# The Party Line: Information Control and Collective Thinking in Democracy*

Olivier Bos†

March 14, 2026

**Abstract**

Political parties are indispensable institutions of modern democracy, enabling coordination, mobilization, and large-scale political participation. At the same time, they are often criticized for suppressing independent judgment and distorting public debate. This paper develops a formal theory of political parties as designers of information. Building on Simone Weil's critique of party organization, I study how discipline, hierarchy, and organizational incentives shape collective thinking. In a benchmark persuasion model, parties may optimally garble or pool information. Extending the analysis to organizational environments, I show how meetings, hierarchy, and strategic ignorance can suppress judgment even among informed actors. At the same time, parties can improve collective accuracy by lowering participation and communication costs. The normative implication is conditional: parties are essential to collective thinking when participation dominates control, and undermine it when control dominates participation.

*Keywords:* Political parties; information design; persuasion; collective judgment; democratic institution.

*I thank Antoinette Baujard and Michael Greinecker for comments and thoughtful discussions.

†Université Paris-Saclay, ENS Paris-Saclay, Centre for Economics at Paris-Saclay. E-mail: olivier.bos@ens-paris-saclay.fr.

# 1 Introduction

Democratic theory assigns a central epistemic role to political institutions. Elections, deliberation, and representation are commonly justified not only as mechanisms for aggregating preferences but also as means of aggregating dispersed information into collectively sound judgments. A functioning democracy is thus expected to transform private signals, expertise, and local knowledge into public decisions that are, on average, correct. Yet recent political developments have raised serious doubts about this epistemic capacity.

In the United States, the rise of Donald Trump revealed not only deep polarization among voters, but also a striking pattern of elite coordination. Elected officials and party leaders repeatedly endorsed claims they privately knew to be false, ranging from allegations of electoral fraud to assessments of public health and foreign policy. These episodes were not limited to marginal actors or uninformed constituencies, but involved highly educated and well-informed political elites. Such patterns are difficult to explain solely by voter misinformation, media bias, or individual irrationality. They instead suggest that the internal logic of political parties and partisan organizations may itself generate incentives to suppress truthful communication.[1]

Similar dynamics are visible across Europe, albeit in different institutional forms. In the United Kingdom, the Brexit process illustrated how party organizations compressed a highly complex and uncertain policy environment into rigid and coordinated public narratives, even as internal disagreement among experts and legislators was widely acknowledged. Strong party discipline and centralized leadership limited the public expression of nuanced positions, producing a sharp alignment between party affiliation and publicly articulated beliefs. In France, and more broadly in parliamentary systems with disciplined party voting, centralized control over political messaging similarly constrains the articulation of dissent on issues such as European integration, immigration, fiscal policy, and institutional reform. In these contexts, informed representatives often face strong incentives to align publicly with party positions, even when private assessments are more complex or uncertain.[2]

[1]The patterns described here are documented in a large political science literature on party discipline, elite cueing, and partisan conformity. In the United States, research shows that party affiliation and organizational incentives strongly shape elite behavior, often dominating private information or expertise (e.g., Barber and McCarty, 2015, Barber and Pope, 2019, Grossman and Hopkins, 2016). Studies of elite communication further document the persistence of publicly endorsed claims that conflict with available evidence, even among highly informed actors (e.g., Nyhan and Reifler, 2015, Bolsen et al., 2020). These findings motivate the theoretical mechanisms developed in this paper.

[2]Comparable dynamics are documented across European parliamentary systems. In the United Kingdom, research on the Brexit and post-Brexit period shows exceptionally high party discipline and the compression of complex and uncertain policy environments into unified elite narratives, despite acknowledged internal disagreement (e.g., Cowley, 2015, Cowley and Stuart, 2023, Spirling and McLean, 2024). In France, recent work emphasizes strong centralization of party leadership and systematic incentives for legislators to align

Taken together, these developments point to a common organizational puzzle: why do informed political elites repeatedly align on simplified, and sometimes misleading, narratives, even when they privately possess more accurate or nuanced information? This puzzle shifts attention away from voter irrationality and toward the internal incentives of political organizations.

This paper studies a simple but underexplored question: *do political parties undermine society's capacity for collective thinking, despite their central role in democratic coordination*? By collective thinking, I mean the ability of a political system to aggregate dispersed private information into accurate collective judgments.

The motivating intuition behind this question was articulated by Simone Weil in a short essay written in 1943 and published posthumously in 1950 (Weil, 1957), *Note sur la suppression générale des partis politiques*.[3] Weil advances a radical critique of political parties, arguing that they are organizations whose primary objective is their own expansion rather than truth-seeking or the resolution of public problems. According to Weil, once organizational growth becomes the dominant objective, internal conformity and collective passion come to dominate independent reasoning, and political speech is organized to sustain the party rather than to discover what is correct. As a result, both party members and citizens are prevented from forming reflective judgments about fundamental political issues.

Despite its influence in philosophical and normative debates, Weil's critique has not, to my knowledge, been formalized within modern political science or political economy, nor analyzed through the lens of organizational incentives and information design. I interpret Weil's claim not as a purely philosophical or moral critique, but as an organizational hypothesis about incentives and information. While Weil does not frame her argument in strategic or informational terms, her analysis can be understood as identifying a misalignment between organizational objectives and epistemic performance: internal discipline, moral pressure to conform, and the suppression of inconvenient truths emerge endogenously when parties prioritize their own development. Rather than assuming that parties simply transmit information, I model them as strategic organizations that actively shape what can be said, how it can be said, and what remains unsaid through sanctions, career incentives, and control over communication channels.

The central argument is that parties affect collective thinking through two distinct margins. On the *participation margin*, parties can strengthen democracy by lowering communication and mobilization costs, bringing more informative voices into the public sphere and

publicly with party positions, even when private assessments are more nuanced or conflicting (e.g., Costa and Frenkiel, 2023, Gougou and Persico, 2024). These references are illustrative rather than exhaustive.

[3]An English translation is available in Weil (2013).

improving inference from dispersed signals. On the *control margin*, the same organizational capacities can be deployed to discipline speech and design information: parties may rationally garble or pool public messages, impose correlation through caucuses and meetings, concentrate informational power via hierarchy, and even induce strategic ignorance. The paper characterizes conditions under which each margin dominates, yielding a conditional answer to Weil's critique.

Before turning to the analysis, I briefly relate my work to the formal literature that models political parties as solutions to coordination and collective action problems in elections and legislatures. Important contributions include Austen-Smith and Banks (1988) on elections and legislative outcomes with parties and coalitions, Morelli (2004) on endogenous party formation, Besley (1997) on candidate entry, and Snyder and Ting (2002), who provide an informational rationale for party organization. These works highlight the benefits of party organization for coordination, entry, and policy aggregation. My focus differs. Conditional on parties' existence, I study how internal organizational incentives shape information transmission, discipline, and the aggregation of beliefs, and how the capacity of parties to coordinate speech and information creates both epistemic benefits and epistemic failures. In this sense, parties are indispensable institutions for democratic coordination at scale, even before any epistemic concerns are considered. This paper therefore does not model parties primarily as vote aggregators or electoral coalitions, but as organizations that design the informational environments within which democratic judgments are formed.[4]

The paper proceeds as follows. Section 2 develops a baseline model in which political parties shape public beliefs by controlling how internally available information is communicated. This model isolates the control margin: even absent internal conflict or strategic disagreement, parties may optimally garble or pool information. Section 3 builds on this baseline by modeling parties as organizations composed of multiple members. It shows how correlated communication, hierarchy, and judgment collapse can suppress the use or expression of information that remains available inside the organization, while strategic ignorance operates at an earlier stage by discouraging the acquisition of information itself. Section 4 studies the welfare and institutional implications of these forces. It highlights a countervailing participation margin through which parties can improve collective accuracy by lowering communication and mobilization costs, and derives conditions under which regulation, rather than suppression, of parties restores collective thinking. Section 5 concludes. All proofs are

[4]The analysis abstracts from inter-party competition and focuses on the internal informational logic of a single party organization. This is not because electoral competition is irrelevant, but because informational control in modern democracies is primarily exercised within party organizations. Even in competitive party systems, politicians, activists, and candidates typically operate within a single organizational structure that governs communication, career incentives, and the expression of dissent.

collected in the Appendix (Section 6).

## 2 Benchmark: Party Control of Public Beliefs

This section introduces a deliberately simple benchmark that isolates the most basic mechanism emphasized by Simone Weil (1957): political parties may distort public understanding not because information is unavailable, but because organizations control how information is communicated. This channel corresponds to what I call the *control margin*: the capacity of party organization to shape, filter, and discipline political speech and information. The benchmark is intended not for realism but for clarity. It shows that even in the absence of internal dissent, strategic agents, or ideological conflict, party control over messaging can undermine collective thinking.

**Political environment.** There is a binary state of the world $\theta \in \{0, 1\}$ representing the truth about a fundamental political issue. For concreteness, $\theta = 1$ may represent that a given policy is in fact desirable, that an election result is legitimate, or that a public threat is real. The value $\theta = 0$ represents the opposite. Prior to any communication, the public assigns probability $\mu_0 \in (0, 1)$ to $\theta = 1$.

The public is modeled as a representative decision-maker who ultimately takes a binary action $a \in \{0, 1\}$. This action captures a collective stance: supporting or rejecting a policy, accepting or rejecting a claim, or coordinating on one of two political positions. The public's payoff is

$$u_R(a, \theta) = \mathbb{1}_{\{a=\theta\}},$$

so the public wishes to take the action consistent with the true state of the world. Consequently, after observing any message, the public chooses $a = 1$ if and only if its posterior belief that $\theta = 1$ is at least one half.

**Information and organizational control.** Inside the political system, truthful but imperfect information about $\theta$ exists. This information is summarized by a signal $s \in \{0, 1\}$ satisfying

$$\mathbb{P}(s = \theta \mid \theta) = q, \qquad q \in (1/2, 1).$$

The signal can be interpreted as expert knowledge, internal reports, or informed assessments available to political elites. Importantly, the public does not directly observe $s$.

A political party controls how this information is translated into public discourse. Formally, the party chooses a communication strategy $\pi(m \mid s)$ that maps the internal signal $s$

into a public message $m \in M$. Messages should be interpreted broadly: official statements, press releases, talking points, or endorsed narratives.

**Beliefs and Bayes plausibility.** Each message $m$ induces a posterior belief $\mu(m) = \mathbb{P}(\theta = 1 \mid m)$ for the public. Because the party does not observe $\theta$ directly and only controls communication, the distribution of posteriors must satisfy Bayes plausibility:

$$\sum_{m \in M} \mathbb{P}(m)\, \mu(m) = \mu_0.$$

This constraint reflects Bayes' rule. While the party can induce dispersion in posterior beliefs across messages, the average posterior, taken with respect to the distribution of messages, must equal the prior. Formally,

$$\mathbb{E}[\mu(M)] = \mathbb{P}(\theta = 1) = \mu_0.$$

That is, the party can shift beliefs across messages, but cannot change beliefs on average. Given a message $m$, the public chooses

$$a(m) = \begin{cases} 1 & \text{if } \mu(m) \geq 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

Throughout, probabilities and expectations over messages, actions, and public beliefs are taken with respect to the joint distribution induced by the party's information policy $\pi$. For notational simplicity, I omit the subscript $\pi$ except when comparing different policies in the proofs. Primitive probabilities, such as signal accuracy, are defined directly by the model.

**Party objectives.** The party values two things. First, it benefits from persuading the public to take action $a = 1$, which may correspond to electoral success, policy adoption, or reputational gain. Second, the party may value organizational cohesion, discipline, or message simplicity. I capture this by the objective

$$U_P(\pi) = \mathbb{P}(a = 1) + \lambda \cdot C(\pi),$$

where $\lambda \geq 0$ and $C(\pi) \equiv -H(M)$ is the negative entropy of the public message distribution with $H(M) = -\sum_{m \in M} \mathbb{P}(m) \log \mathbb{P}(m)$ denotes the Shannon entropy of the induced public message distribution (Shannon, 1948). Lower entropy corresponds to more unified, simpler, and more disciplined messaging.[5]

**Results.** Let us begin by setting $\lambda = 0$, so the party cares only about persuading the public. Let

$$v(\mu) = \mathbb{1}_{\{\mu \geq 1/2\}}$$

denote the party's payoff from inducing posterior belief $\mu$, reflecting that the public chooses $a = 1$ whenever $\mu \geq 1/2$. The party's problem is therefore to choose a distribution over posterior beliefs that maximizes $\mathbb{E}[v(\mu)]$ subject to Bayes plausibility.

**Theorem 1** (Optimal persuasion without cohesion.)**.** *Fix* $\lambda = 0$. *Consider the party's problem of choosing an information policy* $\pi(m \mid s)$ *that induces a distribution of posterior beliefs* $\mu(m) = \mathbb{P}(\theta = 1 \mid m)$ *for the public and maximizes* $\mathbb{P}(a = 1)$*, where the public chooses* $a = 1$ *iff* $\mu(m) \geq 1/2$*. Then:*

*(i) (*Reduction to posteriors*) The party's expected payoff depends on* $\pi$ *only through the induced distribution of posteriors* $\mu(m)$*.*

*(ii) (*Two-posteriors sufficiency*) There exists an optimal policy that induces at most two distinct posterior beliefs.*

*(iii) (*Strict garbling when $\mu_0 < 1/2$*) If* $\mu_0 < 1/2$*, any fully revealing policy (one that makes the posterior depend nontrivially on the signal) is weakly dominated by a policy that pools some information. In particular, full revelation is not optimal.*

Intuitively, because the public's decision rule features a single threshold at $\mu = 1/2$, the party has no incentive to induce more than two distinct posterior beliefs: one below and one above the threshold. Figure 1 illustrates this result.

Theorem 1 characterizes the structure of optimal belief manipulation when the party's sole objective is persuasion. Even in this benchmark, full revelation is generically suboptimal. Indeed, the party prefers to garble information in order to concentrate posterior beliefs just above the receiver's decision threshold. Proposition 1 shows how this logic is amplified once organizational cohesion enters the objective. When parties value unified messaging

[5] I formalize organizational cohesion as the negative entropy of the public message distribution. Entropy provides a parsimonious, reduced-form measure of message dispersion: it is minimized when the party communicates through a single, perfectly unified message (full pooling), and increases as communication becomes more fragmented or differentiated. Using Shannon entropy (Shannon, 1948) is natural in this context, as it is the canonical measure of uncertainty and dispersion over discrete distributions. The objective does not assume that parties explicitly optimize entropy, rather, it captures in reduced form the idea that parties value simple, unified, and disciplined messaging.

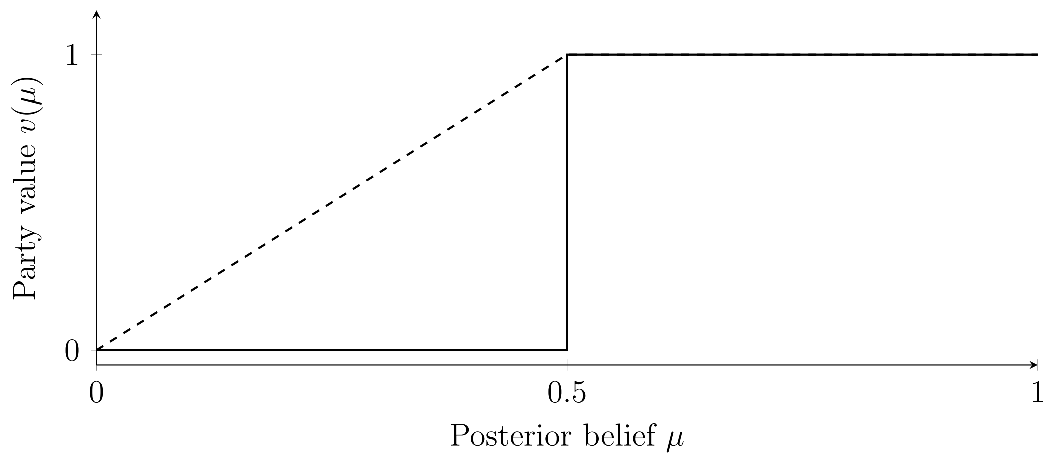


Figure 1: Benchmark persuasion: $v(\mu) = \mathbb{1}_{\{\mu \geq 1/2\}}$ (solid) and its concave envelope (dashed). The optimal signal places mass on at most two posteriors (tangency points).

sufficiently strongly, the optimal policy collapses to full pooling, eliminating public learning altogether. This illustrates the *control margin*: organizational incentives can dominate informational considerations and lead parties to suppress information even when it is available and informative.

When $\lambda > 0$, the party faces a tradeoff between persuading the public and maintaining organizational cohesion. Importantly, political communication is implemented through discrete and observable interventions (speeches, press releases, official posts), rather than through an abstract continuous device. We formalize this feature as follows.

**Assumption 1** (Finite implementation)**.** *The party communicates through $T < \infty$ discrete public interventions.*

Under this implementation structure, a communication strategy specifies how these $T$ interventions are allocated across messages. Hence any public message used with positive frequency must occur at least once, and therefore has probability at least $1/T$. Consequently, the set of admissible message distributions is finite.

**Proposition 1** (Pooling and the collapse of public learning)**.** *Under Assumption 1, there exists $\bar{\lambda} > 0$ such that for all $\lambda \geq \bar{\lambda}$, any optimal party policy is fully pooling (uses a single message with probability one). Consequently, public learning collapses: $\mu(m) = \mu_0$ for all on-path messages.*

Figure 2 illustrates the two canonical forms of information control in the benchmark. In the left panel, the party garbles information by mapping each private signal $s$ into public messages with noise level $\varepsilon \in (0, 1/2)$, so that each message equals the true signal with

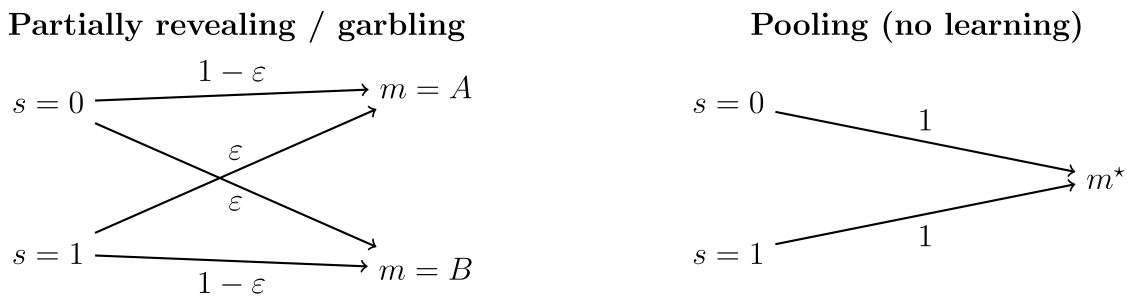


Figure 2: Two forms of information control: garbling (partially revealing communication) and pooling.

probability $1-\varepsilon$ and is flipped with probability $\varepsilon$, preserving some learning. In the right panel, the party fully pools information by sending a single message $m^\star$ regardless of the signal, eliminating public learning. When organizational cohesion is sufficiently valuable, the latter can be optimal even when private signals are informative. Figure 2 thus illustrates the extreme form of the control margin: when cohesion is sufficiently valuable, the party optimally sacrifices all public learning in exchange for maximal message discipline.

**Remark 1** (Robustness without discrete implementation)**.** *The finite-implementation assumption delivers a sharp threshold above which full pooling is exactly optimal. However, the underlying mechanism does not depend on discreteness. In a fully continuous communication environment, arbitrarily small message frequencies remain feasible, and a finite threshold need not exist. Yet, the force of the control margin survives: as the weight placed on cohesion increases, optimal communication becomes arbitrarily concentrated on a single message, and public learning becomes arbitrarily small. Thus the suppression of public learning is not an artifact of discreteness. It is a structural implication of organizational incentives.*[6]

This benchmark captures the simplest version of Weil's concern. Truthful information exists and is known inside the political system, yet organizational incentives lead the party to garble or even fully suppress information. Public discourse is shaped not to convey truth, but to maximize organizational objectives. Importantly, this failure arises even in the absence of internal conflict, strategic agents, or ideological polarization. Organization alone suffices to undermine collective thinking.

[6]The Appendix, Section 7, formalizes this robustness claim and establishes the convergence result under continuous communication.

# 3 Organization: Collective Belief Formation

Section 2 showed that even without internal conflict, a party that controls public messaging may optimally garble or pool information. This benchmark isolates the pure control margin in a persuasion environment. Section 3 extends the analysis by modeling parties not as single senders, but as organizations. Once organizational structure is introduced, the control margin expands beyond standard persuasion logic. The epistemic distortions identified in the benchmark become more severe when the party is modeled as a structure that (i) coordinates members' actions and speech, (ii) concentrates information in leadership, and (iii) can affect whether members become informed in the first place. These are precisely the organizational channels emphasized by Weil (1957), through which parties shape collective beliefs and conformity, and, as I show below following Arendt (1971, 1978), can ultimately suppress the expression of individual judgment even when information remains available.

As in Section 2, the binary state $\theta \in \{0, 1\}$ represents the truth about a fundamental political issue. The section begins with an environment in which party members receive symmetric private signals about $\theta$. The subsequent subsections then relax this information structure: hierarchy concentrates information in the leader, while strategic ignorance endogenizes whether members acquire information in the first place. The key difference is that the party does not merely choose a mapping from one signal to one message anymore. It chooses an *organizational information policy* that governs members' actions and public messaging, subject to feasibility constraints requiring obedience (members must be willing to follow what the party asks them to do).

**Bayes correlated equilibrium as a language for organizational feasibility.** I model the party as choosing an information-and-recommendation device that privately recommends actions to members and produces a public message. This representation follows the Bayes correlated equilibrium (BCE) concept defined by Bergemann and Morris (2016). BCE is useful here because it provides a compact way to describe the joint distribution over states, recommendations, and public messages that are implementable subject to obedience: each member must find it optimal to follow the recommendation he receives. In this sense, BCE serves as a language of organizational feasibility, abstracting from the details of internal communication while capturing the scope of implementable coordination. As emphasized by Taneva (2019), allowing for correlation across recommendations enlarges the set of implementable outcomes, making correlation an organizational instrument rather than a behavioral assumption.[7]

**Players and information.** There is a binary state $\theta \in \{0,1\}$ with prior $\mathbb{P}(\theta = 1) = \mu_0$. There are $n$ party members $i = 1, \ldots, n$. Each member privately observes a signal $s_i \in \{0,1\}$ with accuracy $q > 1/2$:

$$\mathbb{P}(s_i = \theta \mid \theta) = q,$$

independently across $i$ conditional on $\theta$.

After observing a public message $m \in M$, the public forms posterior $\mathbb{P}(\theta = 1 \mid m)$ and chooses $a \in \{0,1\}$ to maximize $u_R(a,\theta) = \mathbb{1}_{\{a=\theta\}}$. Hence $a = 1$ iff $\mathbb{P}(\theta = 1 \mid m) \geq 1/2$.

**Organizational actions and party policy.** Each member chooses an organizational action $x_i \in X$ (finite). Think of $x_i$ as: publicly endorsing the party line, dissenting, remaining silent, campaigning, etc. The party commits *ex ante* to an information policy

$$\varphi(r, m \mid \theta, s_1, \ldots, s_n) \in \Delta(X^n \times M),$$

where $r = (r_1, \ldots, r_n)$ is a profile of private recommendations to members (recommended actions), and $m$ is the public message. After observing $(s_i, r_i)$, each member chooses whether to obey $(x_i = r_i)$ or deviate. The party cannot force obedience; it can only choose policies that make obedience optimal.

In the organizational environments studied below, the public message $m$ is not an independent instrument chosen directly by the party. Rather, it is generated from members' obedient actions (for example, the total number of endorsements or the majority of expressed messages). The notation $\varphi(r, m \mid \theta, s_1, \ldots, s_n)$ describes the joint distribution induced by the policy and members' obedience, but the public only observes such aggregates of $x = (x_1, \ldots, x_n)$.

The party does not observe the state $\theta$ directly. The presence of $\theta$ in the argument of $\varphi$ reflects the fact that the joint distribution over $(\theta, s_1, \ldots, s_n)$ determines the induced distribution over recommendations and messages, but the party's policy can condition only on the information available within the organization. In particular, in environments with-

[7] I use BCE as a feasibility and implementability concept rather than as a behavioral equilibrium prediction. In the sense of Bergemann and Morris (2016), a BCE characterizes a joint distribution that can arise as a Bayes-Nash equilibrium outcome under some information structure, but BCE itself is not a Nash equilibrium of a fixed game. In contrast to applications that solve Bayes-Nash information design problems and rely on equilibrium selection (see, e.g., Taneva (2019) and Mathevet et al. (2020)), my analysis uses BCE only to characterize which patterns of information, speech, and public beliefs are organizationally implementable subject to obedience.

out a leader or internal communication channel, the policy cannot condition on any signal correlated with $\theta$.

The party does not observe the state $\theta$ directly. The presence of $\theta$ in the argument of $\varphi$ reflects the fact that the joint distribution over $(\theta, s_1, \ldots, s_n)$ determines the induced distribution over recommendations and messages, but the party's policy can condition only on the information available within the organization. In particular, in environments without a leader or internal communication channel, the policy cannot condition on any signal correlated with $\theta$.

**Obedience (BCE feasibility).** Let member $i$'s payoff be $u_i(x_i, r_i, s_i, \theta, r_{-i})$, which captures career rewards for conformity, sanctions for dissent, and moral costs of endorsing against one's signal. The policy $\varphi$ is feasible if every member prefers to obey.

**Definition 1** (BCE-feasible party policy)**.** *A policy $\varphi$ is BCE-feasible if for each member $i$, each on-path $(s_i, r_i)$, and each deviation $x_i' \in X$,*

$$\mathbb{E}[u_i(r_i, r_i, s_i, \theta, r_{-i}) - u_i(x_i', r_i, s_i, \theta, r_{-i}) \mid s_i, r_i] \geq 0.$$

The party values persuading the public and organizational cohesion/simplicity:

$$U_P(\varphi) = \mathbb{P}(a = 1) + \lambda \cdot C(\varphi) - \kappa \cdot S(\varphi).$$

Here $\lambda \geq 0$ measures the importance the party assigns to organizational cohesion, while $\kappa \geq 0$ captures the marginal cost of organizational complexity. The cohesion term $C(\varphi)$ is defined as in Section 2, where it plays a central role in generating full pooling when message discipline is sufficiently valuable. The complexity term $S(\varphi)$ represents reduced-form organizational or cognitive costs of implementing rich or finely contingent communication schemes, such as administrative, coordination, or attention costs. Its role is purely to rule out arbitrarily complex designs. For example, absent such a cost, the party could implement recommendation policies that condition on the full profile of private signals and assign a distinct public message to each signal realization, effectively using an arbitrarily large message space. While such schemes are feasible in the BCE concept, they are implausible as organizational practices. None of the qualitative results below hinge on the specific functional form of $S(\cdot)$.

In Propositions 2, 3, and 4, I deliberately abstract from cohesion and complexity by setting $\lambda = \kappa = 0$ (or by comparing policies with identical message structures). This isolates the party's persuasion capacity within the organization that arises purely from organizational features, such as correlation, hierarchy, and information acquisition, so that differences in outcomes are driven entirely by the persuasion term $\mathbb{P}(a = 1)$.

**Three organizational channels.** I now formalize three organizational channels that extend the benchmark and capture distinct aspects of party discipline emphasized by Weil. The first two channels alter not the availability of information, but the way information is aggregated, coordinated, and translated into public expression. Strategic ignorance operates differently: it affects whether information is acquired in the first place.

First, *meetings and caucuses* impose correlation across members' actions and speech. By coordinating recommendations, the party can create higher-order beliefs: each member knows that others have received the same instruction, which relaxes individual incentive constraints and facilitates collective alignment. Second, *hierarchy* concentrates informational control. Leaders observe privileged information and transmit simplified directives downstream, allowing the organization to act on information without revealing it fully. Third, *strategic ignorance* affects whether information is acquired in the first place: when being informed generates moral conflict or threatens cohesion, members may rationally choose not to know. The following propositions show that each of these mechanisms strictly enlarges the set of outcomes the party can implement, and does so in a direction that increases persuasion while potentially reducing informational content.

In many political organizations, members do not act or speak independently. Party caucuses, meetings, and coordinated messaging ensure that members know that others are receiving the same instruction. From an incentive perspective, this correlation matters whenever actions are strategically complementary or when endorsing the party line is costly in isolation but attractive when others do the same.

The key question is whether such correlation merely facilitates coordination, or whether it fundamentally alters the party's ability to shape public beliefs. The next proposition shows that correlation is not an innocuous feature: it strictly expands the party's persuasive power relative to any policy that treats members independently. Formally, call a recommendation policy *independent* (or iid) if, conditional on the state and private signals, recommendations to different members are statistically independent. Such policies treat members in isolation and rule out coordination through correlated instructions. Note that since I isolate the persuasion component $\Pr(a = 1)$, the correlated policy can be constructed to use the same message alphabet and complexity as iid policies, so the comparison is driven by persuasion alone.[8]

[8]Proposition 2 is an existence result. Appendix 6.3 provides a fully specified construction with explicit payoffs and obedience constraints. The argument extends to any $n \geq 2$ by parameter separation: some

**Proposition 2** (Meetings as coordination)**.** *For any $n \geq 2$, there exist primitives (a signal structure and member payoffs with moral costs and strategic complementarities) such that:*

*(i) Under any BCE-feasible independent (iid) recommendation policy, the party's persuasion success probability is bounded away from* 1*.*

*(ii) Under a BCE-feasible* correlated meeting *policy, the party can strictly increase* $\mathbb{P}(a = 1)$ *by inducing a more favorable distribution of public posteriors.*

The result shows that meetings and caucuses are not simply communication devices; they are instruments for manufacturing higher-order beliefs. By ensuring that members know others are also endorsing, correlation allows the party to recommend actions that would be infeasible under independent communication. Persuasion increases not because information improves, but because coordination substitutes for independent judgment.

Political parties are rarely flat organizations. Leadership positions concentrate information and authority, allowing a small number of actors to observe evidence and issue directives that others follow. Hierarchy thus separates the possession of information from the public expression of beliefs.

The next proposition shows that hierarchy does more than simplify decision-making. By filtering information through a leader, the party can implement persuasive outcomes that are impossible under flat, symmetric communication (even when followers have no private information of their own). In the flat communication case, the party does not observe the leader's private signal and cannot condition recommendations or public messages on it. Under hierarchy, by contrast, the leader's signal becomes available through an internal communication channel. Formally, the flat communication case restricts the policy space to information structures that do not condition on the leader's private signal.

**Proposition 3** (Hierarchy as informational filtering)**.** *Suppose that a party has a hierarchical structure in which a leader observes private information about the state while followers do not. Then:*

*(i) Under flat communication (no leader-to-follower message), no BCE-feasible policy can induce the public to choose $a = 1$ with positive probability.*

*(ii) Under delegated hierarchical communication, the party can implement a BCE-feasible policy that induces $a = 1$ with strictly positive probability.*

members' obedience constraints bind under independent recommendations but are relaxed under correlation, while others satisfy obedience under both regimes.

The two-follower hierarchy is chosen for simplicity. The logic extends directly to larger organizations by adding further followers with the same complementarity structure, without altering the feasibility comparison between flat and hierarchical communication. The result therefore does not rely on large populations: even minimal hierarchy suffices to suppress public learning by filtering how information is translated into collective expression.

Hierarchy functions as an information technology. It allows the party to act on information while limiting what is publicly revealed, thereby intensifying belief manipulation. This formalizes Weil's intuition that organizational domination suppresses judgment not by eliminating information, but by concentrating control over how it is used.

So far, information has been treated as exogenously available to members. Yet Weil emphasized that parties also discourage truth seeking itself: members learn that knowing too much is dangerous when it conflicts with loyalty or moral conformity. Therefore, ignorance can be privately optimal.

The next proposition shows that this intuition can be formalized as a stable organizational outcome under obedience constraints. When being informed creates moral or psychological costs, the party may strictly prefer to induce ignorance, thereby making obedience and unified messaging easier to sustain. Here the party's objective differs from the benchmark persuasion objective. Rather than maximizing the probability that an external public takes action $a = 1$, the party maximizes the probability that the member endorses the party line. Formally, the payoff becomes

$$U_P \equiv \Pr(x = 1).$$

This captures organizational success in terms of internal endorsement rather than persuasion of an external audience.

**Proposition 4** (Strategic ignorance)**.** *Suppose that acquiring information imposes private costs on party members when it conflicts with organizational discipline. Then:*

*(i) Under any BCE-feasible policy that induces full information acquisition, the party's endorsement success is bounded.*

*(ii) Under a BCE-feasible policy that induces members not to acquire information, the party can strictly increase the probability of endorsement.*

The single-member environment is chosen for clarity. The logic extends directly to larger organizations whenever individual members face moral or psychological costs from endorsing against their information. In such cases, inducing ignorance remains a feasible way for the party to sustain obedience and uniform messaging, independently of coordination or aggregation concerns.

Weil's critique is not only that parties distort public messages, but also that they discourage truth seeking within the organization itself. Strategic ignorance formalizes this. When being informed makes dissent psychologically or morally costly, members may rationally prefer not to know. In this case, the party benefits because obedience becomes easier to sustain and public messaging becomes more uniform. Ignorance is not assumed, it emerges endogenously as an organizational outcome that is feasible under obedience and preferred by the party.

Together, the results show that moving from "party as message" to "party as organization" strengthens the epistemic critique. Even if truthful information exists inside the party, organizational tools (meetings, hierarchy, and ignorance) allow the party to manufacture collective beliefs and suppress judgment. This is a formal counterpart to Weil's claim that parties oppose the possibility of thinking.

**Judgment Collapse under Party Discipline.** I now refine the organizational environment by distinguishing not only what information party members possess, but how they are allowed to speak. The previous results showed that parties can manipulate public beliefs through coordination, hierarchy, and ignorance. Following Arendt (1971, 1978), this raises a sharper and conceptually distinct question: even when members are informed and deception is absent, do party organizations suppress the expression of judgment itself?

*Speech within the organizational environment.* I continue with the organizational setting introduced above. Party members observe private signals about the state $\theta$, and public decisions are taken collectively. The new element introduced in this subsection is that members now choose *how* to speak: they may either express their private judgment or adhere to a party-line message.

To fix ideas, I assume that public decisions aggregate expressed messages by majority rule (ties broken randomly). For any $x \in [0, 1]$, let $\alpha(x)$ denote the probability that the public chooses the correct action when a fraction $x$ of agents speak judgmentally (truthfully), while the remaining fraction send an uninformative party-line message. For $q > 1/2$, $\alpha(x)$ is strictly increasing in $x$ for large $N$.

*Judgment vs. party line.* Each member chooses one of two communication modes:

- *Judgmental speech:* reporting the private assessment $m_i = s_i$, which triggers an organizational sanction $d \geq 0$;
- *Party-line speech:* issuing a standardized message independent of $s_i$, incurring no sanction.

Following Arendt's distinction between judgment and conformity, this subsection introduces a different choice variable, whether to express independent judgment or adopt party-line speech, and therefore specifies payoffs directly over speech modes. Members' payoff are given by

$$u_i = r \cdot \mathbb{1}_{\{\text{party-line speech}\}} - d \cdot \mathbb{1}_{\{\text{judgmental speech}\}} + \beta \cdot \mathbb{1}_{\{a=\theta\}},$$

where $r \geq 0$ captures the private return to conformity (career or loyalty incentives), $d \geq 0$ represents an organizational sanction or penalty for expressing independent judgment, and $\beta \geq 0$ captures concern for correct collective decisions (e.g. professional norms, civic motivation, or reputational costs from wrong public outcomes). This specification can be viewed as a restriction of the BCE framework in which the party's design determines whether judgmental speech is rewarded or sanctioned.

To study when judgment is sustained or suppressed, we need a way to measure the epistemic performance of collective decisions as a function of how many agents speak judgmentally. I therefore make explicit how individual judgments aggregate into a public decision. Throughout, let us assume that public decisions are determined by majority rule over the expressed messages. Party-line messages are uninformative about the state, while judgmental messages convey independent private assessments.

The next definition formalizes the resulting probability that the collective decision is correct as a function of the number of judgmental speakers. Because judgmental messages are independent conditional on the state and party-line messages are uninformative, collective accuracy under majority rule depends only on the number of judgmental speakers and follows a binomial aggregation logic.

**Definition 2** (Collective accuracy)**.** *Let $N$ be odd (to avoid tie-breaking issues). Suppose $k$ out of $N$ agents speak judgmentally, each sending an independent signal that equals the true state with probability $q > 1/2$, while the remaining agents send uninformative party-line messages. Let $\alpha(k/N)$ denote the probability that majority rule over the $N$ messages selects the correct action. Then*

$$\alpha\left(\frac{k}{N}\right) = \mathbb{P}\left(\mathit{Binomial}(k, q) \geq \frac{k+1}{2}\right).$$

For odd $N$, define the marginal epistemic contribution of one additional judgmental speaker by

$$\Delta(q) \equiv \alpha(1) - \alpha\left(\frac{N-1}{N}\right) > 0.$$

This quantity measures the increase in collective decision accuracy when the last remaining party-line speaker switches to judgmental speech. All expressions are exact for any finite odd $N$. In the characterization below, I focus on a large-$N$ environment in which individual members are effectively non-pivotal, so that a single member's speech choice does not affect the public decision rule.

This marginal accuracy gain determines when the civic benefit of expressing judgment outweighs the private incentives for conformity and silence. I can now characterize when organizational incentives lead to the suppression of judgment, and when truthful judgmental speech can be sustained.

**Proposition 5** (Judgment collapse threshold)**.** *Let $\alpha(1) > 1/2$ denote the accuracy under full judgmental speech as defined in Definition 2. Then:*

*(i) There exists a symmetric organizational outcome with full party-line speech (judgment collapses).*[9]

*(ii) If $\beta\Delta(q) \geq r + d$, there exists a symmetric organizational outcome with full judgmental speech.*

*Moreover, collective decision accuracy is strictly lower under judgment collapse than under full judgmental speech, since $\alpha(0) = 1/2$ while $\alpha(1) > 1/2$.*

Proposition 5 characterizes two possible organizational regimes. When conformity incentives dominate (high $r$ or $d$), members privately prefer to adhere to the party line even when informed, and judgment disappears from public discourse. Conversely, when epistemic motivations are sufficiently strong ($\beta\Delta(q) \geq r + d$), judgmental speech can be sustained despite organizational sanctions. This highlights that the disappearance of judgment need not reflect ignorance. Information may remain privately available to members, yet organizational incentives make its public expression privately dominated. This provides a formal analogue to Arendt's idea that political failure may arise from the erosion of judgment rather than from a lack of information.

Proposition 5 is robust to organizational size. Part (i) becomes unconditional under non-pivotality, while for finite $N$ it requires the threshold stated in footnote. Part (ii) does not rely on non-pivotality and holds for any finite odd $N$. Thus, whether the organization is finite or large, sufficiently strong conformity incentives can eliminate the public expression of independent judgment. In large parties, where individual members are effectively non-pivotal, this collapse emerges even more starkly, as epistemic motivations lose individual impact.

[9]For finite $N$, a unilateral deviation from party-line to judgmental speech increases collective accuracy from $\alpha(0) = 1/2$ to $\alpha(1) = q$, yielding an additional benefit $\beta(q - 1/2)$. In that case, full party-line speech is sustainable whenever $d \geq r + \beta(q - 1/2)$. The qualitative result is unchanged: sufficiently strong conformity incentives lead to judgment collapse, but the threshold depends on signal accuracy and epistemic motivation.

The Appendix, Section 8, shows that judgment collapse can also be expressed directly in the Bayes correlated equilibrium language.

## 4 Welfare, Participation, and Institutional Design

The previous sections identified a systematic wedge between party optimal and socially optimal information aggregation arising from what I have called the *control margin.* Through discipline, coordinated communication, hierarchy, and strategic ignorance, party organizations can shape how information and judgment enter public discourse. These mechanisms may suppress collective thinking not because information is unavailable, but because organizational incentives favor cohesion and control.

This section examines the welfare implications of these forces and the institutional responses they imply. I evaluate welfare using a minimal epistemic criterion: the probability that collective decisions match the true state of the world. Under this epistemic criterion, party organization has two opposing effects. On the one hand, the control margin can distort beliefs and suppress judgment, generating epistemic failure. On the other hand, party organization can also operate along a *participation margin* by lowering the costs of political participation and communication, thereby expanding the number of informative signals aggregated by the public. The analysis shows that when the participation margin dominates the control margin, party organization improves collective accuracy.

Throughout this section, welfare is measured by $\mathbb{P}(a = \theta)$, the probability that collective decisions match the true state. This criterion corresponds to the epistemic role of democracy emphasized in the introduction: the capacity of political institutions to aggregate dispersed information into accurate collective judgments. Under this epistemic criterion, socially optimal information structure is one that maximizes $\mathbb{P}(a = \theta)$, regardless of which organization benefits.

**Institutional levers and epistemic accountability.** The analysis highlights two margins along which institutions affect welfare.

First, institutions can reduce the *returns to discipline and pooling.* Limits on sanctions for dissent, protections for intra-party whistleblowing, or requirements of message transparency directly weaken the party's ability to suppress informative speech. In the terms of Section 3, such institutions reduce the effective payoff from conformity ($r$) or raise the cost of pooling and garbling, thereby expanding the set of environments in which informative communication and judgmental speech are privately sustainable.

Second, institutions can increase *accountability for epistemic performance.* If parties are penalized (electorally, reputationally, or institutionally) for systematically wrong decisions, then party utility becomes more aligned with collective accuracy. Relative to Section 2, this corresponds to increasing the weight placed on accuracy relative to cohesion. Relative to Section 3, it weakens the incentives to impose correlation, hierarchy, or ignorance. Regulation therefore operates by shifting the party's objective closer to the social objective, not by eliminating organization.

The results do not imply that parties should be abolished. On the contrary, parties can improve collective thinking when they lower the costs of participation and communication. The pathology identified in Sections 2 and 3 arises only when the party's objective places excessive weight on control over speech relative to epistemic performance. The relevant normative implication is therefore conditional: regulate informational power rather than suppress political organization per se.

**Parties as epistemic infrastructure: the participation margin.** I now formalize a countervailing channel that favors parties. Even if party organization introduces some informational distortion, it may nonetheless improve welfare by expanding participation. When communication is costly, many informed citizens remain silent. Parties reduce these costs (through mobilization, media access, and coordination) thereby increasing the number of informative signals entering public inference. This *participation margin* can dominate the control margin identified earlier.

**Setup and result.** There is a binary state $\theta \in \{0,1\}$ with prior $\mathbb{P}(\theta = 1) = 1/2$. There are $N$ potential citizen-informants. Each informant $i$ receives an independent private signal $s_i \in \{0,1\}$ with $\mathbb{P}(s_i = \theta) = q > 1/2$. Each informant chooses whether to publicly communicate. Communication is costly. Without party organization, each informant communicates independently with probability $p_0 \in (0,1)$. With party organization, mobilization and coordination reduce communication costs, so participation rises to $p_1 > p_0$.

However, party organization may distort messages. If an informant participates through the party, his public message $\hat{s}_i$ equals the true signal $s_i$ with probability $1-\varepsilon$ and is flipped with probability $\varepsilon \in [0, 1/2)$.

The public observes all communicated messages and chooses $a \in \{0,1\}$ by majority rule among received messages (ties broken randomly). Let $\mathcal{P}(p, \varepsilon)$ denote the probability that $a = \theta$ given participation rate $p$ and garbling rate $\varepsilon$.

**Theorem 2** (Participation margin)**.** *Assume $q \in (1/2, 1)$ and $\varepsilon \in [0, 1/2)$ so that the effective signal quality*

$$q^{\star}(\varepsilon) = (1-\varepsilon)q + \varepsilon(1-q) > 1/2.$$

*Then:*

(i) *For fixed $\varepsilon$, $\mathcal{P}(p, \varepsilon)$ is strictly increasing in $p$ for all sufficiently large $N$.*

(ii) *Consequently, there exist parameters $(N, q, p_0, p_1, \varepsilon)$ such that*

$$\mathcal{P}(p_1, \varepsilon) > \mathcal{P}(p_0, 0).$$

*That is, party organization can strictly improve collective accuracy by increasing participation, even when it introduces message garbling at the individual level.*

Sections 2 and 3 identify a control margin: party organization enables belief manipulation, coordination without judgment, and even ignorance. The participation margin identifies an opposing force: by lowering communication costs, parties expand the set of informative voices entering public inference. The welfare implications are therefore conditional. Democratic institutions should preserve the participation benefits of parties while constraining their capacity to suppress information and judgment. Regulation, rather than suppression, is the appropriate response.

# 5 Concluding Discussion

This paper formalizes a classic but neglected concern: that party organization, long understood as essential for coordination and collective action in democratic politics (e.g., Aldrich, 1995), may also undermine democracy's epistemic function. In my benchmark and organizational models, informational failure does not arise from voter irrationality or a lack of expertise. It arises from incentives internal to political organizations. Parties that value cohesion and growth can rationally garble or pool public communication, impose correlation through caucuses and meetings, concentrate informational control via hierarchy, and even induce strategic ignorance among members. In these environments, collective thinking collapses despite abundant private information.

At the same time, the analysis yields a pro-party channel. By lowering communication and mobilization costs, parties can expand participation and increase the volume of informative signals entering public inference. When this participation margin dominates the control margin, parties improve the probability of correct collective decisions even if some distortion remains. The normative implication is therefore conditional: the democratic challenge is not the existence of parties per se, but the regulation of their informational power.

Two complementary perspectives help interpret these results. Weil's critique highlights the organizational logic by which parties subordinate truth to loyalty and self-development, a logic that my model endogenize. Arendt's reflections on judgment emphasize that political failure can take the form of judgment disappearing even among informed actors; Proposition 5 provides a formal analogue by showing how discipline can make judgment privately dominated (Arendt, 1971, 1978). Finally, accounts that stress the transformation or weakening of traditional party organizations need not imply a decline in parties' epistemic influence: even as parties become less socially embedded, centralization and professionalization may preserve or intensify their capacity to discipline speech (Martinache and Sawicki, 2020).

Weil does not stop at diagnosis. At the end of her essay, she briefly sketches an alternative to party organization based on loose intellectual affiliations, such as journals or circles of writing, to which individuals could be close without formal adherence. The aim is to preserve freedom of judgment by eliminating discipline and collective passion. While this proposal captures an important intuition emphasized by my analysis, the epistemic costs of organizational control, it also highlights a limitation. Removing discipline may restore judgment, but removing organization altogether may undermine participation, coordination, and the aggregation of dispersed information. My results therefore suggest that the relevant normative tradeoff is not between parties and no parties, but between different institutional designs governing organizational power. In this sense, the paper complements coordination based accounts of parties by showing how the same organizational capacities that solve collective action problems can also generate epistemic failure.

Taken together, the analysis shows that the epistemic effects of party organization operate through two opposing margins: a control margin, which can suppress information and judgment, and a participation margin, which can expand the informational basis of collective decisions. Electoral institutions may also shape the strength of these informational forces. While the analysis abstracts from inter-party competition, electoral systems plausibly affect the incentives parties face. Majoritarian systems may increase pressures for message discipline and internal cohesion, whereas more proportional systems may shift informational diversity toward competition across parties rather than dissent within them. Understanding how electoral rules interact with internal party informational control remains an open question. Overall, the results point to an institutional design agenda: preserve the participation benefits of party organization that make parties indispensable coordination devices. At the same time, institutions should constrain mechanisms that enable informational suppression, including sanctions for dissent, excessive centralization of messaging, and organizational incentives that reward cohesion over accuracy.

# 6 Proofs

## 6.1 Proof of Theorem 1

The proof proceeds in five steps. Step 1 characterizes the public's optimal action given a posterior belief. Step 2 establishes the Bayes plausibility constraint linking posterior beliefs to the prior, completing the reduction in part (i) of the theorem. Step 3 reformulates the party's problem as a one dimensional optimization over posterior distributions. Step 4 shows that at most two posterior beliefs are sufficient for optimality, establishing part (ii). Step 5 shows that when $\mu_0 < \frac{1}{2}$, full revelation is weakly dominated by a garbling (pooling) policy, establishing part (iii).

For convenience, in the proof I adopt standard signaling game terminology, referring to the party as the sender and the public as the receiver.

**Step 1: Receiver best response.** Given a public message $m$, the receiver forms posterior belief $\mu(m) = \mathbb{P}(\theta = 1 \mid m)$. Since the receiver's payoff is $u_R(a, \theta) = \mathbb{1}_{\{a=\theta\}}$, the expected payoff from choosing $a = 1$ equals $\mu(m)$, while the expected payoff from choosing $a = 0$ equals $1 - \mu(m)$. Hence the receiver chooses

$$a(m) = \begin{cases} 1 & \text{if } \mu(m) \geq \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, under any information policy $\pi$, the party's payoff can be written as

$$U_P(\pi) = \mathbb{P}_\pi(a = 1) = \sum_{m \in M} \mathbb{P}(m)\, \mathbb{1}_{\{\mu(m) \geq 1/2\}}.$$

Define $v(\mu) \equiv \mathbb{1}_{\{\mu \geq 1/2\}}$. Then

$$U_P(\pi) = \sum_{m \in M} \mathbb{P}(m)\, v(\mu(m)) = \mathbb{E}[v(\mu(M))].$$

**Step 2: Bayes plausibility (mean posterior equals prior).** For any policy $\pi$, the induced posterior beliefs satisfy the Bayes plausibility constraint

$$\sum_{m \in M} \mathbb{P}(m)\, \mu(m) = \mu_0. \tag{1}$$

This follows directly from the law of iterated expectations:

$$\mathbb{E}[\mu(M)] = \mathbb{E}[\mathbb{P}(\theta = 1 \mid M)] = \mathbb{P}(\theta = 1) = \mu_0.$$

Thus, any information policy induces a distribution over posterior beliefs supported on $[0, 1]$ with mean $\mu_0$.

Moreover, because the party can always refine the message space (by splitting messages with identical posteriors) without affecting incentives or outcomes, we may equivalently treat the party's problem as directly choosing a distribution over posterior beliefs satisfying (1).[10]

**Step 3: The party's problem as a one-dimensional optimization.** The party's problem can therefore be written as

$$\max_{\{(\alpha_k,\mu_k)\}_{k=1}^K} \sum_{k=1}^K \alpha_k\, v(\mu_k) \quad \text{s.t.} \quad \sum_{k=1}^K \alpha_k \mu_k = \mu_0, \quad \sum_{k=1}^K \alpha_k = 1, \quad \alpha_k \geq 0, \quad \mu_k \in [0, 1].$$

Since $v(\mu) \in \{0, 1\}$, the objective equals the total probability mass assigned to posterior beliefs at or above the decision threshold $\frac{1}{2}$.

**Step 4: Sufficiency of two posterior beliefs.** I show that there exists an optimal solution supported on at most two posterior values.

Consider any feasible distribution over posterior beliefs with support size $K \geq 3$ and mean $\mu_0$. Because the only nontrivial constraint on posterior beliefs is the linear mean constraint (1), the set of feasible posterior distributions is convex and effectively one-dimensional. By Carathéodory's theorem in $\mathbb{R}$, any point in the convex hull of $[0, 1]$ can be expressed as a convex combination of at most two points.[11]

More explicitly, let $\alpha \equiv \mathbb{P}(\mu \geq \frac{1}{2})$ under the original distribution, let $\mu_H$ be the conditional expectation of $\mu$ given $\mu \geq \frac{1}{2}$, and let $\mu_L$ be the conditional expectation of $\mu$ given $\mu < \frac{1}{2}$. Then

$$\mu_0 = (1 - \alpha)\mu_L + \alpha\mu_H.$$

Now consider the two-point distribution that assigns probability $1 - \alpha$ to $\mu_L$ and probability $\alpha$ to $\mu_H$. This distribution satisfies the same mean constraint and yields the same objective

[10] Formally, any distribution over posterior beliefs with mean $\mu_0$ can be implemented by an information structure (experiment) in the sense of Blackwell (1951, 1953). In Bayesian persuasion problems, this observation underlies the standard reduction in which the designer directly chooses a Bayes-plausible distribution over posteriors; see Kamenica and Gentzkow (2011).

[11] In one dimension, any point in the convex hull of a set can be written as a convex combination of two extreme points. Here the relevant "point" is the mean posterior $\mu_0$, and the set is $[0, 1]$.

value $\alpha$, since $v(\mu_L) = 0$ and $v(\mu_H) = 1$. Hence any feasible payoff can be achieved by a distribution with at most two posterior beliefs, and an optimal solution exists with this property.

Hence any feasible payoff can be achieved by a distribution with at most two posterior beliefs, and an optimal solution exists with this property, which establishes part (ii) of the theorem.

**Step 5: Full revelation is not optimal when $\mu_0 < \frac{1}{2}$.** Assume $\mu_0 < \frac{1}{2}$. Under full revelation of the signal, the induced posterior distribution assigns positive probability to posteriors both above and below $\frac{1}{2}$, and the party's payoff equals the probability mass $\bar{\alpha}$, where $\bar{\alpha} \equiv \mathbb{P}(\mu(m) \geq \frac{1}{2} \mid \text{full revelation}) \in (0, 1)$, on posteriors exceeding the threshold.

I show that full revelation is generically weakly dominated by a pooling (garbling) policy. Consider a two-point posterior distribution with $\mu_H = \frac{1}{2} + \varepsilon$ for arbitrarily small $\varepsilon > 0$ and $\mu_L \in [0, \frac{1}{2})$ chosen to satisfy the mean constraint

$$\mu_0 = (1 - \alpha)\mu_L + \alpha\mu_H.$$

Solving yields

$$\alpha = \frac{\mu_0 - \mu_L}{\mu_H - \mu_L}.$$

In the limiting case $\mu_L = 0$ and $\varepsilon \to 0$, this yields $\alpha \to 2\mu_0$. Since $\mu_0 < \frac{1}{2}$, we have $2\mu_0 > \mu_0$, and for a wide class of signal structures $2\mu_0$ strictly exceeds the persuasion success probability $\bar{\alpha}$ induced by full revelation.

The intuition is standard in Bayesian persuasion: because the receiver's decision rule features a discontinuity at $\frac{1}{2}$, full revelation generally allocates too much probability mass to posteriors far from the threshold, rather than concentrating mass just above it. By pooling information so as to shift posterior mass toward the decision threshold, the party can weakly increase the probability of inducing action $a = 1$.

Thus, when $\mu_0 < \frac{1}{2}$, full revelation is not generically optimal, and there exists a pooling or garbling policy that weakly improves the party's persuasion objective. This establishes part (iii) of the theorem. □

## 6.2 Proof of Proposition 1

**Notation.** When comparing different information policies, I temporarily reinstate the subscript on probabilities to indicate the policy under which the distribution is evaluated.

The proof proceeds in five steps. Step 1 bounds the persuasion component of the party's objective. Steps 2 and 3 show that cohesion is uniquely maximized by pooling. Step 4 establishes the existence of a uniform threshold $\bar{\lambda}$ above which pooling dominates any non-pooling policy. Step 5 shows that pooling eliminates public learning.

Fix a finite message space $M$ and any party policy $\pi$. Recall that the party's objective is

$$U_P(\pi) = \mathbb{P}_\pi(a = 1) + \lambda\, C(\pi), \qquad C(\pi) \equiv -H(M),$$

where

$$H(M) = -\sum_{m \in M} \mathbb{P}(m) \log \mathbb{P}(m)$$

is the Shannon entropy of the induced public message distribution (Shannon, 1948).

**Step 1: Bounding the persuasion term.** For any policy $\pi$, the persuasion component satisfies

$$0 \leq \mathbb{P}_\pi(a = 1) \leq 1,$$

since it is a probability.

**Step 2: Cohesion is maximized by pooling.** Shannon entropy satisfies $H(M) \geq 0$, with equality if and only if the message distribution is degenerate, i.e. $\mathbb{P}(m^\star) = 1$ for some message $m^\star$. Hence

$$C(\pi) = -H(M) \leq 0,$$

with maximum value 0, achieved uniquely by fully pooling policies.

Let $\pi^\star$ denote any fully pooling policy. Then

$$C(\pi^\star) = 0.$$

**Step 3: Any non-pooling policy incurs a strictly negative cohesion payoff.** If $\pi$ is not fully pooling, its induced message distribution assigns positive probability to at least two distinct messages. Since $M$ is finite, this implies

$$H(M) > 0 \quad \text{and hence} \quad C(\pi) = -H(M) < 0.$$

**Step 4: Existence of a uniform threshold $\bar{\lambda}$.** Under Assumption 1, a communication policy specifies an allocation of the $T$ interventions across messages. Hence the induced message distribution assigns probabilities that are multiples of $1/T$, and the set of admissible message distributions is finite. The subset of non-pooling distributions is therefore finite as well. Since Shannon entropy is continuous, it attains a strictly positive minimum over this set. Define

$$\underline{H} \equiv \min\{H(M) : \pi \text{ is not fully pooling}\}.$$

Then $\underline{H} > 0$. Now compare any non-pooling policy $\pi$ to the pooling policy:

$$U_P(\pi^\star) - U_P(\pi) = \big[\mathbb{P}_{\pi^\star}(a=1) - \mathbb{P}_\pi(a=1)\big] + \lambda\big[0 - C(\pi)\big].$$

The first bracket is bounded below by $-1$, while the second term satisfies

$$0 - C(\pi) = H(M) \geq \underline{H}.$$

Hence

$$U_P(\pi^\star) - U_P(\pi) \;\geq\; -1 + \lambda\,\underline{H}.$$

Define

$$\bar{\lambda} \equiv \frac{1}{\underline{H}}.$$

Then for all $\lambda \geq \bar{\lambda}$, we have

$$U_P(\pi^\star) \geq U_P(\pi)$$

for every non-pooling policy $\pi$. Therefore, for sufficiently large $\lambda$, any optimal policy is fully pooling.

**Step 5: Learning collapses under pooling.** Under a fully pooling policy, the public observes the same message regardless of the underlying signal or state. Hence the message is uninformative about $\theta$, and Bayes' rule implies

$$\mu(m) = \mathbb{P}(\theta = 1 \mid m) = \mu_0$$

for all on-path messages $m$. Thus public learning collapses. □

### 6.3 Proof of Proposition 2

This proof provides a concrete $n \geq 2$ construction establishing Proposition 2. The argument is by existence: I exhibit primitives for which correlated recommendation policies strictly dominate all independent policies in terms of persuasion. The proof is constructive and proceeds in four steps: (i) specification of payoffs and obedience constraints; (ii) characterization of BCE-feasible independent (iid) policies; (iii) construction of a correlated "meeting" policy and verification of feasibility; (iv) demonstration of a strict persuasion gain.

Throughout, Bayes correlated equilibrium (BCE) is used as a feasibility notion: a recommendation policy is feasible if all agents optimally obey the recommendations they receive.

**Notation.** Throughout the proof, probabilities are taken with respect to the joint distribution over $(\theta, s_1, \ldots, s_n, x_1, \ldots, x_n)$ induced by the party's recommendation policy and the agents' obedience. For clarity, I use superscripts to distinguish policies: $\mathbb{P}^{\text{iid}}(\cdot)$ denotes probabilities induced by an independent (iid) recommendation policy, while $\mathbb{P}^{\text{co}}(\cdot)$ denotes probabilities induced by the correlated "meeting" policy constructed below. The party's payoff corresponding to these different policies use the same overscript notation, $U_P^{co}$ and $U_P^{iid}$.

**Environment.** There is a binary state $\theta \in \{0, 1\}$ with prior $\mathbb{P}(\theta = 1) = \mu_0 < 1/2$. There are $n \geq 2$ party members indexed by $i = 1, \ldots, n$. Each member observes a private signal $s_i \in \{0, 1\}$ with accuracy $q \in (1/2, 1)$:

$$\mathbb{P}(s_i = 1 \mid \theta = 1) = q, \qquad \mathbb{P}(s_i = 1 \mid \theta = 0) = 1 - q,$$

independently across $i$ conditional on $\theta$.

Each member chooses an action $x_i \in \{0, 1\}$, interpreted as publicly endorsing the party line ($x_i = 1$) or not ($x_i = 0$). The public observes the aggregate endorsement count

$$m = \sum_{i=1}^{n} x_i$$

and chooses $a = 1$ if and only if $\mathbb{P}(\theta = 1 \mid m) \geq 1/2$. The party's payoff is

$$U_P = \mathbb{P}(a = 1).$$

**Member payoffs.** Members' utilities capture strategic complementarities in endorsement and moral costs of endorsing against one's signal:

$$u_i(x_i, x_{-i}, s_i) = B\, x_i \cdot \Big( \frac{1}{n-1} \sum_{j \neq i} x_j \Big) - c_i\, x_i\, \mathbb{1}_{\{s_i = 0\}},$$

where $B > 0$ the strength of complementarities (endorsement is privately valuable when others endorse) and $c \geq 0$ captures the moral cost of endorsing when the member's signal is unfavorable.

We distinguish two types of members:

- Members $i = 1, 2$ have moral cost $c_i = c$ (high).
- Members $i \geq 3$ have moral cost $c_i = \underline{c}$ (low).

Fix parameters satisfying

$$B\rho < c \leq B \qquad \text{and} \qquad \underline{c} \leq B, \tag{2}$$

where

$$\rho \equiv \mathbb{P}(s_j = 1) = \mu_0 q + (1 - \mu_0)(1 - q) \in (0, 1).$$

**Policies and obedience.** A recommendation policy specifies, for each signal profile, a joint distribution over recommended actions $r = (r_1, \ldots, r_n) \in \{0, 1\}^n$. We focus on obedience: on path, $x_i = r_i$.

If agent $i$ is recommended $r_i = 1$, obedience requires

$$B\,\mathbb{E}\left[\frac{1}{n-1}\sum_{j \neq i} r_j \,\middle|\, s_i, r_i = 1\right] \;\geq\; c_i\,\mathbb{1}_{\{s_i = 0\}}. \tag{3}$$

**Lemma 1** (Non-binding members)**.** *For all members $i \geq 3$, obedience holds under both independent (iid) and correlated recommendation policies.*

*Proof.* Fix any $i \geq 3$ and any information set at which $r_i = 1$. The left-hand side of (3) is at most $B$ under iid policies and equals $B$ under fully correlated policies. Since $c_i = \underline{c} \leq B$, the obedience constraint holds regardless of the recommendation structure. □

Thus members $i \geq 3$ are active (their actions affect $m$) but do not constrain feasibility under either policy class.

**Step 1: Restricting to environments where persuasion requires full endorsement.** Fix $\mu_0 \in ((1 - q)^2, 1/2)$ and choose primitives so that the public chooses $a = 1$ if and only if $m = n$.[12] This can be ensured by taking $\mu_0$ in that range and choosing primitives so that the maximal likelihood ratio achievable for any message $m < n$ under the binary-signal structure and any recommendation policy considered below remains below $(1 - \mu_0)/\mu_0$, and hence

$$\mathbb{P}(\theta = 1 \mid m) < \tfrac{1}{2}$$

[12]This restriction is imposed only for the constructive existence argument and serves to isolate the persuasive value of full endorsement.

for all $m < n$. Since Proposition 2 is an existence result, such a choice of primitives is without loss of generality.

Call a recommendation policy *iid* if recommendations factorize conditional on signals:

$$\mathbb{P}(r_1, \ldots, r_n \mid s_1, \ldots, s_n) = \prod_{i=1}^{n} \mathbb{P}(r_i \mid s_i).$$

Write

$$p_i^1 = \mathbb{P}(r_i = 1 \mid s_i = 1), \qquad p_i^0 = \mathbb{P}(r_i = 1 \mid s_i = 0).$$

Under obedience we have $x_i = r_i$, and the public message is $m = \sum_{i=1}^{n} x_i$.

If an iid policy induces $m$ independent of $\theta$ (e.g. $r_i \equiv 1$ for all $i$), then $\mathbb{P}(\theta = 1 \mid m) = \mu_0 < 1/2$ for all on-path $m$ and thus $U_P^{iid} = 0$. Hence only iid policies that make $m = n$ sufficiently informative can achieve persuasion.

**Step 2: iid policies are bounded away from full persuasion.** Fix a constant $\delta \in (0, 1)$ such that

$$\big(q + (1 - q)\delta\big)^2 < 1.$$

Consider any BCE-feasible iid policy.

*Case 1: $p_i^0 \geq \delta$ for some high-cost member $i \in \{1, 2\}$.* Then $(s_i = 0, r_i = 1)$ is an on-path information set. Obedience requires

$$B \cdot \mathbb{E}\left[ \frac{1}{n-1} \sum_{j \neq i} r_j \,\middle|\, s_i = 0, r_i = 1 \right] \geq c.$$

Let $h \in \{1, 2\} \setminus \{i\}$ denote the other high-cost member. Using the decomposition of the average endorsement of others into the other high-cost member and the $n - 2$ low-cost members, we obtain

$$\mathbb{P}(r_h = 1 \mid s_i = 0, r_i = 1) \geq \eta, \qquad \eta \equiv \frac{n-1}{B} c - (n - 2) > 1 - q.$$

Let

$$w \equiv \mathbb{P}(\theta = 1 \mid s_i = 0, r_i = 1).$$

Under iid recommendations, conditional on $\theta$ the recommendations are independent across agents, so

$$\mathbb{P}(r_h = 1 \mid s_i = 0, r_i = 1) = w \, \mathbb{P}(r_h = 1 \mid \theta = 1) + (1 - w)\mathbb{P}(r_h = 1 \mid \theta = 0),$$

which implies

$$\mathbb{P}(r_h = 1 \mid \theta = 0) \geq \frac{\eta - w}{1 - w}.$$

Since

$$w \leq \bar{w} \equiv \mathbb{P}(\theta = 1 \mid s_i = 0) = \frac{\mu_0(1-q)}{\mu_0(1-q) + (1-\mu_0)q},$$

we obtain the uniform bound

$$\mathbb{P}(r_h = 1 \mid \theta = 0) \geq \frac{\eta - \bar{w}}{1 - \bar{w}}.$$

For member $i$, since $p_i^0 \geq \delta$,

$$\mathbb{P}(r_i = 1 \mid \theta = 0) = (1-q)p_i^1 + q\, p_i^0 \geq q\delta.$$

Now consider the likelihood ratio associated with full endorsement:

$$\frac{\mathbb{P}(m = n \mid \theta = 1)}{\mathbb{P}(m = n \mid \theta = 0)} = \prod_{j=1}^{n} \frac{\mathbb{P}(r_j = 1 \mid \theta = 1)}{\mathbb{P}(r_j = 1 \mid \theta = 0)}.$$

For member $i$,

$$\frac{\mathbb{P}(r_i = 1 \mid \theta = 1)}{\mathbb{P}(r_i = 1 \mid \theta = 0)} \leq \frac{1}{q\delta}.$$

For member $h$,

$$\frac{\mathbb{P}(r_h = 1 \mid \theta = 1)}{\mathbb{P}(r_h = 1 \mid \theta = 0)} \leq \frac{1 - \bar{w}}{\eta - \bar{w}}.$$

For each low-cost member $j \geq 3$,

$$\frac{\mathbb{P}(r_j = 1 \mid \theta = 1)}{\mathbb{P}(r_j = 1 \mid \theta = 0)} \leq \frac{q}{1-q},$$

since

$$\mathbb{P}(r_j = 1 \mid \theta = 1) = qp_j^1 + (1-q)p_j^0, \qquad \mathbb{P}(r_j = 1 \mid \theta = 0) = (1-q)p_j^1 + qp_j^0.$$

Therefore

$$\frac{\mathbb{P}(m = n \mid \theta = 1)}{\mathbb{P}(m = n \mid \theta = 0)} \leq \frac{1}{q\delta} \cdot \frac{1 - \bar{w}}{\eta - \bar{w}} \cdot \left(\frac{q}{1-q}\right)^{n-2}.$$

The right-hand side depends only on $(q, B, c, n, \delta, \mu_0)$, not on the policy. Choose primitives so that

$$\frac{1 - \mu_0}{\mu_0} > \frac{1}{q\delta} \cdot \frac{1 - \bar{w}}{\eta - \bar{w}} \cdot \left(\frac{q}{1-q}\right)^{n-2}.$$

Such a choice is feasible. For $q$ sufficiently close to $1/2$ and $c$ sufficiently close to $B$, the right-hand side is finite and can be made smaller than

$$\sup_{\mu_0 \in ((1-q)^2, 1/2)} \frac{1-\mu_0}{\mu_0} = \frac{1-(1-q)^2}{(1-q)^2}.$$

Hence one can choose $\mu_0 \in ((1-q)^2,\, 1/2)$ satisfying the inequality. Then Bayes' rule implies

$$\mathbb{P}(\theta = 1 \mid m = n) < \tfrac{1}{2},$$

so $m = n$ cannot induce persuasion. Hence any iid policy in Case 1 yields $U_P^{iid} = 0$.

*Case 2: $p_1^0 < \delta$ and $p_2^0 < \delta$.* Under any iid policy,

$$\mathbb{P}(m = n) \leq \mathbb{P}(r_1 = 1)\mathbb{P}(r_2 = 1),$$

since the remaining factors are at most 1. Moreover, since $\rho = \mu_0 q + (1-\mu_0)(1-q) \leq q$,

$$\mathbb{P}(r_i = 1) = \rho\, p_i^1 + (1-\rho)p_i^0 \leq q + (1-q)\delta \qquad \text{for } i = 1, 2.$$

Therefore

$$\mathbb{P}^{iid}(m = n) \leq \big(q + (1-q)\delta\big)^2 < 1.$$

Since persuasion requires $m = n$, it follows that

$$U_P^{iid} \leq \big(q + (1-q)\delta\big)^2 < 1.$$

Combining the two cases, there exists $\varepsilon_0 > 0$ such that under any BCE-feasible iid policy,

$$U_P^{iid} \leq 1 - \varepsilon_0.$$

This establishes part (i).

**Step 3: A correlated "meeting" policy.** Define the following correlated recommendation policy:

- If $s_1 = s_2 = 1$, recommend $r_i = 1$ to all $i$.
- If exactly one of $s_1, s_2$ equals 1, recommend $r_i = 1$ to all $i$ with probability $\alpha \in (0, 1)$, and $r_i = 0$ otherwise.
- Otherwise, recommend $r_i = 0$ to all $i$.

If $r_i = 1$ is recommended, all members infer that all others are also endorsing. For members $i = 1, 2$, obedience for $s_i = 0$ requires $B \geq c$, which holds by (2). For members $i \geq 3$, feasibility holds by Lemma 1. Hence the policy is BCE-feasible.

**Step 4: Strict persuasion gain (part (ii)).** Under the correlated policy, the event $m = n$ occurs not only when $s_1 = s_2 = 1$ but also with positive probability when exactly one of $(s_1, s_2)$ equals 1. In particular,

$$\mathbb{P}^{co}(m = n) = q^2 + \alpha\, 2q(1 - q).$$

By Step 2, any iid policy satisfies

$$\mathbb{P}^{iid}(m = n) \leq q^2.$$

Hence

$$\mathbb{P}^{co}(m = n) > \mathbb{P}^{iid}(m = n)$$

for any $\alpha > 0$.

It remains to verify that, for sufficiently small $\alpha$, the posterior associated with $m = n$ remains above the public decision threshold. At $\alpha = 0$, the event $m = n$ occurs if and only if $s_1 = s_2 = 1$, so

$$\mathbb{P}(\theta = 1 \mid m = n) = \mathbb{P}(\theta = 1 \mid s_1 = s_2 = 1).$$

By Bayes' rule,

$$\mathbb{P}(\theta = 1 \mid s_1 = s_2 = 1) = \frac{\mu_0 q^2}{\mu_0 q^2 + (1 - \mu_0)(1 - q)^2}.$$

This posterior exceeds 1/2 whenever

$$\mu_0 > (1 - q)^2.$$

Fix primitives satisfying this condition. Since the probabilities of all signal events and recommendations vary continuously in $\alpha$, the posterior $\mathbb{P}(\theta = 1 \mid m = n)$ under the correlated policy is continuous in $\alpha$. Therefore there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$,

$$\mathbb{P}(\theta = 1 \mid m = n) > \tfrac{1}{2}.$$

Hence the public continues to choose $a = 1$ whenever $m = n$. Since Step 1 implies that persuasion can arise only through the event $m = n$, we have

$$U_P^{iid} = \mathbb{P}^{iid}(a = 1) \leq \mathbb{P}^{iid}(m = n).$$

Hence, for all sufficiently small $\alpha > 0$,

$$U_P^{co} = \mathbb{P}^{co}(a = 1) = \mathbb{P}^{co}(m = n) > \mathbb{P}^{iid}(m = n) \geq U_P^{iid}.$$

This establishes part (ii) and completes the proof. □

## 6.4 Proof of Proposition 3

I provide a constructive proof showing that delegated hierarchy can strictly expand the set of BCE-feasible persuasive outcomes relative to flat communication.

**Notation.** Throughout the proof, $\mathbb{P}^F(\cdot)$ and $\mathbb{P}^H(\cdot)$ denote probabilities induced by the flat and hierarchical communication policies, respectively, together with agents' obedience. The party's payoff corresponding to these different policies uses the same overscript notation, $U_P^F$ and $U_P^H$.

**Environment.** The state $\theta \in \{0, 1\}$ has prior $\mathbb{P}(\theta = 1) = \mu_0 < 1/2$. A leader $L$ observes a private signal $s_L \in \{0, 1\}$ with accuracy $q \in (1/2, 1)$:

$$\mathbb{P}(s_L = 1 \mid \theta = 1) = q, \qquad \mathbb{P}(s_L = 1 \mid \theta = 0) = 1 - q.$$

Two followers $i \in \{1, 2\}$ have no private signals and choose actions $x_i \in \{0, 1\}$ (endorse if 1). The public message is $m = x_1 + x_2 \in \{0, 1, 2\}$. The public chooses $a = 1$ if and only if $\mathbb{P}(\theta = 1 \mid m) \geq 1/2$.

Followers' payoffs exhibit strategic complementarities:

$$u_i(x_i, x_j) = B\, x_i x_j, \qquad B > 0.$$

The leader sends an internal message $\tau \in \{0, 1\}$ and receives payoff

$$u_L(\tau, s_L) = \begin{cases} R & \text{if } \tau = s_L, \\ -\varepsilon & \text{if } \tau \neq s_L, \end{cases}$$

with $R > \varepsilon > 0$, ensuring truth-telling is strictly optimal.

The party's objective is persuasion:

$$U_P = \mathbb{P}(a = 1).$$

**Flat communication cannot induce persuasion.** In the flat communication case, the party's information structure contains no signal correlated with $\theta$: there is no leader and no internal communication channel through which state-correlated information becomes available to the organizational center. Formally, the policy space is restricted to recommendation rules that do not condition on $\theta$ or on any signal correlated with it. Hence any recommendation policy induces a distribution over public messages $m$ that is independent of the state. Therefore, for every on-path message $m$,

$$\mathbb{P}^F(\theta = 1 \mid m) = \mathbb{P}(\theta = 1) = \mu_0 < 1/2,$$

and the public chooses $a = 0$ with probability one. Thus, $U_P^F = 0$. This establishes part (i) of the proposition.

**Delegated hierarchy induces persuasion.** Consider the following hierarchical protocol:

- The leader sends $\tau = s_L$.
- The party recommends $x_1 = x_2 = \tau$ to the followers.

*Public persuasion.* Under the protocol, $m = 2$ occurs if and only if $\tau = 1$, which occurs if and only if $s_L = 1$. Thus

$$\mathbb{P}^H(\theta = 1 \mid m = 2) = \mathbb{P}(\theta = 1 \mid s_L = 1) = \frac{\mu_0 q}{\mu_0 q + (1 - \mu_0)(1 - q)}.$$

The public chooses $a = 1$ after $m = 2$ if and only if $\mu_0 \geq 1 - q$. Since $q > 1/2$, the interval $[1 - q, 1/2)$ is nonempty, and we may choose $\mu_0$ in this range. Finally,

$$\mathbb{P}^H(m = 2) = \mathbb{P}(s_L = 1) > 0,$$

so $U_P^H = \mathbb{P}^H(a = 1) > 0$. Hence, $U_P^H > U_P^F$, establishing part (ii) and completing the proof. □

## 6.5 Proof of Proposition 4

The proof proceeds in two steps. First, I construct an organizational policy under which the party induces ignorance and achieves full public endorsement. This establishes part (ii) of the proposition. Second, I show that under any policy that induces full information acquisition, endorsement must fail with positive probability, so the party's persuasion success is strictly bounded below one. This establishes part (i). Comparing the two outcomes yields a strict preference for inducing ignorance.

**Notation.** For any organizational policy, let $U_P^{\text{policy}}$ denote the party's persuasion success under that policy, defined as the probability of public endorsement induced by agents' optimal responses. In particular, I write $U_P^{\text{IG}}$ for the ignorance-inducing policy and $U_P^{\text{FI}}$ for any policy inducing full information acquisition.

**Environment.** There is a single party member and a binary state $\theta \in \{0, 1\}$ with prior $\mathbb{P}(\theta = 1) = \mu_0 < 1/2$. The member first chooses an attention decision $\ell \in \{0, 1\}$:

- if $\ell = 1$, he observes a private signal $s \in \{0, 1\}$ with accuracy $q \in (1/2, 1)$, i.e. $\mathbb{P}(s = \theta \mid \theta) = q$;
- if $\ell = 0$, he observes no signal.

Attention incurs cost $k > 0$ if $\ell = 1$.

After choosing $\ell$ and possibly observing $s$, the member chooses a public action $x \in \{0, 1\}$, where $x = 1$ denotes publicly endorsing the party line and $x = 0$ denotes not endorsing. The public message is $m = x$. The party's payoff is

$$U_P \equiv \mathbb{P}(x = 1),$$

i.e. the party values public endorsement.[13]

**Member preferences.** The member's payoff is

$$u(x, \ell, s) = -c \cdot \mathbb{1}_{\{x=1,\ \ell=1,\ s=0\}} - k\ell,$$

where $c > 0$ is a moral (or psychological) cost of endorsing when the member is informed and his information indicates the unfavorable state ($s = 0$). If $x = 0$, the payoff is $-k\ell$.

[13] This reduced-form objective isolates the organizational channel: the party values public unity and endorsement, so the behavior of a strategic receiver is not part of the result. If one were to reintroduce an external receiver and the persuasion objective $\Pr(a = 1)$, a trade-off would arise between organizational unity and external persuasion. In particular, when $\mu_0 < 1/2$, uniform endorsements may reduce persuasion relative to more informative endorsement behavior. The point here is therefore only that the strategic-ignorance mechanism can still operate when receiver reactions are introduced, not that uniform messages always strengthen persuasion incentives.

**An ignorance-inducing organizational policy.** Consider the organizational policy that recommends endorsement $x = 1$ regardless of the state and regardless of the attention decision. (The party cannot observe $\ell$.)

I first show that, under this policy, remaining uninformed is optimal for the member.

**Claim 1** (Attention optimality.)**.** *Under the ignorance-inducing policy, choosing $\ell = 0$ strictly dominates choosing $\ell = 1$.*

*Proof.* If the member chooses $\ell = 0$, he pays no attention cost and, upon receiving the recommendation $x = 1$, endorses. Since the moral cost applies only when $\ell = 1$, his utility is $u = 0$.

If the member chooses $\ell = 1$, he pays $k$ and with probability $\mathbb{P}(s = 0) > 0$ observes $s = 0$. If he follows the recommendation $x = 1$, then whenever $s = 0$ he incurs moral cost $c$. Therefore his expected utility from choosing $\ell = 1$ and then endorsing is

$$\mathbb{E}[u \mid \ell = 1, x = 1] = -k - c\,\mathbb{P}(s = 0),$$

which is strictly negative since $k > 0$ and $\mathbb{P}(s = 0) > 0$. Hence $\ell = 0$ strictly dominates $\ell = 1$. □

We then verify that, conditional on remaining uninformed, obeying the party's recommendation is optimal.

**Claim 2** (Obedience given ignorance.)**.** *Conditional on $\ell = 0$, choosing $x = 1$ is optimal.*

*Proof.* If $\ell = 0$, endorsing yields utility 0, while deviating to $x = 0$ also yields utility 0. Thus endorsing is weakly optimal. □

Combining Claims 1 and 2, the ignorance-inducing policy leads to an outcome in which the member remains uninformed and endorses with probability one. Therefore the party achieves $U_P^{\mathrm{IG}} = 1$.

**Limits of full information acquisition.** Now consider any organizational policy under which the member acquires information, i.e. chooses $\ell = 1$ with probability one. Under full information acquisition, with positive probability the member observes $s = 0$. In that event, endorsing ($x = 1$) yields payoff $-c - k$, while not endorsing ($x = 0$) yields payoff $-k$. Hence, conditional on $(\ell = 1, s = 0)$, the member strictly prefers $x = 0$.

It follows that under any policy inducing full information acquisition,

$$\mathbb{P}(x = 1 \mid s = 0) = 0.$$

Because $\mathbb{P}(s=0) > 0$ under the signal structure, we obtain

$$U_P^{\mathrm{FI}} = \mathbb{P}(x=1) \leq \mathbb{P}(s=1) < 1.$$

Therefore no policy that induces full attention can yield party payoff 1.

**Strict comparison.** I have shown that $U_P^{\mathrm{IG}} = 1$ under the ignorance-inducing policy, while under any policy inducing full information acquisition, $U_P^{\mathrm{FI}} < 1$. Hence the party strictly prefers (and can induce) an organizational outcome with ignorance, establishing Proposition 4. □

## 6.6 Proof of Proposition 5

The proof proceeds in six steps. Steps 1 and 2 characterize collective accuracy as a function of the number of judgmental speakers and establish that each additional judgmental speaker strictly improves accuracy. Step 3 derives individual deviation incentives. Steps 4 and 5 establish the existence of two symmetric organizational outcomes (full party-line speech and full judgmental speech) under different parameter regimes. Step 6 compares their epistemic performance.

**Step 1: Define accuracy as a function of the number of judgmental speakers.** Fix $k \in \{0, 1, \ldots, N\}$, the number of agents who speak judgmentally. A judgmental message equals the agent's signal and is correct with probability $q > 1/2$. Party-line messages are uninformative (independent of $\theta$) and therefore do not contribute to correct inference. Under the maintained aggregation rule, the public effectively takes a majority vote over the $k$ informative messages. Hence

$$\alpha\left(\frac{k}{N}\right) = \Pr\left(\text{Binomial}(k, q) \geq \frac{k+1}{2}\right).$$

In particular, when $k = 0$, the public has no informative message and accuracy is $\alpha(0) = 1/2$. When $k = N$, the public aggregates $N$ informative messages, so $\alpha(1) > 1/2$ by the standard Condorcet jury logic for independent signals with accuracy $q > 1/2$.

**Step 2: The marginal contribution of the last judgmental speaker.** Recall that

$$\Delta(q) \equiv \alpha(1) - \alpha\left(\frac{N-1}{N}\right),$$

the accuracy gain from replacing the last party-line message by a judgmental one. I show that

$$\Delta(q) > 0$$

for $q > 1/2$ and odd $N$.

Let $N = 2m+1$ be odd. Compare the cases with $N-1 = 2m$ and $N = 2m+1$ judgmental speakers. Let

$$Y \sim \text{Binomial}(2m, q)$$

denote the number of correct signals among the first $2m$ speakers, and let

$$Z \sim \text{Bernoulli}(q)$$

denote the correctness of the $(2m+1)$-th speaker, independent of $Y$.

With $2m$ speakers, the majority threshold is $m+1$. With $2m+1$ speakers, the majority threshold is also $m+1$. Therefore

$$\alpha(1) = \Pr(Y + Z \geq m+1) \qquad \text{and} \qquad \alpha\left(\frac{N-1}{N}\right) = \Pr(Y \geq m+1).$$

Observe that

$$\{Y + Z \geq m+1\} = \{Y \geq m+1\} \cup \{Y = m \text{ and } Z = 1\}.$$

Hence

$$\Pr(Y + Z \geq m+1) = \Pr(Y \geq m+1) + \Pr(Y = m)\Pr(Z = 1),$$

so

$$\Delta(q) = \Pr(Y + Z \geq m+1) - \Pr(Y \geq m+1) = q\,\Pr(Y = m).$$

Since $q > 1/2$, we have $q > 0$, and since $Y$ has full support on $\{0, 1, \dots, 2m\}$,

$$\Pr(Y = m) > 0.$$

Therefore

$$\Delta(q) > 0.$$

**Step 3: Payoffs and deviation incentives.** Each agent chooses between: - party-line speech (no sanction) and - judgmental speech (sanction $d$). The payoff difference from choosing judgmental speech rather than party-line speech is:

$$\big[-d+\beta\cdot\mathbb{1}_{\{a=\theta\}}\big]-\big[r+\beta\cdot\mathbb{1}_{\{a=\theta\}}\big]=-(d-r)+\beta\cdot(\text{change in accuracy}).$$

**Step 4: Existence of a stable party-line organizational outcome.** If all agents choose party-line speech, then $k=0$ and accuracy is $\alpha(0)=1/2$. Under the large-$N$ non-pivotal environment described above, a unilateral deviation does not affect the distribution of the public decision and therefore does not change the expected value of the epistemic term $\beta\cdot\mathbb{1}_{\{a=\theta\}}$. The deviator therefore compares payoff $r$ (party-line) to $-d$ (judgment). Since $r\geq 0$ and $-d\leq 0$, deviating to judgment is never profitable. Hence full party-line speech is always a (weakly) stable symmetric organizational outcome under non-pivotality.

**Step 5: Existence of a stable judgmental organizational outcome.** If all agents choose judgmental speech, then $k=N$ and accuracy is $\alpha(1)$. Consider a unilateral deviation to party-line speech, which reduces the number of informative messages from $N$ to $N-1$. The deviator gains $r$ and avoids paying $d$, but reduces collective accuracy by exactly $\Delta(q)=\alpha(1)-\alpha\big(\frac{N-1}{N}\big)$. Hence deviating is unprofitable if $-d+\beta\alpha(1)\geq r+\beta\alpha\big(\frac{N-1}{N}\big)$, which is equivalent to $\beta\Delta(q)\geq r+d$. Under this condition, full judgmental speech is a symmetric organizational outcome.

**Step 6: Accuracy comparison.** Under full party-line speech, $k=0$ and accuracy is $\alpha(0)=1/2$. Under full judgmental speech, $k=N$ and accuracy is $\alpha(1)>1/2$. Hence collective decision accuracy is strictly lower under judgment collapse than under full judgmental speech. □

## 6.7 Proof of Theorem 2

The proof proceeds in two parts. Steps 1-2 establish part (i): for fixed $\varepsilon$, collective accuracy $\mathcal{P}(p,\varepsilon)$ is strictly increasing in the participation rate $p$ when $N$ is sufficiently large. Steps 3-5 establish part (ii): there exist parameters for which higher participation with garbling strictly dominates truthful communication with low participation.

**Setup and binomial structure.** Each of the $N$ potential informants communicates independently with probability $p$, so the number of communicated messages satisfies $K\sim\text{Binomial}(N,p)$. Conditional on $K=k$, the $k$ received messages are independent and correct with probability $q^\star(\varepsilon)>1/2$, so the number of correct messages is $\text{Binomial}(k,q^\star(\varepsilon))$.

**Step 1: Accuracy conditional on $k$.** Given $k$ informative messages, let $X \sim \text{Binomial}(k, q^\star(\varepsilon))$ denote the number of correct messages, where each message is independently correct with probability $q^\star(\varepsilon) > 1/2$. Under majority rule with random tie-breaking, collective accuracy is

$$\mathbb{P}(X > k/2) + \tfrac{1}{2}\mathbb{P}(X = k/2).$$

For $q^\star(\varepsilon) > 1/2$, this expression is weakly increasing in $k$. Intuitively, adding one additional independent signal with accuracy above one half cannot reduce the probability of a correct majority decision. Moreover, whenever the additional signal is pivotal (i.e., when the first $k$ signals produce a tie), the probability of a correct majority strictly increases. Hence accuracy is weakly increasing in $k$, with strict increases at values of $k$ for which tie events occur with positive probability.

**Step 2: Monotonicity in the participation probability $p$.** Unconditional accuracy is

$$\mathcal{P}(p, \varepsilon) = \sum_{k=0}^{N} \mathbb{P}(K = k) \cdot \Big[\mathbb{P}(X_k > k/2) + \tfrac{1}{2}\,\mathbb{P}(X_k = k/2)\Big],$$

where $X_k \sim \text{Binomial}(k, q^\star(\varepsilon))$.

As $p$ increases, the distribution of $K$ shifts upward in the sense of first-order stochastic dominance. Since the conditional accuracy term in brackets is weakly increasing in $k$ (by Step 1), it follows that $\mathcal{P}(p, \varepsilon)$ is weakly increasing in $p$. Moreover, for sufficiently large $N$, an increase in $p$ induces a nontrivial upward shift in the distribution of $K$ over values of $k$ at which conditional accuracy is strictly increasing, implying that $\mathcal{P}(p, \varepsilon)$ is strictly increasing in $p$. This establishes part (i).

**Step 3: Truthful but low participation yields accuracy close to $1/2$.** Fix $\varepsilon = 0$ (truthful communication) and choose $p_0 > 0$ small. Then $K$ is typically small, and in particular $\mathbb{P}(K = 0)$ is close to 1. When $K = 0$ the public receives no informative messages and must effectively guess, yielding accuracy $1/2$. Hence $\mathcal{P}(p_0, 0) \to \frac{1}{2} \quad \text{as} \quad p_0 \to 0$.

**Step 4: Higher participation with garbling yields accuracy close to 1.** Fix any $\varepsilon \in [0, 1/2)$ such that $q^\star(\varepsilon) > 1/2$, and choose $p_1 > p_0$. For fixed $p_1$, as $N \to \infty$ the law of large numbers implies $K/N \to p_1$ in probability. Conditional on $K = k$, majority accuracy converges to 1 as $k \to \infty$ when $q^\star(\varepsilon) > 1/2$. More precisely, by Hoeffding's inequality (Hoeffding, 1963),

$$\mathbb{P}\big(\text{Binomial}(k, q^\star(\varepsilon)) < \tfrac{k}{2}\big) \leq \exp\Big(-2k\big(q^\star(\varepsilon) - \tfrac{1}{2}\big)^2\Big),$$

which converges to 0 exponentially fast in $k$. Hence, $\mathcal{P}(p_1, \varepsilon) \to 1 \quad \text{as} \quad N \to \infty$.

**Step 5: Strict comparison.** Combining Steps 3 and 4, we can choose parameters $(N, q, p_0, p_1, \varepsilon)$ such that

$$\mathcal{P}(p_1, \varepsilon) > \mathcal{P}(p_0, 0).$$

In particular, higher participation under party organization can strictly improve collective accuracy even when communication is garbled ($\varepsilon > 0$), relative to truthful but low participation ($\varepsilon = 0$). This establishes part (ii). □

# 7 Appendix: Robustness of the Control Margin

For each $\lambda > 0$ let $\pi^\lambda$ denote an optimal party policy.

**Proposition 6** (Approximate pooling without discrete implementation)**.** *Suppose communication is not subject to discrete implementation constraints. Then for every $\varepsilon > 0$ there exists $\lambda(\varepsilon) > 0$ such that for all $\lambda \geq \lambda(\varepsilon)$, $H(M_{\pi^\lambda}) < \varepsilon$.*

*Consequently, as $\lambda$ increases, optimal communication becomes arbitrarily concentrated and public learning becomes arbitrarily small.*

*Proof.* Because actions and states are binary, party persuasion payoffs are bounded. Hence there exists $B > 0$ such that for any policy $\pi$, $|U_P(\pi)| \leq B$. For any policy,

$$U_P(\pi) - \lambda H(\pi) \leq B - \lambda H(\pi).$$

Pooling yields payoff at least $-B$ since $H = 0$. Fix $\varepsilon > 0$. If $H(\pi) \geq \varepsilon$, then for $\lambda > 2B/\varepsilon$,

$$B - \lambda\varepsilon < -B,$$

so such a policy cannot be optimal. Therefore, for sufficiently large $\lambda$, any optimal policy $\pi^\lambda$ must satisfy $H(M_{\pi^\lambda}) < \varepsilon$. □

Since entropy vanishes only when message distributions become degenerate, posterior beliefs induced by optimal policies converge to the prior.

# 8 Appendix: Judgment Collapse in a BCE Formulation

This appendix provides an equivalent Bayes correlated equilibrium (BCE) formulation of the judgment collapse logic developed in Section 3. The purpose is not to introduce new assumptions, but to show that the main-text channels can be expressed as an implementability result: when conformity incentives dominate, certain action profiles cannot be supported by any organizational information policy that satisfies obedience. Throughout, I use BCE as a feasibility and implementability concept (Bergemann and Morris, 2016).

A party chooses an information and recommendation policy that induces a joint distribution over the state and recommended actions. Each agent chooses an action $x_i \in \{J, P\}$, interpreted as *judgmental speech* ($J$) or *party-line speech* ($P$). An agent's utility is

$$u_i(x_i, \theta, a) = r \cdot \mathbb{1}_{\{x_i = P\}} - d \cdot \mathbb{1}_{\{x_i = J\}} + \beta \cdot \mathbb{1}_{\{a = \theta\}},$$

with parameters satisfying $d > r \geq 0$ and $\beta \geq 0$, as in Section 3.

**Non-pivotal assumption (large-$N$ limit).** I assume each individual agent is non-pivotal for the public decision, in the sense that conditional on the recommendation received, the agent's action does not affect the distribution of $(\theta, a)$. Equivalently, the public-accuracy term $\beta \cdot \mathbb{1}_{\{a=\theta\}}$ is identical under $J$ and $P$ given the agent's information.[14]

**Proposition 7** (Judgment exclusion under BCE feasibility)**.** *Under the non-pivotal assumption, if $d > r$, then in any BCE-feasible recommendation scheme the probability that an agent is recommended (and plays) $J$ is zero. Hence judgmental speech cannot appear in the support of any feasible party design.*

*Proof.* In a BCE-feasible scheme (Bergemann and Morris, 2016), recommendations must satisfy obedience: conditional on the information an agent receives and the recommendation, the agent must weakly prefer obeying the recommendation to deviating.

Fix any information set $\mathcal{I}_i$ for agent $i$ at which the party recommends $J$ with positive probability. Obedience requires that following the recommendation yields at least as high expected utility as deviating to $P$:

$$\mathbb{E}[u_i(J, \theta, a) \mid \mathcal{I}_i] \;\geq\; \mathbb{E}[u_i(P, \theta, a) \mid \mathcal{I}_i].$$

[14] This is the standard large-population approximation. A single agent's action has negligible influence on the public outcome. It mirrors the large-$N$ logic used in Proposition 5.

By the non-pivotal assumption, the distribution of $(\theta, a)$ conditional on $\mathcal{I}_i$ is the same regardless of whether the agent chooses $J$ or $P$. Therefore, the public-accuracy term cancels from both sides, yielding

$$-d + \beta\,\mathbb{E}[\mathbb{1}_{\{a=\theta\}} \mid \mathcal{I}_i] \;\geq\; r + \beta\,\mathbb{E}[\mathbb{1}_{\{a=\theta\}} \mid \mathcal{I}_i] \quad \Longleftrightarrow \quad -d \geq r.$$

This contradicts the assumption $d > r$. Hence no information set can recommend $J$ with positive probability under any obedient (BCE-feasible) scheme. It follows that in any BCE-feasible party policy, judgmental speech is excluded from the support. □

# References


Aldrich, J. H. (1995), *Why Parties? The Origin and Transformation of Political Parties in America*, University of Chicago Press, Chicago.

Arendt, H. (1971), 'Thinking and moral considerations: A lecture', *Social Research* **38**(3), 417–446.

Arendt, H. (1978), *The Life of the Mind*, Harcourt Brace Jovanovich, New York.

Austen-Smith, D. and Banks, J. S. (1988), 'Elections, coalitions, and legislative outcomes', *American Political Science Review* **82**(2), 405–422.

Barber, M. and McCarty, N. (2015), *Causes and Consequences of Polarization*, Cambridge University Press.

Barber, M. and Pope, J. C. (2019), 'Does party trump ideology? disentangling party and ideology in america', *American Political Science Review* **113**(1), 38–54.

Bergemann, D. and Morris, S. (2016), 'Bayes correlated equilibrium and the comparison of information structures in games', *Theoretical Economics* **11**(2), 487–522.

Besley, T. (1997), 'A political economy model of representative democracy', *Quarterly Journal of Economics* **112**(1), 85–114.

Blackwell, D. (1951), 'Comparison of experiments', *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* pp. 93–102.

Blackwell, D. (1953), 'Equivalent comparisons of experiments', *The Annals of Mathematical Statistics* **24**(2), 265–272.

Bolsen, T., Druckman, J. N. and Cook, F. L. (2020), 'The influence of partisan elite messages on public trust in science', *Political Behavior* **42**(4), 919–945.

Costa, O. and Frenkiel, É. (2023), 'Centralization and the decline of parliamentary deliberation in france', *West European Politics* **46**(5), 945–968.

Cowley, P. (2015), 'The rebels: How blair mislaid his majority', *British Journal of Political Science* **45**(1), 1–25.

Cowley, P. and Stuart, M. (2023), 'Party discipline in the uk parliament after brexit', *Parliamentary Affairs* **76**(2), 215–238.

Gougou, F. and Persico, S. (2024), 'Party control and the expression of policy disagreement in the french national assembly', *European Journal of Political Research* **63**(1), 112–131.

Grossman, M. and Hopkins, D. A. (2016), 'Asymmetric politics: Ideological republicans and group-interest democrats', *Perspectives on Politics* **14**(3), 593–617.

Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American Statistical Association* **58**(301), 13–30.

Kamenica, E. and Gentzkow, M. (2011), 'Bayesian persuasion', *American Economic Review* **101**(6), 2590–2615.

Martinache, I. and Sawicki, F., eds (2020), *La fin des partis ?*, La Vie des idées, Presses Universitaires de France, Paris.

Mathevet, L., Perego, J. and Taneva, I. (2020), 'On information design in games', *Journal of Political Economy* **128**(4), 1370–1404.

Morelli, M. (2004), 'Party formation and policy outcomes under different electoral systems', *Review of Economic Studies* **71**(3), 829–853.

Nyhan, B. and Reifler, J. (2015), 'The effect of fact-checking on elites', *American Journal of Political Science* **59**(3), 628–641.

Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell System Technical Journal* **27**(3), 379–423.

Snyder, J. M. and Ting, M. M. (2002), 'An informational rationale for political parties', *American Journal of Political Science* **46**(1), 90–110.

Spirling, A. and McLean, I. (2024), ‘Elite coordination and the language of party unity’, *Political Analysis* **32**(1), 1–18.

Taneva, I. (2019), ‘Information design’, *American Economic Journal: Microeconomics* **11**(4), 151–185.

Weil, S. (1957), Note sur la suppression générale des partis politiques, *in* ‘Ecrits de Londres et dernières lettres’, Gallimard, Paris. Essay written in 1943 and first published posthumously in 1950.

Weil, S. (2013), *On the Abolition of All Political Parties*, New York Review Books, New York.