideal policy is $x = 1$. In the equilibrium on the right, a similar error occurs in the opposite state.

This large scope for miscoordination suggests that it may be optimal to delegate the decision fully to the principal—i.e., to allow her to choose *any* policy, regardless of the agents' collective action. However, such a process renders the agents' actions cheap talk and leads to equilibria without *any* information transmission. Intuitively, for information transmission to occur, the process must provide some incentives in the form of commitments from the principal.

Our next result (Theorem 3) addresses the latter issue: It characterizes the processes of partial commitment that ensure information aggregation; that is, they induce sufficient information transmission for the principal to learn the state from

[3] As we show, even a coin flip between $x = 0$ and $x = 1$ has a higher payoff guarantee than the candidate process, for most priors.

[4] For any process of partial commitment, there is always a range of information structures with such "approximately truthful equilibria."

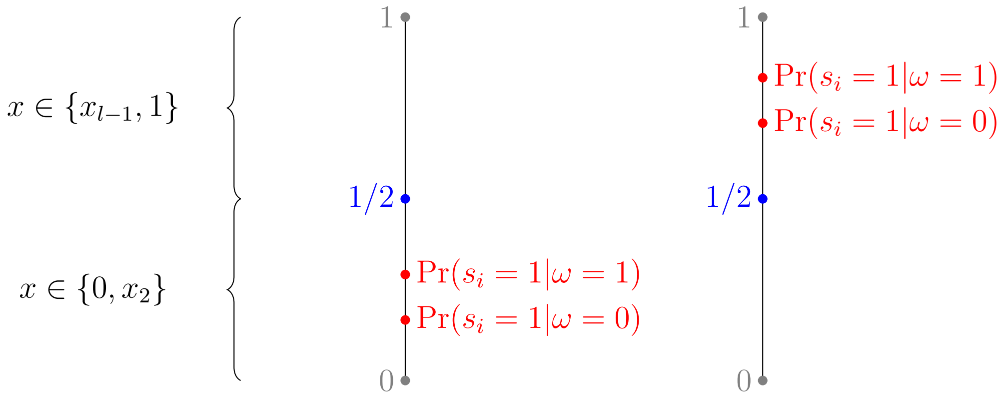

Figure 1: *Approximately truthful* equilibria. A random agent chooses 1 with approximately the likelihood of signal 1, $\Pr(s_i = 1|\omega)$, in each state $\omega \in \{0, 1\}$. In the left panel, the majority cutoff is not reached in either state; in the right panel, it is reached in both.

the agents' collective action (almost surely, as $N \to \infty$). A critical property of these processes is an imbalance condition that requires an unequal split of authority between the principal and the agents, in an appropriate sense. Theorem 3 will imply that vague commitments, which grant only minimal decision authority to the agents, provide sufficient incentives for information aggregation.

Jointly, Theorems 2 and 3 will imply that the processes of Theorem 1—majority votes over exclusion of the maximal policy—minimize possible choice errors across all information scenarios. The principal learns the state, and she fails to achieve the full-information payoff only when she is constrained by the agents to choose $x = x_{l-1}$ instead of $x = 1$ in the state where $x = 1$ is optimal (cf. Figure 1).

Our final main result (Theorem 4) points out that processes of partial commitment may even provide higher payoff guarantees than full commitment, if we relax one of the regularity conditions needed to apply the Condorcet jury theorem. The condition we drop is that the expected share of partisans of each type must be less than $\frac{1}{2}$. When a majority of the agents are partisans of the same type, under a full-commitment mechanism they will enforce their preferred policy, leading to outcomes that are not full-information equivalent. In contrast, under partial commitment, a majority vote over exclusion of the maximal policy will always ensure information aggregation and near full-information outcomes, regardless of the share of partisans.

The results in Section 1 (Theorems 1–4) highlight the interaction of two fundamental problems a principal may face with a large group of agents: coordination and information aggregation. When there are many agents, their inability to coordinate becomes the central issue. "Vague" commitments on the part of the principal simultaneously defuse the coordination problem and provide sufficient incentives for information aggregation. Consequently, the optimal processes are extreme and give almost full decision authority to the principal.

In Section 2 we consider some extensions addressing the limitations of our baseline model. The first limitation is that the baseline model is purely one of information aggregation: All of the players have common ex-post preferences, and the question is simply which processes effectively aggregate their private opinions and information. In Section 2.1, we expand the model to include a preference aggregation problem, allowing groups of agents to have opposing preferences depending on the state. Such preferences have been used to study distributive politics, in which the state of the world determines which group will benefit from a policy.[5] We find that the same processes remain principal-optimal, provided a monotonicity condition adapted from the "strong preference monotonicity" of Bhattacharya (2013). We also provide sufficient conditions for the principal-optimal processes to maximize the agents' ex-ante payoff guarantee. In Section 2.2, we relax the assumption of constant marginal cost and benefit and consider settings where the players' ex-ante preferences are single-basin over the policy space. This captures settings in which intermediate choices or compromises between two extremes are inefficient, such as decisions about the provision of public goods involving economies of scale.

Section 3 contains further results concerning our baseline model. We give a precise condition for our results to hold for arbitrary finite policy sets (rather than only those of the form specified in Section 1.1). We state a result about the existence of efficient equilibria, and we identify certain inefficient equilibria that arise from differences between the principal's and the agents' priors. Finally, we provide conditions under which *all* robust optimal processes are vague, i.e., they never exclude more than one policy.

In Section 4 we discuss the relationship of this paper to the broader literature on delegation and on information aggregation in politics. Here, we summarize the key contributions of our framework of partial-commitment mechanisms. These mechanisms are significant for a host of reasons and have not previously been considered in the study of collective choice problems.

In the information aggregation literature, there has been a large and influential body of work studying elections under the assumption of full commitment power, with a focus on supermajority voting between two alternatives; see, e.g., Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1997), and Krishna and Morgan (2012). Also, a recent stream of the literature has considered (cheap-talk)

[5] See Fernandez and Rodrik (1991), Ali, Mihm and Siga (2025), and Bhattacharya (2018).

models without any commitment power in order to model information aggregation in protests, polls, and informal politics more generally; see, e.g., Battaglini (2017).

Partial commitment adds a complementary positive theory of both standard collective choice mechanisms and informal politics. First, it addresses frictions in commitment, possibly minimal ones. Second, it captures those mechanisms of which partial commitment is a deliberate feature. For example, public referenda are often designed to have some binding implications while leaving details open (e.g., a constitutional referendum may decide whether to update a nation's constitution, but delegate the detailed specification of a new one to a constitutional convention). Similarly, firms use worker feedback polls to inform workplace policies, but delegate the design and implementation of the policies to other entities. Third, the optimality of vague commitments suggests a strategic role for vagueness, rooted in coordination issues. This observation suggests an explanation for the perception that political decision-makers communicate vaguely, have malleable agendas, and often break promises (cf. Shepsle (1972) and Page (1976)).

Our model of partial commitment also provides a benchmark for studying delegation when there is a large set of privately informed agents. The literature has largely focused on the opposite case of a single agent; see, e.g., Holmström (1978), Alonso and Matouschek (2008), and Dessein (2002). In that setting, full delegation to the agent is trivially optimal if the agent knows the state and preferences are common. The results in this paper provide a contrasting benchmark showing that if knowledge about the state is dispersed among many agents, ceteris paribus, minimal delegation to them is optimal. Our benchmark applies, as discussed above, to the question of how to distribute authority between elected officials and the populace in a democracy, as well as to similar delegation questions in large organizations with many departments. A particularly topical application may be to the use of an artificial intelligence (AI) system (corresponding to the principal in our model) to process and interpret inputs from workers at a firm.

More broadly, the concept of vague (or minimal) commitments provides a new perspective on ways for leaders to involve others (e.g., citizens or employees) in decision-making. Minimal commitments might be a useful feature for methods such as town halls, polls, open-door policies, or public referenda.

1 Processes of Partial Commitment

1.1 Model

A policy x needs to be chosen from a finite set of options $\mathcal{P} = \{x_1, \ldots, x_l\}$ with $x_1 < x_2 < \ldots < x_l$. 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 $\omega \in \{0, 1\}$, i.e., the players' payoff from x in ω is $x(\omega - \frac{1}{2})$.

There is a set of agents $\{1, \ldots, N\}$ who hold private information about the state. Each agent i receives a binary private signal $s_i \in \{0, 1\}$ satisfying $0 < \Pr(s_i = 1 \mid \omega = 0) < \Pr(s_i = 1 \mid \omega = 1) < 1$; so, signal 1 is an indication for state 1 and signal 0 an indication for state 0. The signals s_i are independent conditional on the state and are all drawn from the same distribution H. Each agent i also holds a private prior belief $p_i \in [0, 1]$ about the likelihood of $\omega = 1$; p_i is called agent i's *type*. Types are drawn independently from a distribution F on $[0, 1]$ satisfying certain regularity conditions from the literature (Bhattacharya, 2013): It exhibits "rich heterogeneity," meaning it has full support and atoms at 0 and 1, each with mass less than $\frac{1}{2}$. (The atoms correspond to partisan types that prefer either policy $x = 0$ or policy $x = 1$.) Furthermore, it is differentiable on $(0, 1)$.

In addition to the agents, there is a principal who has a commonly known prior $\frac{1}{2} < \Pr(\omega = 1) \leq 1 - \varepsilon$.[6]

The timing is as follows. First, a *process of partial commitment* P (defined below) is announced. Next, each agent i observes his private information and takes a binary action $a_i \in \{0, 1\}$. 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 $P(m)$ determined by the process P.

We define a *process* as a left-continuous mapping from $[0, 1]$ to the set of all subsets of $\mathcal{P}$, with at most finitely many discontinuities, and none at 0. A process P thus maps a collective action m to a policy set $P(m)$.[7] Every process P takes the form of a step function; that is, there exist a finite number of cutoffs $0 < m_1 < \ldots < m_R < m_{R+1} = 1$ such that $P(m)$ is constant on $[0, m_1]$ and on $(m_j, m_{j+1}]$ for $j = 1, \ldots, R$. A *process of partial commitment* is a process P such that $P(m)$ contains at least two policies for every m.

[6] The right constraint ensures that the payoffs from an equilibrium without any information transmission cannot be arbitrarily close to the full-information payoffs.

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

Given a process P, a principal's strategy is a mapping taking each $m \in [0,1]$ to a (possibly random) policy $\tilde{x} \in \Delta\Big(P(m)\Big)$. A symmetric agents' strategy is a mapping $\sigma : [0,1] \times \{0,1\} \to [0,1]$, where $\sigma(p,s)$ represents the likelihood that an agent with prior p and signal s chooses action 1. The analysis that follows focuses on weak perfect Bayesian equilibria in symmetric agents' strategies.

We impose the following tie-breaking rule. Unless otherwise indicated, we consider only equilibria in which, when a partisan of type 1 is indifferent, she chooses a random action distinct from that of a type-0 partisan, i.e., $\sigma(0,1) = \sigma(0,0) \neq \sigma(1,1) = \sigma(1,0) \in \{0,1\}$; the reverse holds for type-0 partisans. The key implication of this refinement is that for each action, there is a mass of agent types who will choose it with positive probability regardless of their signal. Essentially all our results hold for any "noise" refinement implying the same property; for example, they hold if we instead introduce non-strategic agent types that always choose prescribed actions, as has been done in the literature (see, e.g., Damiano, Li and Suen, 2025). The property implies that the mean action for each state is interior, i.e., $q(\omega';\sigma) := \mathrm{E}\Big(\sigma(p,s) \mid \omega = \omega'\Big) \in (0,1)$ for $\omega' \in \{0,1\}$, and we note that this implies trembling-hand perfection (Selten, 1988); see the online appendix.

1.2 Best Responses

Upon observing that k of the N agents have chosen action 1, the principal makes a Bayesian inference and best-responds. Her posterior is (assuming $q(0;\sigma) \in (0,1)$)[8]

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

The players' common utility from policy x in state ω is $x(1-c) = \frac{x}{2}$ if $\omega = 1$ and $-xc = -\frac{x}{2}$ if $\omega = 0$. Hence the principal's best response is $x = 1$ if $\frac{1}{2} < \Pr(\omega = 1 \mid k; \sigma, N)$ and $x = 0$ if $\frac{1}{2} > \Pr(\omega = 1 \mid k; \sigma, N)$. The principal is indifferent if $\frac{1}{2} = \Pr(\omega = 1 \mid k; \sigma, N)$, which can only happen for at most one realized collective action. This allows us to characterize the best reponse in terms of a cutoff $\bar{k} + 1$ as follows: If $q(1;\sigma) \geq q(0;\sigma)$, the posterior $\Pr(\omega = 1 \mid k; \sigma, N)$ is increasing in k. Then, given $\Pr(\omega = 1) < \frac{1}{2}$, either $\frac{1}{2} < \Pr(\omega = 1 \mid k; \sigma, N)$ 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[9]

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

[8] 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.

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

Analogously, if $q(1;\sigma) < q(0;\sigma)$, the posterior $\Pr(\omega = 1 \mid k; \sigma, N)$ is decreasing in k. Then either $\frac{1}{2} < \Pr(\omega = 1 \mid k; \sigma, N)$ for all $0 \leq k \leq N$—in which case we set $\bar{k} = N$—or there is a maximal $\bar{k}$ with $0 \leq \bar{k} < N$ such that $\Pr(\omega = 1 \mid \bar{k}; \sigma, N) > \frac{1}{2} \geq \Pr(\omega = 1 \mid \bar{k} + 1; \sigma, N)$.

The principal can only be indifferent at $k = \bar{k}+1$. Any mixed best response of the principal is thus fully described by the cutoff $\bar{k}$ and a distribution $\tilde{x} \in \Delta\Big(P(\bar{k}+1)\Big)$ with realizations denoted by $\bar{x} \in \mathrm{supp}(\tilde{x})$.

We turn to the agents' best response. Fix a process P with cutoffs $(m_1, \ldots, m_{R+1})$, a strategy profile, and an agent i. Let k_{-i} denote the realized number of other agents choosing action 1. Then agent i's choice affects the policy outcome x only if a *pivotal event* occurs.

Specifically, let $\mathrm{piv}_{0,k}$ denote the event that $k_{-i} = k$, and let $\mathrm{piv}_0 = \cup_{k \in \{\bar{k}, \bar{k}+1\}} \mathrm{piv}_{0,k}$. For $j > 0$, let piv_j denote the event that $\mathrm{piv}_{j'}$ does not hold for $j' = 0, \ldots, j-1$ and $k_{-i} = \lfloor m_j \cdot N \rfloor$.[10] The events $\mathrm{piv}_{0,\bar{k}}$ and $\mathrm{piv}_{0,\bar{k}+1}$ are the only ones in which agent i's choice possibly affects the principal's preference (for 0 versus 1), and piv_j for $j > 0$ means agent i's choice changes the policy set available to the principal. In any other event, agent i's choice does not affect the policy outcome. Let $\mathrm{piv} = \cup_{j=0,\ldots,R}\, \mathrm{piv}_j$ (and note that, by definition the pivotal events piv_j are mutually exclusive).

Now, if agent i has signal $s_i = s$ and type $p_i = p$, then he prefers action 1 if

$$\Pr_i(\omega = 1 \mid p_i = p, s_i = s)\; U(1;\eta) - \Pr_i(\omega = 0 \mid p_i = p, s_i = s)\; U(0;\eta) > 0, \tag{2}$$

where $U(\omega';\eta)$ is the *average effect* on the policy outcome x of an additional agent's choosing action 1 in state ω',

$$\begin{aligned} U(\omega';\eta) := &\sum_{j=0,\ldots,R;\bar{x}\in\mathrm{supp}(\tilde{x})} \Pr(\tilde{x} = \bar{x}) \Pr(\mathrm{piv}_j \mid \omega = \omega'; \eta, N) \\ &\cdot \Big(\mathrm{E}(x \mid \mathrm{piv}_j, a_i = 1; \tilde{x} = \bar{x}; \eta, N) - \mathrm{E}(x \mid \mathrm{piv}_j, a_i = 0; \tilde{x} = \bar{x}; \eta, N)\Big), \end{aligned} \tag{3}$$

given a strategy profile $\eta = (\sigma, \bar{k}, \tilde{x})$.

1.3 Robust Optimal Processes of Partial Commitment

In this section we present our first main result, a characterization of processes of partial commitment that maximize the principal's payoff guarantee. We call such processes *robust optimal.* The principal's payoff guarantee is the proportion of

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

the full-information payoff that the principal obtains in the worst-case scenario, as $N \to \infty$. Formally, for a process P, the principal's payoff guarantee is defined as

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

where we take the infimum over all equilibrium sequences $(\eta_N)_{N\in\mathbb{N}}$ and all pairs of distributions $\pi = (F, H)$, which we refer to as *agents' information structures.*[11] [12] Since the principal and the agents share the same ex-post preferences, it turns out that maximizing the principal's ex-ante payoff guarantee is the same as maximizing the agents', provided the agents' mean prior satisfies the same extremeness bound as the principal's, $\mathrm{E}_F(p_i) \le 1 - \varepsilon$. We show this formally in Section 1.7.

Note that without the noise refinement imposed earlier, the worst-case analysis would be moot: for many processes, the worst case would be achieved by a trivial equilibrium, in which all agent types choose the same action.

Theorem 1 identifies a simple class of processes as being robust optimal: *majority votes over exclusion of the maximal policy.* These are processes P of the form

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

for some $m_1 \in (0, 1)$. They achieve near full-information payoffs.

Theorem 1. *Any majority vote over exclusion of the maximal policy is robust optimal and has a payoff guarantee of* $1 - \varepsilon$.

Majority votes over exclusion of the maximal policy pass the Wilson doctrine (Wilson, 1985): Their construction is detail-free, meaning it does not utilize knowledge about the agents' information. Moreover, their performance requires only that the principal know the mean actions $q(\omega; \sigma)$ (these are a sufficient statistic for her best response; see (1)); she needs no further knowledge about the agents' strategy or information.

The central observations driving Theorem 1 are laid out in Sections 1.4–1.6. These sections provide results about inefficient equilibria, information aggregation,

---

[11] Here, lim inf 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 $G(P)$.

[12] If no equilibrium sequence exists for some π, we set $G(P) = -\infty$. In Section 1.6 we will establish a general existence result that rules out $G(P) = -\infty$ for all monotone processes, defined as those where $\max P(m)$ and $\min P(m)$ are weakly increasing in m.

and equilibrium existence. The formal proof of Theorem 1 is in Section 1.7. Here we discuss the idea of the proof, as well as some alternative candidates for robust optimal processes.

The key property of the optimal processes of Theorem 1 is "vagueness": They allow the agents to exclude at most one policy through their collective action. The benchmark case in which the principal has unlimited commitment power provides a sharp contrast. There, the Condorcet jury theorem (Bhattacharya, 2013; Feddersen and Pesendorfer, 1997) implies that full-information payoffs are achieved by a simple majority vote between the two extreme policies, $x = 0$ and $x = 1$—a process that is the opposite of vague, as all policies but one are excluded. In view of this benchmark, it would be natural to suppose that under partial commitment, another candidate for an optimal process would be an approximation of the latter process: a simple majority vote between the most extreme policy sets possible under partial commitment, namely $\{0, \varepsilon\}$ and $\{1 - \varepsilon, 1\}$.

However, as mentioned in the introduction, it turns out that these approximations are suboptimal, owing to a coordination problem of the agents: As we show in Section 1.4, for any process P of partial commitment and any associated policy set $P(m)$, there are inefficient equilibria in which the principal is constrained to choosing from $P(m)$ with probability approaching 1, as $N \to \infty$ (Theorem 2). In the case of the mentioned approximations, this means there are equilibria in which the highest policy the principal can choose is $x = \varepsilon$, even in state 1 (where $x = 1$ would be optimal), implying a payoff guarantee of at most ε. This is even lower than the payoff guarantee of a random choice between $x = 0$ and $x = 1$, for most values of the principal's prior.[13]

The observation that *any* policy set arising from the process can become a certain constraint on the principal's choice suggests another candidate for optimality: a process P that never excludes any policies ($P(m) = \mathcal{P}$ for all m). Under such a process, the agents' actions are cheap talk. However, it is easy to see that cheap talk is not optimal. For information transmission to occur in all equilibria (that survive the tie-breaking refinement), the principal's ex-ante optimal policy $x = 1$ must be excluded after some collective action. Otherwise, there is an "uninformative" equilibrium without any information transmission: All non-partisans choose the same action, and each partisan matches his action to his type, independent of his signal. The principal learns nothing about the state from observing the collective action; she chooses her ex-ante optimal policy $x = 1$ independent of her observation. This

[13] The payoff guarantee from the random choice is $\frac{1}{2} - \frac{\Pr(\omega=0)}{\Pr(\omega=1)} \cdot \frac{1}{2}$, which exceeds ε if $\Pr(\omega = 1) > \frac{1}{2(1-\varepsilon)}$.

makes all agent types indifferent between both actions, rationalizing their uninformative behavior.[14]

Conversely, for the processes defined by (4) (majority votes over exclusion of the maximal policy), Theorem 3 in Section 1.5 shows that in any equilibrium sequence $(\sigma_N)_{N\in\mathbb{N}}$, *information aggregates*: the principal learns the state almost surely as $N \to \infty$, i.e., $\lim_{N\to\infty} \Pr(\omega = 1 \mid \sigma_N, N) = 1_{\omega=1}$.

Taken together, our results establish that any process P has a lower payoff guarantee than the processes defined by (4). This is because, for such a process, either $x = 1$ is never excluded from the available policies but then information transmission fails in some equilibrium completely, or $P(m) \subseteq \mathcal{P}\backslash\{1\}$ for some $m \in (0,1)$ but then there is an equilibrium where the principal's policy choice is weakly more constrained than in any equilibrium of the processes (4), by Theorem 2. In fact, if $P(m) \subsetneq \mathcal{P}\backslash\{1\}$, the process has a *strictly* lower payoff guarantee.

1.4 Miscoordination on Approximately Truthful Behavior

The Condorcet jury theorem (as in Bhattacharya (2013)) implies that when the principal has unlimited commitment power, simple majority voting between $x = 0$ and $x = 1$ implies efficient outcomes in all equilibrium sequences as $N \to \infty$. Theorem 2, by contrast, shows that partial commitment implies the existence of inefficient equilibrium sequences. Specifically, given any process of partial commitment P and any of its cutoffs m_j, we construct an equilibrium sequence in which the chosen policy set is $P(m_j)$ with probability converging to 1 as $N \to \infty$. Thus, whenever $P(m_j)$ excludes an ex-post optimal policy, i.e., either $x = 0$ or $x = 1$, the principal is constrained to choose a suboptimal policy in at least one of the states.[15]

Theorem 2. *Consider any process of partial commitment with cutoffs* $0 < m_1 < \ldots < m_R < m_{R+1} = 1$*. For any* $j = 1, \ldots, R+1$*, there exist an agents' information structure and a sequence of equilibrium strategies* $(\sigma_N)_{N\in\mathbb{N}}$ *for which*

$$\lim_{N\to\infty} \Pr\Big(m \in P^{-1}(P(m_j)) \mid \sigma_N, N\Big) = 1.$$

As we will show in Section 3, many processes have efficient equilibrium sequences for *all* agents' information structures. So Theorem 2 shows that partial commit-

[14] The literature provides various results in the same spirit for discrete-type settings, showing that binary cheap talk with many senders implies the existence of equilibria with little or no information transmission; see, e.g., Battaglini (2017) and Chen (2025).

[15] On the other hand, if P never excludes both $x = 1$ and $x = 0$, then there is always an (inefficient) equilibrium without any information transmission, as argued in the preceding Section 1.3.

ment creates a coordination problem: The agents may miscoordinate to achieve an inefficient equilibrium instead of an efficient one.

The formal proof of Theorem 2 is in the appendix. Here, we provide a sketch. The idea of the proof is to construct a sequence of equilibria in "approximately truthful" strategies. Formally, for $\delta > 0$, an agents' strategy σ is *(δ-)approximately truthful* if, for any given realized signal, a share of at least $1 - \delta$ types match their actions to the signal. Equilibria in approximately truthful strategies (which we call approximately truthful equilibria) exist for all small δ and some agents' information structure with the following two properties. First, the agents' signals are sufficiently uninformative:

$$\Pr(s_i = 1|\omega = 1) - \Pr(s_i = 1|\omega = 0) \leq \delta. \tag{5}$$

Second, the agents' priors are relatively close to the principal's prior:

$$\Pr_F\Big(p_i \in [\underline{p}(1), \Pr(\omega = 1)]\Big) > 1 - \frac{\delta}{2}, \tag{6}$$

for a certain bound $\underline{p}(1)$.

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. Below we show that, for some fixed information structures with the properties (5) and (6), the agents' incentives given any approximately truthful strategy sufficiently approximate those in the common-value game. This will imply the existence of an equilibrium in approximately truthful strategies.

Figure 2 shows an example process and an agents' information structure satisfying (5) and (6). 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 $\Pr(s_i = 1|\omega)$. They lie in between m_3 and m_4, so that, for any approximately truthful strategy, the realized policy set is almost surely $P(m_3)$ as $N \to \infty$, by an application of the law of large numbers.

A key statistic in our analysis is a measure of the distance between the mean action in each state, $q(\omega'; \sigma_N) = \mathrm{E}\Big(\sigma_N(s)|\omega = \omega'\Big)$, on the one hand, and each cutoff of the process or the principal's cutoff $\bar{k}$, on the other. This distance is given by

$$\mathrm{KL}\Big(m_j, q(\omega'; \sigma_N)\Big)$$

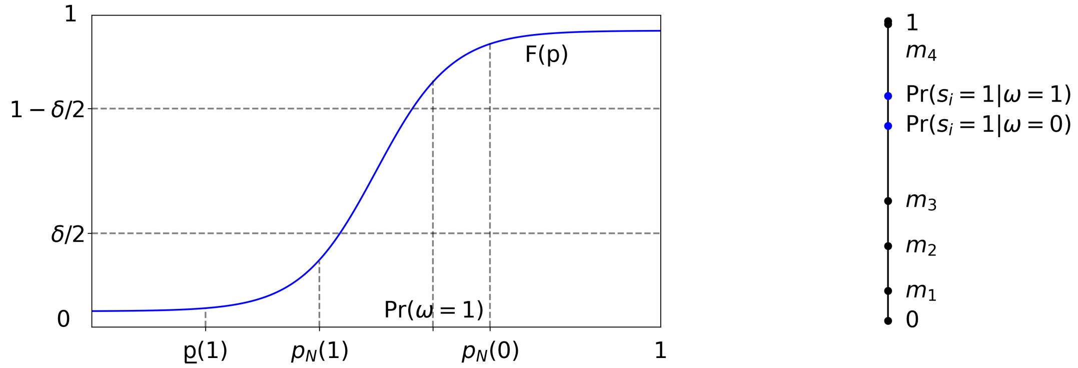

Figure 2: The agents' information structure, given by the prior distribution F (left) and the signal probabilities $\Pr(s_i = 1|\omega = \omega')$ (right). The approximately truthful equilibrium is given by the types $p_N(0)$ and $p_N(1)$ that are indifferent after signals 0 and 1, respectively: An agent i with signal $s \in \{0, 1\}$ chooses $a_i = 1$ if and only if $p_i \geq p_N(s)$.

for $j = 0, \ldots, R$, $\omega' \in \{0, 1\}$, with $m_0 = \lim_{N\to\infty} \frac{\bar{k}}{N}$ denoting the principal's limit cutoff.[16] Here, $\mathrm{KL}(\gamma, q) := \gamma \log\left(\frac{\gamma}{q}\right) + (1-\gamma)\log\left(\frac{1-\gamma}{1-q}\right)$ is the Kullback–Leibler divergence. 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 the realized collective action equals the cutoff is more likely. In the appendix, we employ large deviation theory to show that, in either state, the likelihood of this pivotal event is exponential in $\mathrm{KL}\Big(m_j, q(\omega'; \sigma_N)\Big)$ and in the number N of agents:

$$\Pr(\mathrm{piv}_j|\omega = \omega'; \sigma_N, N) = \exp\Big(-(N-1)\mathrm{KL}\Big(m_j, q(\omega'; \sigma_N)\Big) + o(N)\Big). \quad (7)$$

This large deviation result allows us to precisely understand the agents' incentives. First, it implies that the principal's limit cutoff lies in between the mean actions, i.e.,

$$\lim_{n\to\infty} q(0; \sigma_N) < m_0 < \lim_{n\to\infty} q(1; \sigma_N).$$

If this were not the case, then we would have $\mathrm{KL}\Big(m_0, q(0; \sigma_N)\Big) \neq \mathrm{KL}\Big(m_0, q(1; \sigma_N)\Big)$ as $N \to \infty$,[17] so that (7) would imply that the principal's inference from observing a collective action of $\frac{\bar{k}}{N}$ is unbounded. However, this cannot be, given the definition of $\bar{k}$; see (1). Second, for any process and cutoff m_j, if the signal precision parameter $\delta > 0$ is sufficiently low, there are signal probabilities in between m_j and m_{j+1} that

[16]It is sufficient to show information aggregation for any subsequence where $\frac{\bar{k}}{N}$ converges given that the values of $\frac{\bar{k}}{N}$ are in the compact set $[0, 1]$. We identify the subsequence with the original sequence to omit the subsequence notation.

[17]Note here that σ_N being δ- approximately truthful implies $\lim_{n\to\infty} q(0; \sigma_N) < \lim_{n\to\infty} q(1; \sigma_N)$ for δ sufficiently small.

are much closer to each other, and thus to the principal's cutoff m_0, than to any cutoff of the process; cf. Figure 2. We can choose δ small enough and the signal probabilities so that there is $\gamma > 0$ and

$$\gamma + \mathrm{KL}\Big(m_0, q(\omega'; \sigma_N)\Big) < \mathrm{KL}\Big(m_j, q(\omega'; \sigma_N)\Big) \text{ for all } j > 0 \text{ and } \omega' \in \{0, 1\}, \tag{8}$$

given any approximately truthful strategy σ_N. By (7), the agents are then almost certain to influence the principal's preference for low versus high policies (and not the policy range he chooses from), conditional on being pivotal:

$$\lim_{N\to\infty} \mathrm{Pr}_i(\mathrm{piv}_0|\mathrm{piv}; \sigma_N, N) = 1. \tag{9}$$

In other words, the agents' incentives are almost the same as if their actions were cheap talk. Now, as previously mentioned, if all of the players share a common prior, then the cheap-talk game has an equilibrium in which *all* types are truthful. The property (6) ensures that the agents' prior distribution is sufficiently close to a common prior to guarantee that an *approximately* truthful equilibrium exists in the cheap-talk game. By (9), this equilibrium will extend to our main game. In the appendix we construct the equilibrium formally using a fixed point argument.

The left panel of Figure 2 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 a high and a low signal. 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.

1.5 Information Aggregation

In this section we derive a critical condition for information aggregation that is satisfied by the candidate processes (4). This condition is that decision-making power must not be divided in a "balanced" way between the principal and the agent body. Formally, a process P with a single cutoff $m_1 \in (0, 1)$ has *no balance* if

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

and it has *balance* otherwise. For example, a process given by

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

has balance, with $\frac{1}{2} = \max P(0) = \min P(1)$. It turns out that balance implies the existence of informative equilibria in which the principal's choice does not depend on the observed collective action, although she learns about the state from it.

The logic of these equilibria is simple; we illustrate it for the example process above. The agents follow a strategy σ_N under which the mean action is higher in state 0 than in state 1 ($0 < q(1;\sigma_N) < q(0;\sigma_N) < 1$), below $\frac{1}{2}$ ($q(0;\sigma_N) < \frac{1}{2}$), and such that the principal is indifferent when the majority threshold is just met. That is, if the principal observes $\bar{k} + 1 = \lfloor \frac{N}{2} \rfloor + 1$ actions 1, her posterior is

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

(One can show that strategies with such mean actions $q(\omega'; \sigma_N)$ exist whenever the number of agents is large enough.[18]) If she observes fewer than $\bar{k}+1$ actions 1, then she prefers high policies but can choose at most $x = \frac{1}{2}$. If she observes $\bar{k} + 1$ or more 1-actions, then she prefers low policies but has to choose at least $x = \frac{1}{2}$. Thus it is optimal for her to always choose $x = \frac{1}{2}$. This constant best response makes each agent indifferent between all strategies; in particular, the strategy σ_N that we started with is a best response for the agents.

In sequences of equilibria of this kind, information does not aggregate, as the mean actions are smaller than $\frac{1}{2}$, i.e. $q(0;\sigma_N) < \frac{1}{2}$ for all N. This way, in state 0, the realized number k of 1-actions is smaller than $\bar{k} + 1 = \lfloor \frac{N}{2} \rfloor + 1$ with a non-vanishing probability and the principal's posterior greater than $\frac{1}{2}$ since it is monotone decreasing in k. So the principal does not learn the state.

In these equilibria, the principal's choice exactly nullifies any effect of the agents' collective action, as in a deadlock between two opposed parties. This deadlock arises from the "balanced" split of decision-making power. Perversely, the deadlock occurs even when all players have a common preference and prior.

To conclude, no balance is necessary for information aggregation. The following theorem shows it is also a sufficient condition together with some other properties. The theorem gives a complete characterization of information aggregation for

---

[18]Take any $m_1 \in (0,1)$. Let $m' = \frac{\lfloor m_1 N \rfloor + 1}{N}$, and note that $\frac{\Pr(k=m'N|\omega=1)}{\Pr(k=m'N|\omega=0)} = \exp\Big(-N\Big(\mathrm{KL}(m', q(1)) - \mathrm{KL}(m', q(0))\Big)\Big)$; cf. (7). Consider any pair of mean actions $\mathbf{q} = \Big(q(0), q(1)\Big)$ with $0 < q(1) \leq q(0) < m_1$. If $q(1) = q(0)$, the principal learns nothing from her observations and $\Pr(\omega = 1|k = m'; \mathbf{q}, N) = \Pr(\omega = 1) > \frac{1}{2}$. 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 $\lim_{N\to\infty} \Pr(\omega = 1|k = m'; \mathbf{q}, N) = 0$. 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 $\Pr(\omega = 1|k = m'; \mathbf{q}, N) = \frac{1}{2}$. For any such $q(0)$ and $q(1)$, we can always find a strategy σ that induces these mean actions and where additionally the type-1 partisans choose the 1-action with a different probability than the type-0 partisans.

monotone processes with a single cutoff. A process is monotone if both $\min P(m)$ and $\max P(m)$ are weakly increasing.

Theorem 3. *Consider any monotone process of partial commitment P with a single cutoff $0 < m_1 < 1$ and any agents' information structure. Information aggregates in all equilibrium sequences if and only if the process has no balance and the maximum of its policy sets is non-constant, i.e.,* $\max P(0) < \max P(1)$.

The formal proof of Theorem 3 is in the appendix. Regarding the necessity of the two conditions, we have just explained how a violation of non-balance implies a failure of information aggregation, and in Section 1.3 we explained how a violation of the second property likewise implies a failure.We now sketch why, conversely, not violating the two conditions ensures information aggregation.

First, the condition $\max P(0) < \max P(1)$ rules out the possibility of "uninformative" equilibria, which we define as those having the same mean action in both states, $q(0;\sigma) = q(1;\sigma)$. To see this, recall from Section 1.1 that the tie-breaking rule implies $q(0;\sigma), q(1;\sigma) \in (0,1)$. Hence all collective actions $m \in \{0, \frac{1}{N}, \ldots, 1\}$ are on path. In an uninformative equilibrium, the principal learns nothing from observing the realized collective action; therefore, at any $m > m_1$, she chooses $x = \max P(1)$ since this is closest to her ex-ante optimal policy $x = 1$. At any $m \leq m_1$, she chooses $x(m) = \max P(0) < \min P(1)$. Thus the agents are pivotal in only one event, namely piv_1, which corresponds to m_1. The uninformativeness of the equilibrium implies that piv_1 has the same probability in both states, so that any agent with a uniform prior is indifferent before observing his signal. His best response (and that of nearby types) is then to match his action to his signal. But this means the equilibrium is informative after all. We conclude that there are no uninformative equilibria.

Second, non-balance causes the logic of the "deadlock equilibrium" to fail. The deadlock equilibrium is supported by the indifference of all agent types; each type is indifferent because his action does not affect the policy outcome. Non-balance eliminates this indifference: It implies that an agent's average effect on the policy outcome, given by (3), is non-zero and has the same sign in both states, i.e., in any equilibrium η_N,

$$\begin{aligned} &\text{either } U(\omega';\eta_N) > 0 \text{ for all } \omega' \in \{0,1\}, \\ &\quad\text{or } U(\omega';\eta_N) < 0 \text{ for all } \omega' \in \{0,1\}. \end{aligned}$$

We prove this assertion in two steps. First we argue that the average effect is non-zero in at least one state, i.e.,

$$U(0;\eta_N) \neq 0 \text{ or } U(1;\eta_N) \neq 0. \tag{10}$$

The idea of the argument is that, although in principle the average effect in either state may be zero (e.g., when the effects in various pivotal events may cancel out exactly), the informativeness of the agents' strategy prevents this from happening.

To begin, consider the agents' strategy in the deadlock equilibrium (which had just one pivotal event, piv_1). Here, given that $\max P(0) < \max P(1)$, the principal's best response is no longer constant; it increases at the cutoff m_1 of the process. So $U(\omega;\eta_N) > 0$. Now consider the more general case in which there are two pivotal events. For the purposes of illustration, suppose these are when either $k_{-i} = \bar{k}$ or $k_{-i} = \lfloor m_1 N \rfloor$ with $\lfloor m_1 N \rfloor < \bar{k}$ (where k_{-i} is the number of other agents choosing action 1, from the point of view of agent i). Further suppose that the average effect across the two events is zero in state 0, i.e., $U(0;\eta_N) = 0$. Informativeness means the mean action is higher in one state—say $q(1;\sigma_N) < q(0;\sigma_N)$—and so the same is true for the relative likelihood of the two pivotal events: $\frac{\Pr(k_{-i}=\bar{k}|\omega=1;\sigma_N,N)}{\Pr(k_{-i}=\lfloor m_1 N\rfloor|\omega=1;\sigma_N,N)} < \frac{\Pr(k_{-i}=\bar{k}|\omega=0;\sigma_N,N)}{\Pr(k_{-i}=\lfloor m_1 N\rfloor|\omega=0;\sigma_N,N)}$.[19] The differing likelihood ratios imply different average effects in the two states; thus $U(1;\eta_N) \neq 0$.

Second, we observe that $U(0;\eta_N)$ and $U(1;\eta_N)$ must have the same sign. Otherwise the agents' best response would be uninformative, given (2) and the tie-breaking rule.

Now, since $U(0;\eta_N)$ and $U(1;\eta_N)$ are non-zero and have the same sign, their ratio pins down for each signal s a *unique* type $0 < p_N(s) < 1$ that is indifferent after observing s:

$$\frac{U(0;\eta_N)}{U(1;\eta_N)} = \frac{\Pr(\omega = 1 \mid p_i = p_N(s), s_i = s)}{\Pr(\omega = 0 \mid p_i = p_N(s), s_i = s)}. \tag{11}$$

For generic agents' prior distributions, information aggregation then follows from showing that the limit indifferent types are interior, i.e.,

$$0 < \lim_{N\to\infty} p_N(1) < \lim_{N\to\infty} p_N(0) < 1. \tag{12}$$

This means that, as $N \to \infty$, the mean action differs across signals and thus across the two states. Since the realized collective action is almost surely close to the mean action in each state, the principal learns the state from observing it.

To conclude, we sketch the proof of (12). The basic argument is that, under par-

[19] The ordering $q(1;\sigma_N) < q(0;\sigma_N)$ implies that the posterior likelihood ratio $\frac{\Pr(k_{-i}=k|\omega=1;\sigma_N,N)}{\Pr(k_{-i}=k|\omega=0;\sigma_N,N)}$ is strictly decreasing in k and thus this inequality, given $\lfloor m_1 N\rfloor < \bar{k}$.

tial commitment, a failure of information aggregation generically implies that, conditional on being pivotal, an agent becomes certain (as $N \to \infty$) that he is pivotal to the principal's preference for low versus high policies, i.e., $\lim_{N\to\infty} \Pr_i(\text{piv}_0|\text{piv};\eta_N,N) = 1$. However, the principal's updating from piv_0 is bounded: It changes her prior to a belief close to $\frac{1}{2}$, the indifference point. Hence, the agents' updating from piv_0 is bounded as well. The bounded updating implies bounds on the limit indifferent types, and thus (12).[20]

In the non-generic knife-edge scenarios, equilibrium sequences with non-interior limit indifferent types may exist. For these cases, we provide an alternative argument, based on the knife-edge condition and large deviation results that we derive in Appendix A.

1.6 Equilibrium Existence

We now show that for the processes of interest in this paper, equilibria satisfying the tie-breaking rule always exist.

Proposition 1. *Take any non-constant, monotone process and any agents' information structure. For any N, there is an equilibrium that satisfies the tie-breaking rule.*

We cannot readily prove Proposition 1 by applying a standard fixed-point theorem, since we need to ensure that the tie-breaking rule holds. Instead, we identify a set of candidate strategy profiles that adhere to this rule and construct equilibria as fixed points in this set. (As usual, the details of the proof are given in the appendix.)

The candidate strategy profiles are as follows. The agents use "monotone" strategies σ, i.e., strategies under which the mean action is (weakly) higher in state 1 than in state 0, and where the partisans choose their actions in accordance with their types:

$$q(0;\sigma) \leq q(1;\sigma), \text{ and } \sigma(y,s) = y \text{ for all } s \in \{0,1\} \text{ and all } y \in \{0,1\}. \quad (13)$$

The principal mixes over "monotone" strategies, defined as those in which the policy choice $x = x(k)$ is weakly increasing in the number of observed 1-actions k. That is, $x(k) \geq x(k')$ for all $k, k' \in \{0,\dots,N\}$ with $k > k'$.

[20] Precisely, $\lim_{N\to\infty} \Pr_i(\text{piv}_0|\text{piv};\eta_N,N) = 1$ implies $\frac{U(0;\eta_N)}{U(1;\eta_N)} \approx \frac{\Pr_i(\text{piv}_0|\omega=0;\eta_N,N)}{\Pr_i(\text{piv}_0|\omega=1;\eta_N,N)}$, and the bounded upating from piv_0, i.e., $\lim_{N\to\infty} \frac{\Pr_i(\text{piv}_0|\omega=0;\eta_N,N)}{\Pr_i(\text{piv}_0|\omega=1;\eta_N,N)} \in (0,\infty)$ then implies (12).

For the fixed-point argument, we consider a modification of the best-response correspondence in which the set values are truncated to the set of candidate strategy profiles, denoted by Σ. We verify that this modification meets the requirements of Kakutani's fixed point theorem. Note that it is non-trivial to show that it has non-empty values; this holds essentially because the monotonicity of the process and the players' strategies implies that monotone best responses exist. The principal has a monotone best response because (13) implies that her posterior is weakly increasing in the number of 1-actions. In the formal proof we show that the agents have a best response satisfying (13), through a detailed case analysis of their best-response characterization (11).

Next, we lift the fixed-point problem to a finite-dimensional space. We represent mixtures over the principal's monotone strategies by vectors $v = (v_1, \dots, v_{|\mathcal{P}\times(N+1)|}) \in [0,1]^{|\mathcal{P}\times(N+1)|}$ with $\sum_l v_l = 1$ (note that the dimension of these vectors equals the number of monotone strategies). We then observe that $\Big(q(0;\sigma), q(1;\sigma), v\Big)$ is a sufficient statistic for the best-response correspondence, given (1), (2), and (3). Hence we can understand the modified best-response correspondence as a self-map on the space of vectors $(\mathbf{q}, v)$ satisfying (13). This representation thus allows a direct application of Kakutani's fixed point theorem to show the existence of an equilibrium.

1.7 Payoff Guarantees: Proof of Theorem 1

In this section we put together the results of Sections 1.4–1.6 to prove Theorem 1. We show that all majority votes over exclusion of the maximal policy (i.e., the candidate processes defined by (4)) have a payoff guarantee of $1-\varepsilon$, and that $1-\varepsilon$ is an upper bound on the payoff guarantee of any other process of partial commitment.

For any process P^* satisfying (4), equilibria satisfying the tie-breaking rule generally exist (Proposition 1), ruling out a payoff guarantee of $G(P^*) = -\infty$. Since the principal learns the state under P^* (Theorem 3), the only way she will deviate from the full-information choice is by choosing the policy $1-\varepsilon < 1$ in state $\omega = 1$ if constrained to do so. Consequently, a lower bound for the payoff guarantee is

$$G(P^*) \geq \frac{\Pr(\omega = 1)\cdot \frac{1}{2}(1-\varepsilon)}{\Pr(\omega = 1)\cdot \frac{1}{2}} = 1-\varepsilon.$$

This is also an upper bound for $G(P^*)$, since we have constructed equilibria in which the ε-error is made (Theorem 2).

The same argument (Theorem 2) shows that $1-\varepsilon$ is an upper bound on $G(P)$ for any process P that excludes the policy $x = 1$ for some m. For any process P that never excludes $x = 1$, as argued in Section 1.3, there is an uninformative equilibrium in which $x = 1$ is chosen in both states. Hence the payoff guarantee of P has an even lower upper bound of

$$G(P) \leq \frac{\Pr(\omega = 1) \cdot \frac{1}{2} - \Pr(\omega = 0) \cdot \frac{1}{2}}{\frac{1}{2}\Pr(\omega = 1)} = 1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} < 1 - \varepsilon.$$

(Here the last inequality holds because we assumed the principal's prior is not extreme, $\Pr(\omega = 1) \leq 1 - \varepsilon$; see Section 1.1.)[21]

The processes defined by (4) also maximize the agents' ex-ante payoff guarantee, provided their mean prior satisfies the same upper bound as the principal's prior, i.e., $\mathrm{E}_F(p_i) \leq 1 - \varepsilon$. This is simply because the linearity of the expected utility in the prior means we can evaluate the agents' payoff guarantee by replacing the principal's prior with their mean prior in the definition of $G(P)$. Doing so, we find that the two inequalities above still hold (given that $\mathrm{E}_F(p_i) \leq 1 - \varepsilon$), which implies that the processes defined by (4) are also agent-optimal.

1.8 When Partial Commitment Outperforms Full Commitment

In some situations, processes of partial commitment outperform full commitment to precise single policies. Our next theorem presents one instance in which this is true, namely, when there is no constraint on the expected share of partisan types. We briefly discuss other possible instances at the end of this section.

Formally, in Theorem 4 we revisit our baseline model but drop the assumption that $\Pr_F(p_i = y) < \frac{1}{2}$ for each $y \in \{0, 1\}$. The theorem contradicts the usual intuition, which is that the principal always benefits from full commitment power and will not forgo it intentionally.

Theorem 4. *When the expected share of partisan types* $y \in \{0, 1\}$ *can be any number in* $(0, 1)$, *any robust optimal process is a process of partial commitment. Any majority vote over exclusion of the maximal policy (i.e., any process satisfying* (4)*) is robust optimal and has a payoff guarantee of* $1 - \varepsilon$.

The following sketches the proof. Under the earlier constraint on the share of partisans, a version of the modern Condorcet jury theorem held and full commitment was optimal: Committing to the precise policy $x = 0$ if $m \leq \frac{1}{2}$ and to $x = 1$ otherwise had a payoff guarantee of 1 (Bhattacharya, 2013). By contrast, when there is no constraint on the share of partisans, if the principal is committed to

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

singleton policies, she may be forced to choose the same policy in both states on account of the partisans, implying a failure of the Condorcet jury theorem and a relatively low payoff guarantee.

To be precise, take any process P with singleton policy sets, and suppose the expected share of partisans exceeds the process's largest cutoff, e.g., $\Pr_F(p_i = 1) > m_R$. In all equilibrium sequences, the type-1 partisans choose their dominant strategy, which is to match their action to their type. Hence the principal is required to choose a particular policy x_h (the sole element of $P(1)$) with probability converging to 1 as $N \to \infty$. The payoff guarantee of the process is thus bounded above by $x_h - \frac{\Pr(\omega=0)}{\Pr(\omega=1)} x_h$, which is strictly smaller than $1 - \varepsilon$, since $x_h \leq 1$ and $\Pr(\omega = 1) < 1 - \varepsilon$.[22] A similar argument shows that any process that commits to a singleton policy set after *some* observation has a payoff guarantee lower than $1 - \varepsilon$.[23]

On the other hand, the payoff guarantee of the processes defined by (4) remains $1-\varepsilon$, as in Theorem 1, regardless of the share of partisans, simply because the proof of Theorem 1 did not use the constraint on the share of partisans. The constraint was imposed merely to compare our results to the Condorcet jury theorem in the most transparent manner.

Let us conclude with a critical insight about information aggregation. It is well known that the equilibrium sequences just described, in which the partisans enforce the singleton commitment x_h, do *not* aggregate information. Somewhat surprisingly, partial commitment restores information aggregation. The reason was laid out in Section 1.5, in the sketch of the proof of Theorem 3: Given the partial commitment in the processes defined by (4), a failure of information aggregation generically implies that, conditional on being pivotal, an agent becomes certain (as $N \to \infty$) that he is pivotal to the principal's preference for low versus high policies. However, this in turn implies information aggregation, as explained.

In future research, it may be worthwhile to investigate other instances in which partial commitment outperforms full commitment. The literature on information aggregation has provided settings in which all equilibrium sequences for binary majority elections aggregate information, meaning that an outside observer could learn the state from observing the vote margin; however, the outcomes are not

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

[23] Suppose that $P(m_j) = \{x_h\}$ for some $x_h \in \mathcal{P}$ and $j = 1, \ldots, R+1$, and let $\gamma = m_j - m_{j-1}$ if $j > 1$ and $\gamma = m_j$ if $j = 1$. Suppose the expected share of non-partisans is smaller than $\frac{\gamma}{4}$ and the expected share of 1-partisans is in $[m_{j-1} + \frac{\gamma}{4}, m_{j-1} + \frac{\gamma}{2}]$ if $j > 1$ and in $[0, \frac{\gamma}{2}]$ if $j = 1$. Then, the principal is required to choose the policy x_h with probability converging to 1 as $N \to \infty$. The same calculation as before implies a payoff guarantee smaller than $1 - \varepsilon$.

full-information equivalent (see, e.g., Ekmekci and Lauermann, 2022). In such settings, partial commitment might improve efficiency in a robust sense by allowing for outcomes that are nearly full-information equivalent.

2 Extensions with More General Preferences

2.1 Heterogeneous Ex-Post Preferences

In the baseline model of Section 1, all of the players agree unanimously on the best policy when the state is known. However, such unanimity is often unrealistic. For instance, a reform with distributive consequences may benefit different groups of voters depending on the state of the world; see, e.g., Fernandez and Rodrik (1991) or Ali *et al.* (2025). In this section we extend our model to capture such scenarios, by allowing for more general state-dependent preferences: In addition to partisans who always prefer $x = 1$ or $x = 0$, and agents who prefer $x = 1$ in $\omega = 1$ and $x = 0$ in $\omega = 0$, we now consider agents who prefer $x = 0$ in $\omega = 1$ and $x = 1$ in $\omega = 0$. An agent's type is now given by a prior belief $p_i \in [0, 1]$ and a pair $\boldsymbol{t_i} = \Big(t_i(0), t_i(1)\Big) \in [0, 1]^2$ which describes the agent's constant marginal benefit from the policy choice in each state. Types with $t_i(0) > \frac{1}{2} > t_i(1)$ prefer $x = 0$ in $\omega = 1$ and and $x = 1$ in $\omega = 0$.[24]

We can study the distribution of these more general types in terms of a fundamental object Φ: Given any possible average effects $U(0; \eta)$ and $U(1; \eta)$ and any signal likelihood ratio $l := \frac{\Pr(s_i=s|\omega=0)}{\Pr(s_i=s|\omega=1)}$, Φ maps the triple $\Big(U(0; \eta), U(1; \eta), l\Big)$ to the likelihood that a randomly drawn type with signal s prefers the 1-action. We call this likelihood the mean preferred action). The map Φ is the type distribution's fundamental, since the set of equilibrium outcomes depends on the distribution solely through Φ. We prove this and all results in this section in the online appendix.

The arguments for information aggregation and robust optimality given in Section 1 can be extended to the present context under a monotonicity condition for Φ, an adaptation of "strong preference monotonicity" (Bhattacharya, 2013). This condition is defined as follows. Observe that whenever $U(1; \eta) \neq 0$, Φ depends only on $z_1 = \frac{U(0;\eta)}{U(1;\eta)} \cdot l$ and the sign z_2 of $U(1; \eta)$. To be precise, when $U(1, \eta) > 0$, a type

[24]We assume the pairs t_i are drawn from an absolutely continuous distribution and independently from priors, signals, and the state, as well as independently across agents.

with signal s prefers the 1-action if and only if

$$p_i\Big(t_i(1) - \frac{1}{2}\Big) \cdot \frac{\Pr(s_i = s|\omega = 1)}{\Pr(s_i = s)} \cdot U(1;\eta)$$
$$- (1 - p_i)\Big(\frac{1}{2} - t_i(0)\Big) \cdot \frac{\Pr(s_i = s|\omega = 0)}{\Pr(s_i = s)} \cdot U(0;\eta) \geq 0,$$

which occurs if and only if

$$z_1 \cdot (1 - p_i)\Big(\frac{1}{2} - t_i(0)\Big) \leq p_i\Big(t_i(1) - \frac{1}{2}\Big). \tag{14}$$

When $U(1;\eta) < 0$, the same statement holds with the inequality reversed. We say Φ is *monotone* if it has a continuous derivative $\frac{\partial \Phi}{\partial z_1}$ that has the same non-zero sign for all $z_1 \in (0, \infty)$, given any fixed $z_2 \in \{-1, 1\}$.

If we assume monotone type distributions, the conclusion of Theorem 1 continues to hold in this setting: The processes (4) remain robust optimal, and they maximize the principal's payoff guarantee across all agents' information structures and monotone type distributions. The key implication of the monotonicity condition is that the agents have informative best responses when $\frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0, \infty)$, i.e., the best response satisfies $q(0;\sigma_N) \neq q(1;\sigma_N)$; this is an immediate consequence of (14). The formal proof of the optimality result relies on this observation at critical points. Otherwise, it closely follows the proof for the baseline model; see the online appendix.

Note that monotonicity is satisfied in the baseline model, as all non-partisans have the same preference type. One can show that for some parameters of the model (i.e., for some agents' information structures and majority cutoffs), non-monotonicity implies a failure of information aggregation for the processes (4), in the same way that it implies a failure of the Condorcet jury theorem (Bhattacharya, 2013).

Finally we remark that, as in the baseline model, the processes (4) maximize the agents' ex-ante payoff guarantee as well as the principal's, given two conditions in addition to monotonicity. The first is that the mean marginal benefit exceeds the marginal cost in state 1 but not in the state 0: $0 \leq \mathrm{E}\Big(t_i(0)\Big) < c < \mathrm{E}\Big(t_i(1)\Big) \leq 1$. This implies that the "mean" agent and the principal have the same preference ranking of policies when the state is known. The first condition is satisfied, for example, when the principal is a social planner who maximizes the agents' utilitarian welfare. These incentives may arise from political economy forces under a broad set of conditions, cf. the literature on political agency and electoral accountability.[25]

[25]For the political agency literature, see Barro (1973); Ferejohn (1986) among others; for the

The second condition is that

$$\frac{1-\mathrm{E}_F(p_i)}{\mathrm{E}_F(p_i)}\cdot\frac{\frac{1}{2}-\mathrm{E}\Big(t_i(0)\Big)}{\mathrm{E}\Big(t_i(1)\Big)-\frac{1}{2}}\geq\varepsilon.$$

The relevant implication of this condition is that if the policy outcome is $x=1$ in both states, the agents' mean payoff is not too close to their mean full-information payoff.

2.2 Single-Basin Preferences and Monotone Equilibria

We now relax the assumption of constant marginal cost and benefit in our baseline model; we suppose instead that the players' ex-ante preferences are single-basin over the policy space. This condition is intended to capture settings in which intermediate policy choices are inefficient. Examples include public infrastructure decisions that involve economies of scale such that, depending on public demand, the optimal choice is either the status quo or a "full solution." Similarly, compromises between opposing policies are often inherently inefficient. For instance, consider a company debating whether to replace its current management. A complete overhaul would enable a fresh start, whereas a partial replacement might fuel internal conflicts or prolong existing divisions.

Formally, we maintain the assumptions of the baseline model, but we allow the players to have any common state-dependent payoffs $u(x,\omega)$ with $u(x,0)<u(x',0)$ and $u(x,1)>u(x',1)$ for all $x>x'$. Letting $c(x)=-u(x,0)$ and $b(x)=u(x,1)-u(x,0)$, we can express the payoffs as

$$u(x,\omega)=-c(x)+b(x)\omega$$

for $\omega\in\{0,1\}$. We assume that $\frac{c'(x)}{b'(x)}$ is constant or strictly decreasing in x. (Note that the baseline model can be recovered as the linear case of this model, with $c(x)=\frac{x}{2}$ and $b(x)=x$.) The expected utility given a fixed prior p, which we denote by $u(x,p)=-c(x)+b(x)p$, has negative derivative if and only if $\frac{c'(x)}{b'(x)}\geq p$. Thus, any player's ex-ante expected utility is single-basin with the set of minima, the basin, decreasing in p.

Theorem 5 establishes a version of our prior optimality result (Theorem 1) for this generalized setting. For tractability, we focus on monotone processes P and equilibria in monotone strategies, as defined in Section 1.6; see (13) and thereafter.

literature on electoral accountability, see, e.g., the survey in Ashworth (2012). Relatedly, Battaglini (2017) provides an excellent discussion, with several explicit examples.

Monotone processes capture the idea that elected officials receive "mandates" from the populace to make decisions, with more extreme decisions being feasible when the official has more support.[26] Non-monotone processes and strategies are less intuitive and may be hard to implement.[27]

For the statement of the theorem, note that the assumptions on the preferences imply a $\bar{p} \in (0,1)$, so that all players prefer the maximal policy (i.e., $1 = \arg\max_{x \in \mathcal{P}} u(x,p)$) if holding a belief $p > \bar{p}$ and the minimal one (i.e., $0 = \arg\max_{x \in \mathcal{P}} u(x,p)$) if holding a belief $p < \bar{p}$.

Theorem 5. *Let $\frac{c'(x)}{b'(x)}$ be constant in x or strictly decreasing in x and $\Pr(\omega) > \bar{p}$. Then any majority vote over exclusion of the maximal policy, i.e., any process of the form (4), has the following properties:*

1. *It has a payoff guarantee of $1-\varepsilon$ across all agents' information structures and all monotone equilibria.*

2. *It is robust optimal among all monotone processes of partial commitment.*

The proof is in the online appendix. Here we describe the relevance of the assumptions in the theorem about the principal's prior and the players' preferences.

The assumption $\Pr(\omega) > \bar{p}$ and the restriction to monotone equilibria together imply that any monotone equilibrium of a process of the form (4) is informative and has strictly positive average effects:[28]

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

This is because the principal's equilibrium strategy is not only monotone, i.e., weakly increasing, but also non-constant: At $m = 1$, her belief exceeds the prior (which exceeds $\bar{p}$), so she chooses $x(m) = 1$. At $m = 0$, her choice is constrained to be strictly smaller than $x = 1$. Since the tie-breaking assumption implies interior mean actions (cf. Section 1.1), all observations m are on path, including

[26]The idea that elected officials have mandates to govern, and that these mandates are stronger for officials with greater support, has been explored previously; see, e.g., Herrera, Llorente-Saguer and McMurray (2019) and Damiano *et al.* (2025).

[27]There is a growing literature on voting and communication games that focuses on monotone equilibria; see, e.g., Krishna and Morgan (2001), Chen, Kartik and Sobel (2008), and Dekel and Piccione (2000).

[28]We abuse notation slightly here: Strictly speaking, our original definition is consistent with $U(1;\eta) = 2 \cdot \Big(\mathrm{E}\Big(u(x,1)|a_i = 1; \mathrm{piv}, \eta, N\Big) - \mathrm{E}\Big(u(x,1)|a_i = 0; \mathrm{piv}, \eta, N\Big)\Big)$ and $U(0;\eta) = -2\Big(\mathrm{E}\Big(u(x,0)|a_i = 1; \mathrm{piv}, \eta, N\Big) - \mathrm{E}\Big(u(x,0)|a_i = 0; \mathrm{piv}, \eta, N\Big)\Big)$.

the jump points where $x(m) < x(m + \frac{1}{N})$, and we conclude that the average effects are strictly positive. Since a type with signal s prefers the 1-action if $\Pr_i(\omega = 1|p_i = p, s_i = s)U(1;\eta) - \Pr_i(\omega = 0|p_i = p, s_i = s)U(0;\eta) > 0$, this implies some types best-respond informatively.

The assumption that $\frac{c'(x)}{b'(x)}$ is decreasing ensures a monotone comparative statics result. While $\arg\max_{x\in\mathcal{P}} u(x,p)$ is increasing even without the assumption, the same is not true for $\arg\max_{x\in P(m)} u(x,p)$—that is, when taking into account the constraints imposed by the process P on the principal's choice. We show that the assumption ensures that the correspondence $\arg\max_{x\in P(m)} u(x,p)$ has a monotone selection given any monotone process and strategy profile. In other words, the principal has a monotone best response.

Given these two observations, one can prove Theorem 5 by mimicking the proof of Theorem 1 (in fact, at some points the latter can be shortened).

3 Further Results for the Baseline Model

Here we discuss several more results concerning the baseline model of Section 1.

Arbitrary Finite Policy Sets. Majority votes over exclusion of the maximal policy (the processes defined by (4)) are robust optimal for any policy set $\mathcal{P} = \{0, x_2, \ldots, x_{l-1}, 1\}$ with $0 < x_2 < \ldots < 1$ and $l > 2$,[29] provided that

$$1 - \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \leq x_{l-1},$$

i.e., that the principal's prior is not too large relative to the second-largest policy. This is because the proof of robust optimality in Theorem 1 used the specific properties of the policy set only in the evaluation of payoff guarantees (in the two inequalities in Section 1.7). In those calculations, all that mattered was that $1 - \frac{\Pr(\omega=0)}{\Pr(\omega=1)} \leq 1 - \varepsilon$. This is exactly the condition above, since in our original model we assumed $x_{l-1} = 1 - \varepsilon$.

Efficient Equilibria. Many processes with a single cutoff have efficient equilibrium sequences for *any* agents' information structure. In the appendix we prove a sufficient condition for this property, which is that the minimum and maximum policies are increasing, i.e., $\min P(0) < \min P(1)$ and $\max P(0) < \max P(1)$, and that

[29] We assume $l > 2$ because if $\mathcal{P}$ contains only one or two policies, these are not well-defined processes of partial commitment.

the ex-post optimal policies are not excluded, i.e., $\min P(0) = 0$ and $\max P(1) = 1$ (Theorem 10).

Miscoordination: Other Inefficient Equilibria. Differences between the principal's and the agents' priors can give rise to inefficiencies. Theorem 6, below, shows that if the principal is more pessimistic about the possible benefits from the policy than a sufficient mass of agents, there are equilibria in which the agents miscoordinate on "counteracting" the principal's bias: They constrain the principal to the highest policy range available, given the announced process, with probability close to 1.

Formally, define $\bar{p}$ by

$$\frac{\Pr_i(\omega = 1 \mid p_i = \bar{p}, s_i = 0)}{\Pr_i(\omega = 0 \mid p_i = \bar{p}, s_i = 0)} = \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)}. \tag{15}$$

That is, the principal's posterior conditional on a 1-signal equals the type $\bar{p}$'s posterior conditional on a 0-signal. Clearly, $\Pr(\omega = 1) < \bar{p}$.

Theorem 6. *Consider any non-constant, monotone process with cutoffs* $0 < m_1, \ldots < m_R < m_{R+1} = 1$. *There is some* $\underline{q} \in (m_R, 1)$ *such that, when* $1 - F(\bar{p}) > \underline{q}$, *there is a sequence of equilibrium strategies* $(\sigma_N)_{N \in \mathbb{N}}$ *for which* $\lim_{N \to \infty} \Pr(m_R < m \mid \sigma_N, N) = 1$.

The proof of Theorem 6 is in the appendix. For the converse scenario, where the priors are close, we showed in Section 1.4 that there are often inefficient equilibria in which agents act approximately truthfully. Taken together, these results show that partial commitment gives rise to inefficient equilibria, independent of the scenario.

When Is Vagueness Necessary for Robust Optimality? Theorem 1 establishes that *some* robust optimal processes are *vague*—they exclude at most one policy from the policy space for every collective action. This result can be strengthened under two conditions: (1) if we restrict to processes map each m to connected sets, and (2) if the principal's prior is sufficiently close to $\frac{1}{2}$. Then, *all* robust optimal processes must exhibit vagueness, making it a necessary condition for robust optimality.

To understand this necessity, consider the case where the prior is arbitrarily close to $\frac{1}{2}$ and suppose a process maps some m to a connected policy set $P(m) = \{x_j, \ldots, x_k\}$ where either $j > 1$ or $k < l - 1$ (i.e., the set excludes at least two policies). Given $x_2 = 1 - x_{l-1} = \varepsilon$, Theorem 2 implies that such a process achieves a strictly suboptimal payoff guarantee of less than $1-\varepsilon$. By contrast, the vague processes of Theorem 1 achieve a payoff guarantee of $1-\varepsilon$.

4 Related Literature

We now discuss our results in the context of the wider literature.

Information aggregation. Our paper offers several new insights in relation to the existing work on information aggregation in collective choice—in particular, the influential literature around the Condorcet jury theorem.

First, the paper provides a comprehensive comparison of mechanisms, including those considered in previous work (majority elections with full commitment and cheap talk). Theorems 1 and 5 show that, because of the interplay between a coordination problem and an information aggregation problem, a mechanism involving vague (that is, almost minimal) commitment is robust optimal. In particular, full-commitment mechanisms are not optimal. By contrast, Battaglini (2017) stresses the value of commitment: Based on the Condorcet jury theorem, he argues that full commitment is valuable since it implies better information aggregation properties than the no-commitment cheap-talk process (Battaglini, 2017, Proposition 6). The stark difference between our results and those in Battaglini arises because our comparison includes processes of partial commitment and because our setting differs substantially from his. In particular, we drop the constraint on the proportion of partisans that underpins the Condorcet jury theorem, and we consider a continuous agent type instead of a discrete one.[30]

Second, the literature on majority elections has identified various deviations from the benchmark model of the Condorcet jury theorem that imply a failure of coordination among the agents; see, e.g., Ekmekci and Lauermann (2020), Mandler (2012), Bouton and Castanheira (2012), and Feddersen and Pesendorfer (1997). We add the observation that deviations from full commitment power, even minimal ones, imply a coordination failure (Theorem 2). Most strikingly, the approximation of a majority vote between $x=0$ and $x=1$ has a payoff guarantee lower than random choice between $x=0$ and $x=1$; cf. the discussion in Section 1.3.

Third, our results highlight that the selection of the collective choice mechanism is critical in settings with multiple pivotal events: In our setting, even when all

[30] The result in Battaglini relies on the observation that cheap talk does not aggregate information if there is a large gap between the support of the agents' prior and the principal's prior, relative to the informativeness of the agents' signals. Our assumption of full support rules out the possibility of a gap; see Section 1.8.

players have approximately the same preferences and priors, the information aggregation properties of the mechanism depend on the exact process chosen. Under the vague commitment processes (4), the multiple pivotal events ensure and restore information aggregation, relative to the full-commitment majority rules with a single pivotal event (see Theorem 5). Conversely, given processes with balance, the multiple pivotal events imply a failure of information aggregation, even in all of the scenarios in which the full-commitment majority rules aggregate information (see Theorem 3). Previous works have examined various models of majority elections exhibiting multiple pivotal events. Different types of pivotal events have different implications for the alignment of the players' preferences, and thus different implications for information aggregation: Ahn and Oliveros (2012) and Razin (2003) document aggregation failures due to preference conflicts, while Damiano *et al.* (2025) document instances in which an additional pivotal event aligns the voters' preferences and promotes information aggregation. Our Theorem 6 has the same flavor: It identifies inefficient equilibria that are driven by preference conflicts between the agents and the principal.

Numerous more distantly related models have been used to study information aggregation in politics under other conditions; see, e.g., Lohmann (1994), Bond and Eraslan (2010), Barelli, Bhattacharya and Siga (2022), Ekmekci and Lauermann (2022), and Chen (2025).

Delegation. Following Holmström (1978) and Alonso and Matouschek (2008), there has been a large literature studying the choices made by a principal who can delegate a decision to a single privately informed agent. However, none of this work has considered settings with a large number of agents, as our paper does. To our knowledge, the only exception is Alonso, Dessein and Matouschek (2008).[31]

The political science literature has used the single-agent framework in several important applications that actually feature many agents. The assumption of a single agent is typically justified by appealing to the idea of a "representative agent"; however, it precludes any analysis of the agents' coordination issues. Our results provide a new perspective on several of these applications, by highlighting the relevance of coordination issues. For example, Kartik, Van Weelden and Wolton (2017) apply the single-agent framework to study the so-called trustee-versus-delegate trade-off, where the goal is to understand how much autonomy should be granted to elected representatives. They consider an electoral competition model in which elected

[31]Their paper considers a quite different notion of coordination, where the utilities of the two agents directly depend on the proximity of their individual decisions d_1 and d_2.

candidates have superior information but also have preference conflicts relative to voters. For this model, they show that partial autonomy arises in equilibrium.[32] By contrast, in our baseline model, the agents hold superior information and there is no ex-post preference conflict. Here we find that it is optimal to grant almost full autonomy to the principal, owing to a coordination problem of the agents.

The single-agent framework has also been used to study the "congressional control problem"—that is, the question of how elected officials can control privately informed bureaucrats, such as national security agencies, while also extracting their information. See Gailmard and Patty (2012) for the most recent review and Antic and Iaryzcower (2020) for a newer contribution.[33]

Sincere Voting. 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, sincere voting is generically not an equilibrium when the electorate is large (see, e.g., Austen-Smith and Banks, 1996). In contrast, Theorem 2 and its proof show that, under partial commitment, sincere voting by approximately all agents is an equilibrium for a nontrivial range of information structures. Similar 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).

5 Conclusion

In this paper, we proposed and analyzed a model of partial commitment in a collective choice problem. Our model complements existing ones that have been used to study information aggregation in politics (Austen-Smith and Banks, 1996; Battaglini, 2017; Feddersen and Pesendorfer, 1997; Krishna and Morgan, 2012), as well as problems of delegation (Alonso and Matouschek, 2008; Holmström, 1978). In relation to information aggregation, our work addresses frictions in commitment power and explores a novel set of choice mechanisms. The framework we establish applies to many mechanisms featuring partial commitment that are widely used in the real world, such as public referenda and customer feedback polls (as mentioned

[32] Fox and Shotts (2009) also study the trustee-versus-delegate trade-off but do not invoke the framework of Holmström (1978).

[33] Clearly, the questions described here go beyond delegation: They concern the structure of democratic representation in general. As such, they relate to a massive literature on representation in political science to which I cannot do justice here; see Urbinati and Warren (2008) for a review.

in the introduction). In relation to delegation, we depart from the existing literature in considering a principal facing not just one agent but a large group of agents. We discuss natural applications to politics and the management of large organizations.

Our main results (Theorems 1 and 5) highlight "vagueness" (minimal commitment) as a valuable tool for managing large groups of privately informed agents. When there is uncertainty about the information environment, a mechanism under which the principal is minimally bound by the input of the agents allows her to robustly aggregate the agents' private information while retaining flexibility in her choice.

The starkness of our main result may be thought-provoking: What additional factors might limit the usefulness of vague commitments? As an example, in Section 2.1 we provided some initial observations for scenarios in which the players' ex-post preferences are heterogeneous. In future work, it may be interesting to explore the comparative statics of preference conflicts further, or to identify other critical factors.

Our results suggest several broader questions about governance that may also inspire future work. First, to what extent are rules (i.e., commitments to mechanisms) useful in highly uncertain environments? Intuitively, there seems to be a tension between adherence to fixed rules and robustness to uncertainty. Theorem 5 touches upon this issue by showing that the flexibility of partial commitment may offer advantages over precise commitment, in the face of uncertainty about the number of partisans among the agents.[34]

Second, our results point to a trade-off related to coordination and information transmission. A principal facing a group of agents can incentivize them to transmit information by promising to delegate decision-making power to them (i.e., by making commitments); however, this will be costly if the agents cannot coordinate on a good decision. Our results show that in large groups, the trade-off becomes extreme, so that minimal delegation to the group is optimal. Intuitively, the larger the group of agents, (a) the more information they hold and the easier it is to learn from them, and (b) the harder it is to coordinate among them. Several questions for future research naturally ensue; for example, how should a principal use commitments to manage a group of intermediate size? While in this paper we assumed one-way communication from the agents to the principal, how should the communication between the players be organized in general?

Finally, our results may inspire experimental work on partial commitment in

[34]To study this question, it may be helpful to adopt a version of the "robust mechanism design" approach (Bergemann and Morris, 2005; Carroll, 2019).

collective choice. Although much experimental work has explored standard mechanisms such as simple majority or unanimity voting rules (cf. the surveys in Palfrey, 2009 and Palfrey, 2016), there has been very little work on alternative mechanisms. Given the recent interest of policymakers around the globe in exploring new democratic institutions (see, e.g., OECD, 2020), further empirical exploration seems desirable.

Appendix

A Mathematical Preliminaries

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[35]

$$\Pr\left(X_n = mn\right) = \exp\Big(- n\mathrm{KL}(m,q) + o(n)\Big), \tag{16}$$

where KL denotes the Kullback–Leibler divergence,

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

The idea of the proof of this fact, due to Cramér (1938), is to perform a change of measure, the so-called Escher transform (Escher, 1932). Consider the binomial distribution under which the event is not rare but rather typical, $Z_n \sim \mathcal{B}(n,m)$. Then (16) follows from observing that

$$\frac{\Pr\left(X_n = mn\right)}{\Pr(Z_n = mn)} = \exp\Big(- n\mathrm{KL}(m,q)\Big) \text{ and } \Pr(Z_n = mn) = \exp\Big(o(n)\Big). \tag{17}$$

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

For the equation on the left, note that

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

The equation on the right holds because the probability density function (PDF) 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 $m \approx q$.[36] 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:

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

[36]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

$$\mathrm{KL}(m,q)\approx\frac{(m-q)^2}{2q(1-q)}\ \text{ for }\ m\approx q. \tag{18}$$

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:

$$\begin{aligned}\mathrm{KL}(m,q_1)-\mathrm{KL}(m,q_2)&\approx\frac{\partial}{\partial q}\mathrm{KL}(m,q)_{|q=\bar{q}}(q_1-q_2)\\&=\left(\frac{1-m}{1-\bar{q}}-\frac{m}{\bar{q}}\right)(q_1-q_2).\end{aligned}\tag{19}$$

A.3 Monotonicity Properties of Binomial Distributions

In this section we make two useful observations about the PDFs of binomial distributions. The first is that the binomial distribution $X_n\sim B(n,q)$ has an inverse-U-shaped PDF, with a unique mode at $\lfloor (n+1)q\rfloor$ if $(n+1)q$ is not an integer, and otherwise with two modes given by $(n+1)q$ and $(n+1)q-1$; see, e.g., page 112 in Chapter 3.4 of Johnson, Kemp and Kotz (2005).

Claim 1. *Take a binomial distribution $X_n \sim \mathcal{B}(n,q)$. If $(n+1)q \in \mathbb{N}$, then*

$$\begin{aligned}
&\Pr(X_n = k) < \Pr(X_n = k') \ \textit{for all } k < k' \leq (n+1)q - 1,\\
&\Pr(X_n = k) = \Pr(X_n = k') \ \textit{for } k = (n+1)q - 1 \ \textit{and } k' = (n+1)q,\\
&\Pr(X_n = k) > \Pr(X_n = k') \ \textit{for all } k' > k \geq (n+1)q.
\end{aligned}$$

If $(n+q)q \notin \mathbb{N}$, then

$$\begin{aligned}
&\Pr(X_n = k) < \Pr(X_n = k') \ \textit{for all } k < k' \leq \lfloor (n+1)q \rfloor,\\
&\Pr(X_n = k) > \Pr(X_n = k') \ \textit{for all } k' > k \geq \lfloor (n+1)q \rfloor.
\end{aligned}$$

The second observation establishes a monotonicity property enabling us to compare the PDFs of binomial distributions with different success probabilities.

Claim 2. *For $X_n \sim \mathcal{B}(n,q)$ and $Y_n \sim \mathcal{B}(n,p)$ with $p < q$, the following holds: For any $k, k' \in \{0, \ldots, n\}$, if $k < k'$, then $\frac{\Pr(Y_n = k')}{\Pr(Y_n = k)} < \frac{\Pr(X_n = k')}{\Pr(X_n = k)}$.*

Proof. For a binomial distribution $Z_n \sim \mathcal{B}(n,\theta)$,

$$\frac{\Pr(Z_n = k')}{\Pr(Z_n = k)} = \frac{\binom{n}{k'}}{\binom{n}{k}} \cdot \left(\frac{\theta}{1-\theta}\right)^{k'-k}.$$

Comparing the likelihood ratios, we have

$$\frac{\Pr(Y_n = k')/\Pr(Y_n = k)}{\Pr(X_n = k')/\Pr(X_n = k)} = \left(\frac{p(1-q)}{q(1-p)}\right)^{k'-k}. \tag{20}$$

The ordering $p < q$ implies $\frac{p(1-q)}{q(1-p)} < 1$. Since $k' > k$, it follows that the ratio (20) is strictly smaller than 1, establishing the claim. □

The relevance of the mathematical preliminaries presented in this appendix for our collective choice model derives from the fact that, for any symmetric strategy σ of the agents, the number of 1-actions taken within a group of $N-1$ agents follows a binomial distribution with success probability $q(\omega';\sigma) = \mathrm{E}\Big(\sigma(s)|\omega = \omega'\Big)$. The pivotal events in our model thus correspond to point events of a binomial distribution, and (16) provides a suitable approximation of their likelihood. To be precise, if we let $q = q(\omega', \sigma)$ and $m = \frac{\lfloor m_j N \rfloor}{N}$ for $j > 0$, then (16) becomes

$$\Pr(\mathrm{piv}_j|\omega = \omega';\sigma, N) = \exp\Big(-(N-1)\mathrm{KL}\Big(\frac{\lfloor m_j N\rfloor}{N}, q(\omega';\sigma)\Big) + o(N)\Big). \tag{21}$$

Similarly, if we let $q = q(\omega', \sigma)$ and $m = \frac{k}{N}$ for $k \in \{\bar{k}, \bar{k}+1\}$, then (16) becomes

$$\Pr(\text{piv}_{0,k}|\omega = \omega'; \sigma, N) = \exp\Big(-(N-1)\text{KL}\Big(\frac{k}{N}, q(\omega'; \sigma)\Big) + o(N)\Big). \tag{22}$$

B Proof of Theorem 2

Take any process of partial commitment with cutoffs $m_1, \ldots, m_{R+1}$. Fix $j \in \{1, \ldots, R+1\}$. The following proof identifies an information structure and a corresponding equilibrium sequence $(\sigma_N)_{N\in\mathbb{N}}$ for which

$$\lim_{N\to\infty} \Pr\Big(m \in P^{-1}\Big(P(m_j)\Big)|\sigma_N, N\Big) = 1. \tag{23}$$

B.1 The Information Structures and Candidate Equilibrium Strategies

For $j = 2, \ldots, R+1$, we take an agents' information structure satisfying (5), (8), (6) for the bound $\underline{p}(1)$ given by

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

and

$$m_j < \Pr(s_i = 1|\omega = 0) < \Pr(s_i = 1|\omega = 1) < m_{j+1}. \tag{24}$$

When $j = 1$, instead of the condition (24), we take an information structure satisfying

$$0 < \Pr(s_i = 1|\omega = 0) < \Pr(s_i = 1|\omega = 1) < m_j.$$

For any $x \in (0,1)$, denote by p_x the (generalized) x-quantile of the prior distribution, $p_x = \inf\{p \in [0,1] : F(p) \geq \delta\}$. Note that $0 < p_{\frac{\delta}{2}}$ and $p_{1-\frac{\delta}{2}} < 1$, given (6).

We consider the following candidate strategies. The principal follows a mixed strategy randomizing between the pure strategy where she chooses

$$\begin{aligned} &\min P(m_j + 1) \;\text{ if }\; k \leq k^*, \\ &\max P(m_j + 1) \;\text{ if }\; k > k^*, \end{aligned}$$

and the pure strategy where she chooses

$$\begin{aligned}
&\min P(m_j+1) \text{ if } k \le k^*+1,\\
&\max P(m_j+1) \text{ if } k > k^*+1,
\end{aligned}$$

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 agents follow approximately truthful strategies $\sigma_{\mathbf{p}}$ under which, after observing signal $s_i \in \{0,1\}$, agent i agent chooses $a_i = 1$ if and only if $p_i \ge p(s_i)$, for some $p(s_i) \in (0,1)$. We identify $\sigma_{\mathbf{p}}$ with $\mathbf{p} := \Big(p(0), p(1)\Big)$ and consider the set

$$D(\delta) = \{\mathbf{p} : p_{\frac{\delta}{2}} \le p(1) \le p_\delta,\ \Pr(\omega=1) \le p(0) < 1\}.$$

For any $\mathbf{p} \in D(\delta)$, the strategy $\sigma_{\mathbf{p}}$ is δ-approximately truthful since $\Pr(\omega = 1) > p_{1-\frac{\delta}{2}}$, by (6). 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.[37] (Often we drop the arguments $(\mathbf{p}, z)$ from the notation.)

B.2 The Equilibrium Construction

The first step in the construction of the equilibrium is to fix a cutoff $p(0) \in [\Pr(\omega = 1), 1)$ and construct $p(1) = p^*\Big(1; p(0)\Big)$ so that the principal is indifferent (Sections B.2.1 and B.2.2). The second step is to show that an appropriate mixing z yields a cutoff $p^*(0)$ such that the agents' strategy given by $p^*(0)$ and $p^*\Big(1; p^*(0)\Big)$ is a best response to itself, i.e.,

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

for $\mathbf{p}^* = \Big(p^*(0), p^*\Big(1; p^*(0)\Big)\Big)$ (Sections B.2.3 and B.2.4). In both steps, we will identify an upper bound on δ and a lower bound on N that are required for the arguments to hold.

B.2.1 The Principal's Indifference Cutoff

The principal will be indifferent if she observes that $k = k^*(p_0) + 1$ out of N agents have chosen the 1-action. For any fixed $p(0) \in [\Pr(\omega = 1), 1)$, we define $k^*(p_0) + 1$ as the minimal observed number of 1-actions such that the principal weakly prefers

[37] The best-response cutoffs are well-defined since the properties of $\mathbf{p} \in D(\delta)$ and z imply $U(\omega') > 0$ for $\omega' \in \{0,1\}$.

$x = 1$ given $p(1) = p_\delta$, i.e.,

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1|\omega = 1; p(1) = p_\delta, p(0), N\Big)}{\Pr\Big(k = k^*(p_0) + 1|\omega = 0; p(1) = p_\delta, p(0), N\Big)} \geq 1. \tag{25}$$

Note that $k^*(p_0)$ exists in $\{0, \ldots, N-1\}$ whenever N is sufficiently large and δ sufficiently small. To see this, suppose $\delta \leq \frac{1}{4}$. This implies $p_\delta < p_{1-2\delta}$; thus, there is a uniform lower bound $\gamma > 0$ such that

$$q(1; \mathbf{p}) - q(0, \mathbf{p}) \geq \gamma \text{ for all } \mathbf{p} \in D(\delta) \text{ with } \delta \leq \frac{1}{4}, \tag{26}$$

given the full support and differentiability of the distribution of the agents' priors and the different likelihood ratios of the two signals in the two states. The uniform bound (26) implies that there is some N_1 such that for all $N \geq N_1$ and $\mathbf{p} \in D(\delta)$,

$$\frac{\Pr\Big(\omega = 1|k = 0; \mathbf{p}\Big)}{\Pr\Big(\omega = 0|k = 0; \mathbf{p}\Big)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \left(\frac{1 - q(1; \mathbf{p})}{1 - q(0; \mathbf{p})}\right)^N < 1.$$

The bound N_1 guarantees the uniform existence of $k^*(p_0) \in \{0, \ldots, N-1\}$ for all $p_0 \in [\Pr(\omega = 1), 1)$. Moreover, it guarantees that the best-response cutoff $\bar{k}$ as in (1) lies in $\{0, \ldots, N-1\}$ for all $\mathbf{p} \in D(\delta)$, not just those with $p(1) = p_\delta$.

B.2.2 Ensuring the Principal's Indifference

Claim 3 says that, for any candidate cutoff $p(0)$, we can find a cutoff $p^*\Big(1; p(0)\Big)$ such that, given the agents' strategy $\mathbf{p} = \Big(p(0), p^*\Big(1; p(0)\Big)$, the principal becomes indifferent at $k^*(p_0) + 1$, the cutoff defined in the preceding section.

Claim 3. *There exists $\delta_1 > 0$ such that for all $\delta \leq \delta_1$ there exists $N(\delta) \in \mathbb{N}$ and for all $N(\delta) \leq N$, there is a continuous function that maps each $p(0) \in [\Pr(\omega = 1), 1)$ to a number $p^*\Big(1; p(0)\Big) \in [p_{\frac{\delta}{2}}, p_\delta]$ satisfying*

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1|\omega = 1; p(1) = p^*\Big(1; p(0)\Big), p(0), N\Big)}{\Pr\Big(k = k^*(p_0) + 1|\omega = 0; p(1) = p^*\Big(1; p(0)\Big), p(0), N\Big)} = 1. \tag{27}$$

Proof. We assume $\delta \leq \frac{1}{4}$ throughout and proceed in two steps. In the first step, we show that there is some $N_2 \in \mathbb{N}$ and some $\gamma_1 > 0$ such that

$$\frac{\bar{k}}{N} \in \Big(q(0; \mathbf{p}) + \gamma_1, q(1; \mathbf{p}) - \gamma_1\Big) \tag{28}$$

for all $N \geq N_2$, $\mathbf{p} \in D(\delta)$, and $\delta \leq \frac{1}{4}$, where $\bar{k}$ is the best-response cutoff as in (1). We start by applying the first equation of (17) to obtain

$$\begin{aligned}
&\frac{\Pr\Big(k|\omega = 1; \mathbf{p}\Big)}{\Pr\Big(k|\omega = 0; \mathbf{p}\Big)} \\
&\qquad = \exp\Big(- N\Big(\mathrm{KL}\Big(\frac{k}{N}, q(1; \mathbf{p})\Big) - \mathrm{KL}\Big(\frac{k}{N}, q(0; \mathbf{p})\Big)\Big)\Big). \qquad (29)
\end{aligned}$$

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

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

which is strictly positive for $\gamma_1 > 0$ sufficiently small by (26) and since the Kullback–Leibler divergence has bounded partial derivatives on any compact interval.

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

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

Now, for any $\kappa > 0$, there is some $N(\kappa) \in \mathbb{N}$ such that for all $N \geq N(\kappa)$,

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

and the same bounds apply to the posterior likelihood ratio $\frac{\Pr(\omega=1|k;\mathbf{p},N)}{\Pr(\omega=0|k;\mathbf{p},N)}$ for any $\frac{k}{N} \leq q(0; \mathbf{p}) + \gamma_1$ and for any $\frac{k}{N} \geq q(1; \mathbf{p}) - \gamma_1$ respectively. Finally, we argue that we can choose $\kappa > 0$ small enough so that (28) holds uniformly, that is, for all $N \geq N_2 := N(\kappa)$, $\mathbf{p} \in D(\delta)$, and $\delta \leq \frac{1}{4}$. Since each private signal realization is boundedly informative, we can choose $\kappa > 0$ small enough so that $\frac{\Pr(\omega=1|\bar{k};\mathbf{p},N)}{\Pr(\omega=0|\bar{k};\mathbf{p},N)} < \kappa$ implies $\frac{\Pr(\omega=1|\bar{k}+1;\mathbf{p},N)}{\Pr(\omega=0|\bar{k}+1;\mathbf{p},N)} < 1$ and $\frac{\Pr(\omega=1|\bar{k};\mathbf{p},N)}{\Pr(\omega=0|\bar{k};\mathbf{p},N)} > \frac{1}{\kappa}$ implies $\frac{\Pr(\omega=1|\bar{k}-1;\mathbf{p},N)}{\Pr(\omega=0|\bar{k}+1;\mathbf{p},N)} > 1$ uniformly across all such parameters. This way, if (28) would not hold for some parameters, we would obtain a contradiction to the minimality of $\bar{k}$.

In the second step, we argue for the uniform existence of a number $p^*\Big(1; p(0)\Big) \in [p_{\frac{\delta}{2}}, p_\delta]$ that solves (27). For this we establish the existence of some $\delta_1 > 0$ such that for each $\delta \leq \delta_1$ there is some $N(\delta) \in \mathbb{N}$ and

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr\Big(k = k^*(p_0) + 1|\omega = 1; p(1) = p_{\frac{\delta}{2}}, p(0), N\Big)}{\Pr\Big(k = k^*(p_0) + 1|\omega = 0; p(1) = p_{\frac{\delta}{2}}, p(0), N\Big)} < 1 \tag{30}$$
$$\text{for all } N \geq N(\delta),\ \ \delta \leq \delta_1,\ \text{ and } p(0) \in [\Pr(\omega = 1), 1).$$

(note that we fix $p(1) = p_{\frac{\delta}{2}}$). Combining (25) and (30) and applying the intermediate value theorem then yields a $p^*\Big(1; p(0)\Big) \in (p_{\frac{\delta}{2}}, p_\delta]$ such that the principal is indifferent given $p^*\Big(1; p(0)\Big)$—i.e., (27) holds. Now, δ_1 and $N(\delta)$ for $\delta \leq \delta_1$ exist by the following argument. First, the minimality of $k^*(p_0) + 1$ implies a uniform bound $\gamma_2 > 0$ such that

$$\Pr(\omega = 1|k = k^* + 1; \mathbf{p}, N) - \frac{1}{2} \leq \gamma_2 \tag{31}$$

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 $p(1) = p_\delta$. So, the first step of this proof implies

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

for all $\mathbf{p} \in D(\delta)$ with $p(1) = p_\delta$, any $\delta \leq \frac{1}{4}$, and any $N \geq N_2$; cf. (28). Third, note that $q(0; \mathbf{p})$ and $q(1; \mathbf{p})$ are both strictly decreasing in $p(1)$. Given the second observation and the properties of the prior and signal distribution, there is $\delta_1 > 0$ 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 $\mathbf{p} \in D(\delta)$, $\delta \leq \delta_1$, and $N \geq N_2$. Jointly, these observations and (29) 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, the derivative is bounded from below by an arbitrarily large number if N is arbitrarily large. Recalling (31) and (25), we see that for any fixed $\delta \leq \delta_1$, there is $N(\delta) \in \mathbb{N}$ such that (30) holds. As argued, (30) implies the uniform existence of $p^*\Big(1; p(0)\Big) \in [p_{\frac{\delta}{2}}, p_\delta]$ solving (27).

Finally, we note that $p^*\Big(1; p(0)\Big)$ is unique and continuous. It is unique because the posterior is strictly decreasing. The continuity of $p^*\Big(1; p(0)\Big)$ in $p(0)$ follows from an application of the implicit function theorem. □

B.2.3 Ensuring the First Fixed Point Equation

Claim 4 says that, for any candidate cutoff $p(0)$, we can find a principal's mixed strategy $z^*\Big(p(0)\Big)$ such that

$$\hat{p}(1; \mathbf{p}, z^*) = p^*\Big(1; p(0)\Big). \tag{32}$$

Claim 4. *There exists $\delta_2 > 0$ such that for all $\delta \leq \delta_2$ there is $N_2(\delta) \in \mathbb{N}$ 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 $z^*\Big(p(0)\Big) \in [0, 1]$ such that (32) holds.*

Proof. We fix $p(1) = p^*\Big(p(0)\Big)$ for the duration of the proof of Claim 4. The proof leverages the principal's indifference between the pure strategies with cutoffs $k^*(p_0)$ and $k^*(p_0) + 1$. 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,

$$\hat{p}(1; \mathbf{p}, z) < p^*\Big(1; p(0)\Big) \text{ if } z = 0, \text{ and} \tag{33}$$

$$\hat{p}(1; \mathbf{p}, z) > p^*\Big(1; p(0)\Big) \text{ if } z = 1. \tag{34}$$

Since $\hat{p}(1; \mathbf{p}, z)$ is continuous in the probability z—see (3) and (11)—an application of the intermediate value theorem then implies the existence of a principal's mixed strategy $z^* \in (0, 1)$ such that $\hat{p}(1, \mathbf{p}, z^*) = p^*\Big(1; p(0)\Big)$.

In the following, we first establish that the inequalities (33) and (34) hold in the limit as $N \to \infty$ and then $\delta \to 0$, i.e., that $\lim_{\delta\to 0} \lim_{N\to\infty} \hat{p}(1; \mathbf{p}, z) < p^*\Big(1; p(0)\Big)$ if $z = 0$ and $\lim_{\delta\to 0} \lim_{N\to\infty} \hat{p}(1; \mathbf{p}, z) > p^*\Big(1; p(0)\Big)$ 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 $1 - z = \Pr(\tilde{k} = k^* + 1) = 1$, so that

$$\Pr(\text{piv}_0 | \omega = \omega'; \mathbf{p}, N) = \Pr(k_{-i} = k^* + 1 | \omega = \omega'; \mathbf{p}) \text{ for } \omega' \in \{0, 1\}, \tag{35}$$

where $\Pr(k_{-i} = k^* + 1 | \omega = \omega'; \mathbf{p})$ is the posterior conditional on $k^* + 1$ out of

$N-1$ agents choosing the 1-action. The principal is indifferent if she observes that $k = k^* + 1$ out of N agents have chosen the 1-action, so

$$\frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} = \frac{\Pr\left(k = k^* + 1 | \omega = 0; \mathbf{p}, N\right)}{\Pr\left(k = k^* + 1 | \omega = 1; \mathbf{p}, N\right)}. \tag{36}$$

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

$$1 - q(\omega'; \mathbf{p}) \to \Pr(s_i = 0 | \omega = \omega'), \tag{37}$$

so that

$$\begin{aligned} \frac{\Pr\left(k_{-i} = k^* + 1 | \omega = 0; \mathbf{p}, N\right)}{\Pr\left(k_{-i} = k^* + 1 | \omega = 1; \mathbf{p}, N\right)} &= \left(\frac{q(0; \mathbf{p})}{q(1, \mathbf{p})}\right)^{k^*+1} \left(\frac{1 - q(0; \mathbf{p})}{1 - q(1; \mathbf{p})}\right)^{N-k^*-2} \\ \to \frac{\Pr\left(k = k^* + 1 | \omega = 0; \mathbf{p}, N\right)}{\Pr\left(k = k^* + 1 | \omega = 1; \mathbf{p}, N\right)} &\cdot \frac{\Pr(s_i = 0 | \omega = 1)}{\Pr(s_i = 0 | \omega = 0)}. \end{aligned} \tag{38}$$

Recall the property (8). It implies (9), i.e., that in the limit as $N \to \infty$, an agent cares only about piv_0. Hence, the conditions (11) for the indifference cutoffs $p(s)$ of the best response imply

$$\lim_{N \to \infty} \frac{\hat{p}(s; \mathbf{p}, z)}{1 - \hat{p}(s; \mathbf{p}, z)} = \frac{\Pr\left(\text{piv}_0 | \omega = 0; \mathbf{p}, N\right)}{\Pr\left(\text{piv}_0 | \omega = 1; \mathbf{p}, N\right)} \cdot \frac{\Pr(s_i = s | \omega = 0)}{\Pr(s_i = s | \omega = 1)}. \tag{39}$$

If we combine (35)–(38) and take $\delta \to 0$, this indifference condition becomes $\lim_{\delta \to 0} \lim_{N \to \infty} \hat{p}(1; \mathbf{p}, z) = \underline{p}(1)$ for $s = 1$, with

$$\frac{\underline{p}(1)}{1 - \underline{p}(1)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr(s_i = 0 | \omega = 1)}{\Pr(s_i = 0 | \omega = 0)} \cdot \frac{\Pr(s_i = 1 | \omega = 0)}{\Pr(s_i = 1 | \omega = 1)}. \tag{40}$$

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

$$\Pr(\text{piv}_0 | \omega = \omega'; \mathbf{p}) = \Pr(k_{-i} = k^* | \omega = \omega'; \mathbf{p}) \text{ for } \omega \in \{0, 1\}. \tag{41}$$

As $\delta \to 0$, we have

$$\frac{\Pr\Big(k_{-i} = k^*|\omega = 0; \mathbf{p}, N\Big)}{\Pr\Big(k_{-i} = k^*|\omega = 1; \mathbf{p}, N\Big)} = \left(\frac{q(0;\mathbf{p})}{q(1,\mathbf{p})}\right)^{k^*} \cdot \left(\frac{1-q(0;\mathbf{p})}{1-q(1;\mathbf{p})}\right)^{N-k^*-1}$$
$$\to \frac{\Pr\Big(k = k^*+1|\omega = 0; \mathbf{p}, N\Big)}{\Pr\Big(k = k^*+1|\omega = 1; \mathbf{p}, N\Big)} \cdot \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)}. \tag{42}$$

If we combine (36), (41), and (42), and take $\delta \to 0$, the indifference condition (11) becomes $\lim_{\delta\to 0} \lim_{N\to\infty} \hat{p}(0; \mathbf{p}, z) = \bar{p}(1)$ for $s = 1$, with

$$\frac{\bar{p}(1)}{1-\bar{p}(1)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)} \cdot \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)} \cdot \frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)} = \frac{\Pr(\omega = 1)}{\Pr(\omega = 0)}. \tag{43}$$

Now, we combine the approximations (40) and (43) with the requirement (6) on the prior distribution to argue that (33) and (34) hold "in the limit." The requirement (6) implies

$$\underline{p}(1) < p_{\frac{\delta}{2}} \text{ and } p_{1-\frac{\delta}{2}} < \Pr(\omega = 1). \tag{44}$$

Combining this with $\Pr(\omega = 1) \leq \bar{p}(1)$ and $p_{\frac{\delta}{2}} < p^*\Big(1; p(0)\Big) \leq p_\delta$ shows that $\underline{p}(1) < p^*\Big(1; p(0)\Big) < \bar{p}(1)$. Hence, (40) and (43) imply (33) and (34) in the limit, i.e., as $N \to \infty$ and then $\delta \to 0$. Note that we used Claim 3 here, which guarantees the existence of $p^*\Big(1; p(0)\Big)$ for $\delta \leq \delta_1$ and $N \geq N(\delta)$ (this also explains the order of limits).

Next we show that there are uniform bounds $\delta_2 > 0$ and $N_2(\delta) \in \mathbb{N}$ for $\delta \leq \delta_2$ such that (33) and (34) hold not just in the limit, but for any $\delta \leq \delta_2$, $N_2(\delta) \leq N$, and $\mathbf{p} \in D(\delta)$ with $p(1) = p^*\Big(1; p(0)\Big)$. The limit analysis for $\hat{p}(1; \mathbf{p}, z)$ used two approximations, (37) and (39), 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 (37), this is obvious for some δ_2 small enough. For (39), this follows from observing that the convergence here is exponential in the difference $\min_{\omega'\in\{0,1\}} \mathrm{KL}\Big(m_0, q(\omega'; \mathbf{p})\Big) - \min_{\omega'\in\{0,1\}} \mathrm{KL}\Big(m_1, q(\omega'; \mathbf{p})\Big)$, and this difference is uniformly bounded from below, given (8). 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 $N_2(\delta)$. This way, the bounds analogous to (44) hold uniformly—$\hat{p}(1; \mathbf{p}, z) < p_{\frac{\delta}{2}}$ for $z = 0$ and $p_{1-\frac{\delta}{2}} < \hat{p}(1; \mathbf{p}, z)$ for $z = 1$—and imply (33) and (34), given $p_{\frac{\delta}{2}} < p^*\Big(1; p(0)\Big) \leq p_\delta$.

Finally, since (33) and (34) hold for any $\delta < \delta_2$, $N \geq N_2(\delta)$, and $\mathbf{p} \in D(\delta)$

with $p(1) = p^*\Big(1; p(0)\Big)$, for any $p(0) \geq \Pr(\omega = 1)$ (and $\delta \leq \delta_2$, $N \geq N_2(\delta)$), an application of the intermediate value theorem yields a $z^*(p_0)$ such that (32) 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 $p(0)$, z^* is also continuous in $p(0)$. □

B.2.4 Ensuring the Second Fixed Point Equation

We use a fixed point argument to establish the existence of $p^*(0)$ such that the agents' strategy given by $\Big(p^*(0), p^*\Big(1; p^*(0)\Big)\Big)$ 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 Claim 4. This ensures that the correspondence that maps any $p(0) \geq \Pr(\omega = 1)$ to the projection

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

with $\mathbf{p} = \Big(p(0), p^*\Big(1; p(0)\Big)\Big)$, is well-defined. The correspondence is continuous in $p(0) \in [\Pr(\omega = 1), 1)$ because $\hat{p}\Big(0; \mathbf{p}, z^*(p(0))\Big)$ is continuous in $p(0)$. (The latter holds because $p^*\Big(1; p(0)\Big)$ and $z^*(p_0)$ are continuous in $p(0)$ and all three parameters $p(0)$, $p^*\Big(1; p(0)\Big)$ 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 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 $p^*_N(0)$.

We argue that any fixed point $p^*_N(0)$ is interior, i.e., it is strictly greater than $\Pr(\omega = 1)$. To show this, we use (40). The best-response cutoff $\hat{p}(0; \mathbf{p}, z^*)$ relates to $\hat{p}(1; \mathbf{p}, z^*) = p^*\Big(1; p(0)\Big)$ via the following equation:

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

Comparing this to (40), we see that $\underline{p}(1) < p^*\Big(1; p(0)\Big)$ (which holds since N and δ satisfy the uniform bounds of Claim 4; cf. 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.

We can now finish the proof of Theorem 2. By definition, the agents' strategy σ^*_N given by an interior fixed point $p^*_N(0)$ and $p^*_N(1) = p^*\Big(1, p^*_N(0)\Big)$ is a best response to itself, given $z^*\Big(p_N^*(0)\Big)$. The sequence $(\sigma_N^*)_{N\in\mathbb{N}}$ is a sequence of equilibrium strategies. Since $p_N^*\Big(1; p_N^*(0)\Big) < p_\delta$ and $p_N^*(0) \geq \Pr(\omega = 1) > p_{1-\frac{\delta}{2}}$, by construction, it is a sequence of δ-approximately truthful strategies. For any $\delta > 0$ sufficiently small, an application of the weak law of large numbers implies (23). For $j \geq 2$, this follows from (24); for $j = 1$, this follows from the analogous condition thereafter. This proves Theorem 2.

C Proof of Theorem 3: Sufficiency

C.1 Unique Interior Cutoff Types

Take any equilibrium of a monotone process with a single cutoff, with no balance, and for which $\max P(0) < \max P(1)$ holds. We show that, for any signal realization $s \in \{0,1\}$, there are unique types $0 < p_N(s) < 1$ that are indifferent after observing s. As shown in the main text, for this it is sufficient to establish that $U(\omega) \neq 0$ for some $\omega \in \{0,1\}$, i.e., (10). Then $U(\omega)$ has non-zero and equal sign in both states, and the unique indifferent types are pinned down by (11). We prove (10) by contradiction, in two steps (Claim 5 and Claim 6).

Claim 5. *Take any monotone process with a single cutoff and no balance, for which* $\max P(0) < \max P(1)$ *holds. Take any equilibrium* $\eta = (\sigma, \bar{k}, \tilde{x})$*. If* $U(0;\eta) = U(1;\eta) = 0$*, then*

(i) $0 < q(1;\sigma) < q(0;\sigma) < 1$*, and*

(ii) $\lfloor m_1 N \rfloor \in \{\bar{k}, \bar{k}+1\}$.

Proof. For item (i), clearly $q(0;\sigma) \neq q(1;\sigma)$ since otherwise the equilibrium is uninformative, and we have shown in the main text that no equilibrium is uninformative. Next, $q(0;\sigma) \in \{0,1\}$ would imply $q(0;\sigma) = q(1;\sigma)$, which we have just excluded. If $0 < q(0;\sigma) < q(1;\sigma) < 1$, a 1-action is more likely in state 1. Thus, the principal's posterior and her policy choice $x(k)$ are increasing in the observed ratio $m = \frac{k}{N}$ of 1-actions. We claim this implies that the average effect of an additional 1-action on x is positive; that is, $U(\omega) > 0$. To see this, note that all observations m are on path since $0 < q(\omega;\sigma) < 1$. Bayesian consistency implies that $\Pr(\omega = 1|m) > \frac{1}{2}$ for some m. Thus, the principal chooses $\max P(m)$ after some m. Either she chooses $\max P(m)$ for all m, or she chooses $\max P(m)$ for high m only and a policy smaller than $\max P(m)$ otherwise; in either case, the claim follows from the fact that $\max P(0) < \max P(1)$. We conclude that $U(1;\eta) = U(0;\eta) = 0$ implies $0 < q(1;\sigma) < q(0;\sigma) < 1$.

For item (ii), we introduce some notation:

$$
\begin{aligned}
p_1(\omega') &:= & \Pr(k_{-i} = \bar{k} | \omega = \omega'; \sigma, N), \\
p_2(\omega') &:= & \Pr(k_{-i} = \bar{k} + 1 | \omega = \omega'; \sigma, N),
\end{aligned}
$$

and

$$
\begin{aligned}
A_1 &:= & \mathrm{E}(x | a_i = 1, \mathrm{piv}_{0,\bar{k}}; \eta, N) - \mathrm{E}(x | a_i = 0, \mathrm{piv}_{0,\bar{k}}; \eta, N), \\
A_2 &:= & \mathrm{E}(x | a_i = 1, \mathrm{piv}_{0,\bar{k}+1}; \eta, N) - \mathrm{E}(x | a_i = 0, \mathrm{piv}_{0,\bar{k}+1}; \eta, N).
\end{aligned}
$$

Now, suppose the counterstatement is true, i.e., $\lfloor m_1 N \rfloor \notin \{\bar{k}, \bar{k}+1\}$. Then

$$
\bar{k} < \bar{k} + 1 < \lfloor m_1 N \rfloor, \quad \text{or} \tag{45}
$$

$$
\lfloor m_1 N \rfloor < \bar{k} < \bar{k} + 1. \tag{46}
$$

We argue that

$$
\sum_{h=1,2} p_h(\omega') A_h < 0 \tag{47}
$$

for all $\omega' \in \{0, 1\}$, so that $U(\omega') = 0$ is equivalent to

$$
1 + \frac{\Pr(\mathrm{piv}_1 | \omega = \omega'; \eta, N)}{\sum_{h=1,2} p_h(\omega') A_h} \Big(\mathrm{E}(x | a_i = 1, \mathrm{piv}_1; \eta, N) - \mathrm{E}(x | a_i = 0, \mathrm{piv}_1; \eta, N) \Big) = 0, \tag{48}
$$

For this we make some preliminary observations, labeled (a)–(c). First, since $0 < q(\omega'; \sigma) < 1$, each action is taken with positive probability in each state. Thus, there is a positive probability that k_{-i} takes the value $\lfloor m_1 N \rfloor$. This yields the observation (a): $\Pr(\mathrm{piv}_1 | \omega'; \eta, N) > 0$. Furthermore, since all possible observations are on path, the principal chooses the lowest possible policy if the number of 1-actions she observes is more than $\bar{k} + 1$, and the highest possible policy if the number of 1-actions she observes is less than $\bar{k} + 1$. If the number of 1-actions she observes is exactly $\bar{k} + 1$, then she is indifferent between all policies in $P(\frac{\bar{k}+1}{N})$. Thus we have the observation (b):

$$
\begin{aligned}
&\mathrm{E}(x | a_i = 1, \mathrm{piv}_1; \eta, N) - \mathrm{E}(x | a_i = 0, \mathrm{piv}_1; \eta, N) \\
&\quad = \begin{cases} \min P(1) - \min P(0) & \text{if (45) holds,} \\ \max P(1) - \max P(0) & \text{if (46) holds.} \end{cases}
\end{aligned}
$$

We also have (c): $\sum_{h=1,2} p_h(\omega') A_h \le 0$.

We now apply these observations in a case analysis. We start with the boundary cases where $\bar{k} \in \{N-1, N\}$. For these values of $\bar{k}$, (46) holds. Given the assumption that $\max P(1) - \max P(0) > 0$, this means $\mathrm{E}(x|a_i = 1, \mathrm{piv}_1; \eta, N) - \mathrm{E}(x|a_i = 0, \mathrm{piv}_1; \eta, N) = \max P(1) - \max P(0) > 0$. But then (47) must hold, since otherwise $U(\omega) = \Pr(\mathrm{piv}_1|\omega'; \eta, N)\Big(\max P(1) - \max P(0)\Big) > 0$, given (a) and (c).

We turn to the case where $\bar{k} \notin \{N-1, N\}$. Here, $0 < q(\omega'; \sigma) < 1$ implies there is a positive probability that k_{-i} takes each of the values $\bar{k}$ and $\bar{k}+1$. In our notation, $p_h(\omega') > 0$ for $h = 1, 2$. The optimality of the principal's equilibrium strategy implies

$$A_1 + A_2 = \min P(0) - \max P(0) < 0 \ \text{ if (45) holds,}$$
$$A_1 + A_2 = \min P(1) - \max P(1) < 0 \ \text{ if (46) holds.}$$

Taken together, these observations imply (47).

Some final observations imply a contradiction to the assumption $U(0; \eta) = U(1; \eta) = 0$ made in Claim 5. First, (47) implies that $p_h(\omega') > 0$ for some $h \in \{1, 2\}$, given $\omega' \in \{0, 1\}$. Second, if $p_h(\omega') > 0$ for some $\omega' \in \{1, 2\}$, the same is true for all $\omega' \in \{1, 2\}$. Third, Claim 2 implies that for any $h \in \{1, 2\}$ with $p_h(0) > 0$ and $p_h(1) > 0$,

$$\frac{\Pr(\mathrm{piv}_1|\omega = 1; \eta, N)}{p_h(1)} < \frac{\Pr(\mathrm{piv}_1|\omega = 0; \eta, N)}{p_h(0)} \ \text{ if (45) holds,}$$
$$\frac{\Pr(\mathrm{piv}_1|\omega = 1; \eta, N)}{p_h(1)} > \frac{\Pr(\mathrm{piv}_1|\omega = 0; \eta, N)}{p_h(0)} \ \text{ if (46) holds.}$$

Inspecting (48), we see that $U(\omega')$ cannot be zero for both states ω', a contradiction. □

Claim 6. *Take any monotone process with a single cutoff and no balance for which* $\max P(0) < \max P(1)$ *holds. In any equilibrium* $\eta = (\sigma, \bar{k}, \tilde{x})$*, either* $U(0; \eta) \neq 0$ *or* $U(1; \eta) \neq 0$.

To prepare for the proof, we decompose $U(\omega; \eta)$. Let $O(P) = P(0) \cap P(1)$,

$$r_1(\omega') = \Pr\left(\mathrm{piv}_{0,\bar{k}}|\omega = \omega'; \eta, N\right) \Pr\left(\tilde{x} \notin O(P)\right),$$
$$r_2(\omega') = \Pr\left(\mathrm{piv}_{0,\bar{k}}|\omega = \omega'; \eta, N\right) \Pr\left(\tilde{x} \in O(P)\right),$$
$$r_3(\omega) = \Pr\left(\mathrm{piv}_{0,\bar{k}+1}|\omega = \omega'; \eta, N\right) \Pr\left(\tilde{x} \notin O(P)\right),$$
$$r_4(\omega) = \Pr\left(\mathrm{piv}_{0,\bar{k}+1}|\omega = \omega'; \eta, N\right) \Pr\left(\tilde{x} \in O(P)\right),$$

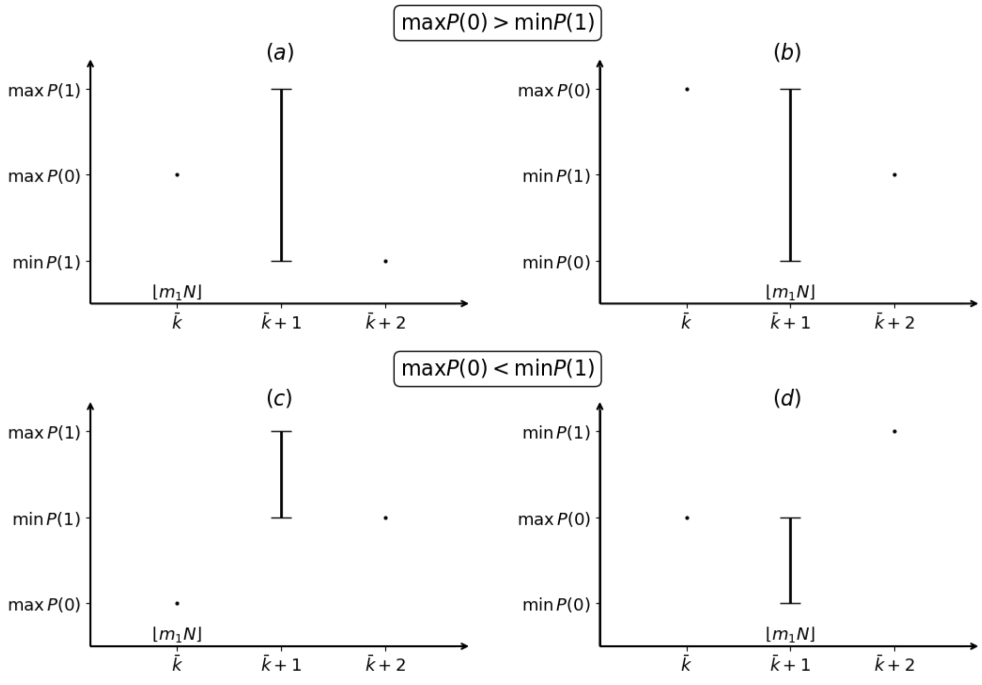

Figure 3: The principal's best-response correspondence after observing $k \in \{\bar{k}, \bar{k}+1, \bar{k}+2\}$ 1-actions, in all possible scenarios.

and

$$\begin{aligned}
B_1 &= \mathrm{E}\Big(x|a_i = 1, \mathrm{piv}_{0,\bar{k}}, \bar{x} \notin O(P); \eta, N\Big) - \mathrm{E}\Big(x|a_i = 0, \mathrm{piv}_{0,\bar{k}}, \bar{x} \notin O(P); \eta, N\Big), \\
B_2 &= \mathrm{E}\Big(x|a_i = 1, \mathrm{piv}_{0,\bar{k}}, \bar{x} \in O(P); \eta, N\Big) - \mathrm{E}\Big(x|a_i = 0, \mathrm{piv}_{0,\bar{k}}, \bar{x} \in O(P); \eta, N\Big), \\
B_3 &= \mathrm{E}\Big(x|a_i = 1, \mathrm{piv}_{0,\bar{k}+1}, \bar{x} \notin O(P); \eta, N\Big) - \mathrm{E}\Big(x|a_i = 0, \mathrm{piv}_{0,\bar{k}+1}, \bar{x} \notin O(P); \eta, N\Big), \\
B_4 &= \mathrm{E}\Big(x|a_i = 1, \mathrm{piv}_{0,\bar{k}+1}, \bar{x} \in O(P); \eta, N\Big) - \mathrm{E}\Big(x|a_i = 0, \mathrm{piv}_{0,\bar{k}+1}, \bar{x} \in O(P); \eta, N\Big).
\end{aligned}$$

We carry out the proof by analyzing four cases, based on whether $\bar{k} = \lfloor m_1 N \rfloor$ or $\bar{k}+1 = \lfloor m_1 N \rfloor$, and whether $\max P(0) > \min P(1)$ or $\max P(0) < \min P(1)$ (the assumption of no balance rules out the case where $\max P(0) = \min P(1)$). These cases are labeled (a)–(d) as illustrated in Figure 3.

Proof. Note that $\lfloor m_1 N \rfloor \in \{\bar{k}, \bar{k}+1\}$ implies $\mathrm{piv}_1 = \emptyset$ by definition. Thus, $U(\omega') = \sum_h r_h(\omega') B_h$ in all cases. We now assume for the sake of contradiction that $U(0;\eta) = U(1;\eta) = 0$.

We define numbers j^* and k^* as follows: Set $j^* = 1$ and $k^* = 3$ in cases (a) and (d); set $j^* = 3$ and $k^* = 1$ in cases (b) and (c). Inspection of Figure 3 shows that

$U(\omega'; \eta) = 0$ is equivalent to

$$r_{j^*}(\omega') \cdot |B_{j^*}| = \sum_{j \neq j^*} r_j(\omega') \cdot |B_j|,$$

or equivalently

$$|B_{j^*}| = \sum_{j \neq j^*} \frac{r_j(\omega')}{r_{j^*}(\omega')} \cdot |B_j|, \tag{49}$$

and implies $r_{j^*}(\omega') \cdot |B_{j^*}| > 0$, which in turn implies $r_{k^*}(\omega') \cdot |B_{k^*}| > 0$.

Since $0 < q(1;\sigma) < q(0;\sigma) < 1$ by Claim 5, an application of Claim 2 yields $\frac{\Pr(k_{-i}=\bar{k}+1|\omega=1;\eta,N)}{\Pr(k_{-i}=\bar{k}|\omega=1;\eta,N)} < \frac{\Pr(k_{-i}=\bar{k}+1|\omega=0;\eta,N)}{\Pr(k_{-i}=\bar{k}|\omega=0;\eta,N)}$. Hence

$$\sum_{j \neq j^*} \frac{r_j(1)}{r_{j^*}(1)} \cdot |B_j| < \sum_{j \neq j^*} \frac{r_j(0)}{r_{j^*}(0)} \cdot |B_j| \text{ in cases (a) and (d),} \tag{50}$$

$$\sum_{j \neq j^*} \frac{r_j(1)}{r_{j^*}(1)} \cdot |B_j| > \sum_{j \neq j^*} \frac{r_j(0)}{r_{j^*}(0)} \cdot |B_j| \text{ in cases (b) and (c).} \tag{51}$$

Combining (49) with (50)–(51), we conclude that, in every case, either $U(0;\eta) \neq 0$ or $U(1;\eta) \neq 0$, which contradicts our initial assumption. $\square$

C.2 Information Aggregation

Claim 7. *Take any monotone process with a single cutoff and no balance for which* $\max P(0) < \max P(1)$ *holds. Then any equilibrium sequence aggregates information.*

Proof. Take any equilibrium sequence $(\eta_N)_{N\in\mathbb{N}}$ of a process meeting the conditions in the claim. In Section C.1 we established the existence of unique indifferent types $0 < p_N(s) < 1$ for any signal s and any N.

Suppose that the limit indifferent types are interior, i.e., $0 < \lim_{N\to\infty} p_N(1) < \lim_{N\to\infty} p_N(0) < 1$.[38] Then the limit of the mean action differs across signals and thus across the two states, i.e., $0 < \lim_{N\to\infty} q(0;\sigma_N) \neq \lim_{N\to\infty} q(0;\sigma_N) < 1$. By an application of the law of large numbers, the realized share of 1-actions is almost surely close to the mean action in each state, implying that the principal learns the state from observing it. Thus, information aggregates.

In the following, we consider the scenario in which the limit indifferent types

[38] It suffices to derive a contradiction for any subsequence along which $p_N(0)$ and $p_N(1)$ converge, since the values of $p_N(0)$ and $p_N(1)$ lie in the compact set $[0,1]$. For simplicity of notation, we replace the original sequence with such a subsequence.

are not interior,[39]

$$\lim_{N\to\infty} p_N(0) = \lim_{N\to\infty} p_N(1) \in \{0,1\}. \tag{52}$$

The arguments are based on a detailed analysis of point events for the realized number k_{-i} of 1-actions 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 (16) 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 (17) to obtain

$$\frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} = \frac{\Pr(\omega = 1)}{1 - \Pr(\omega = 1)} \cdot \exp\Big((N-1)\Big(\mathrm{KL}\Big(m_N, q(0,\sigma_N)\Big) - \mathrm{KL}\Big(m_N, q(1,\sigma_N)\Big)\Big)\Big). \tag{53}$$

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 $k_N \neq N$ for all N large enough, we have

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in (\kappa, \frac{1}{\kappa}) \text{ for } m_N \in \{m''_N, m'''_N\}, \tag{54}$$

for some $\kappa > 0$, by the defining property (1) of $\bar{k}_N$ (whenever we apply (54), we will rule out the case $k_N = N$). Second, (52) implies

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in \{0, \infty\} \text{ for } m_N = m'_N. \tag{55}$$

Otherwise the inference from piv_1 would be bounded as $N \to \infty$. Then, for any N large enough, we would have either $k_N = N$, so that only piv_1 would be relevant for the agents' best response, or, by (54), the inference from piv_0 would also be uniformly bounded. In either case, the indifferent types would be bounded away from 0 and 1.[40] Third, (52) implies $\lim_{N\to\infty} q(0;\sigma_N) = \lim_{N\to\infty} q(1;\sigma_N)$. We set

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

[39] It suffices to derive a contradiction for any subsequence along which $p_N(0)$ and $p_N(1)$ converge, since the values of $p_N(0)$ and $p_N(1)$ lie in the compact set $[0,1]$. For simplicity of notation, we replace the original sequence with such a subsequence.

[40] The indifferent types are pinned down by (11). Given (54) and (55), boundedness of the indifferent type would follow from setting $a = \Pr(\mathrm{piv}_0|\omega = 0; \eta_N, N)$, $b = \Pr(\mathrm{piv}_0|\omega = 1; \eta_N, N)$, $c = \Pr(\mathrm{piv}_1|\omega = 0; \eta_N, N)$, $d = \Pr(\mathrm{piv}_1|\omega = 1; \eta_N, N)$, and using the fact that for any $u, v, a, b, c, d > 0$, we have $\min(\frac{a}{b}, \frac{c}{d}) \leq \frac{au+cv}{bu+dv} \leq \max(\frac{a}{b}, \frac{c}{d})$.

Case 1: $m_1 \neq q^*$.[41]

In this case we derive a contradiction to (52). 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 (19) applies, so that

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

For $m_N = m_N'$, the unbounded inference (55) implies

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

We show case by case that

$$\lim_{N\to\infty} m_N'' = \lim_{N\to\infty} m_N''' = q^*. \tag{56}$$

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 $\gamma > 0$,

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

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 (56) 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 (54), this implies interior limit cutoffs (cf. (3) and (11)), contradicting the initial assumption (52).

[41] In the main text we assert that this is a generic case. This is true because (52) implies $q^* \in \{F(0), F(1^-), 1-F(0), 1-F(1^-)\}$.

Case 2: $m_1 = q^*$.

This case can be broken down into several subcases, all of which are analogous; we present only one. Consider an equilibrium sequence such that $\lim_{N\to\infty} \Pr(\omega = 1|\text{piv}_1; \sigma_N, N) = 1$, and such that, for any N, a type chooses the 1-action if and only if $p_i \leq p_N(s)$. Then, since $p_N(1) < p_N(0) \in (0,1)$, it holds that $F(0) < q(1;\sigma_n) < q(0;\sigma_N) < F(1^-)$. Recall that $p_N(s) \to 0$ or $p_N(s) \to 1$ by the assumption (52). If $p_N(s) \to 1$, then $q^* = F(1^-)$. However, then $\lim_{N\to\infty} \Pr(\omega = 1|\text{piv}_1; \sigma_N, N) = 1$ cannot hold. Thus, $p_N(s) \to 0$, which implies $q^* = F(0)$.

We now carefully examine the mean actions in each state,

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

Using simple algebra,[42] 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

$$\lim_{N\to\infty} \frac{q(\omega';\sigma_N) - m_1}{-\Delta_N} \in (0,\infty) \tag{57}$$

for all ω'.[43] Using the approximation $q(\omega';\sigma_N) \approx q^*$, we restate the quadratic approximation (18) for $m = m_N$ and $q = q(\omega';\sigma_N)$ as follows:

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

This approximation yields the following difference in divergences:

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

For $m_N = m'_N$, the unbounded inference (55) together with (57) then implies that $\Delta_N^2 N \to \infty$. Applying the central limit theorem and denoting by $\frac{k_N}{N}$ the realized

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

[43] The inequalities hold for any subsequence along which $\frac{q(1;\sigma_N)-m_1}{q(0;\sigma_N)-m_1}$ converges. For simplicity of notation, we replace the original sequence by such a subsequence.

share of 1-actions 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 almost surely

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

hence, almost surely

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

Given (53), this means the principal learns the state almost surely: $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N, N) = 1$ if $\omega = 1$ and $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N, N) = 0$ if $\omega = 0$. $\square$

D Proof of Proposition 1: Omitted Parts

We verify that the modified best-response correspondence from the main text satisfies the requirements of Kakutani's fixed point theorem—namely, that it has convex and non-empty values and a closed graph.

First, the modified correspondence inherits certain properties of the best-response correspondence: Its graph is the intersection of the graph of the best-response correspondence, which is closed, with $\Sigma \times \Sigma$, which is also closed (recall that Σ is the closed set of candidate strategy profiles). In addition, the modified correspondence is convex-valued since the monotonicity properties (i.e., $x(k) \geq x(k')$ for all $k' > k$ and (13)) are preserved by mixtures.

To establish the non-emptiness, we verify that the principal has a monotone best response and that the agents have a best response satisfying (13). First, observe that (13) implies that the principal's posterior is weakly increasing in the number of 1-actions. Thus, she has a monotone best response. Second, the monotonicity of the principal's strategy implies that the average effect of an additional 1-action is positive in each state, i.e., $U(0;\eta) \geq 0$ and $U(1;\eta) \geq 0$. Given the best-response characterization (11), this immediately implies that it is a best response for each partisan to match his action to his type. Therefore there is generally a best response satisfying the right condition of (13). A case analysis shows that in addition we can meet the left condition of (13):

- If $U(0;\eta) > 0$ and $U(1;\eta) = 0$, then all types except the type-1 partisans choosing $x = 0$ is a best response with $q(0;\sigma) = q(1;\sigma)$.

- If $U(0;\eta) = 0$ and $U(1;\eta) > 0$, then all types except the type-0 partisans choosing $x = 1$ is a best response with $q(0;\sigma) = q(1;\sigma)$.

- If $U(1;\eta) > 0$ and $U(0;\eta) > 0$, then (11) pins down the (essentially) unique best response, which is given by the cutoffs $0 < p_N(1) < p_N(0) < 1$, with types choosing $x = 1$ if and only if $p_i \geq p_N(s)$. This best response satisfies $q(0;\sigma) < q(1;\sigma)$.

- If $U(0;\eta) = U(1;\eta)$, then every agents' strategy is a best response. In particular, there is a best response meeting both conditions of (13).

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., Dessein, W. and Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, **98** (1), 145–179.

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

Antic, N. and Iaryzcower, M. (2020). *Beyond delegation: Agency discretion when budgets matter*. Working paper.

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.

Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, pp. 1771–1813.

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.

Bond, P. and Eraslan, H. (2010). Strategic voting over strategic proposals. *The Review of Economic Studies*, **77** (2), 459–490.

Bouton, L. and Castanheira, M. (2012). One person, many votes: Divided majority and information aggregation. *Econometrica*, **80** (1), 43–87.

Burke, E. (1774). *The works of the right honourable Edmund Burke*, vol. 5. Wells and Lilly.

Carroll, G. (2019). Robustness in mechanism design and contracting. *Annual Review of Economics*, **11** (1), 139–166.

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

Chen, Y., Kartik, N. and Sobel, J. (2008). Selecting cheap-talk equilibria. *Econometrica*, **76** (1), 117–136.

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.

Dekel, E. and Piccione, M. (2000). Sequential voting procedures in symmetric binary elections. *Journal of political Economy*, **108** (1), 34–55.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies*, **69** (4), 811–838.

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

— and — (2022). Information aggregation in poisson elections. *Theoretical Economics*, **17** (1), 1–23.

Escher, F. (1932). On the probability function in the collective theory of risk. *Skand. Aktuarie Tidskr.*, **15**, 175–195.

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.

— and — (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political science review*, **92** (1), 23–35.

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.

Fox, J. and Shotts, K. W. (2009). Delegates or trustees? a theory of political accountability. *The Journal of Politics*, **71** (4), 1225–1237.

Gailmard, S. and Patty, J. W. (2012). Formal models of bureaucracy. *Annual review of political science*, **15** (1), 353–377.

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

Holmström, B. R. (1978). *On Incentives and Control in Organizations*. Stanford University.

Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate discrete distributions*. John Wiley & Sons.

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. (2001). A model of expertise. *The Quarterly Journal of Economics*, **116** (2), 747–775.

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

Ledyard, J. O. and Palfrey, T. R. (2002). The approximation of efficient public good mechanisms by simple voting schemes. *Journal of Public Economics*, **83** (2), 153–171.

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

Madison, J., Hamilton, A. and Jay, J. (1788). The federalist papers, 1987 edition.

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

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

Page, B. I. (1976). The theory of political ambiguity. *American Political Science Review*, **70** (3), 742–752.

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

— (2016). Experiments in political economy. *Handbook of experimental economics*, **2**, 347–434.

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.

Shepsle, K. A. (1972). The strategy of ambiguity: Uncertainty and electoral competition. *American Political science review*, **66** (2), 555–568.

Urbinati, N. and Warren, M. E. (2008). The concept of representation in contemporary democratic theory. *Annu. Rev. Polit. Sci.*, **11** (1), 387–412.

Wilson, R. (1985). *Game-theoretic analyses of trading processes*. Institute for Mathematical Studies in the Social Sciences, Stanford University.

Online Appendix

E Heterogeneous Ex-Post Preferences

We state an analog of Theorem 1 for a model in which the voters have a private preference type and payoffs can depend on the state in a general way (Theorem 7). Except for this additional type dimension, the baseline model from Section 1.1 is unchanged.

Formally, an agent's private type is now a prior $p_i \in [0,1]$ and a pair $\boldsymbol{t_i} = \Big(t_i(0), t_i(1)\Big) \in [-1,1]^2$, describing the type's constant marginal benefit from the policy choice in the two states. Types t_i are drawn from an absolutely continuous distribution G and independently from priors, signals, and the state, and across voters. Types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) < 0$ and $(1-p_i)\Big(t_i(1) - \frac{1}{2}\Big) > 0$ prefer $x = 0$ in $\omega = 0$ and $x = 1$ in $\omega = 1$. Types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) > 0$ and $(1-p_i)\Big(t_i(1)-\frac{1}{2}\Big) < 0$ have opposed preferences and prefer $x = 1$ in $\omega = 0$ and $x = 0$ in $\omega = 1$. Types for which $p_i\Big(t_i(0)-\frac{1}{2}\Big) \leq 0$ and $(1-p_i)\Big(t_i(1)-\frac{1}{2}\Big) \leq 0$ weakly prefer $x = 0$ in each state; types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) \geq 0$ and $(1-p_i)\Big(t_i(1) - \frac{1}{2}\Big) \geq 0$ weakly prefer $x = 1$. We denote the likelihood of these "partisans" by $\rho(0)$ and $\rho(1)$ respectively. We assume that the mass of types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) = 0$ *and* $(1-p_i)\Big(t_i(1) - \frac{1}{2}\Big) = 0$ is zero and ignore these types in the following, without loss of generality. Note that $\rho(0)$ and $\rho(1)$ are then well-defined. As before, we maintain $\rho(0) > 0$ and $\rho(1) > 0$, and the tie-breaking rule for the partisans. Finally, we generalize the notion of an agents' information structure π to mean the pair of a signal and a type distribution.

Before stating the analogous theorem, we argue that the equilibrium set in this generalized setting depends on the type distribution only through the function

$$\begin{aligned} & \Phi\Big(U(0;\eta), U(1;\eta), l\Big) \\ = \quad & \Pr\Big(\{(p_i, \boldsymbol{t_i}) : p_i\Big(t_i(1) - \frac{1}{2}\Big)U(1;\eta) \leq (1-p_i)\Big(\frac{1}{2} - t_i(0)\Big) \cdot l \cdot U(0;\eta)\}\Big) \end{aligned}$$

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 $U(\omega; \eta)$;[44] and the average effects are a sufficient statistic for the agents' best reply, given that a type $\Big(p_i, t_i(0), t_i(1)\Big)$ prefers the 1-action if and only if

$$\begin{aligned} p_i\Big(t_i(1) - \frac{1}{2}\Big) \frac{\Pr(s_i = s|\omega = 1)}{\Pr_i(s_i = s)} U(1; \eta) \\ -(1 - p_i)\Big(\frac{1}{2} - t_i(0)\Big) \frac{\Pr(s_i = s|\omega = 0)}{\Pr_i(s_i = s)} U(0; \eta) \geq 0. \end{aligned} \tag{58}$$

In conclusion, $\mathbf{q}(\sigma)$ and $(\bar{k}, \tilde{x})$ are a sufficient statistic for the best reply correspondence, which yields the claimed equiibrium characterization.

Second, multiplying (58) by $\frac{\Pr(s_i = s)}{\Pr(s_i = s|\omega = 0)}$ 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.

For the statement of the analogous theorem, recall the definition of a monotone type distribution from the main text; cf. (14).

Theorem 7. *Any majority vote over excluding the maximal policy has a payoff guarantee of* $1 - \varepsilon$ *across all agents' information structures with monotone type distributions and is robust optimal.*

Proof. The proof structure follows that of Theorem 1's proof in Section 1.7. The previous proof combined Theorems 2 and 3, and Proposition 1. The identical proof applies, given the analogs of these three results, which we will derive in the following. Theorem 2 showed existence of an information structure and a corresponding inefficient equilibrium sequence. It remains true since we only expanded the set of feasible information structures. The analog of Proposition 1 is stated and proven in Subsection E.1, and the analog of Theorem 3 is stated and proven in Subsection E.2. □

E.1 Analog of Proposition 1

We state Proposition 1's analog for the setting with ex-post heterogeneous preferences.

Proposition 2. *Take any non-constant, monotone process and any agents' information structure with a monotone type distribution. For any* N*, there is an equilibrium that satisfies the tie-breaking rule.*

[44] Cf. (1) and (3).

Proof. We mimic the original proof up to an adjustment of the argument showing that the agents have a monotone best reply given any monotone strategy profile: The proof identifies a set of candidate strategy profiles that ensure the partisans tie-breaking rule and constructs equilibria as fixed points in this set.

The candidate strategy profiles are as follows. The agents use "monotone" strategies σ, i.e., strategies under which the mean action is (weakly) higher in state 1 than in state 0, and where the partisans choose their actions in accordance with their types:

$$q(0;\sigma) \leq q(1;\sigma), \text{ and } \sigma(y,s) = y \text{ for all } s \in [0,1] \text{ and all } y \in \{0,1\}. \tag{59}$$

The principal mixes over "monotone" strategies, defined as those in which the policy choice $x = x(k)$ is weakly increasing in the number of observed 1-actions k. That is, $x(k) \geq x(k')$ for all $k, k' \in \{0, \ldots, N\}$ with $k' > k$.

For the fixed-point argument, we consider a modification of the best-response correspondence in which the set values are truncated to the set of candidate strategy profiles, denoted by Σ. We verify that this modification meets the requirements of Kakutani's fixed point theorem.

First, the modified correspondence inherits certain properties of the best-response correspondence: Its graph is the intersection of the graph of the best-response correspondence, which is closed, with $\Sigma \times \Sigma$, which is also closed (recall that Σ is the closed set of candidate strategy profiles). In addition, the modified correspondence is convex-valued since the monotonicity properties (i.e., $x(k) \geq x(k')$ for all $k' > k$ and (59)) are preserved by mixtures.

To establish the non-emptiness, we verify that the principal has a monotone best response and that the agents have a best response satisfying (59). First, observe that (59) implies that the principal's posterior is weakly increasing in the number of 1-actions. Thus, she has a monotone best response. Second, the monotonicity of the principal's strategy implies that the average effect of an additional 1-action is positive in each state, i.e., $U(0;\eta) \geq 0$ and $U(1;\eta) \geq 0$. Given the best-response characterization (58), this immediately implies that it is a best response for each partisan to match his action to his type. Therefore there is generally a best response satisfying the right condition of (59). A case analysis shows that in addition we can meet the left condition of (59):

- If $U(0;\eta) > 0$ and $U(1;\eta) = 0$, all types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) < 0$ choosing $x = 0$, all types for which $p_i\Big(t_i(0) - \frac{1}{2}\Big) > 0$ choosing $x = 1$, and all partisans choosing their weak preference is a best response with $q(0;\sigma) = q(1;\sigma)$.

- If $U(0;\eta) = 0$ and $U(1;\eta) > 0$, all types for which $(1-p_i)\Big(t_i(1) - \frac{1}{2}\Big) < 0$ choosing $x = 0$, all types for which $(1-p_i)\Big(t_i(1) - \frac{1}{2}\Big) > 0$ choosing $x = 1$, and all partisans choosing their weak preference is a best response with $q(0;\sigma) = q(1;\sigma)$.

- If $U(1;\eta) > 0$ and $U(0;\eta) > 0$, all partisans choosing their weak preference, and all non-partisans choosing according to (58) is a best response satisfying $q(0;\sigma) < q(1;\sigma)$, given the type distribution's monotonicity.

- If $U(0;\eta) = U(1;\eta)$, all agents' strategies are a best response. In particular, there is a best response meeting both conditions of (59), including $q(\sigma;0) \leq q(\sigma;1)$.

□

E.2 Analog of Theorem 3

Theorem 3's statement about information aggregation holds in this setting, with the generalized notion of the agents' information structures as pairs of signal and type distributions.

Theorem 8. *Consider any monotone process of partial commitments with a single cutoff and any agents' information structure with a monotone type distribution. Information aggregates in all equilibrium sequences if and only if the process has no balance and* $\max P(0) < \max P(1)$.

The proof is below. Some parts of it are verbatim to the original proof, and therefore not repeated, only highlighted.

Proof. To prove Theorem 3's statement in the generalized setting, the first three proof parts (the arguments in the main text, Claim 5, and Claim 6) are essentially identical and not repeated here.

They establish the necessity of the theorem's conditions for information aggregation. They also show that any equilibrium η of a process satisfying the conditions is (a) informative, (b) either $U(\omega';\eta) > 0$ for all $\omega \in \{0,1\}$ or $U(\omega';\eta) < 0$ for all $\omega \in \{0,1\}$. Thus, $\frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0,\infty)$.

The last part of Theorem 3's proof, Claim 7, starts from $\frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0,\infty)$ and establishes that any equilibrium sequence of a process satisfying the conditions aggregates information. We prove the following analog of Claim 7:

Claim 8. *Take any monotone process with a single cutoff and no balance for which* $\max P(0) < \max P(1)$*. Take any agents' information structure with a monotone type distribution. Any equilibrium sequence aggregates information.*

The detailed version of the proof is below. Here is a quick summary: The previous proof established interior limits for the indifferent agent types. We establish the equivalent claim that $\lim_{N\to\infty} \frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0,\infty)$; cf. (11). This observation yields strictly different mean actions, i.e. $q(0;\sigma) \neq q(1;\sigma)$ whenever the type distribution is monotone, cf. (14). Since the realized collective action is almost surely close to the mean action in each state, the principal learns the state from observing it. That is, information aggregates.

Proof. Take any equilibrium sequence $(\eta_N)_{N\in\mathbb{N}}$ of a process meeting the conditions in the claim.

We just argued in the section above that $\frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0,\infty)$ for all N and that $\lim_{N\to\infty} \frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0,\infty)$ implies information aggregation. In the following, we consider the counter-scenario, where[45]

$$\lim_{N\to\infty} \frac{U(0;\eta_N)}{U(1;\eta_N)} \in \{0,\infty\}. \tag{60}$$

The arguments are based on a detailed analysis of point events for the realized number k_{-i} of 1-actions 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 (16) 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 (17) to obtain

$$\begin{aligned}\frac{\Pr(\omega = 1 | k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0 | k_{-i} = m_N N; \sigma_N, N)} &= \frac{\Pr(\omega = 1)}{1 - \Pr(\omega = 1)} \\ &\cdot \exp\Big((N-1)\Big(\mathrm{KL}\Big(m_N, q(0,\sigma_N)\Big) - \mathrm{KL}\Big(m_N, q(1,\sigma_N)\Big)\Big)\Big).\end{aligned} \tag{61}$$

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 pre-

[45] It is sufficient to derive a contradiction for any subsequence where $\frac{U(0;\eta_N)}{U(1;\eta_N)}$ is converging to 0 in the extended reals, and for any subsequence where $\frac{U(0;\eta_N)}{U(1;\eta_N)}$ is converging to ∞. We identify the subsequence with the original sequence to omit the subsequence notation.

liminary observations: First, as long as $k_N \neq N$ for all N large enough, we have

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in (\kappa, \frac{1}{\kappa}) \text{ for } m_N \in \{m''_N, m'''_N\}, \tag{62}$$

for some $\kappa > 0$, by the defining property (1) of $\bar{k}_N$ (whenever we apply (54), we will rule out the case $k_N = N$). Second, (60) implies

$$\lim_{N\to\infty} \frac{\Pr(\omega = 1|k_{-i} = m_N N; \sigma_N, N)}{\Pr(\omega = 0|k_{-i} = m_N N; \sigma_N, N)} \in \{0, \infty\} \text{ for } m_N = m'_N. \tag{63}$$

Otherwise the inference from piv_1 would be bounded as $N \to \infty$. Then, for any N large enough, we would have either $k_N = N$, so that only piv_1 would be relevant for the agents' best response, or, by (62), the inference from piv_0 would also be uniformly bounded. In either case, the ratio $\frac{U(0;\eta_N)}{U(1;\eta_N)}$ would be bounded away from 0 and 1.[46] Third, (60) 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: $m_1 \neq q^*$.[47]

In this case we derive a contradiction to (60). 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 (19) applies, so that

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

For $m_N = m'_N$, the unbounded inference (63) implies

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

We show case by case that

$$\lim_{N\to\infty} m''_N = \lim_{N\to\infty} m'''_N = q^*. \tag{64}$$

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

[46] The average effect ratio can be expressed via (11). Given (62) and (63), boundedness of the ratio 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 $\min(\frac{a}{b}, \frac{c}{d}) \leq \frac{au+cv}{bu+dv} \leq \max(\frac{a}{b}, \frac{c}{d})$.

[47] In the main text we assert that this is a generic case. This is true because (52) implies $q^* \in \{F(0), F(1^-), 1 - F(0), 1 - F(1^-)\}$.

see that for any $\gamma > 0$,

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

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 (64) 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 (62), this implies interior limits, $\lim_{N\to\infty} \frac{U(0;\eta_N)}{U(1;\eta_N)} \in (0, \infty)$ (cf. (3) and (11)), contradicting the initial assumption (60).

Case 2: $m_1 = q^*$.

This case can be broken down into several subcases, all of which are analogous; we present only 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, $U(1; \eta_N) < 0$. In the previous section, we remarked that the same proofs as before establishes $\frac{U(0;\eta_N)}{U(1;\eta_n)} \in (0, \infty)$, so that (58) pins down a unique the agents' best response. Recall that $\rho(0)$ is the likelihood of an agent being a 0-partisan and $\rho(1)$ the likelihood of an agent being a 1-partisan. We see that $U(1; \eta_n) < 0$ and monotonicity imply $\rho(0) < q(1; \sigma_n) < q(0; \sigma_N) < 1 - \rho(1)$. Recall that $\frac{U(0;\eta_N)}{U(1;\eta_N)} \to 0$ or $\frac{U(0;\eta_N)}{U(1;\eta_N)} \to 1$ by assumption (60). If $\frac{U(0;\eta_N)}{U(1;\eta_N)} \to 1$, then $q^* = 1 - \rho(1)$. However, then $\lim_{N\to\infty} \Pr(\omega = 1|\mathrm{piv}_1; \sigma_N, N) = 1$ cannot hold. Thus, $\frac{U(0;\eta_N)}{U(1;\eta_N)} \to 0$, which implies $q^* = \rho(0)$.

We now carefully examine the mean actions in each state,

$$q(\omega'; \sigma_N) - m_1 = \sum_{s=0,1} \Pr(s_i = s|\omega = \omega')\Big(\Phi\Big(z_1(s), z_2\Big) - \rho(0)\Big),$$

with $z_1(s) = \frac{U(0;\eta_N)\Pr(s_i=s|\omega=0)}{U(1;\eta_n)\Pr(s_i=s|\omega=1)}$ and Z_2 the sign of $U(1, \eta)$. Using simple algebra,[48]

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

we see that

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

which implies

$$\lim_{N\to\infty} \frac{q(\omega';\sigma_N)-m_1}{-\Delta_N} \in (0,\infty) \tag{65}$$

for all ω'.[49] Using the approximation $q(\omega';\sigma_N) \approx q^*$, we restate the quadratic approximation (18) 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:

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

For $m_N = m'_N$, the unbounded inference (63) together with (65) 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 1-actions 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 almost surely

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

hence, almost surely

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

$\max\left(\frac{a}{b}, \frac{c}{d}\right)$.

[49]The inequalities hold for any subsequence along which $\frac{q(1;\sigma_N)-m_1}{q(0;\sigma_N)-m_1}$ converges. For simplicity of notation, we replace the original sequence by such a subsequence.

Given (61), this means the principal learns the state almost surely: $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N,N) = 1$ if $\omega = 1$ and $\lim_{N\to\infty} \Pr(\omega = 1|k_N;\sigma_N,N) = 0$ if $\omega = 0$. $\square$

E.3 Agent-Optimal Processes

We provide sufficient conditions for the majority voting over exclusion of the maximal policy (the processes defined by (4)) to be agent-optimal processes. 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

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

The first condition is

$$0 \le \mathrm{E}_G\Big(t_i(0)\Big) < \frac{1}{2}, \text{ and } 1 \ge \mathrm{E}_G\Big(t_i(1)\Big) > \frac{1}{2}$$

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 $\omega = 0$ and higher ones in $\omega = 1$). This means, the agents' full information payoff is

$$\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(1)\Big) - \frac{1}{2}\Big)$$

and the agents' payoff guarantee is

$$\hat{G}(P) := \inf_{(\eta_N)_{N\in\mathbb{N}},\pi} \Big(\lim_{N\to\infty} \inf \mathrm{E}(x \mid \omega = 1; \eta_N) - \frac{\mathrm{E}_F(1-p_i)\Big(\frac{1}{2} - \mathrm{E}_G\Big(t_i(0)\Big)\Big)}{\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(1)\Big) - \frac{1}{2}\Big)} \mathrm{E}(x \mid \omega = 0; \eta_N)\Big)$$

It differs from the principal's payoff guarantee simply by replacing $\Pr(\omega = 0)$ with $\mathrm{E}_F(1-p_i)\Big(\frac{1}{2} - \mathrm{E}_G\Big(t_i(0)\Big)\Big)$ and $\Pr(\omega = 1)$ with $\mathrm{E}_F(p_i)\Big(\mathrm{E}_G\Big(t_i(0)\Big) - \frac{1}{2}\Big)$.

The second condition is

$$\frac{1-\mathrm{E}_F(p_i)}{\mathrm{E}_F(p_i)} \cdot \frac{\frac{1}{2} - \mathrm{E}\Big(t_i(0)\Big)}{\mathrm{E}\Big(t_i(1)\Big) - \frac{1}{2}} \ge \varepsilon. \tag{66}$$

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.[50]

[50] The payoff from choosing $x = 1$ in both states divided by the full-information payoff is

Given the two conditions, mimicking Section 1.7's proof of Theorem 1 shows that the processes (4) maximize the agents' payoff guarantee. □

F Proof of Theorem 5

The proof structure follows that of Theorem 1's proof in Section 1.7. The previous proof combined Theorems 2 and 3, and Proposition 1. The identical proof applies, given the analogs of these three results, which we will derive in the following.

Before doing so, we provide some auxiliary results.

Claim 9. *Let $\frac{c'(x)}{b'(x)}$ be weakly decreasing in x.*

1. *If $u(x,p) = u(x',p)$ for some $p \in [0,1]$ and $x < x'$, then*

   i. *$u(x'',p') \leq (<) u(x''',p')$ for any $p' \geq (>) p$, $x''' \geq x'$, and $x \leq x'' < x'''$.*

   ii. *$u(x'',p') \geq (>) u(x''',p')$ for any $p' \leq (<) p$, $x'' \leq x$, and $x'' < x''' \leq x'$.*

2. *If $u(x,p) < u(x',p)$ for some $p \in [0,1]$ and $x < x'$, then, $u(x'',p') < u(x''',p')$ for any $p' \geq p$, $x''' \geq x'$, and $x \leq x'' < x'''$.*

3. *If $u(x,p) > u(x',p)$ for some $p \in [0,1]$ and $x < x'$, then, $u(x'',p') > u(x''',p')$ for any $p' \leq p$, $x'' \leq x$, and $x'' < x''' \leq x'$.*

Proof. The first item: By definition, $u(y,p) = u(y,0)(1-p) + u(y,1)p$ for all $y \in \mathcal{P}, p \in [0,1]$. Since $u(y,0) > u(y',0)$ and $u(y,1) < u(y',1)$ for all $y < y'$, the equality $u(x,p) = u(x',p)$ implies

$$
\begin{aligned}
&u(x,p') < u(x',p') \ \text{ for } \ p' > p, \\
&u(x,p') > u(x',p') \ \text{ for } \ p' < p.
\end{aligned}
$$

Then, the first claim follows from the single-basin property of the preferences $u(y,p')$. The second and third item: The inequality $u(x,p) < u(x',p)$ implies $u(x,p') < u(x',p')$ for any $p' \geq p$. The second claim follows again from the single-basin property. The inequality $u(x,p) > u(x',p)$ implies $u(x,p') > u(x',p')$ for any $p' \leq p$. The third claim follows again from the single-basin property. □

Based on Claim 9, we establish a critical "monotone comparative statics" result. We show that principal has a monotone best reply to any monotone strategy profile; see Section 1.6 for the definition of monotone strategies:

$1 - \frac{\mathrm{E}_F(1-p_i)\left(\frac{1}{2} - \mathrm{E}_G\left(t_i(0)\right)\right)}{\mathrm{E}_F(p_i)\left(\mathrm{E}_G\left(t_i(1)\right) - \frac{1}{2}\right)}$, which is smaller than $1-\varepsilon$ if (66) holds.

Take any monotone process P and any monotone strategy profile η_N. Note this implies that $\max P(m)$ and $\min P(m)$ and $\Pr(\omega = 1|k; \eta_N)$ are all weakly increasing. Denote $p(k) = \Pr(\omega = 1|k; \eta_N)$, $x(k) = \min P(\frac{k}{N})$, and $x'(k) = \max P(\frac{k}{N})$. An application of Claim 9 yields: Either $u\Big(x(k), p(k)\Big) > u\Big(x'(k), p(k)\Big)$ for all $k \in \{0, \dots, N\}$, or there is $\bar{k} \in \{0, \dots, N\}$ so that $u\Big(x(k), p(k)\Big) > u\Big(x'(k), p(k)\Big)$ for all $k < \bar{k}$ and $u\Big(x(k), p(k)\Big) \leq u\Big(x'(k), p(k)\Big)$ for all $k \geq \bar{k}$. In any case, this means the principal has a best reply where he chooses $x(k)$ if k is so that $u\Big(x(k), p(k)\Big) > u\Big(x'(k), p(k)\Big)$ and $x'(k)$ otherwise.

We note the following cut-off property that will be useful for proving an analog of Theorem 2: When the agents' strategy is informative, i.e., when $q(0; \sigma_N) < q(1; \sigma_N)$, the posteriors $p(k)$ are strictly increasing in k. Thus, when there is k so that $u\Big(x(k), p(k)\Big) = u\Big(x'(k), p(k)\Big)$, the lemma implies strict preferences between $x(k')$ and $x'(k')$ for any $k' \neq k$. Thus, the principal has two cut-off best replies, one where she chooses $x'(k')$ if $k' < k$ and $x(k')$ otherwise and one where she chooses $x'(k')$ if $k' \leq k$ and $x(k')$ otherwise.

Now, we sketch the proof of the analogs of Theorems 2 and 3 and Proposition 1.

Theorem 9. *(Analog of Theorem 3)*
Let $\frac{c'(x)}{b'(x)}$ be constant in x or strictly decreasing in x and $\Pr(\omega) > \bar{p}$. Then, for any process (4) *and any agents' information structure, information aggregates in all equilibrium sequences.*

Proof. In the main text, we argued that, for all monotone equilibria of the processes (4), the average effect $U(\omega; \eta)$ is strictly positive in each state, and thus, that there are unique interior cutoff types $p_N(s) \in (0, 1)$. Based on that, one proves the claim of information aggregation by mimicking Claim 7's proof verbatim. $\square$

Analog of Theorem 2. Theorem 2 shows existence of an information structure and a corresponding inefficient equilibrium sequence. The theorem's conclusion remains true and the identical proof applies, given the above observation about the existence of the two cut-off best replies.

Analog of Proposition 1. To prove Proposition 1, we constructed equilibria in monotone strategies. The proposition's conclusion remains true, in the setting with heterogeneous ex-post preferences, and the identical proof applies, with one adjustment: As before, we consider a modified best-response correspondence: It maps any monotone strategy profile to the intersection of its set of best-response profiles with the set Σ of monotone strategy profiles. The proof then verifies that the requirements of Kakutani's fixed point theorem are met. As before, the modified best response has a closed graph—the graph is the intersection of the best-response correspondence's closed graph with $\Sigma \times \Sigma$, which is also closed; it is convex-valued since the monotonicity of the strategies is preserved by mixtures. Finally, the argument showing it is non-empty has to be adjusted. For this, we have to argue that all players have a monotone best reply. For the principal, we did so above already. For the agents, any best reply is monotone as a consequence of $U(\omega'; \eta) > 0$ for $\omega' = 0, 1$ and the best response characterization (11).

G Existence of Efficient Equilibria for Processes with a Single Cutoff

Theorem 10 demonstrates that many processes with a single cutoff $m_1 \in (0, 1)$ generally have efficient equilibrium sequences—i.e., the equilibrium sequence achieves full-information payoffs, as $N \to \infty$. The following conditions are sufficient: The processes have increasing minimum and maximum policies and do not exclude the ex-post optimal policies.

Theorem 10. *Take any process with a single cutoff. If* $\min P(0) < \min P(1)$, $\max P(0) < \max P(1)$, $\min P(0) = 0$, *and* $\max P(1) = 1$, *given any agents' information structure, there exists an equilibrium sequence* $(\eta_N)_{N \in \mathbb{N}}$ *for which*

$$\lim_{N \to \infty} \Pr(x = \omega | \eta_N, N) = 1.$$

Proof. The equilibrium strategies are found among a parametric set of candidate strategies σ_L where an agent with signal $s \in \{0, 1\}$ chooses $a_i = 1$ if and only if $p_i \geq p_L(s)$ with $p_L(s)$ solving

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

Note that $p_L(s)$ is increasing in L. Hence, the mean action in each state, $q(\omega; \sigma_L)$ is strictly decreasing in L. We consider a compact set of parameters $L \in [\underline{L}_N, \bar{L}_N]$, with the parameter bounds implicitly given by the equations

$$q(0; \sigma_{\underline{L}_N}) = \frac{\lfloor m_1 N \rfloor}{N}, \text{ and} \tag{67}$$

$$q(1; \sigma_{\bar{L}_N}) = \frac{\lfloor m_1 N \rfloor}{N}. \tag{68}$$

A preliminary result prepares the formal fixed-point argument that constructs the equilibrium sequence.

Claim 10. *Take any sequence $(L_N)_{N\in\mathbb{N}}$ with $L_N \in [\underline{L}_N, \bar{L}_N]$ for all $N \in \mathbb{N}$. The sequence of the cutoffs $(\bar{k}_N)_{N\in\mathbb{N}}$ of the principal's best response to $(\sigma_{L_N})_{N\in\mathbb{N}}$ satisfies $\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(0;\sigma_{L_N})\Big) = \lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(1;\sigma_{L_N})\Big) > 0$.*

Proof. The bound $L_N \in [\underline{L}_N, \bar{L}_N]$ implies $0 < \lim_{N\to\infty} q(0;\sigma_{L_N}) < \lim_{N\to\infty} q(1;\sigma_{L_N}) < 1$. An application of the law of large numbers thus shows that information aggregates. In particular, this rules out the boundary case $\bar{k}_N = N$ because the principal's posterior crosses her indifference belief $\frac{1}{2}$ at some collective action $\bar{k}+1$ with $\lim_{N\to\infty} q(0;\sigma_{L_N}) < \lim_{N\to\infty} \frac{\bar{k}_N}{N} < \lim_{N\to\infty} q(1;\sigma_{L_N})$.

Now, we suppose that

$$\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(0;\sigma_{L_N})\Big) \neq \lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}}{N}, q(1;\sigma_{L_N})\Big) \tag{69}$$

and derive a contradiction. Given (69), an application of (16) yields $\lim_{N\to\infty} \Pr(\omega = 1|\bar{k};\sigma_{L_N}, N) \in \{0,1\}$, but this contradicts the minimality of $\bar{k}+1$.

Finally, $\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(0;\sigma_{L_N})\Big) = \lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(1;\sigma_{L_N})\Big)$ and $\lim_{N\to\infty} q(0;\sigma_{L_N}) < q(1;\sigma_{L_N})$ together imply $\lim_{N\to\infty} \mathrm{KL}\Big(\frac{\bar{k}_N}{N}, q(\omega;\sigma_{L_N})\Big) > 0$ for $\omega \in \{0,1\}$. $\square$

We describe the fixed-point correspondence f. It is a mapping on the parameter space $[\underline{L}_N, \bar{L}_N]$: Take any $L \in [\underline{L}, \bar{L}]$, any principal's best reply $(\bar{k}, \tilde{x})$ to σ_L, and the strategy profile $\eta = \Big(\sigma_L, (\bar{k}, \tilde{x})\Big)$. We claim that $U(\omega';\eta) > 0$ for any ω' and large enough N, and prove this momentarily. Given (11), the agents' best reply to η is then the strategy $\sigma_{L'}$ with $L' = \frac{U(0;\eta)}{U(1;\eta)} \in (0,\infty)$. We consider the correspondence f that maps L to the set consisting of the projections $\min\Big(\max(\underline{L}_N, L'), \bar{L}_N\Big)$ of all such best-replies L'.

The proof of the claim $U(\omega';\eta) > 0$ relies on our assumptions for the process. Recall from Claim 10's proof that $0 < \lim_{N\to\infty} q(0;\sigma_{L_N}) < \lim_{N\to\infty} \frac{\bar{k}_N}{N} < \lim_{N\to\infty} q(1;\sigma_{L_N}) < 1$. This implies, (a) all collective actions $k \in \{0,\ldots,N\}$ are on path, (b) the principal's posterior $\Pr(\omega = 1|k;\eta_N, N)$ is weakly increasing, when N is large enough. The assumptions $\min P(0) < \min P(1)$ and $\max P(0) < \max P(1)$ then imply that (c) under any principal's best reply, the choice $x(k)$ is weakly increasing in k, and (d) $x(\bar{k}) < x(\bar{k}+2)$. Together (a), (c), and (d) imply the claim.

The fixed-point argument: Given the defining equation (67) for $\underline{L}_N$, for any N and $L_N = \underline{L}_N$,

$$\mathrm{KL}\Big(\frac{\lfloor m_1 N\rfloor}{N}, q(0;\sigma_{L_N})\Big) = 0 < \mathrm{KL}\Big(\frac{\lfloor m_1 N\rfloor}{N}, q(1;\sigma_{L_N})\Big).$$

Claim 10 together with the expressions (21) and (22) for the pivotal likelihoods implies $\lim_{N\to\infty} \Pr\Big(\omega = 1|\text{piv}; \sigma_L, (\bar{k}_N, \tilde{x}_N), N\Big) = 0$ for any sequence of principal's best replies $(\bar{k}_N, \tilde{x}_N)$ to σ_{L_N}. Consequently,

$$L'_N < \underline{L}_N \text{ for all } L'_N \in f(\underline{L}_N) \tag{70}$$

and N large enough. Conversely, for $L = \bar{L}_N$,

$$\text{KL}\Big(\frac{\lfloor m_1 N \rfloor}{N}, q(1;\sigma_L)\Big) = 0 < \text{KL}\Big(\frac{\lfloor m_1 N \rfloor}{N}, q(0;\sigma_L)\Big).$$

Claim 10 together with (21) and (22) implies $\lim_{N\to\infty} \Pr\Big(\omega = 1|\text{piv}; \sigma_L, (\bar{k}_N, \tilde{x}_N), N\Big) = 1$ for any sequence of principal's best replies $(\bar{k}_N, \tilde{x}_N)$ to σ_L. Consequently,

$$L'_N < \bar{L}_N \text{ for all } L'_N \in f(\bar{L}_N) \tag{71}$$

and N large enough. Finally, an application of Kakutani's fixed point theorem yields a sequence of fixed points $\underline{L}_N < L^*_N < \bar{L}_N$ for which $L^*_N \in f(L^*_N)$.

Finally, we argue that any sequence of fixed points corresponds to an equilibrium sequence that achieves the full-information payoffs. First, given (70) and (71), a continuity argument implies $\lim_{N\to\infty} \underline{L}_N < \lim_{N\to\infty} L^*_N < \lim_{N\to\infty} \bar{L}_N$. This means that any fixed point L^*_N is interior when N is sufficiently large so that the corresponding sequence σ_{L_N} is a sequence of equilibrium strategies. Further, it implies

$$\lim_{N\to\infty} q(0;\sigma_{L^*_N}) < m_1 < \lim_{N\to\infty} q(1;\sigma_{L^*_N}). \tag{72}$$

Second, given (72), information aggregates and the principal's choice is $x = \min P(0) = 0$ in $\omega = 0$ and $x = \max P(1) = 1$ in $\omega = 1$, with probability converging to 1. We conclude that the equilibrium sequence achieves the full-information payoffs. $\square$

H Proof of Theorem 6

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

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

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

The equilibrium strategy is found among a parametric set of candidate strategies σ_L where an agent with signal $s \in \{0,1\}$ chooses $a_i = 1$ if and only if $p_i \geq p_L(s)$

with $p_L(s)$ solving

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

We will consider a compact set of parameters $L \in [\underline{L}, \bar{L}]$ with the parameters given by

$$\underline{L} = \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \frac{\Pr(s_i = 1|\omega = 0)}{\Pr(s_i = 1|\omega = 1)} < \frac{\Pr(\omega = 0)}{\Pr(\omega = 1)} \frac{\Pr(s_i = 1|\omega = 1)}{\Pr(s_i = 1|\omega = 0)} = \bar{L}. \tag{74}$$

These bounds guarantee that

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

for any principal's best reponse $(\bar{k}, \tilde{x})$ to σ_L which one verifies from its characterization (1).

A preliminary result prepares the fixed-point argument that we will use to construct the equilibria.

Claim 11. *There is $\bar{q} > m_R$ so that, for any distribution of the agent's priors with $1-F(\bar{p}) > \bar{q}$, there are $\bar{N}$ and $\bar{\delta} > 0$ for which the following holds: For any $N \geq \bar{N}$, and σ_L with $L \in [\underline{L}-\bar{\delta}, \bar{L}+\bar{\delta}]$, and any principal's best response $(\bar{k}_N, \tilde{x}_N)$ to σ_L, the agents' best response to σ_L and $(\bar{k}_N, \tilde{x}_N)$ is a strategy $\sigma_{L'}$ with $L' \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$.*

Proof. Throughout the proof, we fix an information structure with $1-F(\bar{p}) > \bar{q}$ for some $\bar{q} > m_R$.

Step 1. *There is $\delta > 0$ so that $q(\sigma_L; \omega) > \bar{q}$ for all $\omega \in \{0, 1\}$ and $L \in [\underline{L}-\delta, \bar{L}+\delta]$.*

Take any $L \in [\underline{L} - \delta, \bar{L} + \delta]$. Since the cutoffs $p_L(s)$ are increasing in L,

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

In turn, since $q(0; \sigma_L)$ is strictly decreasing in $p_L(s)$ and since $p_L(1) < p_L(0)$,

$$q(\sigma_L; \omega) \geq 1 - F\Big(p_{\bar{L}+\delta}(0)\Big) \text{ for all } \omega \in \{0, 1\}.$$

Note that the definitions (15) and (74) are such that $p_{\bar{L}}(0) = \bar{p}$. Given $1-F(\bar{p}) > \bar{q}$, there is $\delta > 0$ so that $1 - F(p_{\bar{L}+\delta}) > \bar{q}$, and thus $q(\sigma_L; \omega) > \bar{q}$ for all $\omega \in \{0, 1\}$, as claimed.

Step 2. *Fix $L \in [\underline{L}, \bar{L}]$. For any sequence of principal's best responses $(\bar{k}_N, \tilde{x}_N)$ to σ_L,*

$$q(0; \sigma_L) < \lim_{N \to \infty} \frac{\bar{k}_N}{N} < q(1; \sigma_L) \tag{75}$$

First, $q(0; \sigma_L) < q(1; \sigma_L)$ follows from the definition of σ_L, the full support of the prior distribution and the different likelihood ratios of the private signals in the two states. Second, suppose the inner inequalities do not hold. 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 that (16) would imply $\lim_{N\to\infty} \Pr(\omega = 1 | \bar{k}_N; \sigma_L, N) \in \{0, 1\}$. But $\Pr(\omega = 1 | \bar{k}_N; \sigma_L, N)$ is close to the principal's prior, given (1) and since a single agent's action is boundedly informative about the state. We arrive at a contradiction.

Step 3. *There is $\bar{q} \in \Big(m_R, F(1^-)\Big)$ so that when $q(0; \sigma_L) \geq \bar{q}$, then*

$$\mathrm{KL}\Big(m_R, q(\omega; \sigma_L)\Big) > \mathrm{KL}\Big(m_0, q(0; \sigma_L)\Big) \text{ for any } \omega \in \{0, 1\}. \tag{76}$$

where $m_0 = \lim_{N\to\infty} \frac{\bar{k}_N}{N}$ is the limit of the principal's best reply cutoff $\bar{k}_N$ given by (1).

For $q(0; \sigma_L) \approx F(1^-)$, the ordering (75) implies $\mathrm{KL}\Big(m_0, q(\sigma_L; \omega)\Big) \approx 0$ for $\omega \in \{0, 1\}$. However, $\mathrm{KL}(m_R, q(0; \sigma_L)) > 0$ since $m_R < F(1^-)$. Given the continuity of the Kullback–Leibler divergence, (76) holds if $q(0; \sigma_L) \geq \bar{q}$ for some large enough $\bar{q} \in \Big(m_R, F(1^-)\Big)$.

Step 4. *Take $\delta > 0$ from Step 1 and $\bar{q}$ from Step 3. There is $\bar{N} \in \mathbb{N}$ so that, for any $N \geq \bar{N}$, the following holds : For any $L \in [\underline{L} - \delta, \bar{L} + \delta]$ and any principal's best response $(\bar{k}_N, \tilde{x}_N)$, the agents' best response to σ_L and $(\bar{k}_N, \tilde{x}_N)$ is a strategy $\sigma_{L'}$ with $L' \in [\underline{L} - \delta, \bar{L} + \delta]$.*

Take any $L \in [\underline{L} - \delta, \bar{L} + \delta]$. Our assumptions imply that (76) holds. The key implication of (76) is that the agents' best response is dominated by the incentive to influence the principal's policy preference, i.e.,

$$\lim_{N\to\infty} \Pr(\mathrm{piv}_0 | \mathrm{piv}; \sigma_L, N) = 1 \tag{77}$$

This implication directly follows from the expression for the likelihood of the pivotal events in terms of the Kullback–Leibler divergence, (21) and (22).

Given (77), for any sequence of principal's best responses $(\bar{k}_N, \tilde{x}_N)$ to σ_L, the characterization (11) of the agents' best response implies it is a strategy $\sigma_{L'_N}$ with

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

a $N \to \infty$. Since the bounds $\underline{L}$ and $\bar{L}$ were constructed to contain the posterior likelihood ratio conditional on piv_0, we conclude there is $\bar{N}$ so that for any $N \geq \bar{N}$, $\underline{L} - \delta \leq L'_N \leq \bar{L} + \delta$.

□

Claim 11 establishes a correspondence from $L \in [\underline{L} - \bar{\delta}, \bar{L} + \bar{\delta}]$ to subsets of $L' \in [\underline{L}, \bar{L}]$. This correspondence is the mapping that takes σ_L, sends it to all principal's best responses, and subsequently to all agents' best responses $\sigma_{L'}$ (to σ_L and a given principal's best response). One verifies that this correspondence has a closed graph, and non-empty, compact, convex images. An application of Kakutani's fixed point theorem implies a fixed point L^*_N for any $N \geq \bar{N}$. The corresponding $\sigma_{L^*_N}$ is an equilibrium strategy and satisfies $\bar{q} < q(0; \sigma_{L^*_N}) < q(1; \sigma_{L^*_N})$; cf. Step 1. Since $m_R < \bar{q}$, there is $\gamma > 0$ so that $m_R + \gamma < \bar{q}$ holds. We conclude that we constructed an equilibrium with the desired property (73), which finishes the proof of Theorem 6.

I Interior Mean Actions Imply Trembling-hand Perfection

We show that any equilibrium $\eta = (\sigma, \tilde{x}, k)$ with interior mean actions, $q(\omega; \sigma) \in (0, 1)$ for $\omega \in \{0, 1\}$, is trembling-hand perfect (Selten, 1988). By definition, η is trembling-hand perfect if there exists a sequence of completely mixed strategy profiles $(\eta_k)_{k\in\mathbb{N}}$ with agent strategies $(\sigma_k)_{k\in\mathbb{N}}$ and the following properties:

(i) $(\eta_k)_{k\in\mathbb{N}}$ converges to η,

(ii) η is a best reply to η_k for all k.

It is easy to verify that interior mean actions imply that there is a completely mixed sequence $(\eta_k)_{k\in\mathbb{N}}$ for which (i) and

(iii) $q(\omega; \sigma_k) = q(\omega; \sigma)$ and $U(\omega; \eta_k) = U(\omega; \eta)$ for all $\omega \in \{0, 1\}$,

hold. Since the vector $\Big(q(\omega; \sigma), U(\omega; \eta)\Big)_{\omega\in\{0,1\}}$ is a sufficient statistic for the best-response correspondence if the mean actions in each state are interior, given (1) and (2), (ii) also holds. To conclude, η is perfect.

We already observed in Section 1.1 that the tie-breaking rule implies interior mean actions, $q(\omega; \sigma) \in (0, 1)$ for $\omega \in \{0, 1\}$. Hence, any equilibrium satisfying the tie-breaking rule is, in particular, trembling-hand perfect.