INFORMED COLLECTIVE DECISION MAKING VIA VOTING

By

QISHEN HAN

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Lirong Xia and David Pennock

And approved by

New Brunswick, New Jersey

October 2026

ABSTRACT OF THE DISSERTATION

Informed Collective Decision Making via Voting

by QISHEN HAN

Dissertation Director:

Lirong Xia and David Pennock

This dissertation addresses group decision-making where agents have subjective preferences that vary across social groups and objective preferences that depend on an unknown ground truth. Classical axiomatic social choice studies how to aggregate preferences, and epistemic social choice studies how voting aggregates private information, but few study these two challenges in a unified framework. Moreover, previous work has shown that strategic agents have incentives to deviate from the behavior that aggregates the information. My research builds a framework that takes subjective preferences, objective information, and strategic behavior into account together. We model each agent's utility as a combination of a subjective component (depending only on the alternative) and an objective component (depending also on the ground truth). A "good" decision is the informed majority decision – the one a majority would favor when the ground truth is common knowledge. We then conduct equilibrium analysis to address two questions. (1) What structures do equilibria exhibit under strategic voting? (2) Are strategic behaviors a benefit or a barrier to making informed majority decisions? The first result provides a surprisingly positive answer to the second question: when agents share aligned objective preferences, strategic behavior can actually benefit informed decisions, as every equilibrium in the majority vote reaches the informed majority decision. This contrasts with non-strategic voting, which requires additional restrictive assumptions to reach the informed majority decision. The second result proposes a polling-voting two-round mechanism. Like the majority vote, every equilibrium in the mechanism reaches the informed majority decision. Moreover, it induces a more natural equilibrium, in which agents share information in the polling stage and vote for their updated preferences in the voting stage. The equilibrium is supported by AI-simulated experiments, where LLM agents play this strategy without an explicit hint. The last result studies a more challenging setting in which groups of agents have opposing objective preferences contingent on the ground truth and can coordinate only within limited groups. It characterizes when equilibria reach the informed majority decision in terms of the relative group sizes and the coalition size limit, and identifies the equilibrium strategies.

ACKNOWLEDGMENTS

For nothing is worthy of man as man unless he can pursue it with passionate devotion.

— Max Weber, *Science as a Vocation*

I owe my deepest thanks to my advisor, Professor Lirong Xia. I have learned more from him over these five years than I could hope to put into words. He taught me how to do research with rigor and how to stand on my own as a scholar, but what stayed with me most was the example he set for doing the right thing, even when it was not the easy thing. His judgment, his patience, and his integrity have shaped this dissertation and, more than that, the kind of researcher and person I want to be.

I am equally grateful to my co-advisor, Professor David Pennock. He always encourages me toward the questions I genuinely cared about, and he gave me clear-eyed advice when it came time to think about what comes next. His trust in me meant a great deal.

My thanks go to my committee members, Professor Kangning Wang and Professor Bo Waggoner, who generously agreed to serve and whose careful feedback made this a better dissertation.

I have been lucky to work alongside and learn from a remarkable group of people. I owe a special debt to Professor Xiaotie Deng and Professor Yuqing Kong, who first opened the door to economics and computation for me. I am grateful to Professor Grant Schoenebeck, Professor Biaoshuai Tao, Professor Edith Elkind, and Professor Ulle Endriss, researchers I admire, who shaped my thinking at different points along the way. I also want to thank Ying Wang and Ao Liu, collaborators I could always rely on; we worked hard together, and we grew up together in this field.

Behind every smooth-running day were people whose work too often goes unseen. I am grateful to Tracy and Shannon at RPI, and to Professor Jie Gao, Mychelle, and Nicole at

Rutgers, whose care and dedication spared me countless headaches and let me focus on the work itself.

To my family: thank you for carrying me through this PhD and into the start of my career, each of you in your own way. You are my harbor and my shield. Wherever I go and whatever happens, you are where I return to and the ones who stand behind me. I am proud of you, and I hope I have given you reason to be proud of me. And to my closest friends, Yusong Li, Qifan Ren, Ruizhe He, and Tieqi Yue: we have scattered across the world, but we have stayed close and kept pushing forward together.

Even now, I feel the pull of everything that remains uncertain about the future. But with all of you behind me, I can make informed decisions and take the next steps with more confidence than I would have found on my own.

ACKNOWLEDGMENT OF PREVIOUS PUBLICATIONS

The dissertation comprises the following publications, in which the student is the first author and has made significant contributions. Specifically, Chapter 3 is adapted from P1, Chapter 4 is adapted from P2, and Chapter 5 is adapted from P3.

P1 Qishen Han, Grant Schoenebeck, Biaoshuai Tao, and Lirong Xia. 2023. The Wisdom of Strategic Voting. In Proceedings of the 24th ACM Conference on Economics and Computation (EC '23). Association for Computing Machinery, New York, NY, USA, 885−905.

P2 Qishen Han, Grant Schoenebeck, Biaoshuai Tao, and Lirong Xia. 2026. The Art of Two-Round Voting. In Proceedings of the ACM Web Conference 2026 (WWW '26). Association for Computing Machinery, New York, NY, USA, 135−145.

P3 Qishen Han, Grant Schoenebeck, Biaoshuai Tao, and Lirong Xia. 2026. Aggregating Information and Preferences under Different Coordination Ability. In Proceedings of the ACM Web Conference 2026 (WWW '26). Association for Computing Machinery, New York, NY, USA, 123−134.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND NOTATIONS

BNE	Bayes-Nash Equilibrium *or* Bayesian Nash Equilibrium.
CJT	Condorcet Jury Theorem
LLM	Large Language Models
(ε, k)**-EBSE**	ε-ex-ante Bayesian k-strong equilibrium
n	the total number of agents
$\mathbf{A}, \mathbf{R}$	alternatives
$\mathcal{W} = \{L, H\}$	the set of world states
$\mathcal{S} = \{\ell, h\}$	the set of private signals
P_k	the prior belief on the probability of world state being k
P_{mk}	the probability of receiving signal m under world state k
μ	threshold of the majority vote
$v_i(k, d)$	the utility of agent i for alternative d under world state k
$v_i^S(d)$	the subjective utility of agent i
$v_i^O(k, d)$	the objective utility of agent i
B	the upper bound of the utility function
$\alpha_F, \alpha_U, \alpha_C$	the approximated fraction of three types of agents
σ, Σ	strategy and strategy profile
β_h, β_h	(in a strategy) the probability of voting for $\mathbf{A}$ when receiving signal h (l)
$\{\Sigma_n\}_{n=1}^{\infty}$	a sequence of strategy profiles
$\lambda_k^{\mathbf{A}}(\Sigma), \lambda_k^{\mathbf{R}}(\Sigma)$	the probability that $\mathbf{A}$ ($\mathbf{R}$) becomes the winner under world state k

$A(\Sigma)$	fidelity: probability that Σ leads to the informed majority decision
$u_i(\Sigma)$	the (ex-ante) expected utility of agent i
f^n_H, f^n_L, f^n	the excess expected vote share
$\mathcal{I}$	an instance of a voting game
$\{\mathcal{I}_n\}_{n=1}^{\infty}$	a sequence of instances
$\hat{\sigma}, \hat{\Sigma}$	(Chapter 4) strategy and strategy profile for one-round voting
X	(Chapter 4) number of votes for **A** in the first round.
$\xi = k/n$	(Chapter 5) fractional coordination ability.
α	(Chapter 5) the approximated fraction of majority type agents.
θ	largest fraction of majority agents such that an $o(1)$-strong BNE does not exist.
φ_h, φ_ℓ	(Chapter 5) expected vote share
γ	(Chapter 5) fraction of type-0 minority agents

CHAPTER 1
INTRODUCTION

When groups must make decisions, particularly under time pressure, voting provides a procedurally fair method to aggregate preferences and knowledge. In our polarized, AI-integrated world plagued by misinformation and echo chambers, making fair, equitable, informed, and incentive-aware decisions has become both more critical and more challenging.

In many voting scenarios, agents are embedded in distinct social environments – shaped by cultural background, economic status, or ideological affiliation – and access information through different networks that may be distorted by information bias. This social and informational heterogeneity gives rise to a natural decomposition of preferences: *subjective preferences* that depend on agents' social groups (e.g., young or old, left-wing or right-wing), and *objective preferences* tied to external ground truth. The incomplete and noisy nature of agents' information about these external factors creates a fundamental tension between the preferences they perceive themselves to have and those they would hold under complete information.

Example 1 (COVID Policy)**.** *Suppose the voters vote to decide the policy on the COVID-19 pandemic. The two choices are to accept the more restrictive policy (Accept) and to keep the status quo (Reject). The consequence of the policy depends on the fact that the COVID virus is of high or low risk, and more people tend to accept the policy when COVID is of high risk than when COVID is of low risk. The voters do not know the risk level of the virus directly. Instead, every voter forms a private judgment on the risk level based on his/her own information sources. Voters may have different opinions on whether to accept the policy which may or may not depend on the risk level. Can the voters achieve a good decision via the majority vote?*

Similar challenges arise in political elections, corporate board decisions, expert panels, and human-AI collaboration across domains from healthcare to finance. Motivated by these scenarios, the question we care about is

How can we make an informed, efficient, fair, and incentive-aware decision?

Successfully addressing this problem requires tackling three interconnected dimensions as well as challenges: preferences, aggregation, and incentives. **Preference modeling** involves developing formal representations of agents' subjective preferences alongside their uncertain objective preferences about the unknown ground truth (or, equivalently, *world state*). **Aggregation** concerns defining appropriate notions of efficiency and fairness when combining both preference types, and designing mechanisms that achieve these goals. The **incentive challenge** addresses understanding how strategic behavior affects decision outcomes when agents may misrepresent both their preferences and private information. These challenges are deeply intertwined, as solutions to one dimension constrain and inform approaches to the others.

One might expect these challenges to be readily addressed by extending existing frameworks. The celebrated Condorcet Jury Theorem [1] provides guarantees for aggregation, showing that majority rule will choose the alternative that agents would prefer if they all knew the underlying ground truth. However, the combination of subjective and objective preferences introduces fundamental complications that existing theory cannot easily accommodate. This observation, however, presumes a homogeneous setting in which voters share the same preferences and disagree only about that ground truth. When agents possess different preferences towards the outcomes, there is no unanimous "correct" decision with respect to the agents' preferences. Therefore, instead of revealing the true state of the world, the goal is to achieve *informed majority decision*, which is the decision favored by the majority of the agents if the world state were known to them.

On the other side, social choice theory offers a rich arsenal of axioms that characterize fair and efficient decision procedures [2]. Under two-alternative scenarios, majority voting ensures truthful reporting of preference [3]. However, these axioms presuppose agents have well-defined, complete preferences rather than uncertainty about their own objective preferences. Most critically, when agents have preferences decided by uncertain world states, the surprising result by Austen-Smith and Banks [4] shows that even in binary voting, informative voting may fail to form a Nash equilibrium. The key insight is that an agent's vote makes a difference only when all other votes form a tie, which means that when an agent strategically thinks about his/her vote, effectively he/she gains more information about the ground truth (by assuming that other votes are tied). This is illustrated in the following example.

Example 2. *Consider an instance of the COVID policy problem, where the utility of the agents and signal distribution of different risk levels are shown in the tables below.*

State	*High Signal*	*Low Signal*
High Risk	*0.9*	*0.1*
Low Risk	*0.4*	*0.6*

Table 1.1: Signal distributions.

State	*Accept*	*Reject*
High Risk	*1*	*0*
Low Risk	*0*	*1*

Table 1.2: Agents' utilities.

Suppose all but one agent are informative and the remaining agent is strategic. Informative agents vote for Accept when they receive a high signal and Reject when they receive a low signal. The strategic agent only cares about the pivotal case where exactly half of the informative agents vote for Accept. However, given that agents receive high signals with a probability of 0.9 given the risk level being high, the pivotal case implies a high probability that the risk level is low, and the strategic agent will vote for Reject even after receiving a high signal.

The above deviation of strategic binary voting with preferences endogenously affected by the unobservable world state sharply contrasts with the axiomatic social choice where preferences are exogenously given. In the latter case, the majority rule is strategy-proof, while the former case has attracted a large literature on the binary voting problem in game-theoretic contexts, examining the impact of strategic behavior. When agents have homogeneous preferences, and the goal is to reveal the ground truth, Wit [5] and Myerson [6] show that a selected equilibrium with mixed strategy reveals the world state with high probability. Feddersen and Pesendorfer [7] show the existence of such an equilibrium in any non-unanimous voting, while in unanimous voting strategic voting has a constant probability to make a mistake. And for the informed majority decision, Feddersen and Pesendorfer [8] adopt a model with continuous world states and an asymptotically large number of agents whose preferences are drawn from a distribution with full support on a continuum and show that the equilibrium is unique and always leads to the informed majority decision with high probability. Schoenebeck and Tao [9] propose a mechanism incentivizing informative voting from agents and leads to the informed majority decision with high probability.

Nevertheless, there are three aspects not addressed by previous work. Firstly, previous work (except for Schoenebeck and Tao [9]) focuses on Nash equilibrium, which allows only individual manipulation. In real-world scenarios, on the other hand, such strategic manipulation often occurs in a *coalition* of agents. Coalitional manipulation is more powerful than individual manipulation as it allows multiple agents to coordinate and deviate at the same time. The following example shows that a Nash equilibrium is still prone to a group of manipulators in binary voting.

Example 3. *Consider an instance with three agents, whose utility is shown as follows.*

Agent	(High, Accept)	(High, Reject)	(Low, Accept)	(Low, Reject)
1	1	0	1	0
2	1	0	0	1
3	0	1	0	1

Table 1.3: The utility of three agents under different states and decisions.

The following strategy profile is a Nash equilibrium: agent 1 always votes for Accept, agent 2 votes informatively, and agent 3 always votes for Reject. Here, agent 1 and 3 play their dominant strategies, and, consequently, informative voting is the best strategy for agent 2.

However, this strategy profile is dominated by the profile where all agents vote informatively. Under informative voting, the decision in accordance with the state (Accept in the High state, and Reject in the Low state) is selected with a larger probability, and the utility of agent 2 increases. On the other hand, the overall probability of choosing to Accept or Reject does not change, so agent 1 and 3's utilities remain the same.

Secondly, previous work focuses on the existence of certain equilibria that achieve the goal (revealing the world state or reaching the informed majority decision). However, the existence of multiple equilibria [5], including "bad equilibria" that do not lead to the goal, makes the behavior of strategic agents unpredictable, as it is uncertain which equilibrium agents will play. One response to multiple equilibria is to select an equilibrium that is more "natural" or "reasonable" than others, named *equilibrium selection*. However, equilibrium selection cannot guarantee that agents will play the selected equilibrium, as it is unclear which equilibrium is more "natural" or "reasonable" in many scenarios, and agents may not agree on a "more natural" equilibrium even if it exists.

Finally, most previous work sets limitations on agents' preference profiles, such as homogeneous preferences and other domain restrictions. A more expressive preference model is needed for further analysis.

1.1 Our Contribution

My research develops a unified game-theoretic framework that addresses all three challenges through interconnected theoretical innovations. For preference modelling, the foundation is a combined utility model that formally incorporates both subjective and objective preferences under uncertainty, providing the mathematical structure needed to analyze this complex decision environment. For aggregation, we focus on the majority rule, which is the simplest yet most widely applied aggregation method in real-world practices. We adopt an informed majority decision as the criterion for decisions that would emerge under complete information. Finally, we provide a comprehensive strategic analysis through game-theoretic characterizations of when majority voting achieves informed decisions despite agents' strategic behavior, revealing the conditions under which simple mechanisms can overcome complex incentive problems. This work bridges social choice theory and information economics, establishing theoretical foundations for fair and efficient group decision-making in our information-rich but uncertainty-laden world.

Specifically, my research addresses two central questions: **(1)What structures do equilibria exhibit under strategic voting? (2) Are strategic behaviors a benefit or a barrier to making informed majority decisions?**

Aligned Objective Utility + Majority Rule (Chapter 3). The first part provides a surprisingly affirmative answer to the second question: strategic behavior can actually benefit informed decisions. The work studies a population where agents share aligned objective preferences – for example, agents prefer approving a project when it is truly beneficial and rejecting it when harmful – while their subjective preferences may differ substantially. The main result shows that asymptotically, every equilibrium under majority voting achieves the informed majority decision, and conversely, any outcome that reaches an informed majority decision constitutes an equilibrium. Moreover, such equilibria are guaranteed to exist. This finding sharply contrasts with classical non-strategic voting behaviors [4], where additional restrictive assumptions are required to achieve informed majority decisions. Thus, strategic behavior emerges as a benefit rather than a barrier to informed group decision-making. Regarding the first question, the work characterizes equilibrium structures by analyzing the expected vote shares under different ground truth states, providing a complete answer to how equilibria manifest in this setting.

Majority vote with polling + common objective utility (Chapter 4). The second part shows that adding a polling stage before the vote produces a clearer answer to the first question – equilibrium structures become more natural and intuitive – while maintaining the positive answer to the second question. The key insight is that the first-round poll does not determine the winner but serves as a communication phase where agents share information. Once better informed about the ground truth through observed polling results, agents can make more accurate decisions in the final vote. Following this idea, we demonstrate that the polling-voting mechanism admits a pure, symmetric, and easily computable equilibrium: agents truthfully reveal their signals in the polling round, then vote for their expected preferred outcome in the voting round based on the aggregated information. This characterization provides a much more interpretable answer to what equilibria look like compared to the randomized or asymmetric equilibria in standard majority voting. Moreover, every equilibrium in this mechanism still achieves the informed majority decision, confirming that strategic behavior remains beneficial. To validate these theoretical findings, the work conducts experiments with large language model agents, demonstrating empirically that the two-round mechanism aggregates information more effectively than single-round voting. Notably, LLM agents naturally follow the proposed equilibrium strategy without explicit instruction, suggesting the equilibrium captures intuitive reasoning patterns. This makes the two-round mechanism not only as effective as standard majority voting, but also more transparent and natural in its equilibrium behavior.

Majority Vote + Opposite objective utility + bounded-size strategic behavior (Chapter 5). The third part addresses both questions in a more challenging setting where different groups of agents have opposing objective preferences contingent on the ground truth. Unlike populations with common objective preferences, strong equilibria may not exist when the two groups have similar sizes [10]. Motivated by the observation that real-world agents face practical limits on coordination capabilities, the work restricts attention to equilibria that are stable only against deviations by coalitions of bounded size. When deviators' coordinating power is limited, strategic behavior in the majority vote can still reach efficient and fair outcomes. The work characterizes the boundary of when equilibria achieve informed majority decisions based on the relative sizes of the two groups and the coalition size limit –

answering whether strategic behavior helps (addressing the second question) – and identifies what these equilibrium strategies are (addressing the first question). These results contrast with the earlier work on aligned preferences, where strong equilibria always exist, highlighting both how population preference structures affect the complexity of achieving informed group decisions and the rich interplay between preference heterogeneity, coordination constraints, and decision quality.

CHAPTER 2
PRELIMINARIES AND BACKGROUND

2.1 Setting

Alternatives and World States. n agents vote for two alternatives $\mathbf{A}$ (standing for "accept") and $\mathbf{R}$ (standing for "reject"). There are $K = 2$ possible world states $\mathcal{W} = \{L, H\}$ (standing for "low risk" and "high risk" respectively), where $\mathbf{A}$ is more preferred in H, and $\mathbf{R}$ is more preferred in L. We use k to denote a generic world state. The world state is not directly observable by the agents. Let $P_H = \Pr[W = H]$ and $P_L = \Pr[W = L]$ be the common prior of the world states. We assume $P_H > 0$ and $P_L > 0$.

Private Signals. Every agent receives a signal in $\mathcal{S} = \{l, h\}$. We use m to denote a generic signal, and Ψ_i to denote the random variable representing the signal that agent i receives. We assume the signals agents receive are independent and have identical distributions conditioned on the world state. Let $P_{mk} = \Pr[\Psi_i = m \mid W = k]$ be the probability that an agent receives signal m under world state k. The signal distributions $((P_{hH}, P_{lH}), (P_{hL}, P_{lL}))$ are also common knowledge. We assume that the signals are positively correlated to the world states. Specifically, we have $P_{hH} > P_{hL}$ and $P_{lH} < P_{lL}$. On the other hand, we allow biased signals and DO NOT assume $P_{hH} > P_{lH}$ or $P_{hL} < P_{lL}$.

Majority Vote. This paper considers the majority vote with threshold μ. Each agent i votes for $\mathbf{A}$ or $\mathbf{R}$. If at least $\mu \cdot n$ agents vote for $\mathbf{A}$, $\mathbf{A}$ is announced to be the winner; otherwise, $\mathbf{R}$ is announced to be the winner.

Combined Utility Model. Each agent i has a utility which is a function of the true world state and the outcome of the vote. Formally, we have $v_i : \mathcal{W} \times \{\mathbf{A}, \mathbf{R}\} \to \{0, 1, \ldots, B\}$, where B is the positive integer upper bound. The combined utility functions model both subjective and objective preferences as components of agent utilities. The utility function consists of a subjective and an objective component,

$$v_i(k,d) = v_i^S(d) + v_i^O(k,d).$$

v_i^S is agent i's subjective utility that only depends on winner d and is independent of the ground truth k. v_i^O is agent i's objective utility that depends on both the ground truth k and the alternative d. This is a rich model that naturally extends previous models in axiomatic social choice (where $v_i^O \equiv 0$) and epistemic social choice, where $v_i^S \equiv 0$ and v_i^O's are the same for all agents as in Example 2. While the separation between $u_i^S(d)$ and $u_i^O(k,d)$ appears mathematically redundant, it allows us to clearly model the sources of preferences.

Informed Majority Decision The goal of the voting is to output the informed majority decision, which is the alternative favored by the majority of the agents if the world state were known. The informed majority decision shares the same threshold μ as the majority vote threshold. If $\mathbf{A}$ is preferred by at least $\mu \cdot n$ agents, then $\mathbf{A}$ is the informed majority decision; otherwise, $\mathbf{R}$ is the informed majority decision.

Example 4. *Consider the COVID policy-making scenario. $n = 20$ voters decide whether to accept (denoted as $\mathbf{A}$) or reject (denoted as $\mathbf{R}$) the more restrictive policy. The world state $\{L, H\}$ describes the real risk level of the virus. $W = H$ means high risk level, and $W = L$ means low risk level. The voters' beliefs form a common prior based on some preliminary reports. Suppose $P_H = 0.4$ and $P_L = 0.6$, which means the risk level has a prior probability of 0.4 to be high.*

Every voter receives a private signal l or h from his/her information sources. The signals somehow reflect the risk level but are noisy. Suppose this is a biased scenario (for example, there has been a boost of positive cases in the past week), and members are always more likely to receive the high signal. For example, $P_{hH} = 0.8$ and $P_{hL} = 0.6$, i.e., a voter will receive *an h signal with probability 0.8 if the risk level is high and receive an h signal with probability 0.6 if the risk level is low.*

The majority vote threshold is $\mu = 0.6$. Therefore, $\mathbf{A}$ *is the winner if and only if at least 12 voters vote for it.*

We assume there are three groups of agents, and agents in each group share the same utility function (which may not be true in general), shown in Table 2.1. The informed majority decision depends on the world state: "accept" is the informed majority decision if the world state is H, and "reject" is the informed majority decision if the world state is L.

(World State, Winner)	$(H, \mathbf{A})$	$(L, \mathbf{A})$	$(H, \mathbf{R})$	$(L, \mathbf{R})$
Group 1: 4 agents	8	6	2	4
Group 2: 6 agents	3	1	5	8
Group 3: 10 agents	3	2	1	8

Table 2.1: Utility of agents in Example 4.

Table 2.2 gives a possible partition of subjective and objective utilities for agents. In the objective preferences, all groups prefer $\mathbf{A}$ *in world state H and* $\mathbf{R}$ *in world state L.*

	Subjective Preference		Objective Preference			
	$v^S(\mathbf{A})$	$v^S(\mathbf{R})$	$v^O(H, \mathbf{A})$	$v^O(L, \mathbf{A})$	$v^O(H, \mathbf{R})$	$v^O(L, \mathbf{R})$
Group 1	6	2	2	0	0	2
Group 2	1	5	2	0	0	3
Group 3	2	1	1	0	0	7

Table 2.2: Utility Partition of agents in Table 2.1

Strategy. A (mixed) strategy is a mapping from the agent's private signal to a distribution on $\{\mathbf{A}, \mathbf{R}\}$. For a set S, let $\Delta(S)$ be the set of all possible distributions on S. Formally, an agent i's strategy $\sigma_i : \mathcal{S} \to \Delta(\{\mathbf{A}, \mathbf{R}\})$. A strategy can be represented as a vector $\sigma = (\beta_l, \beta_h)$, where β_m is the probability that the agent votes for $\mathbf{A}$ when receiving signal m. A strategy profile is the vector of strategies of all agents. $\Sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. We call a strategy profile Σ *a symmetric strategy profile induced by strategy* σ if all agents play the same strategy σ in Σ.

Here, we introduce two classical voting behaviors: informative voting and sincere voting. They have been viewed as the most "natural" behavior of non-strategic voters. On the other hand, Austen-Smith and Banks [4] show that they do not necessarily form an equilibrium.

Definition 1. An informative strategy is $\sigma = (0, 1)$, i.e. voting for $\mathbf{A}$ when receiving h and voting for $\mathbf{R}$ when receiving l. A strategy profile is informative when every agent votes informatively.

Definition 2 (Sincere Voting)**.** A sincere agent votes as if he/she is making an individual decision. The agent compares his/her expected utility conditioned on his/her private signal m and on $\mathbf{A}/\mathbf{R}$ becoming the winner, respectively, defined as follows.

$$u_i(\mathbf{A} \mid m) = \Pr[W = L \mid m] \cdot v_i(L, \mathbf{A}) + \Pr[W = H \mid m] \cdot v_i(H, \mathbf{A}).$$
$$u_i(\mathbf{R} \mid m) = \Pr[W = L \mid m] \cdot v_i(L, \mathbf{R}) + \Pr[W = H \mid m] \cdot v_i(H, \mathbf{R}).$$

We call this expected utility the *sincere expected utility* of the agent. If $u_i(\mathbf{A} \mid m) \geq u_i(\mathbf{R} \mid m)$, the agent votes for $\mathbf{A}$; otherwise, the agent votes for $\mathbf{R}$.

Fidelity and Expected Utility. Given a strategy profile Σ, let $\lambda_k^{\mathbf{A}}(\Sigma)$ ($\lambda_k^{\mathbf{R}}(\Sigma)$, respectively) be the (ex-ante, before agents receiving their signals) probability that $\mathbf{A}$ ($\mathbf{R}$, respectively) becomes the winner when the world state is k.

Definition 3 (**Fidelity**)**.** Fidelity is the likelihood that the informed majority decision is reached. In our setting, the fidelity when agents play strategy profile Σ is

$$A(\Sigma) = P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma) + P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma).$$

We use the word *fidelity* to distinguish the notion from *accuracy*, which usually denotes the likelihood that the correct world state is revealed.

The (ex-ante) expected utility of an agent i exclusively depends on $\lambda_k^{\mathbf{A}}(\Sigma)$ and $\lambda_k^{\mathbf{R}}(\Sigma)$:

$$u_i(\Sigma) = P_L(\lambda_L^{\mathbf{A}}(\Sigma) \cdot v_i(L, \mathbf{A}) + \lambda_L^{\mathbf{R}}(\Sigma) \cdot v_i(L, \mathbf{R})) + P_H(\lambda_H^{\mathbf{A}}(\Sigma) \cdot v_i(H, \mathbf{A}) + \lambda_H^{\mathbf{R}}(\Sigma) \cdot v_i(H, \mathbf{R})).$$

Instance and Sequence of Strategy Profiles. We define an instance $\mathcal{I}$ of a voting game on the agent number n, the majority vote threshold μ, the world state prior distribution (P_L, P_H), the signal distributions $((P_{hH}, P_{lH}), (P_{hL}, P_{lL}))$, the utility functions of all the agents $\{v_i\}_{i=1}^n$, and the approximated fraction of each type $(\alpha_F, \alpha_U, \alpha_C)$. Let $\{\mathcal{I}_n\}_{n=1}^{\infty}$ (or $\{\mathcal{I}_n\}$ for short) be a sequence of instances, where each $\mathcal{I}_n$ is an instance of n agents. The instances in a sequence share the same parameters

$$\{\mu, (P_L, P_H), ((P_{hH}, P_{lH}), (P_{hL}, P_{lL})), (\alpha_F, \alpha_U, \alpha_C)\}.$$

We do not regard agents in different instances as related and have no additional assumption on the utility functions of agents.

We define a sequence of strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$ on an instance sequence $\{\mathcal{I}_n\}$. Similarly, we do not have additional assumptions about the agents. Therefore, for different instances in the sequence, the strategies and utility functions of agents can be drastically different. A strategy profile sequence $\{\Sigma_n\}$ is symmetric and induced by strategy σ if every strategy profile Σ_n in the sequence is a symmetric strategy profile induced by σ.

2.1.1 Solution Concept

ε-strong Bayes-Nash Equilibrium. In this paper, we use the solution concept of ε-strong Bayes-Nash Equilibrium, an approximation of strong Bayes-Nash Equilibrium where no group of agents can increase their utilities by more than ε through deviation. A strategy profile $\Sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ is an *ε-strong Bayes-Nash Equilibrium* (ε-strong BNE) if there does not exist a subset of agents D and a strategy profile $\Sigma' = (\sigma_1', \sigma_2', \cdots, \sigma_n')$ such that

1. $\sigma_i = \sigma_i'$ for all $i \notin D$;

2. $u_i(\Sigma') \geq u_i(\Sigma)$ for all $i \in D$; and

3. there exists $i \in D$ such that $u_i(\Sigma') > u_i(\Sigma) + \varepsilon$.

We call such a group D a deviating group and the strategy profile Σ' a deviating strategy. When such D and Σ' exist (so that Σ is not an equilibrium), we say group D deviates successfully.

By definition, when $\varepsilon = 0$, the equilibrium is a strong Bayes-Nash equilibrium in which no group of agents can strictly increase their utilities by deviating. In Section 5, we will discuss the extension of ε-strong BNE into scenarios where agents have different coordination capabilities reflected by the largest deviating group agents can form.

2.2 Related Work

2.2.1 The Condorcet Jury Theorem and Its Generalizations

The celebrated Condorcet Jury Theorem (CJT) [1] initiated the study of epistemic social choice, where voting aims to identify a "correct" decision that agents cannot directly observe; it has "formed the basis for the development of social choice and collective decision-making as modern research fields" [11]. Condorcet considers a scenario in which agents receive independent yet noisy signals about an unknown state and seek to recover that state via majority voting, and shows that if agents honestly reveal their signals (informative voting), and each agent is more likely than not to receive the correct signal, the probability of the majority vote coinciding with the ground truth converges to one as the number of agents increases. A vast literature has followed Condorcet's path and extended the CJT to more general settings [12, 13, 14, 15], including agents with correlated private information [16, 17, 18], heterogeneous competencies [19, 20, 21], and plurality voting with multiple alternatives [22, 23]. Experimental studies have also examined the theorem in real-world scenarios [24, 25, 26].

2.2.2 Strategic Voting in Epistemic Social Choice

The game-theoretical study of the Condorcet Jury Theorem starts from Austen-Smith and Banks [4], the first to cast the problem in a game-theoretical setting. They study a collaborative voting game in which agents share the same preference and each receives a binary signal correlated with an unknown binary state, and show that sincere and informative voting do not always form a Nash equilibrium; informative voting is thus incompatible with the strategic behavior of voters. A large subsequent literature therefore studies the strategic behavior of voters and the existence of (non-informative) equilibria that nonetheless reveal the ground truth. Wit [5], Feddersen and Pesendorfer [7], and Myerson [6] establish mixed-strategy equilibria with this property, and further analyses appear in [27, 28, 29]; Martinelli [30] gives conditions under which large committees aggregate information under the unanimity rule, and Chakraborty and Ghosh [31] characterize efficient equilibria and information aggregation in common-interest voting games. These results focus on revealing the ground truth when agents share a common goal, and their analysis focuses on unilateral strategic behavior under the solution concept of Bayesian Nash Equilibrium (BNE), which does not take coalitional strategic behavior into account.

2.2.3 Preference Structures

While CJT assumes that agents share a common goal, a large body of literature takes the heterogeneous preferences of agents into account under different models. In general, the preference structures in these models can be categorized into three directions. In the first direction, agents have different ordinal preferences over a shared common value [8, 9, 32], typically with three agent types: two with fixed preferences for the two alternatives respectively, and a third whose preference aligns with the ground truth. Our Chapter 3 and 4 follow this direction. In the second direction, agents hold opposite preferences contingent on the state [33, 34, 10]. Our Chapter 5 follows this direction. Finally, a broader line of studies presents more general preference configurations [35, 36]; in particular, Bhattacharya [37] provide a full-spectrum analysis that covers all four ordinal preference types, and Bhattacharya [38] study a spatial model in which the alternatives are located on a policy line.

Across these heterogeneous-preference models, the shared notion of a good decision is the *informed majority decision*, the alternative a majority would favor had all private signals been pooled. Feddersen and Pesendorfer [8] establish this as *full information equivalence*, showing that in the unique large-election equilibrium the elected alternative is unchanged even if all private signals were common knowledge, and a substantial literature characterizes when such aggregation can or cannot be attained under more general preferences [37, 36, 38, 35, 32, 33, 34]. Han et al. [39] sharpen this picture by establishing a surprising equivalence between a strategy profile forming an equilibrium and its reaching the informed majority decision, lending strong support to the view that majority voting reaches a good outcome through strategic behavior alone.

2.2.4 Beyond Majority Voting

A substantial body of work enriches the voting procedure to better aggregate information, whether through auxiliary mechanisms, pre-vote communication, or multiple rounds. The strand introduces a communication or deliberation stage before the vote. Coughlan [40], Schulte [41], and Gerardi [42] show that a communication stage prior to voting improves equilibrium outcomes, and Gerardi and Yariv [43] show that the equilibrium outcomes under "deliberation + different voting rules" coincide yet diverge without deliberation, with Goeree and Yariv [26] confirming empirically that these stable outcomes resemble those under one-round informative voting. For sequential procedures, Dekel and Piccione [44] show that voting in a predetermined order yields no new equilibria relative to simultaneous voting, which implies that sequential voting may not admit a simple equilibrium, and Chopra et al. [45] study the knowledge agents need in order to strategize.

Another strand designs more sophisticated mechanisms that directly reach the informed majority decision. Schoenebeck and Tao [9] leverage the "surprisingly popular" mechanism [46] by asking agents to additionally report their predictions of others' votes, which incentivizes truthful reporting; the surprisingly popular answer is reliable both when agents are sincere [46, 47] and when they are strategic [9]. Deng et al. [10] design a truthful mechanism that reaches an informed majority decision when there is a significant power gap between groups with opposing preferences. Such added structure, however, involves a trade-off between the simplicity of agent behavior and that of the mechanism: although they induce "simpler" equilibria, their mechanisms require supplemental inputs or extended procedures, increasing the complexity of the mechanisms themselves. Monroe and Waggoner [48] proposes a two-stage mechanism that aggregates strategic agents' subjective preferences and noisy signals about each alternative's external welfare impact, and shows that the price of anarchy approaches one in the two-alternative case, though under a utilitarian social welfare objective.

2.2.5 Coalitional Strategic Behavior and Capacitated Coordination Ability.

Whereas the equilibrium concepts above rule out only unilateral deviations, agents may also coordinate, giving rise to *group strategic behavior*. The central solution concept is the strong equilibrium, which requires robustness to deviations by any coalition; it originates from complete-information games [49] and has since been studied in voting [50, 51, 52, 53, 54, 55] and, in the Bayesian setting, across cooperative games [56, 57] and implementation theory [58, 59].

Requiring robustness to coalitions of unlimited size, however, is often too demanding for an equilibrium to exist [60]. Consequently, recent research has highlighted the importance of examining capacitated coalitional behavior and has introduced various solution concepts for this scenario. Guo and Yannelis [61] propose a coalitional interim equilibrium parameterized by an exogenously given set of admissible coalitions; Han et al. [62] propose an (ex-ante) Bayesian k-strong equilibrium family parameterized by the largest capacity k of the deviating group, which serves as the solution concept in our Chapter 5; Abraham et al. [63] propose a k-equilibrium in which the deviators share all their private information within the group; and Abraham et al. [64] study capacitated strategic behavior under a non-Bayesian game. By bounding the coalition size, these concepts interpolate between Nash equilibrium and full robustness to coalitions, making it possible to model bounded coordination.

2.2.6 Other Related Topics

Our work further relates to *information elicitation*, which aims to collect truthful and high-quality information from agents under a noisy information structure, through scoring rules [65, 66], peer-prediction mechanisms [67, 68, 69], the Bayesian Truth Serum [70, 71], and prediction markets [72, 73]. Information elicitation is nonetheless incompatible with our voting scenario for two reasons: it requires agents to be indifferent to the outcome, whereas here agents are incentivized by the outcome of the vote; and it rewards agents through payments, whereas voting carries no monetary rewards.

Our work, especially Chapter 4, is also related to strategic information transmission between informed experts and an uninformed decision maker [74, 75, 32, 76, 77]. Even when experts strategically report their information, the decision maker can still collect enough information to make an optimal decision [78, 32].

CHAPTER 3
THE WISDOM OF STRATEGIC VOTING

In this section, we focus on the simplest majority vote and address the following question

Is the informed majority decision always reached with coalitional strategic agents?

with a surprising confirmatory answer under mild conditions. We show that coalitional strategic behaviors positively impact achieving the informed majority decision and outperform non-strategic voting. We show that every equilibrium leads to the informed majority decision, and every voting profile that leads to the informed majority decision is an equilibrium. On the contrary, non-strategic behaviors lead to the informed majority decision only under certain conditions. Our results give merit to strategic behaviors and extend Feddersen and Pesendorfer's results to settings with coalitional strategic agents.

We study the solution concept of *ε-strong Bayes-Nash Equilibrium*, which precludes groups of agents from reaching higher expected utilities by coordinating. We show the equivalence of a strategy profile being "good" (leading to the informed majority decision with high probability, or, equivalently, of high *fidelity*) and being an ε-strong Bayes-Nash Equilibrium with $\varepsilon = o(1)$ (Theorem 2). We also guarantee the existence of an ε-strong Bayes-Nash equilibrium with $\varepsilon = o(1)$ in any instance (Theorem 3). At a high level, if we view fidelity as a proxy of social welfare, this result implies that the price of anarchy converges to 1 in this voting problem.

On the other hand, we characterize the conditions where strategy profiles succeed and fail to achieve the informed majority decision. Applying these results, we study two common non-strategic behavior – *informative voting*, where agents honestly reflect their private information in their votes, and *sincere voting*, where agents vote as if they are the only decision-maker. We show that (1) informative voting leads to the informed majority decision only when the majority vote threshold is unbiased compared with the signal distribution (Corollary 3), and (2) sincere voting leads to the informed majority decision only when it is also an equilibrium (Corollary 4). These observations indicate that strategic behavior "prevails" over non-strategic behaviors in binary voting!

The technical key for the probability analysis is to compute the *excess expected vote share*, i.e., the amount of expected vote share an alternative attracts that exceeds the threshold, and to upper (or lower) bound the fidelity given different cases of excess expected vote shares. A strategy profile has high fidelity if and only if its excess expected vote share is strictly positive (Theorem 4).

We follow the setting in Schoenebeck and Tao [9], which is an extension of the setting in Austen-Smith and Banks [4], and consider agents with preferences contingent on underlying world states in a single framework. Also, as in Example 1 and previous work, we assume that various constraints in the real world prevent discussion after agents see their signals. Such constraints can be of a time aspect (a quick decision must be made and there is no time for discussion), a procedural aspect (a formal conference that prohibits participants from discussing privately before voting), and/or a societal aspect (it is socially unsuitable to discuss some preferences), etc. Therefore, we consider an *ex-ante* setting where the expected utilities are computed before agents receive their signals.

3.1 Preference Structure

Utility and Types of Agents. Each agent i has a utility which is a function of the true world state and the outcome of the vote. Formally, we have $v_i : \mathcal{W} \times \{\mathbf{A}, \mathbf{R}\} \to \{0, 1, \dots, B\}$, where B is the positive integer upper bound. We assume that $\mathbf{A}$ is more preferable in H than in L, and $\mathbf{R}$ is the opposite: for every agent i, $v_i(H, \mathbf{A}) > v_i(L, \mathbf{A})$ and $v_i(H, \mathbf{R}) < v_i(L, \mathbf{R})$.

The different endogenous preferences of agents are reflected by different utility functions. *Predetermined* agents always prefer the same alternative, and *contingent* agents have preferences depending on the world state. Predetermined agents can be further divided into *friendly* and *unfriendly* agents based on the alternative they prefer. For an agent i, if i is a friendly agent, $v_i(H, \mathbf{A}) > v_i(L, \mathbf{A}) > v_i(L, \mathbf{R}) > v_i(H, \mathbf{R})$; if i is an unfriendly agent, $v_i(L, \mathbf{R}) > v_i(H, \mathbf{R}) > v_i(H, \mathbf{A}) > v_i(L, \mathbf{A})$; and if i is a contingent agent, $v_i(H, \mathbf{A}) > v_i(H, \mathbf{R})$ and $v_i(L, \mathbf{R}) > v_i(L, \mathbf{A})$.

We give the interpretation of such a preference structure under the combined utility model. The objective utility v_i^O of each agent i shares the same ordinal preferences such that $v_i^O(H, \mathbf{A}) > v_i(L, \mathbf{A})$ and $v_i^O(H, \mathbf{R}) < v_i^O(L, \mathbf{R})$. Their subjective preferences differ. If the subjective preference of the agents dominates, the agent is a predetermined agent. On the other hand, the agent is a contingent agent.

Example 5. *Recall the utilities of agents in Example 4. Group 1 agents are friendly agents, group 2 agents are unfriendly agents, and group 3 agents are contingent agents.*

Let α_F, α_U, and α_C be the approximated fraction of each type of agent. Formally, given n agents, $n_F = \lfloor \alpha_F \cdot n \rfloor$ is the number of friendly agents, $n_U = \lfloor \alpha_U \cdot n \rfloor$ is the number of unfriendly agents, and $n_C = n - n_F - n_U$ is the number of contingent agents. α_F, α_U, and α_C are common knowledge and do not depend on n.

Informed Majority Decision In this section, we assume that neither friendly agents nor unfriendly agents can dominate the vote. (That is, $\alpha_F < \mu$ and $\alpha_U < 1 - \mu$.) Otherwise, the informed majority decision does not depend on the state, and one coalition can always enact it via a dominant strategy. As a result, $\mathbf{A}$ is the informed majority decision when the world state is H, and $\mathbf{R}$ is the informed majority decision when the world state is L.

In this section, we will focus on *regular* strategy profiles, defined as follows.

Definition 4. A strategy profile Σ is *regular* if all friendly agents always vote for $\mathbf{A}$, and all unfriendly agents always vote for $\mathbf{R}$ in Σ. A sequence of strategy profiles is regular if every strategy profile in the sequence is a regular profile.

We believe this restriction is mild and natural since "always vote for **A**" is the dominant strategy for a friendly agent, and "always vote for **R**" is the dominant strategy for an unfriendly agent in the majority vote. We will focus our study on regular strategy profiles in Chapter 3, while in Chapter 4 and 5 we look into a wider range of strategies.

3.2 Strong Equilibrium May Not Exist

As a preliminary, we first show that a strong BNE does not always exist. Therefore, we seek ε-strong BNE as an approximation.

Theorem 1. *For any $N_0 \in \mathbb{N}$, there exists an instance of $N > N_0$ agents, in which a strong Bayes-Nash equilibrium does not exist.*

Proof Sketch. For any $N_0 \in \mathbb{N}$, we construct an instance of $N = 2N_0 + 3$ agents. The agents consist of three parts: F is a set of $N_0 + 1$ friendly agents. C is a set of two contingent agents. And U is a set of N_0 unfriendly agents. Agents in the same set share the same utility, which is shown in Table 3.1. The threshold is $\mu = 0.5$. The prior distribution is $P_L = P_H = 0.5$. The signal distribution is $P_{hH} = P_{lL} = 0.8$ and $P_{lH} = P_{hL} = 0.2$.

Agents	$v(H, \mathbf{A})$	$v(L, \mathbf{A})$	$v(L, \mathbf{R})$	$v(H, \mathbf{R})$
F	100	99	1	0
C	90	0	100	0
U	1	0	100	99

Table 3.1: Utility of three groups

Agents	Σ_1	Σ_2	Σ_3
F	50.396	66.14	50.3
C	85.12	75.2	76
U	50.396	34.46	50.3

Table 3.2: Expected utility under three profiles

Consider the following three strategy profiles, under which the expected utility of each group is shown in Table 3.2.

- Σ_1: N_0 agents in F always vote for **A**, and one agent votes informatively. C vote informatively. U always vote for **R**.

- Σ_2: F always vote for **A**. C vote informatively. U always vote for **R**.

- Σ_3: F always vote for **A**. One C agent votes informatively, and the other always votes for **R**. U always vote for **R**.

These three strategy profiles form a cycle $\Sigma_1 \to \Sigma_2 \to \Sigma_3 \to \Sigma_1$ of deviation, where a group of agents has incentives to deviate to the next profile.

For any other strategy profile Σ, there exists a group of agents with incentives to deviate to one of the three profiles. Firstly, F agents and U agents would like to deviate from their dominant strategy of always voting for **A** (**R**, respectively) whenever it can increase the probability that their preferred candidate wins. Given F and U agents play dominant strategies, the best strategy for two C agents is to play the strategy in Σ_3 (one agent votes informatively, the other votes for **R**). Then we know that an F agent and two C agents have incentives to deviate from Σ_3 to Σ_1. Therefore, there does not exist a strong Bayes-Nash equilibrium in this instance. Full proof is available in Appendix A.1. $\square$

3.3 Equivalence between High Fidelity and Strong Equilibrium

In this section, we show that strategic behaviors indeed have a positive impact on leading to the informed majority decision. Theorem 2 states that if the fidelity of a regular (Definition 4) strategy profile sequence $\{\Sigma_n\}_{n=1}^{\infty}$, i.e., $A(\Sigma_n)$, converges to 1 as n goes to infinity, every Σ_n in the sequence will be an ε-strong Bayes-Nash equilibrium where ε converges to 0. On the other hand, if $A(\Sigma_n)$ does not converge to 1, then we can find infinitely many Σ_n that are not ε-strong BNE with a constant ε. Moreover, Theorem 3 guarantees that there always exists a regular strategy profile whose fidelity converges to 1, which leads to an ε-strong BNE with $\varepsilon = o(1)$. The two theorems together indicate that strategic voting leads to the informed majority decision in any sequence of instances.

Theorem 2. *Given an arbitrary sequence of instances and an arbitrary regular strategy profile sequence $\{\Sigma_n\}_{n=1}^{\infty}$, let $\{A(\Sigma_n)\}_{n=1}^{\infty}$ be the sequence of the fidelities of Σ_n.*

- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$*, then for every* n*,* Σ_n *is an* ε*-strong BNE with* $\varepsilon = o(1)$*.*

- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$ *does not hold, then there exist infinitely many* n *such that* Σ_n *is NOT an* ε*-strong BNE for some constant* ε*.*

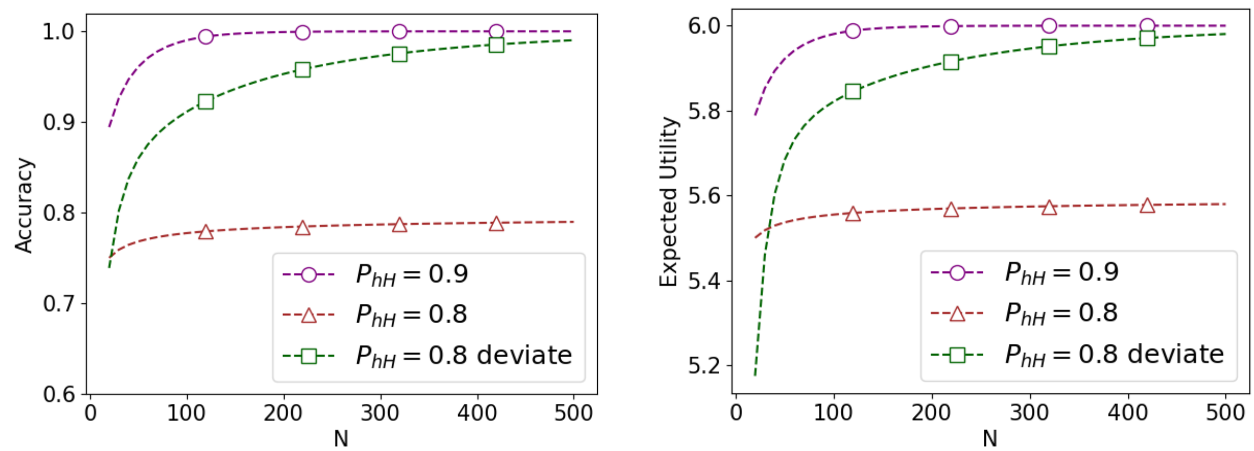
Theorem 3. *Given any arbitrary sequence of instances, there always exists a sequence of regular strategy profiles* $\{\Sigma'_n\}_{n=1}^{\infty}$ *such that* $A(\Sigma'_n)$ *converges to 1.*

We first give a concrete example to illustrate Theorem 2, in which we show an instance for each case in the theorem.

Example 6. *We follow the setting of Example 4 except for two differences. First, there is a series of* $n = 20, 30, \ldots, 500$*. For each* n*, the ratio of friendly (group 1), unfriendly (group 2), and contingent (group 3) agents is fixed at* $2 : 3 : 5$*. Second, we consider two different cases of signal distributions that fall into different cases of Theorem 2. They share the same signal distribution in world state L:* $P_{lL} = 0.8, P_{hL} = 0.2$*, but the signal distribution for H is different. In case (1),* $P_{lH} = 0.1, P_{hH} = 0.9$*; and in case (2),* $P_{lH} = 0.25, P_{hH} = 0.75$*.*

We focus on regular strategy profiles where all contingent agents vote informatively (Definition 1). For case (2), we also consider another sequence of regular strategy profiles Σ'_n *where contingent agents play* $\sigma' = (0.48, 0.96)$*. In Example 8 and Theorem 4 later, we verify that the fidelity of the regular informative voting converges to 1 in case (1) but does not converge to 1 in case (2). On the other hand, the fidelity of the deviating strategy profile* Σ'_n *in case (2) converges to 1. Figure 3.1a illustrates these trends of fidelity.*

The expected utilities of contingent agents in different cases and strategies are shown in Figure 3.1b. Note the maximum expected utility that a contingent agent can get is $0.4 \times 3 + 0.6 \times 8 = 6.0$*. In accordance with the fidelity, the expected utility of* Σ_n *in case (1) converges to the maximum. In case (2), on the other hand,* Σ_n *is dominated by* Σ'_n *by a utility gain of at least 0.4. Therefore, the group of contingent agents has no incentive to deviate in case (1) but has incentives to deviate to* Σ'_n *in case (2).*



(a) Fidelity

(b) Expected utility for contingent agents

Figure 3.1: Fidelity and expected utilities of informative voting.

3.3.1 Proof Sketch of Theorem 2

To show the relationship between $A(\Sigma_n)$ and ε, we have the following lemma.

Lemma 1. *For every n, a regular strategy profile Σ_n is an ε-strong BNE with $\varepsilon = 2B(B+1)(1-A(\Sigma_n))$, where B is the upper bound of utility function v_i.*

Lemma 1 is an extension of Theorem 3.3 in Schoenebeck and Tao [9]. It shows that every Σ_n is an ε-strong BNE with ε proportional to $1 - A(\Sigma_n)$. To prove Lemma 1, we show that, for any other strategy profile Σ'_n, a group of agents with incentives to deviate does not exist.

There are two cases of Σ'_n. In the first case, the fidelity Σ'_n is bounded by the fidelity of Σ_n. More precisely, $(1 - A(\Sigma'_n)) < (B+1) \cdot (1 - A(\Sigma_n))$. Then two profiles do not make a big difference, and no agent can gain more than $\varepsilon = 2B(B+1)(1-A(\Sigma_n))$ after deviation (Claim 1).

In the second case, the fidelity Σ'_n is much worse than the fidelity of Σ_n. Then any contingent agent has no incentives to deviate, as their expected utilities are most positively correlated with fidelity (Claim 2). Next, we show that a deviating group cannot contain both friendly and unfriendly agents, because if the expected utility of one side increases by more than ε, the other side's will decrease (Claim 3). Therefore, a deviating group contains either only friendly agents or only unfriendly agents. Finally, we show that in neither case can the deviation succeed, because pre-determined agents have already played their dominant strategies in a regular profile (Claim 4).

Now we are ready to prove Theorem 2. We will actually use Lemma 1 and Theorem 3 to prove Theorem 2. We will discuss the two cases separately.

When $\lim_{n\to\infty} A(\Sigma_n) = 1$, we apply Lemma 1 to each Σ_n, and get that every Σ_n is an ε-strong BNE where $\varepsilon = 2B(B+1)\cdot(1 - A(\Sigma_n))$. Then ε will converge to 0 as $n \to \infty$.

When $\lim_{n\to\infty} A(\Sigma_n) = 1$ does not hold, there are infinitely many n with Σ_n being of low fidelity. By Theorem 3 there exists a regular strategy profile sequence $\{\Sigma'_n\}$ with fidelity converging to 1. Because of the difference in fidelity, there are infinitely many n such that $\Sigma_n \neq \Sigma'_n$. Then we show that, for all sufficiently large n where Σ_n is of low fidelity, if all contingent agents turn to play Σ'_n from Σ_n, every contingent agent will gain at least a constant amount of extra utility. Therefore, for infinitely many n, Σ_n is NOT an ε-strong BNE for some constant ε. Full proof is available in Appendix A.3.1 and A.3.3.

3.3.2 Proof Sketch of Theorem 3

In the proof of Theorem 3, we construct a strategy σ' and show that the regular strategy profile sequence $\{\Sigma'_n\}$ where all contingent agents play σ' has fidelity that converges to 1. It suffices to construct σ' such that

1. if H is the actual world state, the expected fraction of the voters voting for **A** is more than μ by a constant;
2. if L is the actual world state, the expected fraction of the voters voting for **A** is less than μ by a constant.

If this is true, $A(\Sigma'_n)$ converges to 1 due to the Hoeffding Inequality. It remains to construct σ' such that (1) and (2) hold.

We first construct σ'_μ such that the expected fraction of the voters voting for **A** is exactly μ, where in σ'_μ the contingent voter votes for **A** with a probability that is independent of the signal she receives. This can be done by setting $\sigma'_\mu = (\beta^*, \beta^*)$ where β^* satisfies $\alpha_F + \alpha_C \cdot \beta^* = \mu$ (notice that, given the fraction α_F of the friendly voters who always vote for **A**, the fraction α_U of the unfriendly voters who never vote for **A**, and the fraction α_C of the contingent voters who vote for **A** with probability β^*, the expected fraction of votes for **A** is $\alpha_F + \alpha_C \cdot \beta^*$).

Next, we will adjust σ'_μ to $\sigma' = (\beta_l, \beta_h)$ that satisfies (1) and (2). Naturally, we would like to increase the probability for voting **A** if an h signal is received, and we would like to decrease this probability if l is received. That is, we have $\beta_l = \beta^* - \delta_l$ and $\beta_h = \beta^* + \delta_h$ for some $\delta_l, \delta_h > 0$, and we need to show the existence of δ_l and δ_h that make (1) and (2) hold.

When H is the actual world, comparing with σ'_μ, the probability that each contingent agent votes for **A** is increased by $P_{hH} \cdot \delta_h - P_{lH} \cdot \delta_l$ in σ'. Thus, the total expected fraction of votes for **A** is increased by

$$\alpha_C \cdot \left(P_{hH} \cdot \delta_h - P_{lH} \cdot \delta_l\right).$$

Similarly, when L is the actual world, similar calculations reveal that the total expected fraction of votes for **A** is increased by

$$\alpha_C \cdot \left(P_{hL} \cdot \delta_h - P_{lL} \cdot \delta_l\right).$$

Since the expected fraction of votes for **A** is exactly μ for σ'_μ, we need to choose δ_h and δ_l such that

$$\begin{cases} \alpha_C \cdot (P_{hH} \cdot \delta_h - P_{lH} \cdot \delta_l) > 0 \\ \alpha_C \cdot (P_{hL} \cdot \delta_h - P_{lL} \cdot \delta_l) < 0 \end{cases}.$$

This can always be done due to the positive correlation $P_{hH} > P_{hL}$ and $P_{lH} < P_{lL}$. In particular, if we set $\delta_h = \delta_l \cdot \frac{P_{lH}}{P_{hH}}$, the first inequality would become equality, while the second inequality holds due to the positive correlation. By slightly increasing δ_h, we can make both inequalities hold. During these adjustments, we just need to make sure the two constants δ_h and δ_l are small enough such that β_h and β_l are valid probabilities. Full proof is available in Appendix A.3.2.

Example 7. *In this example, we follow the setting of case (2) in Example 6 to illustrate the construction of the strategy* σ'. *Recall that* $\alpha_F = 0.2, \alpha_U = 0.3$, *and* $\alpha_C = 0.5$. *The signal distribution* $P_{hH} = 0.75$ *and* $P_{lL} = 0.8$. *The threshold* $\mu = 0.6$.

In the first step, let $\sigma'_\mu = (0.8, 0.8)$. *We could verify that* $\alpha_F + \alpha_C \cdot 0.8 = 0.2 + 0.5 \times 0.8 = 0.6 = \mu$.

In the second step, let $\delta_l = 0.3$. *Then* $\delta_h = \delta_l \cdot \frac{P_{lH}}{P_{hH}} = 0.1$. *Then* $\sigma' = (0.5, 0.9)$. *Then we have*

$$P_{hH} \cdot \delta_h - P_{lH} \cdot \delta_l = 0.75 \times 0.1 - 0.25 \times 0.3 = 0.$$

$$P_{hL} \cdot \delta_h - P_{lL} \cdot \delta_l = 0.2 \times 0.1 - 0.8 \times 0.3 = -0.22 < 0.$$

Finally, we increase δ_h *by 0.06. Then* $\delta_l = 0.3$, $\delta_h = 0.16$, *and* $\sigma' = (0.5, 0.96)$. *We have*

$$P_{hH} \cdot \delta_h - P_{lH} \cdot \delta_l = 0.75 \times 0.16 - 0.25 \times 0.3 = 0.05 > 0.$$

$$P_{hL} \cdot \delta_h - P_{lL} \cdot \delta_l = 0.2 \times 0.16 - 0.8 \times 0.3 = -0.208 < 0.$$

Therefore, $\sigma' = (0.5, 0.96)$ *satisfies the condition.*

3.4 Probability Analysis on Fidelity

In this section, we analyze the condition that a strategy profile is of high fidelity and apply the analysis to the most common forms of non-strategic voting: informative voting and sincere voting. We show that neither informative nor sincere voting can lead to the informed majority decision in every instance, and we characterize the conditions where they lead to the informed majority decision. Our results give merit to strategic voting.

In order to characterize the fidelity, we introduce the notion of the *excess expected vote* *share.* Given a world state k, the excess expected vote share is the expected vote share the informed majority decision alternative attracts under state k minus the threshold of the alternative.

Definition 5 (**Excess expected vote share).** Given an instance of n agents, and a strategy profile Σ, let random variable X_i^n be "agent i votes for $\mathbf{A}$": $X_i^n = 1$ if agent i votes for $\mathbf{A}$, and $X_i^n = 0$ if i votes for $\mathbf{R}$. Then the excess expected vote share is defined as follows:

$$f_H^n = \frac{1}{n}\sum_{i=1}^{n} E[X_i^n \mid H] - \mu. \tag{3.1}$$

$$f_L^n = \frac{1}{n}\sum_{i=1}^{n} E[1 - X_i^n \mid L] - (1-\mu). \tag{3.2}$$

Specifically, f_H^n is the excess expected vote share of $\mathbf{A}$ condition on world state H, and f_L^n is the excess expected vote share of $\mathbf{R}$ condition on world state L. For technical convenience, we define $f^n = \min(f_H^n, f_L^n)$.

Our next result shows that we can judge whether the fidelity of a strategy profile sequence converges to 1 with the tendency of its excess expected vote share (or more precisely, the lower limit of $\sqrt{n} \cdot f^n$). If $\sqrt{n} \cdot f^n$ has a lower limit of $+\infty$, then the fidelity of the profiles in the sequence converges to 1. Otherwise, the fidelity is likely not to converge to 1.

Theorem 4. *Given an arbitrary sequence of instances and arbitrary sequence of strategy profiles* $\{\Sigma_n\}_{n=1}^{\infty}$*, let* f^n *be the excess expected vote share for each* Σ_n*.*

- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n = +\infty$*, the fidelity of* Σ_n *converges to 1 , i.e.,* $\lim_{n\to\infty} A(\Sigma_n) = 1$.
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n < 0$ *(including* $-\infty$*),* $A(\Sigma_n)$ *does NOT converge to 1.*
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$ *(not including* $+\infty$*), and the variance of* $\sum_{i=1}^{n} X_i^n$ *is at least proportional to* n*,* $A(\Sigma_n)$ *does NOT converge to 1.*

Remark 1. *Although Theorem 4 does not cover the case when* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$ *and the variance of* $\sum_{i=1}^n X_i^n$ *is not large enough, we argue that this case is very special and rare. In this case,* $(\inf f^n)$ *converges to* 0 *at the rate of* $O(\frac{1}{\sqrt{n}})$*, which means the expected vote share of an alternative is almost equal to the threshold. Moreover, the strategies of the agents have low randomness in total. Therefore, we believe that Theorem 4 covers the most interesting cases of a sequence of strategy profiles.*

Proof Sketch. Recall that $A(\Sigma) = P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma) + P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma)$. Note that $\lambda_H^{\mathbf{A}}(\Sigma)$ is the probability that the total vote share on **A** exceeds the threshold μ when the world state is H. Therefore, we can write $\lambda_H^{\mathbf{A}}(\Sigma)$ using the following formula. $\lambda_L^{\mathbf{R}}(\Sigma)$ can be written using a similar formula.

$$\lambda_H^{\mathbf{A}}(\Sigma_n) = \Pr\left[\sum_{i=1}^n X_i^n \geq \mu \cdot n \mid H\right] = \Pr\left[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E\left[X_i^n \mid H\right] \geq -f_H^n \cdot n \mid H\right]$$

For the first and the second case, we apply the Hoeffding Inequality. For the first case, we show that both $\lambda_H^{\mathbf{A}}(\Sigma)$ and $\lambda_L^{\mathbf{R}}(\Sigma)$ are lower bounded by a function of n that converges to 1. For the second case, we show that either $\lambda_H^{\mathbf{A}}(\Sigma)$ and $\lambda_L^{\mathbf{R}}(\Sigma)$ is upper bounded by a constant smaller than 1.

For the third case, we apply the Berry-Esseen Theorem [79, 80], which bounds the difference between the distribution of the sum of independent random variables and the normal distribution. Therefore, for some constant δ and infinitely many n, $\lambda_H^{\mathbf{A}}(\Sigma_n)$ (or $\lambda_L^{\mathbf{R}}(\Sigma_n)$) will not deviate from $1 - \Phi(\delta)$ too much and is bounded away from 1 by a constant. Φ is the CDF of the standard normal distribution. The requirements for the variance in the third case are from the Berry-Esseen Theorem.

The full proof of Theorem 4 and the detailed definition of the Berry-Esseen Theorem are available in Appendix A.3 $\square$

Theorem 4 provides a criterion for judging whether a strategy profile sequence is of high fidelity. If we apply Theorem 2 to each case of Theorem 4, we directly get a criterion for judging whether a regular strategy profile sequence is an ε-strong equilibrium.

Corollary 1. *Given an arbitrary sequence of instances and an arbitrary regular sequence of strategy profiles* $\{\Sigma_n\}_{n=1}^{\infty}$*, let* f^n *be defined for each* Σ_n*.*

- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n = +\infty$*, then for every* n*,* Σ_n *is an* ε*-strong BNE with* $\varepsilon = o(1)$*.*
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n < 0$ *(including* $-\infty$*), there are infinitely many* n *such that* Σ_n *is NOT an* ε*-strong BNE with constant* ε*.*
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$ *(not including* $+\infty$*), and the variance of* $\sum_{i=1}^n X_i^n$ *is at least proportional to* n*, there are infinitely many* n *such that* Σ_n *is NOT an* ε*-strong BNE with constant* ε*.*

Example 8. *In this example, we use Theorem 4 to bound the fidelity of different cases in Example 6. Note that in each* n *the ratio of friendly, unfriendly, and contingent agents is fixed to be* $2:3:5$*, and agents of the same type play the same strategy for different* n*. Therefore, the excess expected vote share of profile* Σ_n *and* Σ'_n *is independent of* n*.*

Case 1: $P_{lH} = 0.1, P_{hH} = 0.9$**.** *In both world states, we have* $E[X_i^n \mid H] = 1$ *for friendly agents and* $E[X_i^n \mid H] = 0$ *for unfriendly agents. Contingent agents vote for* **A** *with probability* $P_{hH} = 0.9$ *in* H *state and for* **R** *with probability* $P_{lL} = 0.8$ *in* L *state. Therefore,* $f^n > 0$*, and* $\sqrt{n} \cdot f^n$ *goes to* $+\infty$*.*

$$f_H^n = 0.2 + 0.5 \times 0.9 - 0.6 = 0.05 \qquad f_L^n = 0.3 + 0.5 \times 0.8 - 0.4 = 0.3.$$

Case 2: $P_{lH} = 0.25, P_{hH} = 0.75$**.** *For* Σ_n*, we have* $f^n < 0$*, and* $\sqrt{n} \cdot f^n$ *goes to* $-\infty$*.*

$$f_H^n = 0.2 + 0.5 \times 0.75 - 0.6 = -0.025 \qquad f_L^n = 0.3 + 0.5 \times 0.8 - 0.4 = 0.3.$$

And for the deviating strategy profile Σ'_n, we have $f^n > 0$, and $\sqrt{n} \cdot f^n$ goes to $+\infty$.

$$f_H^n = 0.2 + 0.5 \cdot (0.75 \times 0.96 + 0.25 \times 0.48) - 0.6 = 0.02$$
$$f_L^n = 0.3 + 0.5 \cdot (0.8 \times 0.52 + 0.2 \times 0.04) - 0.4 = 0.112.$$

In Case 1, the regular strategy profile Σ_n lies in the first case of Theorem 4, has fidelity converging to 1, and is an ε-strong BNE with $\varepsilon = o(1)$. In Case 2, Σ_n lies in the second case of Theorem 4 and is dominated by the deviating strategy profile Σ'_n. This is in accordance with our observation in Figure 3.1 and Example 6.

Although Theorem 4 (and Corollary 1) do not cover all the strategy profile sequences, the following result provides a dichotomy for symmetric profile sequences to judge fidelity. Given a symmetric strategy profile Σ_n induced by strategy $\sigma = (\beta_\ell, \beta_h)$, we can compute the excess expected vote share of Σ_n. Recall the definition of excess expected vote share:

$$f_H^n = \frac{1}{n} \sum_{i=1}^{n} E[X_i^n \mid H] - \mu.$$

In H state, an agent with signal h votes for **A** with probability β_h, and an agent with signal l votes for **A** with probability β_ℓ. Therefore, $E[X_i^n \mid H] = P_{hH} \cdot \beta_h + P_{lH} \cdot \beta_\ell$. Then we have

$$f_H^n = \frac{1}{n} \sum_{i=1}^{n} (P_{hH} \cdot \beta_h + P_{lH} \cdot \beta_\ell) - \mu = P_{hH} \cdot \beta_h + P_{lH} \cdot \beta_\ell - \mu.$$

With similar reasoning, we can compute f_L^n and f^n:

$$f_L^n = (P_{hL} \cdot (1 - \beta_h) + P_{lL} \cdot (1 - \beta_\ell)) - (1 - \mu), \quad f^n = \min\left(f_H^n, f_L^n\right).$$

An interesting observation for the symmetric strategy profiles is that its excess expected vote share is independent of the number of agents n. This is because when every agent plays the same strategy, the expectation of every X_i^n is the same. For simplicity, given a sequence of symmetric strategy profiles $\{\Sigma_n\}$, we write its excess expected vote share as f_H, f_L, and f.

Corollary 2. *For an arbitrary strategy σ and an arbitrary sequence of instances, let $\{\Sigma_n\}$ be the sequence of symmetric strategy profile Σ_n induced by σ, and f be the excess expected vote share of $\{\Sigma_n\}$.*

- *If $f > 0$, $A(\Sigma_n)$ converges to 1.*
- *If $f \leq 0$, $A(\Sigma_n)$ does not converge to 1.*

Proof Sketch. The proof of Corollary 2 works by showing that each case of a symmetric strategy profile falls into some case of Theorem 4. When $f > 0$, $\sqrt{n} \cdot f^n \to +\infty$. When $f < 0$, $\sqrt{n} \cdot f^n \to -\infty$. When $f = 0$,and $\sqrt{n} \cdot f^n = 0$. The variance requirement is also satisfied. (Otherwise, the strategy σ must always vote for the same candidate. This directly implies $f < 0$, which is a contradiction.) Full proof is available in Appendix A.3. □

3.4.1 Case Study: Informative Voting and Sincere Voting

In this section, we study the two most common non-strategic voting schemes – informative voting and sincere voting under our information structure. We show that both voting schemes lead to the informed majority decision if and only if certain conditions are satisfied.

In informative voting, all agents play the strategy $\sigma = (0, 1)$. When the world state is H, an agent receives signal h and votes for **A** with probability P_{hH}. Therefore, the excess expected vote share in the H state is $f_H = P_{hH} - \mu$. Similarly, the excess expected vote share in the L state is $f_L = P_{lL} - (1 - \mu) = \mu - P_{hL}$. Applying Corollary 2, we get the following statement.

Corollary 3. *For an arbitrary sequence of instances, let $\{\Sigma_n\}$ be the sequence of informative voting profiles. Then, the fidelity $A(\Sigma_n)$ converges to 1 if and only if $P_{hH} > \mu > P_{hL}$.*

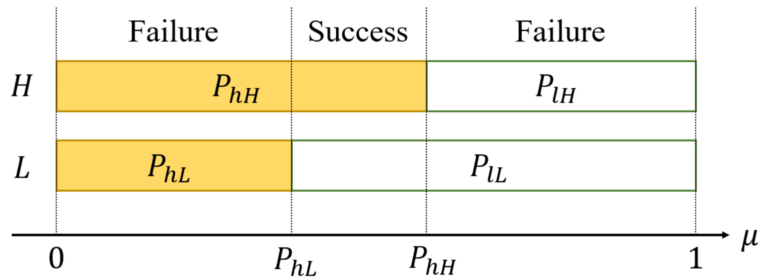


Figure 3.2: Illustration of Corollary 3.

Corollary 3 forms a comparison with Theorem 2. Strategic behavior always leads to the informed majority decision, while non-strategic informative voting achieves the same only when the majority vote threshold is "unbiased" compared with the signal distribution.

In sincere voting, an agent votes as if she is making the decision individually. A sincere agent chooses the alternative that maximizes the expected utility conditioned on the signal. The expected utility of an agent making an individual decision conditioned on signal m is

$$u_i(\mathbf{A} \mid m) = \Pr[W = L \mid m] \cdot v_i(L, \mathbf{A}) + \Pr[W = H \mid m] \cdot v_i(H, \mathbf{A}).$$
$$u_i(\mathbf{R} \mid m) = \Pr[W = L \mid m] \cdot v_i(L, \mathbf{R}) + \Pr[W = H \mid m] \cdot v_i(H, \mathbf{R}).$$

Definition 6. A strategy profile Σ is *sincere* if for any agent i, conditioned that i receives signal m, i votes for $\mathbf{A}$ if $u_i(\mathbf{A} \mid m) > u_i(\mathbf{R} \mid m)$ and votes for $\mathbf{R}$ otherwise.

A sincere strategy profile is not always symmetric, because the sincere behavior of agents not only depends on his/her signal but also on his/her utility. Therefore, sincere agents with different utility functions may play different strategies. Given the assumption of $P_{hH} > P_{hL}$, we have $\Pr[W = L \mid l] > \Pr[W = L \mid h]$ and $\Pr[W = H \mid l] < \Pr[W = H \mid h]$. As a result, $u_i(\mathbf{A} \mid l) < u_i(\mathbf{A} \mid h)$ and $u_i(\mathbf{R} \mid l) > u_i(\mathbf{R} \mid h)$. Therefore, a sincere voter would play one of the five strategies below based on her utility function v_i.

1. If $u_i(\mathbf{A} \mid l) < u_i(\mathbf{R} \mid l)$, and $u_i(\mathbf{A} \mid h) < u_i(\mathbf{R} \mid h)$, an agent always votes for $\mathbf{R}$.

2. If $u_i(\mathbf{A} \mid l) < u_i(\mathbf{R} \mid l)$, and $u_i(\mathbf{A} \mid h) = u_i(\mathbf{R} \mid h)$, an agent votes for $\mathbf{R}$ under signal l and votes arbitrarily under signal h.

3. If $u_i(\mathbf{A} \mid l) < u_i(\mathbf{R} \mid l)$, and $u_i(\mathbf{A} \mid h) > u_i(\mathbf{R} \mid h)$, an agent votes informatively.

4. If $u_i(\mathbf{A} \mid l) = u_i(\mathbf{R} \mid l)$, and $u_i(\mathbf{A} \mid h) > u_i(\mathbf{R} \mid h)$, an agent votes arbitrarily under signal l and votes for $\mathbf{A}$ under signal h.

5. If $u_i(\mathbf{A} \mid l) > u_i(\mathbf{R} \mid l)$, and $u_i(\mathbf{A} \mid h) > u_i(\mathbf{R} \mid h)$, an agent always votes for $\mathbf{A}$.

A sincere profile is also a regular profile, as friendly agents always vote for $\mathbf{A}$, and unfriendly agents always vote for $\mathbf{R}$ in their individual decisions. Therefore, applying Theorem 2, we have the following statement.

Corollary 4. *For an arbitrary sequence of instances, let $\{\Sigma_n\}$ be the sequence of sincere strategy profiles. Then the fidelity $A(\Sigma_n)$ converges to 1 if and only if Σ_n is an ε-strong Bayes-Nash equilibrium with $\varepsilon = o(1)$.*

Corollary 4 tells us that sincere voting performs as well as strategic voting if and only if itself is also strategic. The following example illustrates different behaviors of sincere voters by "manipulating" their utility functions under the same world state and signal distribution and gives examples where sincere voting succeeds and fails.

Example 9. *Consider the following scenario. The world state prior $P_H = P_L = 0.5$. The signal distribution $P_{hH} = P_{lL} = 0.8$, and $P_{lH} = P_{hL} = 0.2$. By the Bayes Theorem, we compute the probability of a world state conditioned on a private signal as follows.*

$$\Pr[W = L \mid l] = 0.8, \qquad \Pr[W = H \mid l] = 0.2$$
$$\Pr[W = L \mid h] = 0.2, \qquad \Pr[W = H \mid h] = 0.8.$$

We assume all agents are contingent and share the same utility function, and consider three different cases as shown in Table 3.3. Suppose Σ is a strategy profile where all agents vote sincerely.

(Winner, World State)	$(\mathbf{A}, H)$	$(\mathbf{A}, L)$	$(\mathbf{R}, H)$	$(\mathbf{R}, L)$
Case 1	*1*	*0*	*0*	*1*
Case 2	*5*	*1*	*0*	*2*
Case 3	*4*	*1*	*0*	*2*

Table 3.3: Utility functions for three cases.

In case 1, $u_i(\mathbf{A} \mid h) = u_i(\mathbf{R} \mid l) = 0.8$, *and* $u_i(\mathbf{A} \mid l) = u_i(\mathbf{R} \mid h) = 0.2$. *Therefore, every sincere voter votes informatively, and* Σ *is also informative voting. By corollary 3,* Σ *leads to the informed majority decision if and only if the threshold* $0.2 < \mu < 0.8$.

In case 2, $u_i(\mathbf{A} \mid l) = 1.8, u_i(\mathbf{R} \mid l) = 1.6$, $u_i(\mathbf{A} \mid h) = 4.2$, *and* $u_i(\mathbf{R} \mid h) = 0.4$. *Therefore, all sincere agents always vote for* $\mathbf{A}$ *even if they are contingent. In this case,* $\mathbf{A}$ *is always the winner, and* $A(\Sigma)$ *does not converge to 1.*

In case 3, $u_i(\mathbf{A} \mid l) = u_i(\mathbf{R} \mid l) = 1.6$, $u_i(\mathbf{A} \mid h) = 3.4$, *and* $u_i(\mathbf{R} \mid h) = 0.4$. *In this case, a sincere agent votes* $\mathbf{A}$ *under signal* h, *and votes arbitrarily under signal* l. *Then for any* $0.2 < \mu < 1$, Σ *induced by strategy* $\sigma = (\beta_\ell, 1)$ *leads to the informed majority decision with high probability, where* $\beta_\ell \geq 0$ *satisfies* $\beta_\ell > 5\mu - 4$ *and* $\beta_\ell < 1.25\mu - 0.25$. *These conditions guarantee the excess expected vote share of* Σ *to be strictly positive.*

As shown in Example 9, sincere agents can have drastically different behaviors in different scenarios. Nevertheless, once we know the strategy of each agent, we can apply Theorem 4 to analyze the probability of a sequence of sincere voting leading to the informed majority decision.

3.5 Non-Binary World States and Non-Binary Signals.

In this section, we discuss how we extend our model and results to a setting with non-binary world states and non-binary signals. We follow the setting of Schoenebeck and Tao [9]. The largest difference in the non-binary setting is that the preferences of agents form a spectrum along multiple world states. Different agents have different thresholds of world states in which they switch the preferred alternatives.

World State. There are K possible world states $\mathcal{W} = \{1, 2, \cdots, K\}$. The higher the world state is, the more $\mathbf{A}$ is preferred to $\mathbf{R}$. We use k to denote a generic world state. Let $P_k = \Pr[W = k]$ be the common prior of the world state. We assume $P_k > 0$ for every $k \in \mathcal{W}$.

Signal. Every agent receives a signal from $\mathcal{S} = \{1, 2, \cdots, M\}$. We use m to denote a generic signal. Signals are i.i.d conditioned on the world state. Let $P_{mk} = \Pr[\Psi_i = m \mid W = k]$ be the probability that an agent receives signal m given world state is k. The assumption of $P_{hH} > P_{hL}$ in the binary state is extended to the following assumption of stochastic dominance, which requires signals to be positively correlated to world states.

Assumption 1 (Stochastic Dominance)**.** *For any agent i and any world states $k_1 > k_2$,*

$$\Pr[\Psi_i \geq m \mid W = k_1] > \Pr[\Psi_i \geq m \mid W = k_2].$$

Utility Every agent i has a utility function $v_i : \mathcal{W} \times \{\mathbf{A}, \mathbf{R}\} \to \{0, 1, \cdots, B\}$, where B is the positive integer upper bound. We assume that $\mathbf{A}$ is more preferable in a higher world state than in a lower state, and $\mathbf{R}$ is the opposite: for any k_1 and k_2 with $k_1 > k_2$, $v_i(k_1, \mathbf{A}) > v_i(k_2, \mathbf{A})$, and $v_i(k_1, \mathbf{R}) < v_i(k_2, \mathbf{R})$. We also assume $v_i(k, \mathbf{A}) \neq v_i(k, \mathbf{R})$ for any k and any i.

Fraction of agents Given a world state k, let $\alpha_k^{\mathbf{A}}$ and $\alpha_k^{\mathbf{R}}$ be the *approximated* fraction of agents prefer $\mathbf{A}$ ($\mathbf{R}$, respectively) in world state k. We assume $\alpha_k^{\mathbf{A}}$ and $\alpha_k^{\mathbf{R}}$ are independent from n. Formally, given n, let $N(k, \mathbf{A}) = \{i \mid v_i(k, \mathbf{A}) > v_i(k, \mathbf{R})\}$ be the set of agents preferring $\mathbf{A}$ in k ($N(k, \mathbf{R})$ defined similarly). We have $|N(k, \mathbf{R})| = \lfloor \alpha_k^{\mathbf{R}} \cdot n \rfloor$ and $|N(k, \mathbf{A})| = n - |N(k, \mathbf{R})|$. Naturally, $\alpha_k^{\mathbf{A}}$ increases, and $\alpha_k^{\mathbf{R}}$ decreases as k increases. We assume that

$\alpha_k^{\mathbf{A}}$ and $\alpha_k^{\mathbf{R}}$ are common knowledge.

Majority vote and informed majority decision We study the majority vote with threshold μ. If at least $\mu \cdot n$ agents vote for $\mathbf{A}$, $\mathbf{A}$ is announced to be the winner; otherwise, $\mathbf{R}$ is announced to be the winner. The informed majority decision is defined on each world state k. Given a world state k, if $\alpha_k^{\mathbf{A}} > \mu$, we say $\mathbf{A}$ is the informed majority decision; otherwise, we say $\mathbf{R}$ is the informed majority decision. We assume that $\alpha_k^{\mathbf{A}} \neq \mu$ for all $k \in \mathcal{W}$, and the rounding between $\alpha_k^{\mathbf{A}}$ and $|N(k, \mathbf{A})|$ ($\alpha_k^{\mathbf{R}}$ and $|N(k, \mathbf{R})|$, respectively) does not flip the informed majority decision.

Types of agents Let $\mathcal{L} = \{k \in \mathcal{W} \mid \alpha_k^{\mathbf{A}} < \mu\}$ and $\mathcal{H} = \{k \in \mathcal{W} \mid \alpha_k^{\mathbf{A}} > \mu\}$ be the sets of world states where $\mathbf{R}$ ($\mathbf{A}$, respectively) is the informed majority decision. We only consider the case where both $\mathcal{L}$ and $\mathcal{H}$ are non-empty. (Otherwise, an alternative is unanimously the informed majority decision, and there is no uncertainty.) There is a threshold partitioning $\mathcal{W}$ into two sets. Let $L = \max\{k \in \mathcal{L}\}$ to be the largest world where $\mathbf{R}$ is the informed majority decision, and $H = \min\{k \in \mathcal{H}\}$ to be the smallest world where $\mathbf{A}$ is the informed majority decision. We have $H = L + 1$.

Similarly, for an agent i, let $\mathcal{L}_i = \{k \in \mathcal{W} \mid v_i(k, \mathbf{R}) > v_i(k, \mathbf{A})\}$ and $\mathcal{H}_i = \{k \in \mathcal{W} \mid v_i(k, \mathbf{R}) < v_i(k, \mathbf{A})\}$. $\mathcal{L}_i$ ($\mathcal{H}_i$, respectively) is the set of world states where $\mathbf{R}$ ($\mathbf{A}$, respectively) is preferred by i. For a agent i, let $L_i = \max\{k \in \mathcal{L}_i\}$ be the largest world where i prefers $\mathbf{R}$, and $H_i = \min\{k \in \mathcal{H}_i\}$ to be the smallest world where i prefers $\mathbf{A}$. Specifically, let $L_i = 0$ if $\mathcal{L}_i = \emptyset$ and $H_i = K + 1$ if $\mathcal{H}_i = \emptyset$. We have $H_i = L_i + 1$.

1. We say an agent i is *(candidate) friendly* if $\mathcal{L} \cap \mathcal{H}_i \neq \emptyset$. This says that there exists a world state k where i prefer $\mathbf{A}$ while the informed majority decision is $\mathbf{R}$.

2. Similarly, an agent i is *(candidate) unfriendly* if $\mathcal{H} \cap \mathcal{L}_i \neq \emptyset$. This says that there exists a world state k where i prefers $\mathbf{R}$ while the informed majority decision is $\mathbf{A}$.

3. Finally, an agent n is *contingent* if $\mathcal{L}_i = \mathcal{L}$. or equivalently, $\mathcal{H}_i = \mathcal{H}$.

Unlike the binary setting, friendly/unfriendly agents in the non-binary setting do not always prefer one alternative. They just have thresholds above or below the majority.

Example 10. *We extend the COVID policy-making scenario in Example 4 to the non-binary setting.* $N = 20$ *voters decide whether to accept or reject the more-restrictive policy. The world states* $\mathcal{W} = \{1, 2, 3\}$ *describe the risk level of the virus, whereas a larger state represents a higher risk. Suppose* $P_1 = P_2 = 0.3$*, and* $P_3 = 0.4$*.*

Every voter receives a private signal from $\mathcal{S} = \{1, 2, 3, 4\}$*. The larger the signal is, the higher the risk is likely to be. Table 3.4 is the signal distribution given the world state.*

World State	*Signal 1*	*Signal 2*	*Signal 3*	*Signal 4*
1	*0.6*	*0.2*	*0.1*	*0.1*
2	*0.4*	*0.2*	*0.2*	*0.2*
3	*0.1*	*0.2*	*0.3*	*0.4*

Table 3.4: Signal distribution.

The majority threshold is $\mu = 0.6$*. Therefore,* **A** *is the informed majority decision if and only if at least 12 agents prefer* **A** *to* **R***. The voters are categorized into four different groups. Each group has five voters, and voters in the same group share the same utility shown in Table 3.5. The larger the group index is, the more voters prefer* **A** *to* **R***.*

Table 3.6 shows the preferences of each group and the informed majority decision under each world state. The preference of each group comes from the comparison of utilities in Table 3.5. The informed majority decision is the aggregation of group preferences. Since each group has five voters, and the majority threshold is 12 agents, **A** *needs to be preferred by at least three groups to become the informed majority decision. Therefore,* **A** *is the informed majority decision only in state 3. By comparing the preferences of each group and the informed majority decision, we know that group 1 voters are unfriendly agents, group 2 voters are contingent agents, and group 3 and 4 voters are friendly agents.*

Winner	**A**			**R**		
World State	*1*	*2*	*3*	*1*	*2*	*3*
Group 1	*1*	*2*	*3*	*8*	*6*	*4*
Group 2	*2*	*3*	*4*	*6*	*4*	*2*
Group 3	*2*	*5*	*8*	*4*	*3*	*2*
Group 4	*4*	*6*	*9*	*3*	*2*	*1*

Table 3.5: Utility functions group

World State	*1*	*2*	*3*
Group 1	**R**	**R**	**R**
Group 2	**R**	**R**	**A**
Group 3	**R**	**A**	**A**
Group 4	**A**	**A**	**A**
Informed Majority	**R**	**R**	**A**

Table 3.6: Preference of each group and the majority.

Strategy. In the non-binary setting, a strategy can be represented as a vector $\sigma = (\beta_1, \beta_2, \cdots, \beta_M)$, where β_m is the probability that the agent votes for **A** when receiving signal m. A strategy profile is the vector of strategies of all agents. $\Sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$.

The definition of the regular strategy profile remains the same as in the binary setting: friendly agents always vote for **A**, and unfriendly agents always vote for **R**.

Fidelity and Expected Utility Given a strategy profile Σ, let $\lambda_k^{\mathbf{A}}(\Sigma)$ ($\lambda_k^{\mathbf{R}}(\Sigma)$, respectively) be the probability that **A** (**R**, respectively) becomes the winner when world state is k. We can define the fidelity and the expected utility in the same manner as in the binary setting.

$$A(\Sigma) = \sum_{k \in \mathcal{L}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma) + \sum_{k \in \mathcal{H}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma).$$
$$u_i(\Sigma) = \sum_{k=1}^{K} P_k(\lambda_k^{\mathbf{A}}(\Sigma) \cdot v_i(k, \mathbf{A}) + \lambda_k^{\mathbf{R}}(\Sigma) \cdot v_i(k, \mathbf{R})).$$

Excess Expected Vote Share Similarly, the excess expected vote share is the expected vote share that an alternative attracts under state k minus the threshold of the alternative. In different instances, the informed majority decision may change in different world states.

Therefore, we define the excess expected vote share for both **A** and **R** in every world state.

$$f^n_{k\mathbf{A}} = \frac{1}{n}\sum_{i=1}^{n} E[X_i^n \mid k] - \mu. \quad f^n_{k\mathbf{R}} = \frac{1}{n}\sum_{i=1}^{n} E[1 - X_i^n \mid k] - (1-\mu).$$

For world states $k \in \mathcal{H}$ where **A** is the informed majority decision, we care about $f^n_{k\mathbf{A}}$; and for states $k \in \mathcal{L}$, we care about $f^n_{k\mathbf{R}}$. Therefore, we define f^n to be the smallest excess expected vote share among those we care about, i.e. $f^n = \min\left(\min_{k\in\mathcal{H}}(f^n_{k\mathbf{A}}), \min_{k\in\mathcal{L}}(f^n_{k\mathbf{R}})\right)$. We use $f_{k\mathbf{A}}, f_{k\mathbf{R}}$, and f to denote excess expected vote share in symmetric profile sequences.

Instance and Sequence of Strategy Profiles. Let $\{\mathcal{I}_n\}_{n=1}^{\infty}$ (or $\{\mathcal{I}_n\}$ for short) be a sequence of instances, where each $\mathcal{I}_n$ is an instance of n agents. The instances in a sequence share the same majority threshold μ, world state prior distribution $\{P_k\}$, signal prior distribution $\{P_{mk}\}$, and approximated type fractions $(\alpha_k^{\mathbf{A}}, \alpha_k^{\mathbf{R}})$. Same to the binary setting, we do not regard agents in different instances as related and have no additional assumption on the utility functions of agents. We define a sequence of strategy profile $\{\Sigma_n\}_{n=1}^{\infty}$ based on the instance sequence, where for each n, Σ_n is a strategy profile in $\mathcal{I}_n$.

Our positive results on strategic voting can be extended to the non-binary setting. Theorem 5 states the equivalence of fidelity converging to 1 and the ε-strong Bayes-Nash equilibrium with $\varepsilon = o(1)$. Theorem 6 guarantees the existence of the regular profile sequence with fidelity converging to 1.

Theorem 5. *Given an arbitrary sequence of instances and an arbitrary regular strategy profile sequence* $\{\Sigma_n\}_{n=1}^{\infty}$*, let* $\{A(\Sigma_n)\}_{n=1}^{\infty}$ *be the sequence of the fidelities of* Σ_n*.*

- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$*, then for every* n*,* Σ_n *is an* ε*-strong BNE with* $\varepsilon = o(1)$*.*
- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$ *does not hold, then there exist infinitely many* n *such that* Σ_n *is NOT an* ε*-strong BNE for some constant* ε*.*

Theorem 6. *Given any arbitrary sequence of instances, there always exists a sequence of regular strategy profiles* $\{\Sigma'_n\}_{n=1}^{\infty}$ *such that* $A(\Sigma'_n)$ *converges to 1.*

Theorem 5 and Theorem 6 preserve the same reasoning as Theorem 2 and Theorem 3. More details, as well as the characterization of the fidelity in the non-binary setting, are available in Appendix A.3

CHAPTER 4
THE ART OF TWO-ROUND VOTING

Chapter 3 reveals that any equilibrium in the majority vote can reach the informed majority decision with high probability. However, as shown in Section 3.4.1, the natural and intuitive behavior of *informative voting*, in which every voter votes for his/her signal, and *sincere voting*, in which every voter votes as if he/she is making an individual decision and vote for his/her preferred alternative in expectation, are not equilibria in many situations. On the other hand, the equilibrium constructed involves carefully calculated mixed strategies.

In analyzing strategic behavior, we emphasize the importance of natural and intuitive equilibria. It is quite unlikely that real-world agents can compute and play counterintuitive mixed or asymmetric strategies, let alone coordinate to the same equilibrium. Therefore, these counterintuitive strategy profiles fail to predict the outcome of strategic agents as they are intended to. Moreover, the simplicity of the mechanism is also a desirable property and a proxy of usability, as the theoretical guarantees of complex mechanisms may not translate into practice.

Can we design *simple* mechanisms with *natural* equilibria
to reach the informed majority decision?

The definition of a natural equilibrium depends on the scenarios and the purposes. In this chapter, we identify the following criteria. Firstly, agents play pure/deterministic strategies. Mixed or randomized strategies are hard for agents to compute and implement and are rarely applied by real-world voters. Secondly, agents play symmetrically, i.e., agents with the same preferences should play the same strategy. Otherwise, the mapping from similar agents to diverse strategies is unclear and unincentivized, and the equilibrium is unlikely to predict the outcome. Moreover, agents can easily compute the strategies. Two examples of intuitive and natural strategies are informative voting and sincere voting, where the first simply follows the signal, and the second follows the preferences in expectation.

Unfortunately, previous one-round voting mechanisms fail to have a natural equilibrium. Moreover, **one round voting game with the majority rule has NO pure symmetric equilibrium** under a large class of games (Proposition 1). The known equilibria are either randomized or asymmetric and are hard to compute.

Another line of work designs mechanisms to achieve the informed majority decision via simple equilibria, yet the mechanism structures are sophisticated. Schoenebeck and Tao [9] propose an information-aggregation mechanism that incentivizes truthful reporting and reaches the informed majority decision with a high probability. However, the mechanism leverages the "surprisingly popular" mechanism [46] and the median-voter theorem [81], resulting in a structure that is quite sophisticated and may not be trusted by the people. Moreover, the mechanism elicits a distribution from each agent, which requires infinite bits to be represented. This further increases the difficulty of implementing the mechanism.

4.1 Overview of the Chapter

In this chapter, we provide a solution to the research question of simple mechanism design—*two-round voting*, which has a polling stage (first round) and a voting stage (second round). Between the two rounds, agents can observe the number of votes cast for each alternative in the first round, and the majority vote in the second round completely determines the winner.

For a sharp contrast, we first show that the one-round voting game with the majority rule has NO pure and symmetric equilibrium under a large class of games (Proposition 1), which also rules out the natural informative voting and sincere voting. We also illustrate the existing mixed or asymmetric equilibria brings difficulties for agents to calculate and coordinate.

On the other hand, the two-round voting mechanism is a simple and powerful mechanism with natural equilibria. Firstly, we prove that the two-round voting mechanism indeed achieves good decisions with strategic agents, as every ε-approximate strong Bayes-Nash equilibrium, in which no group of agents has incentives to deviate, leads to the informed majority decision with high probability (Theorem 7). The probability converges to 1 as ε goes to 0. Secondly, the two-round voting mechanism has the merit of a simple mechanism in both structure-wise and equilibrium-wise. It has a simple structure that involves simply conducting a poll before the actual vote. Such a polling-voting structure has been widely applied in real-world elections.

Moreover, we show a natural ε-approximate strong Bayes-Nash equilibrium that leads to the informed majority decision (Theorem 8). As the number of voters increases, ε converges to 0. The equilibrium combines the most natural and classical informative voting and sincere voting. In this equilibrium, agents whose preferred alternative changes with the world state vote informatively (vote for their signal) in the first round, and vote sincerely (vote for the alternative they preferred in expectation conditioned on the first-round outcome) in the second round. The "informative+sincere" equilibrium follows the intuition of sharing information and choosing the (expected) preferred alternative. We further discover that the "informative+sincere" equilibrium belongs to a class of "informative + threshold" equilibria (Theorem 9), in which agents vote informatively in the first round and vote for the alternative whose first-round vote exceeds a threshold. Such a class derives more natural equilibria, including the "informative + surprisingly popular" equilibrium, which resembles the "surprisingly popular" mechanism.

Furthermore, we show with mild assumptions that every one-round equilibrium gives a two-round equilibrium that is not more "complicated" (Theorem 10). Specifically, agents in the two-round voting mechanism can ignore the polling stage and play the one-round equilibrium in the second round. Therefore, the two-round voting mechanism has a natural equilibrium in every instance, including those where one-round voting fails to have a natural solution, and it can reach an informed majority decision whenever one-round voting can.

We also test the one-round voting and the two-round voting with synthetic experiments where we apply generative AI agents to simulate voters. When the signal distribution is biased, the AI agents powered by the reasoning model reach the informed majority decision more often in the two-round voting than in the one-round voting. We also observe Bayesian updates and a vote switch (in the second round). Our experiments provide insights into future applications for AI-augmented or proxy votes and serve as references for real-world experiments.

4.2 Equilibria in One-Round Voting

In this chapter, we fix the majority winning threshold $\mu = \frac{1}{2}$. In order to distinguish one-round and two-round voting, we use $\bar{\sigma}$ and $\bar{\Sigma}$ to denote a strategy and a strategy profile in one-round voting.

We show scenarios where natural strategy profiles fail to reach high fidelity (and thus are not approximate strong BNEs according to the if-and-only-if characterization in Chapter 3) and how complex the existing equilibria are.

Firstly, in games with only contingent agents with identical utility functions, NO pure symmetric equilibria exist in the one-round voting game in ANY environment with biased signals.

Proposition 1. *Suppose $\alpha_C = 1$. For any sequence of environments $\{\mathcal{I}_n\}$ such that $P_{hL} > 0.5$ or $P_{lH} > 0.5$ and that for each n, all agents share the same utility function v, no sequence pure symmetric strategy profiles $\{\hat{\Sigma}_n\}$ of the one-round voting game satisfies that $\hat{\Sigma}_n$ is an ε-strong BNE for every n with $\varepsilon \to 0$.*

Note that our definition of symmetry (agents with the same utility function play the same strategy) is very broad. Therefore, Proposition 1 excludes a large number of strategy profiles from being equilibria.

Proof. Without loss of generality, we consider the case that $P_{hL} > 0.5$, and the reasoning for the case that $P_{lH} > 0.5$ is similar. There are in total four pure symmetric strategy profiles:

all agents always vote for **A**, always vote for **R**, vote informatively, and vote against their private signals (vote for **A** when receiving signal ℓ and for **R** when receiving signal h). Always voting for an alternative with probability P_H or P_L to miss the informed majority decision cannot be an equilibrium. When $P_{hL} > 0.5$, agents are more likely to receive h than ℓ on world state L. Therefore, by applying the Hoeffding inequality, informative voting does not lead to the informed majority decision in the world state L with high probability and cannot be an equilibrium. Similarly, voting against signals cannot be an equilibrium as $P_{hH} > 0.5$. □

We give more examples where natural strategies fail to be equilibria in the one-round voting, including sincere voting (where an agent chooses the alternative maximizing his/her expected utility when this alternative becomes the winner) and the Bayesian strategy (an agent votes for an alternative with the probability of the Bayesian posterior probability that this alternative is the informed majority decision).

We follow the example where all agents are contingent, and the signal distribution is $P_{hH} = 0.9, P_{lH} = 0.1, P_{hL} = 0.7$, and $P_{lL} = 0.3$. Now we show that sincere voting fails, and the Bayesian strategy is an equilibrium only when common prior is $P_H = P_L = 0.5$,

One natural strategy is to choose the alternative that maximizes the agent's expected utility, assuming this agent is the only voter who decides the outcome. This is called *the sincere strategy*. Suppose the common prior is $P_H = P_L = 0.5$. An agent's posterior belief for world H upon receiving signal h is

$$\Pr(W = H \mid S_i = h) = \frac{\Pr(W = H) \cdot \Pr(S_i = h \mid W = H)}{\Pr(S_i = h)} = \frac{0.5 \times 0.9}{0.5 \times 0.9 + 0.5 \times 0.7} = \frac{9}{16},$$

and the posterior belief for world H upon receiving signal l is

$$\Pr(W = H \mid S_i = l) = \frac{\Pr(W = H) \cdot \Pr(S_i = l \mid W = H)}{\Pr(S_i = l)} = \frac{0.5 \times 0.1}{0.5 \times 0.1 + 0.5 \times 0.3} = \frac{1}{4}.$$

If an agent i receives h, the expected utility of voting for $\mathbf{A}$ is

$$\Pr(W = H \mid S_i = h) \cdot v_i(H, \mathbf{A}) + \Pr(W = L \mid S_i = h) \cdot v_i(L, \mathbf{A}) = \frac{9}{16} v_i(H, \mathbf{A}) + \frac{7}{16} v_i(L, \mathbf{A}),$$

and the expected utility of voting for $\mathbf{R}$ is

$$\Pr(W = H \mid S_i = h) \cdot v_i(H, \mathbf{R}) + \Pr(W = L \mid S_i = h) \cdot v_i(L, \mathbf{R}) = \frac{9}{16} v_i(H, \mathbf{R}) + \frac{7}{16} v_i(L, \mathbf{R}).$$

When voting sincerely, the agent receiving signal h will vote for $\mathbf{A}$ if and only if the expected utility for $\mathbf{A}$ is higher:

$$\frac{9}{16} v_i(H, \mathbf{A}) + \frac{7}{16} v_i(L, \mathbf{A}) > \frac{9}{16} v_i(H, \mathbf{R}) + \frac{7}{16} v_i(L, \mathbf{R}).$$

Similarly, the agent receiving signal l will vote for $\mathbf{A}$ if and only if

$$\frac{1}{4} v_i(H, \mathbf{A}) + \frac{3}{4} v_i(L, \mathbf{A}) > \frac{1}{4} v_i(H, \mathbf{R}) + \frac{3}{4} v_i(L, \mathbf{R}).$$

The utility for a contingent agent satisfies that $v_i(H, \mathbf{A}) > v_i(H, \mathbf{R})$ and $v_i(L, \mathbf{R}) > v_i(L, \mathbf{A})$. However, when agents care more about the outcome under world H and are relatively indifferent about the outcome under world L, they will always vote for $\mathbf{A}$ and it is not favored by them under world L. In our example, if $v_i(H, \mathbf{A}) - v_i(H, \mathbf{R}) > 3 \cdot (v_i(L, \mathbf{R}) - v_i(L, \mathbf{A}))$, agents will vote for $\mathbf{A}$ even when signal l is received. Thus, the sincere strategy profile does not guarantee the informed majority decision, and it is not a strong BNE by the if-and-only-if characterization.

Perhaps one more natural strategy is to vote according to the posterior, i.e., adopt the mixed strategy with probabilities matching the posterior belief: an agent votes for $\mathbf{A}$ with probability $\beta_\ell = \Pr(W = H \mid S_i = l)$ when receiving signal ℓ and probability $\beta_h = \Pr(W = H \mid S_i = h)$ when receiving signal h. In our specific environment, with the prior belief $P_H = P_L = 0.5$, our previous calculations reveal that the strategy is given by $(\beta_l, \beta_h) = (\frac{1}{4}, \frac{9}{16})$. When the actual world is H, the "correct" alternative **A** gets a share of $0.1 \cdot \frac{1}{4} + 0.9 \cdot \frac{9}{16} = 53.125\%$ in expectation; when the actual world is L, the "incorrect" alternative **A** wins a share of $0.3 \cdot \frac{1}{4} + 0.7 \cdot \frac{9}{16} = 46.875\%$ in expectation. This looks good: the "correct" alternative always receives more than half of the votes in expectation. Some calculations reveal that the informed majority decision can be guaranteed (more formally, the fidelity goes to 1 when $n \to \infty$) by this kind of matching-posterior strategy profile if the prior belief is half-half (i.e., $P_H = P_L = 0.5$). However, if the prior belief is significantly biased, say, P_H is close to 1, some calculations reveal that the expected fraction of **A** votes is close to 1 under both world states, in which case the informed majority decision is not secured under the world state L.

In the environments where pure symmetric strategies fail to be equilibria, reaching the informed majority decision requires agents to artificially "shift" the probabilities in the strategy. For example, consider a biased setting where the common prior is $P_H = P_L = 0.5$, and the signal distribution is $P_{hH} = 0.9, P_{lH} = 0.1, P_{hL} = 0.7$, and $P_{lL} = 0.3$. An equilibrium under this setting is a symmetric mixed profile where agents with signal l always vote for **R**, and agents with h vote for **A** with probability 0.6.

In this strategy profile, the informed majority decision receives more than half of the vote in expectation. Alternatively, agents can play a pure asymmetric profile where 40% of agents always vote for **R**.

These "successful" strategy profiles, however, are often too complex for agents to agree on. For the mixed strategy profile, even if agents are clever enough to be aware of a probability shift, different agents may choose different probability shifts, which may still lead to the undesirable outcome. For the pure asymmetric profile, without an intrinsic incentive, it is difficult and unclear for identical agents to coordinate who casts the "opposite vote" **R** when receiving signal h.

4.3 Two-Round Voting and Equilibria

Two-round voting and one-round voting share the same environment, including the number of agents, distributions, the fraction of each type, and utilities. Their difference lies in the game form and the strategies.

Agents play a two-round anonymous voting game where the winner is completely determined by the second round. The five steps are shown in Algorithm 1.

Algorithm 1 Two-Round voting mechanism

1: All the agents receive their signals.
2: **First round** (Polling stage): agents cast the first round vote for **A** or **R**.
3: Agents observe the number of votes for **A** in the first round denoted by $x \in \{0, 1, \cdots, n\}$.

4: **Second Round** (Voting stage): agents cast the second round vote for **A** or **R**.
5: The alternative that gets more than a 0.5 fraction of votes in the second round becomes the winner.

Strategy. A strategy in a two-round voting game σ_i is a pair (σ_i^1, σ_i^2). The first-round strategy σ_i^1 is a mapping from agent i's signal to a distribution on $\{\mathbf{A}, \mathbf{R}\}$ denoting her action in the first round. The second-round strategy σ_i^2 is a mapping from agent i's signal AND the result of the first-round vote x to a distribution on $\{\mathbf{A}, \mathbf{R}\}$ denoting her action in the second round. A strategy profile Σ is the vector of strategies of all agents. The definition of a pure and a symmetric strategy profile follows that of the one-round voting.

We define a sequence of two-round voting strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$ on an environment sequence $\{\mathcal{I}_n\}$. For each n, Σ_n is a profile of a two-round voting game with parameters in $\mathcal{I}_n$.

Expected Utility and Fidelity. Given a strategy profile Σ, let $\lambda_k^{\mathbf{A}}(\Sigma)$ ($\lambda_k^{\mathbf{R}}(\Sigma)$, respectively) be the—ex-ante, before agents receiving their signals—probability that **A** (**R**, respectively) becomes the winner when the world state is k. The expected utility in the two-round voting is in the same form as that in the one-round voting.

$$u_i(\Sigma) = P_L(\lambda_L^{\mathbf{A}}(\Sigma) \cdot v_i(L, \mathbf{A}) + \lambda_L^{\mathbf{R}}(\Sigma) \cdot v_i(L, \mathbf{R})) \\ + P_H(\lambda_H^{\mathbf{A}}(\Sigma) \cdot v_i(H, \mathbf{A}) + \lambda_H^{\mathbf{R}}(\Sigma) \cdot v_i(H, \mathbf{R})).$$

Fidelity is the likelihood that the informed majority decision is reached. In the two-round voting game, the fidelity of a strategy profile Σ is as follows.

$$A(\Sigma) = P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma) + P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma).$$

An ε-strong BNE in the two-round voting has the same definition as that in the one-round voting, except that the strategies and expected utilities are those in the two-round voting.

Some Remarks. The game of two-round voting is not exactly a standard *Bayesian game*, as the strategy of an agent also depends on the first round outcome. Our definitions of strategies, utilities, and equilibria are straightforward extensions of the standard Bayesian games.

Another natural formulation for the two-round voting is an *extensive-form game*. We believe this is an interesting direction for future work. However, standard equilibrium solution concepts such as *subgame perfect Nash equilibrium* and *perfect Bayesian equilibrium* fail to capture the feature of group deviations. The first step in this future direction is to propose a reasonable equilibrium concept that extends the strong Bayes-Nash equilibrium to extensive-form games.

4.3.1 Equilibria in Two-Round Voting

We show that the two-round voting mechanism indeed has natural equilibria leading to the informed majority decision. First, we show that every ε-strong BNE in the two-round voting mechanisms with ε converging to 0 reaches the informed majority decision with high probability.

Theorem 7 (**All ε-BNE are good**)**.** *Let $\{\Sigma_n\}$ be a sequence of profiles in the two-round voting game. If for any n, Σ_n is an ε-strong BNE with $\varepsilon = o(1)$, then $\lim_{n\to\infty} A(\Sigma_n) = 1$.*

Moreover, we show the existence of a natural equilibrium in the two-round mechanism. While the classical informative voting or sincere voting may not be an equilibrium in the one-round voting game, their combination becomes an equilibrium.

"Informative + Sincere" voting All the friendly agents always vote for **A** in both rounds, and all the unfriendly agents always vote for **R** in both rounds. All the contingent agents vote informatively in the first round and vote sincerely based on their observation of the first round outcome, as shown in Definition 7.

Definition 7 (Sincere Voting in the second round.)**.** The agent compares his/her expected utility between the event that **A**/**R** becomes the winner conditioned on his/her private signal m, the first-round outcome x. He/she votes for **A** if and only if $u_i(\mathbf{A} \mid x, m) \geq u_i(\mathbf{R} \mid x, m)$.

$$u_i(\mathbf{A}|x,m) = \Pr[W = L|x,m] \cdot v_i(L,\mathbf{A}) + \Pr[W = H|x,m] \cdot v_i(H,\mathbf{A}).$$
$$u_i(\mathbf{R}|x,m) = \Pr[W = L|x,m] \cdot v_i(L,\mathbf{R}) + \Pr[W = H|x,m] \cdot v_i(H,\mathbf{R}).$$

Theorem 8 (**"Informative + sincere" equilibrium**)**.** *For any environment sequence $\{\mathcal{I}_n\}$, the sequence of the "information + sincere" strategy profiles $\{\Sigma_n^\dagger\}$ satisfies that (1) $\lim_{n\to\infty} A(\Sigma_n^\dagger) = 1$, and (2) for every n, $\Sigma_n^\dagger$ is an ε-strong BNE where ε converges to 0 as n goes to ∞.*

The rationale behind the "informative + sincere" strategy is straightforward—by observing the outcome of the first-round informative voting, every contingent agent knows the private information of all contingent agents anonymously, which informs them of the correct world state with high probability. Therefore, in the second round, they are nearly certain of their preferred alternative and vote accordingly.

While agents may not necessarily perform Bayesian updates in the votes, we show that the intuition of "sharing information and getting updated" derives a large class of "informative + threshold" equilibria (including the "informative + sincere" equilibrium). Therefore, as agents still follow this "informative + threshold" pattern, the informed majority can be reached.

In an "informative + threshold" strategy profile, all the predetermined agents vote for their preferred alternative in both rounds; every contingent agent i votes informatively in the first round and votes for $\mathbf{A}$ in the second round if and only if the first round outcome x exceeds some threshold x_i^*.

For the "informative + sincere" voting, the threshold x^* is determined by the constraint $u_i(\mathbf{A} \mid x, m) \geq u_i(\mathbf{R} \mid x, m)$, as agents apply the Bayes Theorem to update their belief. Specifically, a contingent agent i vote for $\mathbf{A}$ in the second round if and only if $x \geq x_i^*$, where

$$x_i^* = n_F + \frac{\log \frac{P_L \cdot (v_i(L,\mathbf{R}) - v_i(L,\mathbf{A}))}{P_H \cdot (v_i(H,\mathbf{A}) - v_i(H,\mathbf{R}))} + n_C \cdot \log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH} P_{\ell L}}{P_{\ell H} P_{hL}}}.$$

Example 11. *We calculate the "informative + sincere" strategy under the setting in Example 4. In this environment, $n_F = 4$, $n_C = 10$, $P_H = 0.4$ $P_{hH} = 0.8$, and $P_{hL} = 0.6$. Therefore, $x_i^* = 4 + \frac{\log \frac{9}{2} + 10 \times \log 2}{\log \frac{8}{3}} \approx 12.6$. Therefore, the contingent voters who vote sincerely in the second round will vote for* $\mathbf{A}$ *if and only if there are at least 13 votes for* $\mathbf{A}$ *in the first round.*

Additionally, "informative + threshold" class also includes more "natural" strategy profiles, such as the "informative + surprisingly popular" strategy profile.

Example 12. *The "informative + surprisingly popular" strategy shares a similar idea with the "surprisingly popular" mechanism [46]. The threshold of every agent i is the prior expectation on the first-round outcome x, which is $x^* = n_F + n_C \cdot (P_H \cdot P_{hH} + P_L \cdot P_{hL})$. This* *comes from the fact that the n_F friendly agents always vote for* **A***, and the expected share of n_C contingent agents voting for* **A** *is P_{hH} under world state H and P_{hL} under world state L. When the number of votes for* **A** *exceeds the expectation,* **A** *becomes the "surprisingly popular" alternative and is elected the winner in the second round. Otherwise,* **R** *becomes the "surprisingly popular" alternative.*

The "informative + threshold" strategy profiles share the merit that the "informative + sincere" equilibrium has. Every agent votes deterministically. When agents with equal utility functions adopt the same threshold, they are also symmetric. Most importantly, every such strategy profile in which all thresholds lie between the expected first-round outcome conditioned on two world states is an equilibrium that leads to the informed majority decision.

Definition 8. We say a sequence of "informative + threshold" strategies is constantly separable if there exists a $n_0 > 0$ and a constant $\delta > 0$ such that for every $n \geq n_0$ and every agent i, $n_F + n_C \cdot P_{hL} + \delta \cdot n \leq x_i^* \leq n_F + n_C \cdot P_{hH} - \delta \cdot n$. (See Figure 4.1.)

Theorem 9 (**"Informative + threshold" equilibrium**)**.** *Under any environment sequence $\{\mathcal{I}_n\}$, every constantly separable sequence of profiles $\{\Sigma^*\}$ satisfies that (1) $\lim_{n\to\infty} A(\Sigma_n^*) = 1$, and (2) for every n, Σ_n^* is an ε-strong BNE where ε converges to 0 as n goes to ∞.*

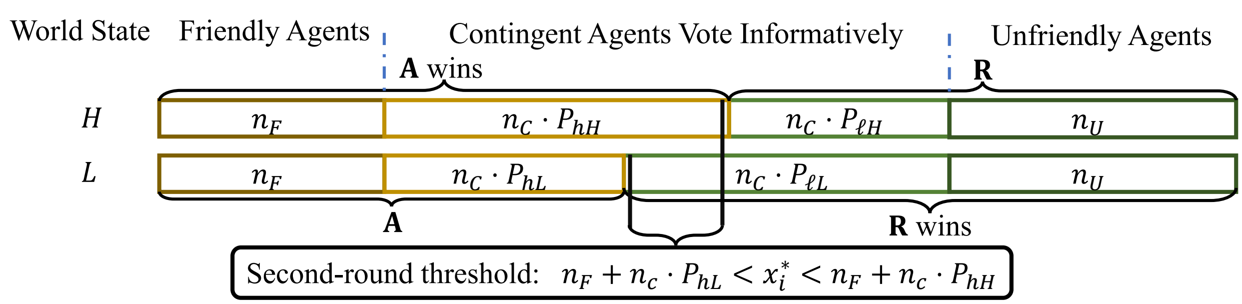


Figure 4.1: Illustration of the range of the threshold for an "informative + threshold" equilibrium.

Remark on Theorem 7. Theorem 7 states that any $o(1)$-strong BNE leads to nearly optimal fidelity. Unlike the one-round voting, the other direction—optimal fidelity implying equilibria—does not hold in the two-round voting game, even when predetermined agents only play their dominant second-round strategies.

For example, friendly and unfriendly agents voting informatively in the first round can help contingent agents infer the world state and lead to an informed majority decision. However, this does not form an equilibrium, as friendly agents have incentives to always vote for **A** in the first round and deceive contingent agents that the world state is H. Similarly, unfriendly agents have incentives to vote for **R**.

4.3.2 Proof Sketch for Theorem 7

Suppose $\{\Sigma_n\}$ is a sequence of strategy profiles whose fidelity does not converge to 1. We show that there exists a different sequence of profiles $\{\Sigma_n'\}$ such that agents have incentives to deviate from Σ_n to Σ_n' for infinitely many n. We first give the proof when $\{\Sigma_n\}$ is a *regular* profile sequence, where all friendly agents always vote for **A** and all unfriendly agents always vote for **R** in the *second* round. This helps us characterize the main idea while reducing complications. Then we extend the proof to non-regular profiles. The full proof is in Appendix B.1.4.

Proof for regular profiles. We take an infinite set $\mathcal{N} \subseteq \mathbb{N}_+$ and a constant $\delta > 0$ such that for all $n \in \mathcal{N}$, the fidelity of Σ_n does not exceed $1 - \delta$. Let the deviating group D be all the contingent agents. If Σ_n' increases the fidelity, contingent agents get higher expected utility and are incentivized to deviate. Therefore, we aim to construct Σ_n' so that contingent agents reveal the world state with high probability in the first round and vote for the (likely) informed majority decision in the second round. Given that predetermined agents vote for their preferred alternative in the second round, the informed majority decision becomes the winner.

Contingent agents can (collectively, with high probability) reveal the world states because signal distributions are different in different states of the world. The expected number of h signals from contingent agents is $P_{hH} \cdot n_C$ in world state H and $P_{hL} \cdot n_C$ in world state L. If all the contingent agents vote informatively in the first round, by the Hoeffding inequality, the number of votes will concentrate to the expectation with high probability. The expected votes for A is higher in world state H than in world state L, as $P_{hH} > P_{hL}$. Therefore, contingent agents can play a "threshold" strategy (between $P_{hH} \cdot n_C$ and $P_{hL} \cdot n_C$) in the second round: vote for **A** together if the first round result x exceeds the threshold and vote for **R** together otherwise.

However, such a strategy still must overcome one difficulty: strategies of predetermined agents in the first round are not guaranteed to be "regular" and may offset the difference in the votes of contingent agents. For example, predetermined agents can vote for **R** more often in world state H than in world state L in Σ_n so that x has a smaller expectation in world state H than in world state L. However, contingent agents can "reverse" their votes and observations accordingly: they vote opposite to their signal in the first round and vote for **A** if x is below a threshold. In this way, they can still distinguish the different world states.

Now we explicitly present the strategy profile Σ'_n. Let $\phi_H(\Sigma'_n)$ and $\phi_L(\Sigma'_n)$ be the expectation of the first-round outcome x of Σ'_n conditioned on the world state H and L respectively.

1. All predetermined agents (they do not deviate) play the same strategy as in Σ_n. This includes that they play their dominant second-round strategies.

2. If more votes for **A** from predetermined agents are expected in world state H than in L in the first round, all contingent agents vote informatively in the first round. For the second round, they vote for **A** if $x \geq \frac{\phi_H(\Sigma'_n)+\phi_L(\Sigma'_n)}{2}$ and vote for **R** otherwise.

3. Otherwise, contingent agents "reverse" the strategy. In the first round, they vote for **A** with l and **R** with h. In the second round, they vote for **A** if $x < \frac{\phi_H(\Sigma'_n)+\phi_L(\Sigma'_n)}{2}$ and **R** otherwise.

In this way, the fidelity $A(\Sigma'_n)$ converges to 1 as n goes to infinity, contingent agents have incentives to deviate, and Σ_n is NOT an ε-strong BNE for some constant ε.

Proof Sketch for non-regular profiles. We adopt the same idea to construct a strategy profile where the deviating group forces the state to be revealed by the first-round vote and votes for the preferred agent in the second round. The difficulty for the non-regular case is that a deviating group with only contingent agents may not work, as there is no assumption on the behavior of predetermined agents in the second round. (For example, if $\alpha_C < 0.5$, and all the predetermined agents always vote for **A**, then the winner will always be **A**.) Fortunately, the following lemma shows that there will always be a type of predetermined agents who have incentives to deviate together to reach the informed majority decision.

Lemma 2. *Let $\{\Sigma_n\}$ be a sequence of profiles such that $A(\Sigma_n)$ does not converge to 1. Then there exist infinitely many n such that for each n, at least one of the following holds for all sequences of profiles $\{\Sigma'_n\}$ with $\lim_{n\to\infty} A(\Sigma'_n) = 1$:*
(1) For every friendly agent i_1, $u_{i_1}(\Sigma'_n) - u_{i_1}(\Sigma_n) \geq 0$. (2) For every unfriendly agent i_2, $u_{i_2}(\Sigma'_n) - u_{i_2}(\Sigma_n) \geq 0$.
Moreover, for a fixed n and Σ_n, either 1 holds for all Σ'_n or 2 holds for all Σ'.

In this way, we construct two deviating strategies. In the first strategy profile Σ_F, all the friendly and contingent agents play an "informative + threshold" strategy. The second strategy profile Σ_U is defined similarly. With similar reasoning to the regular case, both Σ_F and Σ_U have fidelity converging to 1. Applying Lemma 2, there is a type of pre-determined agents that get expected utility increase via deviation in both Σ_F and Σ_U. Contingent agents will also get a constant increase in expected utility due to the increase in fidelity. Therefore, agents have incentives to deviate to either Σ_F or Σ_U, and Σ_n is NOT an ε-strong BNE with constant ε.

4.3.3 Proof Sketch for Theorem 8 and 9.

As the "informative + sincere" strategy profile is also an "informative + threshold" strategy profile, it suffices to prove Theorem 9 and show that the sincere threshold satisfies the constantly separable constraint. The proof of Theorem 9 consists of two parts: Lemma 3 shows that the fidelity converges to 1, and Lemma 4 shows that the profiles are $o(1)$-strong BNE. The full proof is in Appendix B.1.2.

Lemma 3. *For any sequence of instances and any sequence of "informative + threshold" strategy profiles $\{\Sigma_n\}$, if there exists a constant $\delta > 0$ such that for all n and all agents i, $n_F + n_C \cdot P_{hL} + \delta \cdot n \leq x_i^n \leq n_F + n_C \cdot P_{hH} - \delta \cdot n$, then $A(\Sigma_n)$ converges to 1.*

Lemma 3 uses reasoning similar to the proof of Theorem 7 to show the deviating strategy profile has fidelity converging to 1. Firstly, the threshold x^* is guaranteed to be between the expectation of the first-round vote conditioned on the world state being H and being L. Therefore, with probability converging to 1, the first-round outcome is larger than all the thresholds under world state H and lower than all the thresholds under world state L. Therefore, the contingent agents know the world state and vote for the informed majority decision with high probability.

Lemma 4. *Σ_n^* is an ε-strong BNE with $\varepsilon = o(1)$.*

The proof of Lemma 4 follows a similar idea to the proof of Theorem 2. For arbitrary n, and a strategy profile Σ_n, let $\varepsilon = 2B(B+2)(1 - A(\Sigma_n))$, where B is the upper bound of the utility. Then, for any deviating strategy profile Σ'_n, we show that the deviation will not succeed. Therefore, Σ_n is an ε-strong BNE. Since $A(\Sigma_n^*)$ converges to 1, ε converges to 0.

Lemma 5 (Informal)**.** *For any strategy profile Σ, a deviating group D that has an incentive to deviate from Σ to another strategy profile Σ', under the approximation parameter $\varepsilon = 2B(B+2)(1-A(\Sigma))$ for ε-strong BNE, contains either only friendly agents or only unfriendly agents.*

Lemma 5 shares almost the same proof as Lemma 1. The formal statement of Lemma 5 is in Appendix B.1.1.

Then we show that D contains only friendly agents or only unfriendly agents cannot succeed. Without loss of generality, suppose D contains only friendly agents. If friendly agents deviate in the first round, the result of the first-round voting x will be less likely to reach the threshold, and contingent agents are more likely to vote for $\mathbf{R}$. If friendly agents deviate in the second round, there will be strictly fewer expected votes for $\mathbf{A}$, and $\mathbf{A}$ will be less likely to be the winner. Therefore, no deviation succeeds, and Σ_n^* is an ε-strong BNE with ε converging to 0.

Finally, we show that the "informative+sincere" threshold is between the two expectations.

Lemma 6. *The sequence of "informative + sincere" strategy profiles is constantly separable.*

It shares the same idea with Lemma 3. Recall that for any agent i, the threshold x_i^* is the boundary for $u_i(\mathbf{A} \mid x, m) \geq u_i(\mathbf{R} \mid x, m)$, When x is close to the upper bound (expected utility when the world state is H), $\Pr[H \mid x, m]$ is close to 1, and $u_i(\mathbf{A} \mid x, m) \geq u_i(\mathbf{R} \mid x, m)$ always hold (as contingent agents prefer $\mathbf{A}$ in world state H); and when x is close to the lower bound (expected utility when the world state is L), $\Pr[H \mid x, m]$ is close to 0, and $u_i(\mathbf{A} \mid x, m) < u_i(\mathbf{R} \mid x, m)$ always hold. Therefore, the threshold must be in between.

4.4 Two-Round Voting is More Powerful

In this section, we show that in any sequence of environments where there exists a one-round equilibrium (where predetermined agents play their dominant strategy), there always exists a two-round equilibrium that will not be more "complicated" than the one-round equilibrium. Precisely, in this two-round equilibrium, agents disregard the first-round voting and play exactly as in the one-round equilibrium in the second round. By not exploiting the information revelation in the first round, the two-round voting was reduced to a one-round voting game. (A one-round voting strategy profile is *regular* if all predetermined agents play their dominant strategies.)

Therefore, we claim that two-round voting is more powerful than one-round voting, as it can reach an informed majority decision whenever one-round voting can reach it, and it provides a natural equilibrium in every environment including those where one-round voting fails to have a natural solution. The proof resembles that of Theorem 9.

Theorem 10. *Under the same sequence of environments, for any regular one-round voting strategy profile sequence being $o(1)$-strong BNE, any two-round voting profile sequence, where agents play exactly this one-round strategy in the second round, is also an $o(1)$-strong BNE.*

4.5 Experiments

In addition to our theoretical results, we conducted simulated voting experiments with generative AI agents to compare one-round and two-round voting. Specifically, we inject generative AI bots with different preferences and information drawn from given distributions, then run both one-round and two-round voting under various environment parameters. Our findings show that the generative AI agents achieved the informed majority decision more often under the two-round mechanism than under the one-round mechanism.

4.5.1 Experimental Setting

Our experiments employed two generative AI modelsOpenAI GPT-4o and Deepseek R1, a newly released reasoning model. In each voting scenario, all voters were drawn from the same model.

For each vote, we set $n = 40$ agents, including 5 friendly agents, 3 unfriendly agents, and 32 contingent agents. The prior is fixed at $P_H = 0.5$. We run experiments on two different signal distributions, the unbiased distribution $P_{hH} = 0.8, P_{hL} = 0.2$ and the biased distribution $P_{hH} = 0.8, P_{hL} = 0.6$. We also test a wider range of signal distributions with fewer samples and select the distributions that capture different scenarios anticipated by our theoretical findings on the performance of one-round versus two-round voting. Under the unbiased distribution, both mechanisms admit natural equilibria that achieve the informed majority decision, so we expect strong performance from both. In contrast, the biased distribution, as shown by Proposition 1, admits no pure-symmetric equilibria for one-round voting, whereas “informative + sincere” remains viable in the two-round mechanismimplying better performance in that setting.

We run 50 votes for each set of parameters. In each vote, we first sampled the world state and signals for every agent from the designated distribution. Using these same samples, we then ran both a one-round and a two-round voting procedure. Specifically, we create n generative AI agents for the one-round voting and another n agents for the two-round voting. In one-round voting, each agent was asked only once for its vote. In two-round voting, each agent voted twice: we first collected every agents first-round vote simultaneously, then shared the first-round outcome with all agents (including their own first-round response as a history message) when asking for the second-round vote.

The prompt for AI agents contains the following components.

Voting scenario. We chose a company hiring setting: each AI agent plays the role of a human resource specialist and votes on whether to hire a candidate.

World state and signals. The world state is whether the candidate is qualified, which is not revealed to the agents. Each AI agent is told to have a Good or Bad impression of the candidate, representing its private signal.

Prior and signal distribution. AI agents are informed of the prior and signal distribution in two different ways. In the *accurate distribution* setting, every agent is told the exact distribution (P_H, P_L) and $(P_{hH}, P_{hL}, P_{\ell H}, P_{\ell L})$. In the *vague distribution* setting, the information is transferred into a vague description via a Deepseek R1 chatbot. For example, the distribution $P_H = 0.5$, $P_{hH} = 0.8$, and $P_{hL} = 0.6$ is converted the following description.

Example 13 (Vague information in the prompt.)**.** *Before meeting with the candidate, you thought she had a balanced chance to be qualified. If the candidate is qualified, you will likely* *have a good impression after meeting with her with a high probability well above half. If the candidate is not qualified, you will be less likely, though still moderately probable, to have a good impression at a likelihood distinctly lower than the qualified scenario but remaining above the halfway mark.* □

Types and preferences. Every agent is told their preferences based on their type.

Voting Mechanisms. Every agent is informed that it will participate in either a one-round or a two-round voting procedure. For the two-round voting, we clarify that the first round does not determine the winner but is publicly observable.

Vote. In both the one-round voting and each round of the two-round voting, every agent is asked to determine and report its best vote (YES or NO), along with a brief explanation of its reasoning.

4.5.2 Experimental Results

Model	Distribution	Unbiased (0.8, 0.2)		Biased (0.8, 0.6)	
	Information	Accurate	Vague	Accurate	Vague
Deepseek R1	One-round	1	1	**0.56**	**0.52**
	Two-Round	1	1	**0.88**	**0.78**
GPT-4o	One-round	1	1	0.5	0.52
	Two-Round	1	1	0.48	0.5

Table 4.1: The likelihood of the informed majority decision is reached.

In Table 4.1, we compare how often one-round versus two-round voting achieves the informed majority decision under different signal distributions, information conditions, and generative AI models. When signals are unbiased, both voting mechanisms always reach the informed majority decision. However, their performances diverge significantly under biased signals. In particular, Deepseek R1 agents show a marked accuracy gap: one-round voting frequently misses the informed majority decision when the world state is unqualified (even though agents still receive a "Good" signal with probability 0.6), whereas two-round voting retains good performance under both world states. On the other hand, for the non-reasoning model GPT-4o, there is no significant difference between the performance of one-round voting and two-round voting.

In addition to the analysis of the informed majority decision, we also examine the behaviors of the AI agents from their reasoning and analysis and have some interesting observations.

Bayesian Updates with accurate and vague information. Almost every agent performs a Bayesian update to infer the likelihood of each world state based on the signal distribution and the private information it obtains. Even when vague information is given, many agents tend to guess a concrete value for each description and perform a Bayesian update based on the guessed values.

universal switch in the second-round votes where a voter casts a different vote to its first-round vote. Examining the analysis of these voters, they usually find considerable votes in the first round that contradict their beliefs, which drives them to switch. On the other hand, there are also agents who believe more in their beliefs and stick to the same vote. An example of the prompts, answers, and reasoning of agents is in Appendix B.2.

4.5.3 Takeaway Messages

Our experiments illustrate a scenario where two-round voting is more likely to lead to an informed majority decision than one-round voting played by generative AI agents. These findings can guide future applications of generative AI in voting systems, such as AI-augmented or AI-proxy voting. As AI reasoning capabilities continue to improve, two-round voting will likely become even more effective in reaching informed outcomes. Moreover, our experiments serve as a practical reference for designing real-world experiments that compare different voting rules and explore strategic behavior. Because generative AI simulations allow for direct preference injection, they offer a valuable tool for preliminary experimentation and design.

CHAPTER 5
AGGREGATING INFORMATION AND PREFERENCES UNDER DIFFERENT COORDINATION ABILITY

In the previous chapters, we have shown that there always exists a "good" equilibrium that reaches the informed majority decision in the one-round majority vote and the two-round polling-and-voting structure, when agents share the same ordinal direction of objective preferences. In this chapter, we investigate a different preference structure, when two groups of agents have opposite directions of objective preferences, as shown in the following example.

Example 14. *Consider a legislative vote on a mixed economic policy with uncertain consequences. The policy may produce either inflation or deflation. Representatives come from different social backgrounds and have different preferences. Doves prefer inflation to stimulate economic growth, while Hawks prefer deflation to avoid market bubbles. Each representative possesses private information about the policy's likely effects from different sources. They may also coordinate with others sharing similar objectives to increase their collective influence. How can majority voting produce efficient, fair, and informed majority decisions under these conditions?*

The scenario in Example 14 occurs when two groups of stakeholders with conflicting interests make decisions with an uncertain outcome. Under the combined utility model, this implies that the objective preferences of the two groups have opposite directions. This preference structure is different from the model in Chapter 3 and 4 captures, where agents share the same direction of the objective preferences. Under this preference structure, Deng et al. [10] show that a strong equilibrium may not exist when the population of two groups is close to each other.

In this scenario, we argue that it is important to take the *coordination ability*–how large a group of agents can coordinate on their behaviors–into consideration. The coordination ability of agents varies depending on the society of the vote. In practice, coordination occurs within bounded groups, such as in local communities or trusted networks, rather than across entire populations. Theoretical and empirical studies indicate that the magnitude of group coordination is restricted by cognitive burden [82], communication complexity [83], and efficiency loss [84], etc. However, a strong Bayes-Nash equilibrium, which implicitly assumes that agents can form an unlimited-size coordination, is too pessimistic to admit a stable outcome. Standard Bayesian Nash equilibrium, on the other hand, ignores the potential coordination of strategic agents to break an equilibrium, especially in a voting scenario where a single vote can rarely flip the outcome. This choice between "no coordination" and "unbounded coordination" is too coarse to characterize the strategic behavior of agents in societies of different abilities to coordinate. Consequently, the following research question remains largely open.

When and how is the informed majority decision reached under societies with different abilities to coordinate?

To address such a question, a more fine-grained solution concept parameterized by the group coordination ability is required. In this paper, we employ the approximated *ex-ante Bayesian k-strong equilibrium* [62] (abbreviated as k-strong equilibrium) parameterized by the coordination ability parameter k. In a k-strong equilibrium, no coalition of at most k agents can collectively benefit by jointly changing their votes. Such a solution concept encompasses both Bayesian Nash equilibrium ($k = 1$) and strong Bayes-Nash equilibrium ($k = n$, n is the number of agents) as special cases and allows conducting a spectrum of equilibrium analyses on different k, fitting into societies with different coordination ability.

5.1 Overview of the Chapter

We address this question by conducting equilibrium analysis on the ex-ante Bayesian k-strong equilibrium, which enables us to make more refined predictions on the stable outcomes of strategic behaviors. Our research question can be interpreted into the following two questions.

1. When do all k-strong equilibria reach the informed majority decision?

2. When does there exist a k-strong equilibrium reaching the informed majority decision?

The two questions ask whether the worst-case/best-case equilibrium can still reach a good outcome. At a high level, they correspond to the price of anarchy [85] and the price of stability [86], respectively. Under a representative scenario of Example 14 where two groups of agents possess different preferences, we give the closed-form threshold on when each question leads to an affirmative answer, as in Figure 5.1. Consequently, when we have knowledge on the coordination ability of a voting society, our analysis allows us to make a more refined prediction of the outcome of strategic votes in the worst and the best case.

For the first question, the answer is affirmative if and only if the coordination ability $k \geq \frac{n}{2}$, where n is the number of agents (Theorem 11). This threshold is inevitable, as unanimous votes on the same alternative cannot be flipped by any coalition smaller than half of the population and becomes an $\frac{n}{2}$-strong equilibrium. However, unanimous voting ignores all private information and fails to reach informed majority decisions.

The second question, on the other hand, yields a far more complex answer. Whether such a k-equilibrium exists depends on three factors: the coordination ability k, the fraction α of the agents whose preferences are the majority, and the distribution of agents' private information (Theorem 12). As illustrated in Figure 5.1, the boundary between existence and non-existence displays non-linear, non-continuous, and non-strictly monotone segmental behavior across different fractions α, revealing the intricate strategic landscape shaped by different levels of coordination ability. In the order as the majority fraction α decreases, it starts with a range of α where a strong equilibrium exists and reaches the informed majority decision. The threshold drops from this maximal value after a critical point and displays at most four distinct segments. Segment 1 exhibits a constant threshold despite decreasing α.

Segment 2 is a linear decline in the threshold as α decreases. Segment 3 reveals the most complex behavior, where the threshold displays a non-linear relationship with α. When present, Segment 4 returns to linear decline, but the threshold maintains a non-zero value even when α approaches 0.5. Depending on the information structure, certain segments may be absent, and this boundary presents as two or three segments.

Combining the two boundaries yields a key insight: higher coordination ability produces a more selective set of stable outcomes. When coordination ability is high, equilibria either always reach informed majority decisions (if preferences are sufficiently aligned) or fail to exist entirely (if preferences are divergent). Low coordination ability, conversely, tolerates both informed and uninformed equilibria, leading to mixed predictions about outcomes.

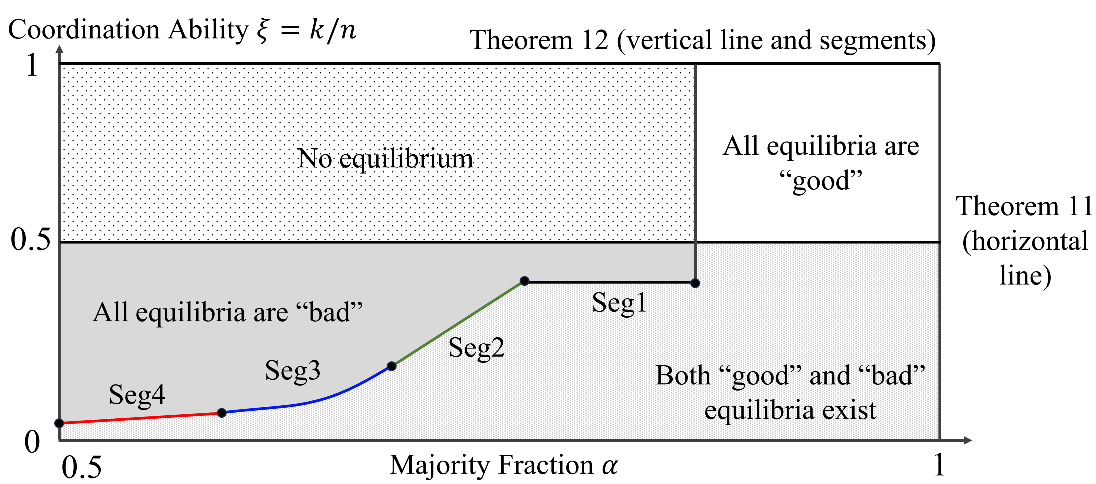


Figure 5.1: Illustration of boundaries characterized in Theorem 11 and Theorem 12 on the coordination ability and the majority fraction. "Good" and "bad" refer to equilibria that reach the informed majority decision and those that do not reach the informed majority decision, respectively.

Moreover, we construct explicitly the k-equilibrium reaching the informed majority decision with the largest possible k for each segment (Section 5.5.4). Like the segmental boundary curve, equilibria under different α exhibit different structures, revealing the non-trivial behavior of strategic agents with different coordination ability. For example, when the majority fraction is large (in Segment 1 and 2), all agents whose preferences are the minority converge on the same alternative in an equilibrium. However, as this fraction approaches $\frac{1}{2}$

(in Segment 4), minority agents split evenly between two alternatives in the equilibrium.

We also analyze the relationship between the shape of the boundaries and the distribution of the private information of the agents (Section 5.3.4). When agents possess more accurate private information, the threshold ξ^* value increases. This observation aligns with the intuition that agents with more precise information are more certain about their preferences, thereby increasing the likelihood of reaching a stable outcome.

Technical Overview. At a high level, the existence of an equilibrium hinges on the fraction of votes each alternative gets in expectation (Equation 5.3 and 5.4). For an equilibrium to exist, the informed majority decision must secure more than half of the expected votes under both ground truth states, even when a fraction $\frac{k}{n}$ of agents deviates from their prescribed strategies. Consequently, the majority agents employ strategies that maximize their margin above the $\frac{1}{2}$ threshold in expected vote share. Minority agents, conversely, work to diminish this margin and push the vote share below the majority threshold. The critical deviation threshold is determined by three competing factors: a majority's ability to build its margin, its requirement to maintain the margin under both ground truth states, and a minority's ability to deviate. This interplay makes finding the threshold a complex analytical challenge.

5.2 Setting

In this chapter, we fix the majority winning threshold $\mu = \frac{1}{2}$.

Let $\Delta = P_{hH} - P_{hL}$ be the difference in the signal distribution under two world states. From the fact that $P_{hH} + P_{\ell H} = P_{\ell L} + P_{hL} = 1$, we have the following observation.

Proposition 2. $\Delta = P_{hH} - P_{hL} = P_{\ell L} - P_{\ell H} = P_{hH} \cdot P_{\ell L} - P_{hL} \cdot P_{\ell H} = P_{hH} + P_{\ell L} - 1.$

Preference Structure. We consider two types of agents. The majority type prefer $\mathbf{A}$ in world state H ($v(H, \mathbf{A}) > v(H, \mathbf{R})$) and $\mathbf{R}$ in world state L ($v(L, \mathbf{A}) < v(L, \mathbf{R})$), and the minority type prefer **R** in world state H ($v(H, \mathbf{A}) < v(H, \mathbf{R})$) and **A** in world state L ($v(L, \mathbf{A}) > v(L, \mathbf{R})$). Let $\alpha \in (0.5, 1]$ be the (approximated) fraction of the majority agents, where the number of majority agents $n_{maj} = \lfloor \alpha \cdot n \rfloor$. In this setting, the informed majority decision follows the majority type agents' preference: the informed majority decision is **A** in world state H and **R** in world state L.

Example 15 (Information Structure and Agent Preferences)**.** *Consider the legislator voting scenario in Example 14. Suppose* $n = 50$ *agents vote on the economic policy. The world state is the outcome of the project, which can be a high inflation rate (H) or a low inflation rate (L). The prior that a high rate occurs is* $P_H = 0.6$ *and* $P_L = 0.4$*. The distribution is of private signals* $P_{hH} = 0.7$*,* $P_{\ell H} = 0.3$*,* $P_{hL} = 0.2$*, and* $P_{\ell L} = 0.8$*. This means that an agent receives an* h *signal with probability* 0.7 *when the world state is* H *and probability* 0.2 *when the state is* L*. The utilities of the agents are illustrated in Table 5.1. In this example, we assume that agents of the same type share an identical utility function, which may not necessarily be true in general.*

Type	$v_i(H, \mathbf{A})$	$v_i(L, \mathbf{A})$	$v_i(H, \mathbf{R})$	$v_i(L, \mathbf{R})$
Majority (Doves)	*4*	*0*	*1*	*2*
Minority (Hawks)	*2*	*3*	*3*	*1*

Table 5.1: Utility of agents.

Non-dominated strategies Our theoretical results concern equilibria in which agents employ only non-dominated strategies. A strategy is dominated if playing a different strategy always brings the agent a higher expected utility.

Definition 9. A strategy σ_i' is an (ex-ante) weakly dominated strategy for agent i if there exists a strategy σ_i such that, regardless of what other agents play Σ_{-i}, $u_i((\sigma_i', \Sigma_{-i})) \leq u_i((\sigma_i, \Sigma_{-i}))$, and there exists a Σ_{-i} such that $u_i((\sigma_i', \Sigma_{-i})) < u_i((\sigma_i, \Sigma_{-i}))$.

The constraint of no agents playing weakly dominated strategies follows the convention in the literature [87, 8, 5], and we believe this restriction is mild and reasonable. Firstly, an agent playing a dominated strategy has incentives to switch to a non-dominated strategy unilaterally, which brings no utility loss in any circumstances and utility gain in some circumstances. This contradicts the intention of an equilibrium that predicts a stable outcome where strategic agents aim to maximize their utilities. Secondly, as shown in the following lemma, whether a strategy is (weakly) dominated can be easily determined, and playing a non-dominated strategy does not bring extra computational cost and difficulties to the agents.

Lemma 7. *For an arbitrary agent i, a strategy $\sigma = (\beta_\ell, \beta_h)$ is NOT a weakly dominated strategy if and only if the following conditions hold.*

- *If i is a majority agent: $\beta_\ell = 0$ or $\beta_h = 1$.*
- *If i is a minority agent: $\beta_\ell = 1$ or $\beta_h = 0$.*

The proof of Lemma 7 is in Appendix C.1.

A strategy profile $\Sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ is an *ε-ex-ante Bayesian k-strong equilibrium* ((ε, k)-EBSE) if there does not exist a subset of agents D (the *deviating group*) with $|D| \leq k$ and a strategy profile $\Sigma' = (\sigma'_1, \sigma'_2, \cdots, \sigma'_n)$ (the *deviating strategy profile*) such that

1. $\sigma_i = \sigma'_i$ for all $i \notin D$;
2. $u_i(\Sigma') \geq u_i(\Sigma)$ for all $i \in D$; and
3. there exists $i \in D$ such that $u_i(\Sigma') > u_i(\Sigma) + \varepsilon$.

By definition, when $\varepsilon = 0$, the equilibrium is an ex-ante Bayesian k-strong equilibrium where no group of at most k agents can strictly increase their utilities through deviation; and when $k = n$, the equilibrium is an ε-ex-ante Bayesian strong equilibrium (named ε-strong Bayes-Nash equilibrium in [9, 39, 10]). We name k the coordination ability of agents, and its fraction version is $\xi = \frac{k}{n}$.

5.3 Main Results: Equilibrium Analysis

The main results of this paper are summarized in Figure 5.1, which partitions sequences of environments by (1) whether all sequences of equilibria reach the informed majority decision and (2) whether there exists a sequence of equilibria that reaches the informed majority decision (a "good" equilibrium). The partition depends on the signal distribution $\mathbf{P}$, majority fraction α, and the coordination ability fraction ξ. Section 5.3.1 characterizes the boundary determining when a "bad" equilibrium exists (so that not all equilibria are "good"), and Section 5.3.2 characterizes the boundary determining when at least one "good" equilibrium exists. Combining the two boundaries yields our full characterization in Figure 5.1, implying that higher coordination ability produces more selective stable outcomes, while low coordination ability leads to more tolerated, mixed outcomes. In Section 5.3.4, we discuss how signal distribution $\mathbf{P}$ affects the boundaries.

We formalize the definition of a "good" or "bad" equilibrium, named as *informed* and *uninformed* equilibrium, respectively.

Definition 10. A sequence of profiles $\{\Sigma_n\}$ is an *informed ξ-strong equilibrium* if it satisfies the following conditions.

1. for all n, there are no weakly dominated strategies in Σ_n,

2. the fidelity $A(\Sigma_n)$ converges to 1, and

3. there exists a sequence $\{\varepsilon_n\}$ such that ε_n converges to 0 and that for all n, Σ_n is an ε_n-ex-ante Bayesian ξn-strong equilibrium.

A sequence of profiles $\{\Sigma_n\}$ is an *uninformed ξ-strong equilibrium* if it satisfies conditions (1) and (3), but the fidelity $A(\Sigma_n)$ does not converge to 1. Specifically, $\xi = 1$ implies the sequence is asymptotically an ε_n-ex-ante Bayesian strong equilibrium.

5.3.1 Worst Case Equilibrium

We first characterize the boundary deciding when all k-strong equilibria reach the informed majority decision. The answer to this question is affirmative if and only if $\xi > \frac{1}{2}$ ($k > \frac{n}{2}$). Below this threshold, there always exists at least one uninformed ξ-strong equilibrium.

Theorem 11. *For any $\xi \in [0,1]$ and any sequence of instances $\{\mathcal{I}_n\}$, an uninformed ξ-strong equilibrium exists if and only if $\xi \leq \frac{1}{2}$.*

Theorem 11 contributes to one of the two partitions in Figure 5.1 (the horizontal $\frac{k}{n} = 0.5$). When $\xi \leq \frac{1}{2}$, the existence of an uninformed equilibrium implies that "all equilibria are good" does not hold. When $\xi > \frac{1}{2}$, once there exists a k-strong equilibrium, it reaches the informed majority decision. At a high level, this observation connects to the Price of Anarchy [85], which measures the ratio between optimal social welfare and the worst equilibrium outcome. Using fidelity as a proxy for social welfare, Theorem 11 implies that the Price of Anarchy approaches one only when society possesses high coordination ability.

5.3.2 Best Case Equilibrium

We then characterize the boundary deciding when there exists a k-strong equilibrium reaching the informed majority decision. The boundary depends on the signal distribution $\mathbf{P}$, the majority fraction α, and the coordination ability fraction ξ, and can be represented by two threshold functions θ and ξ^*. The first threshold θ takes the signal distribution $\mathbf{P}$ as input and outputs a value in $(\frac{1}{2}, 1)$. When the majority fraction α exceeds $\theta(\mathbf{P})$, an informed 1-strong equilibrium (a sequence of "good" n-strong equilibria) exists. The second threshold ξ^* takes both $\mathbf{P}$ and α as inputs and outputs a value in $(0, \frac{1}{2})$. When $\alpha \leq \theta(\mathbf{P})$, an informed ξ-equilibrium exists if and only if $\xi < \xi^*(\mathbf{P}, \alpha)$.

Theorem 12. *There exist threshold functions θ and ξ^* such that for any sequence of instances $\{\mathcal{I}_n\}$ with signal distribution $\mathbf{P}$ and majority fraction α,*

1. *if $\alpha > \theta(\mathbf{P})$, an informed 1-strong equilibrium exists.*

2. if $\frac{1}{2} < \alpha \leq \theta(\mathbf{P})$, an informed ξ-strong equilibrium exists if and only if $\xi < \xi^*(\mathbf{P}, \alpha)$.

Note that the first part of Theorem 12 is proved by [10].

Theorem 12 contributes to the other partition in Figure 5.1, yet it is much more complicated. In the order of α decreasing, this partition starts with a critical point θ where the largest ξ that tolerates an informed equilibrium declines drastically. The boundary then follows with at most four segments, depending on the signal distribution. These segments, ordered from highest to lowest values of α as Segment 1 to 4, are given descriptive names based on their distinct features. The "Flat" Segment 1 begins at θ and maintains a constant value ξ^* as α decreases. In the "Steep" Segment 2, ξ^* decreases linearly with α. Segment 3, the "Non-Linear" segment, is characterized by a non-linear relationship between ξ^* and α. When the signal distribution is inaccurate or symmetric (satisfying $(|P_{\ell L} - P_{hH}| < \min(P_{hL}, P_{\ell H}))$, Segment 3 ends before $\alpha = \frac{1}{2}$, there will be a "Tail" Segment 4; otherwise, the boundary ends with three segments.

A special case when the signal is symmetric, i.e, $P_{hH} = P_{\ell L}$. Under this specific setting, both the "Flat" Segment 1 and the "Non-Linear" Segment 3 reduce to points, and ξ^* consists of only "Steep" Segment 2 and "Tail" Segment 4 as shown in Figure 5.2a.

Example 16. *Here is an example adopting the scenario in Example 15 to illustrate the thresholds θ and ξ^* in Theorem 12. Note that $P_{hH} = 0.7$ and $P_{\ell L} = 0.8$ imply that this instance has all four segments. In this case, the first threshold $\theta = \frac{11}{16}$. For the threshold in the non-linear segment, we give a range of $\Theta(\frac{1}{\alpha})$ rather than the explicit form.*

Segment	Range of α	Threshold ξ^*
$\alpha > \theta$	$(\frac{11}{16}, 1)$	Unbounded
Flat 1	$[\frac{5}{8}, \frac{11}{16}]$	$\frac{5}{16}$
Steep 2	$[0.556, \frac{5}{8})$	$\frac{5}{2}(\alpha - \frac{1}{2})$
Non-linear 3	$[\frac{10}{19}, 0.556)$	$\Theta(\frac{1}{\alpha})$
Tail 4	$(\frac{1}{2}, \frac{10}{19})$	$\frac{1}{4}\alpha$

Table 5.2: Range and Threshold for each segment.

Likewise, Theorem 2 connects to the Price of Stability [86], which measures the ratio between optimal social welfare and the best-case equilibrium outcome. Using fidelity as a proxy for social welfare, Theorem 2 implies that the Price of Stability approaches one if and only if society's coordination ability is below the threshold.

5.3.3 Explicit Form of the Thresholds.

In this section, we give the explicit form of the thresholds θ and ξ^*.

$$\theta = \begin{cases} \frac{1}{2} + \frac{P_{\ell H}}{2P_{\ell L}} & P_{\ell L} \geq P_{hH}, \\ \frac{1}{2} + \frac{P_{hL}}{2P_{hH}} & P_{\ell L} < P_{hH}. \end{cases}$$

For ξ^*, assume without loss of generality that $P_{hH} \leq P_{\ell L}$. In this case, $\theta = \frac{1}{2} + \frac{P_{\ell H}}{2P_{\ell L}}$. Given a fixed information structure $\mathbf{P}$, we explicitly give the formula of ξ^* as a function of α. We will also abuse this notation of $\xi^*(\alpha)$ in the rest of the appendix.

When $\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} > \frac{1}{2}$,

$$\xi^*(\alpha) = \begin{cases} \frac{\Delta}{2P_{\ell L}} & \theta \geq \alpha \geq \frac{1}{2P_{\ell L}} \\ \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}} & \frac{1}{2P_{\ell L}} > \alpha \geq \alpha_{NL} \\ \xi_{NL} & \alpha_{NL} > \alpha > \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} \\ \frac{1}{2} \cdot \Delta \cdot \alpha & \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} > \alpha > \frac{1}{2} \end{cases}$$

When $\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} \leq \frac{1}{2}$,

$$\xi^*(\alpha) = \begin{cases} \frac{\Delta}{2P_{\ell L}} & \theta \geq \alpha \geq \frac{1}{2P_{\ell L}} \\ \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}} & \frac{1}{2P_{\ell L}} > \alpha \geq \alpha_{NL} \\ \xi_{NL} & \alpha_{NL} > \alpha > \frac{1}{2} \end{cases}$$

Here, ξ_{NL} is a non-linear function on α which makes up the "Non-Linear" Segment 3 in the curve. α_{NL} is the changing point between Segment 2 and 3 and comes from solving the inequality $\xi_{NL} \geq \frac{\Delta(\alpha - 1/2)}{P_{hL}}$.

$$\alpha_{NL} = \frac{2P_{hH}P_{\ell L}^2 + P_{hH}\cdot(P_{hH}+3P_{\ell L}) - (3P_{hH}+P_{\ell L}) + 2 - 2P_{hL}\cdot\sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2}. \tag{5.1}$$

$$\xi_{NL} = \frac{\Delta}{4\sqrt{2(1-\alpha P_{\ell H})(1-2\alpha P_{\ell L})P_{\ell H}P_{\ell L}} + 2P_{\ell H} + 4P_{\ell L} - 8\alpha P_{\ell H}P_{\ell L}}. \tag{5.2}$$

It is not hard to see that $\theta > \frac{1}{2}$. For ξ^*, Note that $\Delta = P_{\ell L} - P_{\ell H} < P_{\ell L}$, so the $\xi^* < \frac{1}{2}$ in Segment 1. Given the decreasing nature of ξ^*, it's smaller than $\frac{1}{2}$ for all $\alpha \in (\frac{1}{2}, \theta]$.

5.3.4 Shape Analysis on the Boundary

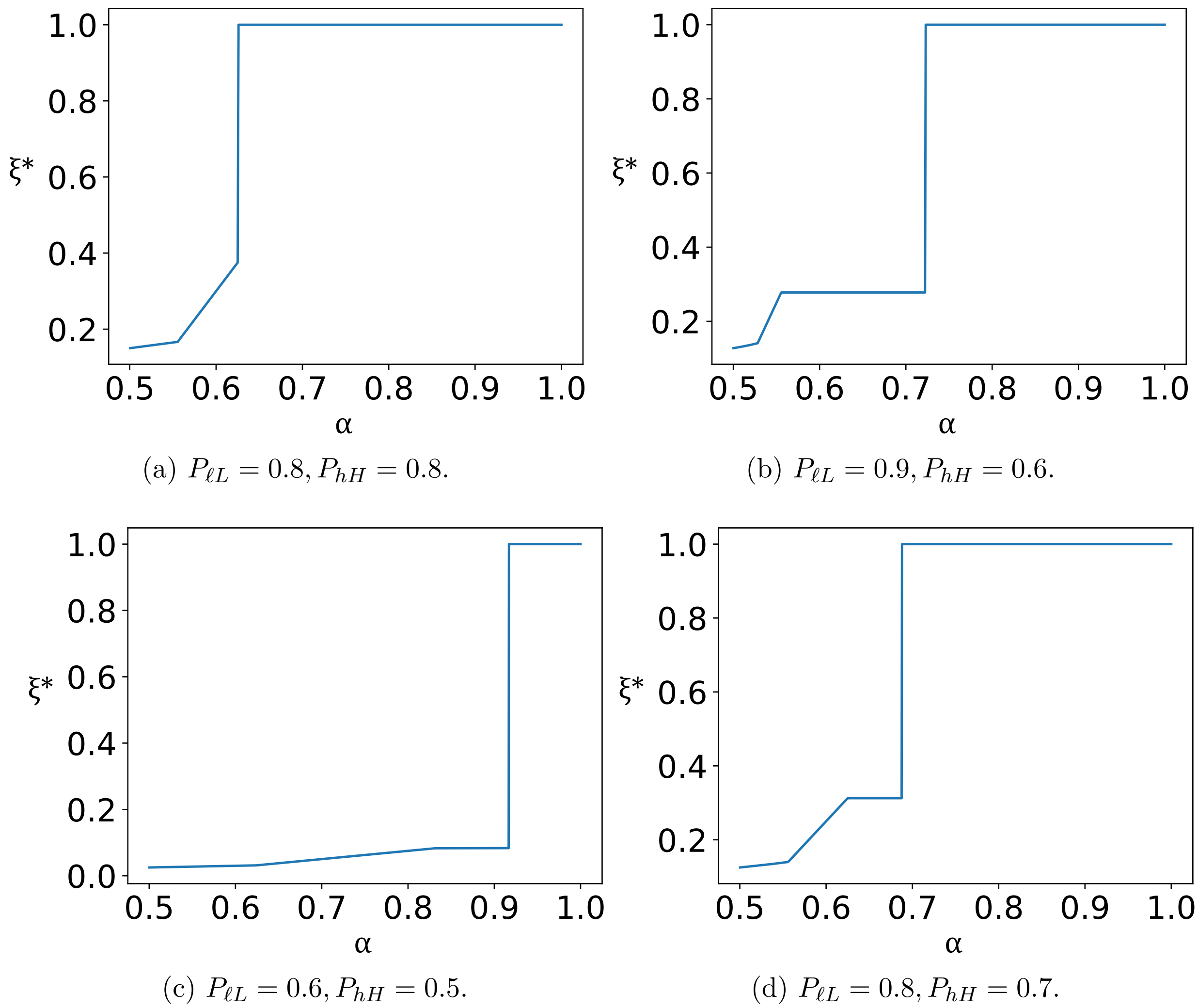
In this section, we discuss how the signal distribution $\mathbf{P}$ affects the shape of the boundary characterized in Theorem 12.

Accuracy. As the signals become more accurate (i.e., P_{hH} and $P_{\ell L}$ increase), two effects emerge in the equilibrium threshold ξ^*. First, the interval $(\frac{1}{2}, \theta]$, where ξ^* takes effect, contracts because θ is negatively correlated with signal accuracy. Second, the equilibrium threshold ξ^* increases within each segment of its piecewise function. This observation aligns with intuition: as agents receive more precise signals about the true state, the majority group can more effectively leverage their information advantage to secure their preferred outcome.

Asymmetry. As the signal asymmetry $P_{\ell L} - P_{hH}$ increases, there are corresponding changes in the intervals of Segments 1 and 4. The length of Flat Segment 1 exhibits a positive linear relationship with the signal asymmetry. Conversely, "Tail" Segment 4 contracts as the signal asymmetry increases, since its left boundary is a decreasing function of $(P_{\ell L} - P_{hH})$ while its right boundary remains fixed at $\frac{1}{2}$.

Non-vanishing. Regardless of the signal distribution, ξ^* does not converge to 0 when α goes to $\frac{1}{2}$. This indicates that, even when the majority only slightly outnumbers the minority, an unorganized minority still yields a stable outcome favoring the 51%.

Figure 5.2 illustrates the shapes of ξ^* under different distributions. For example, Figure 5.2a with a symmetric signal distribution contains only Segment 2 and 4, while Figure 5.2b with an asymmetric signal distribution has a much larger "Flat" Segment 1 and shorter Segment 4. In Figure 5.2c where the signal is inaccurate, the interval of ξ^* is long, yet the value is close to 0.



(a) $P_{\ell L} = 0.8, P_{hH} = 0.8$.

(b) $P_{\ell L} = 0.9, P_{hH} = 0.6$.

(c) $P_{\ell L} = 0.6, P_{hH} = 0.5$.

(d) $P_{\ell L} = 0.8, P_{hH} = 0.7$.

Figure 5.2: The shape of ξ^* under different signal distribution. We set $\xi^* = 1$ when $\alpha > \theta$, indicating the existence of a strong equilibrium.

5.4 Proof Sketch for Theorem 11.

For $\xi \leq \frac{1}{2}$, there is an easy proof for the existence of a k-strong equilibrium not reaching the informed majority decision: the strategy profile where all the agents unanimously vote for **A**.

Under this strategy profile, **A** is elected the winner even under world state L, yet no group of size at most $\frac{n}{2}$ agents can flip the outcome (note that **A** is favored by the tie-breaking).

For $\xi > \frac{1}{2}$, the proof follows the same high-level approach of leveraging expected vote share, but now the deviators are majority agents. With a deviating coalition exceeding half the population, majority agents can construct deviating strategies that reach the informed majority decision. The construction proceeds as follows: one fraction of deviators plays strategies that "complement" the non-deviators, setting the expected vote shares φ'_H and φ'_L both slightly below $\frac{1}{2}$. The remaining fraction votes informatively to maximize $\varphi'_H - \varphi_L$' ensuring φ'_H exceeds $\frac{1}{2}$ while φ'_L remains below it. This deviating strategy achieves fidelity converging to 1, providing majority agents with non-negligible benefits from deviation. The complete proof appears in Appendix C.3.

5.5 Proof Sketch for Theorem 12

The case where $\alpha > \theta$ is shown in [10]. We will solve the case where $\alpha \leq \theta$. Section 5.5.1 introduces our key analytical tool, the expected vote share framework, which connects expected utility, fidelity, and strategic behavior through a max-min perspective on equilibrium analysis. Section 5.5.2 and 5.5.3 apply this framework to prove Theorem 12, transforming the problem into a tractable optimization. Section 5.5.4 extends the proof sketch by exhibiting the equilibria achieving the threshold ξ^* across different segments, each displaying distinct strategic behaviors.

5.5.1 Key Framework: Expected Vote Share.

Given a strategy profile Σ, its expected vote share is the (ex-ante) expected fraction of votes for **A** conditioned on two world states, respectively. Specifically, when each agent i's strategy is $\sigma_i = (\beta_h^i, \beta_\ell^i)$, the expected vote share has the following form.

$$\varphi_H = \frac{1}{n} \cdot \sum_{i=1}^{n} (P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i). \tag{5.3}$$

$$\varphi_L = \frac{1}{n} \cdot \sum_{i=1}^{n} (P_{hL} \cdot \beta_h^i + P_{\ell L} \cdot \beta_\ell^i). \tag{5.4}$$

In Chapter 3, we reveal the asymptotical relationship between the expected vote share and the fidelity for a sequence of profiles: with a few exceptions, the fidelity converges to 1 if and only if $\varphi_H > \frac{1}{2}$ and $\varphi_L < \frac{1}{2}$ holds for all sufficiently large n. At a high level, this is because agents' votes can be viewed as independent Bernoulli variables, and their sum will concentrate around the expectation with high probability as n increases.

Now consider a sequence of strategy profiles $\{\Sigma_n\}$ that satisfies (1) no agents play weakly dominated strategies and (2) the fidelity converges to 1. We establish that only the minority agents need to be considered in potential deviation. This is because any sequence of deviating strategy profiles falls into one of two cases: either its fidelity still converges to 1, making utility changes negligible for all agents, or its fidelity fails to converge to 1, in which case majority agents experience strictly lower utility and have no incentive to deviate.

Consequently, we examine only deviations by minority agents, whose utilities are negatively correlated with fidelity. By the relationship between the fidelity and the expected vote share, a group of minority agents could successfully deviate if and only if they could create a new strategy profile (with expected vote share φ'_H and φ'_L) such that either

$$\varphi'_H \leq \frac{1}{2}, \tag{5.5}$$

leading to the expected majority being $\mathbf{R}$ in state H, or

$$\varphi'_L \geq \frac{1}{2}, \tag{5.6}$$

leading to the expected majority being $\mathbf{A}$ in state L.

At a high level, characterizing the boundary can be reinterpreted as finding the boundary of a maximin game. First, the Max player, aiming to maintain the expected vote share from the border $\frac{1}{2}$, chooses a sequence of profiles $\{\Sigma_n\}$. Then, the Min player, aiming to break the border, chooses a group of $\xi \cdot n$ deviators and their strategies. ξ^* is the smallest ξ such that the Min player wins, and no strategy profiles can tolerate a group of $\xi \cdot n$ deviators and maintain the expected vote share.

For the Min player, the optimal action is to embrace one of two extremes. In the first extreme, the goal is to maximize φ'_L, the expected fraction of **A** votes in world L after deviation. The deviators switch to always vote for **A**, and consist of the minority agents who are initially least likely to vote for **A**. In the second extreme, the goal is to minimize φ'_H, the expected fraction of **A** votes in world H after deviation. Here, the deviators switch to always vote for **R**, and consist of the minority agents who are initially least likely to vote for **R**. If either deviation succeeds, they succeed; and if both fail, then no deviation will succeed, implying the original strategy profile is an equilibrium.

For the Max player, the optimal action is non-trivial. In general, three correlated factors may affect how large a deviating group, ξ, a particular strategy profile can tolerate.

- The first factor is the difference between φ_H and φ_L, which is determined by the quantity $\beta^i_h - \beta^i_\ell$ of each agent i. When $\varphi_H - \varphi_L$ is large, it is at least possible for the threshold of $1/2$ to be between the two values and not too close to either of them.

- Secondly, the strategies of the minority agents. Agents' ability to change φ_H and φ_L depends on their initial strategy. For example, an agent who always votes for **A** cannot help to increase φ_L but can help to increase φ_H by $\frac{1}{n}$. Meanwhile, an agent who votes for **A** with signal ℓ and **R** with signal h can deviate to either increase φ_L by $\frac{P_{hL}}{n}$ or to decrease φ_H by $\frac{P_{\ell H}}{n}$.

- The last factor is balance. Because the deviators have two possible directions to deviate (decrease φ_H and increase φ_L), a deviation-proof strategy profile needs to keep both margins $\varphi_H - \frac{1}{2}$ and $\frac{1}{2} - \varphi_L$ sufficiently large to overwhelm the aforementioned deviating power of minority agents. Sometimes deviations can by neutralized by keeping margin balanced even at the expense of decreasing the difference $\varphi_H - \varphi_L$.

The interplay of these three factors largely contributes to the complexity of the problem.

The rest of the proof proceeds as follows. We will start from the first segment $\theta \geq \alpha \geq \frac{1}{2P_{\ell L}}$ whose proof is somewhat simpler and independent but shares the high-level idea with other parts. Then, we present a unified analysis of Segments 2, 3, and 4. The key step is converting the upper-bound and lower-bound problems into constraints on the expected vote share, which is further transformed into an optimization problem where the goal is to maximize the size of the deviation ξ that can be tolerated. The parameters that yield the optimum can be leveraged to construct a sequence of equilibria that proves the lower bound of the theorem. By case analysis, we can determine which constraints are binding in different settings and also construct the upper bounds.

5.5.2 Segment 1: $\theta \geq \alpha \geq \frac{1}{2P_{\ell L}}$.

We first introduce the strategy profile achieving the lower bound and explain why the threshold ξ^* is constant in this region. Then we present the upper bound proof, which serves as a simplified version of that for other segments.

Proposition 3. *For all $\alpha \geq \frac{\Delta}{1-\theta}$, for any $\xi < \frac{\Delta}{2P_{\ell L}}$, there exists a strategy profile sequence that satisfies all three constraints of the theorem.*

In this strategy profile sequence, all the minority agents always vote for **A**. Recall that the strategies of majority agents need to guarantee that both Equation 5.5 and 5.6 fail to hold. Note that these minority agents cannot increase φ_L because they always vote for **A**. On the other hand, they can decrease φ_H by ξ by switching to always vote for **R**. Therefore, the strategies of the majority agents must satisfy $\varphi_H > \frac{1}{2} + \xi$ and $\varphi_L < \frac{1}{2}$ to defend the possible deviations.

Without loss of generality, we consider a strategy profile where all the majority agents play the same strategy, denoted as $(\beta_h^{maj}, \beta_\ell^{maj})$. In this case, the expected vote share is

$$\varphi_H = \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1-\alpha).$$
$$\varphi_L = \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1-\alpha).$$

We want to find a ξ as large as possible to satisfy the following two constraints. Firstly, β_h^{maj} and β_ℓ^{maj} should be in $[0,1]$. Secondly, $\varphi_H > \frac{1}{2} + \xi$ and $\varphi_L < \frac{1}{2}$ hold.

The important observation that helps us solve the problem is that φ_H is more sensitive to β_h^{maj}, and φ_L is more sensitive to β_ℓ^{maj}. This is because $P_{hH} > P_{hL}$ and $P_{\ell H} < P_{\ell L}$. Therefore, if we could find β_h^* and β_ℓ^* in $(0,1)$ that satisfy $\{\varphi_H = \frac{1}{2} + \xi, \varphi_L = \frac{1}{2}\}$, we can get a valid pair of $(\beta_h^{maj}, \beta_\ell^{maj})$ by increasing β_h^* and decreasing β_ℓ^* at a certain ratio so that φ_H increases and φ_L decreases. The modified $(\beta_h^{maj}, \beta_\ell^{maj})$ will satisfy the two constraints.

Let β_h^* and β_ℓ^* be the solution to the equation set $\{\varphi_H = \frac{1}{2} + \xi, \varphi_L = \frac{1}{2}\}$ with the explicit form.

$$\beta_\ell^* = \frac{1}{\alpha}(\alpha - \frac{1}{2} - \frac{P_{hL}}{\Delta} \cdot \xi).$$
$$\beta_h^* = \frac{1}{\alpha}(\alpha - \frac{1}{2} + \frac{P_{\ell L}}{\Delta} \cdot \xi).$$

Examining the above equations, when $\xi < \frac{\Delta}{2P_{\ell L}}$ we can easily see that both β_ℓ^* and β_h^* are in $(0,1)$. In the second step, we simultaneously increase β_h^* at rate $P_{\ell H} + P_{\ell L}$ and decrease β_ℓ^* at rate $(P_{hH} + P_{hL})$ until the former reaches 1 or the latter reaches 0, at which point we stop. The strategy after the second step is $\beta_\ell^{maj} = \beta_\ell^* - \delta \cdot (P_{hH} + P_{hL})$ and $\beta_h^{maj} = \beta_h^* + \delta \cdot (P_{\ell H} + P_{\ell L})$, where $\delta > 0$ is chosen so that both are between 0 and 1 and one is at the boundary. The strategy of every majority agent is set to $(\beta_h^{maj}, \beta_\ell^{maj})$.

This implies that all the three constraints hold. Because either $\beta_h^{maj} = 1$ or $\beta_\ell^{maj} = 0$ holds, all the majority agents are playing their non-dominated strategy. The two-step construction guarantees that $\varphi_H > \frac{1}{2}+\xi$ and $\varphi_L < \frac{1}{2}$. Thus, the fidelity of the sequence converges to 1. Moreover, because, with high probability, a ξ fraction of minority agents cannot change the outcome, it is an ϵ-ex-ante Bayesian ξn-strong equilibrium, where ϵ converges to 0.

It seems a little counter-intuitive that ξ is constant even while α varies in segment 1. Here we try to give some high-level explanation for this behavior. At a high level, this is a trade-off between the difference $\varphi_H - \varphi_L$ (the first factor that affects ξ) and the balance (the third factor). When α is large, the margin of φ_H is smaller than that of φ_L, and the majority agents' votes have to lean to **A**, which sacrifices $\varphi_H - \varphi_L$ to keep the balance. As α decreases, there are more minority agents always voting for **A**, and the margins of φ_H and of φ_L become more balanced. While the fraction of the majority agents decreases, more balanced margins allow them to play a strategy that leads to a larger $\varphi_H - \varphi_L$. This trade-off keeps the ξ at a flat constant. Eventually, when $\alpha = \frac{1}{2P_{\ell L}}$, the margins are completely balanced, and majority agents play informatively which maximizes $\varphi_H - \varphi_L$. When α further decrease, the margin on φ_L becomes larger than φ_L, and majority agents cannot increase $\varphi_H - \varphi_L$ anymore. Therefore, the threshold ξ^* begins to decrease.

Now we sketch the proof of the upper bound. We take an arbitrary strategy profile sequence satisfying the first two constraints and show that if $\xi \geq \frac{\Delta}{2P_{\ell L}}$, constraint 3 is violated.

Proposition 4. *For all $\alpha \leq \theta$, for any profile sequence $\{\Sigma_n\}$ such that $A(\Sigma_n)$ converges to 1 and no agents play weakly dominated strategies, there is a constant $\varepsilon > 0$ and infinitely many n such that Σ_n is NOT an ε-ex-ante Bayesian $\xi \cdot n$-strong equilibrium, where $\xi = \frac{\Delta}{2P_{\ell L}}$.*

For each environment with n agents, we fix the deviating group D with $\xi = \frac{\Delta}{2P_{\ell L}}$ minority agents and show that no matter what the strategy profile is, the deviator can always succeed by decreasing φ_H or increasing φ_L to $\frac{1}{2}$. This is a stronger statement than the existence of a successful deviation, as it fixes the deviating group in advance.

We first consider the strategies of the non-deviators, including all the majority agents and some minority agents. Since they are not going to deviate, their contribution to φ_H and φ_L

can be represented by the average of all their strategies, denoted by $\bar{\beta}_h$ and $\bar{\beta}_\ell$. Likewise, if the deviation fails, the expected vote share under state H must be greater than $\frac{1}{2}$ even when all the deviators vote for **R**, and the expected vote share under state L must be smaller than $\frac{1}{2}$ even when all the deviators vote for **A**. Formally, this leads to the following constraints:

$$\varphi'_H := (1-\xi)\cdot(\bar{\beta}_h \cdot P_{hH} + \bar{\beta}_\ell \cdot P_{\ell H}) > \frac{1}{2}.$$
$$\varphi'_L := (1-\xi)\cdot(\bar{\beta}_h \cdot P_{hL} + \bar{\beta}_\ell \cdot P_{\ell L}) + \xi < \frac{1}{2}.$$

Therefore, to show the deviation will succeed, it suffices to show that for any $(\bar{\beta}_\ell, \bar{\beta}_h)$, at least one of the inequalities is violated. As before, we utilize the observation that φ'_H is more sensitive to $\bar{\beta}_h$ and φ'_L is more sensitive to $\bar{\beta}_\ell$. First, we solve the equation set $\{\varphi'_H = \frac{1}{2}, \varphi'_L = \frac{1}{2}\}$ and get the solution $\bar{\beta}^*_h$ and $\bar{\beta}^*_\ell$. However, when $\xi = \frac{\Delta}{2P_{\ell L}}$, the solution $\bar{\beta}^*_h \geq 1$. This difference changes the outcome in the opposite direction, as we will show that one of the inequalities is violated.

We consider an arbitrary pair of $\bar{\beta}_h$ and $\bar{\beta}_\ell$ in $[0,1]$. Note that $\bar{\beta}_h \leq \bar{\beta}^*_h$. Therefore, if $\bar{\beta}_\ell$ is not large enough, then $\varphi'_H > \frac{1}{2}$ is violated. On the other hand, if $\bar{\beta}_\ell$ is large enough so that $\varphi'_H > \frac{1}{2}$ holds, it turns out that it is too large. φ'_L is more sensitive to $\bar{\beta}_\ell$, so a large $\bar{\beta}_\ell$ affects φ'_L more than it affects $\bar{\beta}_h$. Therefore, when $\varphi'_H > \frac{1}{2}$, $\varphi'_L < \frac{1}{2}$ will be violated. This implies that one of the two deviations can succeed, and Σ fail to be an equilibrium.

5.5.3 Segment 2, 3, and 4, $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$.

The proof of the rest of the three segments follows the same high-level idea and shares some preliminary steps. According to Lemma 7, we can further categorize minority agents into to two types: Type-0 agents who always vote for **R** when they receive h signal, and type-1 agents who always vote for **A** when they receive ℓ signal. While a minority agent can both always vote for **R** when receiving h signal and always vote for **A** when receiving ℓ signal, in our analysis, we will always partition the agents so that each is assigned to exactly one type.

The fraction of type-0 agents will be denoted by γ, and the fraction of the type-1 agents then $1-\alpha-\gamma$.

This part of proof consists of five steps.

1. We solve the case when γ is biased (either γ or $1-\alpha-\gamma$ is close to 0) and restrict the case when γ is not biased to be "double-symmetric": all type-0 agents play the same strategy, and all type-1 agents play the same strategy.

2. We show that a sequence of strategy profiles satisfies all three constraints if and only if $\xi < \xi_h$ and $\xi < \xi_\ell$. Here ξ_h and ξ_ℓ are two thresholds derived by the constraints on the expected vote share ($\varphi'_H > \frac{1}{2}$ and $\varphi_L < \frac{1}{2}$, like that in the "Flat" Segment 1). Therefore, the problem of finding the upper and lower bounds is converted into maximizing $\min(\xi_h, \xi_\ell)$.

3. We reduce the search space of maximizing $\min(\xi_h, \xi_\ell)$ into two corner cases. In the first case, all type-0 agents always vote for **R**. In the second case, all type-1 agents always vote for **A**. The threshold ξ^* takes the maximum of these two cases and $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$, which derives from case with biased γ.

4. We tackle the two extreme cases respectively. In the case where all type-0 agents always vote for **R** (the first case), the extension of "Steep" Segment 2 and "Tail" Segment 4 serves as an upper bound. On the other hand, the case where all type-1 agents always vote for **A** (the second case) derives the "Non-Linear" Segment 3 and the "Tail" Segment 4.

5. We take the maximum of all the cases and reach the upper and the lower bounds.

5.5.3.1 Preliminaries

We first consider the case when either γ is close to 0 or $1-\alpha-\gamma$ is close to 0, which means that one type of minority agents dominates the minority group. In this case, the ξ fraction of deviators shall exhaust the type with fewer agents and fill up the rest fraction with agents from the other type. Following a similar reasoning to that in Proposition 4 (pre-fixing a group of deviators and showing that at least one of the inequalities is violated), we can get the upper bound for ξ of this case is $\frac{\Delta\dot{(}\alpha-1/2)}{P_{hL}}$.

Proposition 5. *For* $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$*, for any* $\xi' \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$*, NO sequence with infinitely many strategy profiles where* $\gamma \leq \xi'$ *or* $1-\alpha-\gamma \leq \xi'$ *satisfy all three constraints in Theorem 12 with* $\xi = \xi'$*.*

Moreover, $\frac{\Delta\dot{(}\alpha-1/2)}{P_{hL}}$ also serves as a lower bound for ξ.

Proposition 6. *When* $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$*, for any* $\xi < \frac{\Delta(\alpha-1/2)}{P_{hL}}$*, there exists a sequence of strategy profiles* $\{\Sigma\}$ *that satisfies the constraints in Theorem 12 with* $\xi = \frac{\Delta(\alpha-1/2)}{P_{hL}}$

The strategy profiles in Proposition 6 resemble that in Proposition 3. All the minority agents always vote for **A**, and the majority agents play the same strategy constructed by solving the equation for the expected vote share and modifying. The proof also resembles Proposition 3.

From now on, we focus on the strategy profile sequences where type-0 and type-1 agents are not biased, i.e., $\gamma > \xi$ and $1-\alpha-\gamma > \xi$ holds for all sufficiently large n. We assume this assumption holds unless otherwise stated.

Before we start analyzing the expected vote shares and deviations with type-0 and type-1 agents, we further narrow the strategy profiles to be considered to those where all type-0 agents play the same strategy, and all type-1 agents play the same strategy, with the following lemma.

Lemma 8. *If a strategy profile sequence* $\{\Sigma_n\}$ *satisfies all three constraints on* ξ*, then the following strategy profile* $\{\Sigma'_n\}$ *also satisfies all three constraints on* ξ*. For each* n*, (1) each majority agent in* Σ'_n *play the same strategy as in* Σ_n*; (2) all the type-1 minority agents in* Σ'_n *play the (arithmetic) average strategy of type-1 minority agents in* Σ_n*; and (3) all the type-0 minority agents in* Σ'_n *play the average strategy of type-0 minority agents in* Σ_n*.*

The high-level idea of Lemma 8 is that Σ'_n shares the same expected vote with Σ_n (so as the same margin on φ_H and φ_L), but the minority agents have weaker deviating power. For example, when the deviation aims to decrease φ_H, the deviators (switching to always vote for $\mathbf{R}$) in Σ_n can be the type-1 agents with largest β_h^i. This brings a larger decrease to φ_H compared to a ξ fraction of type-1 agents who play the average β_h among all type-1 agents. Therefore, if deviations cannot succeed in Σ_n, it cannot succeed in Σ'_n as well.

5.5.3.2 Transition to Maximization Problem.

In this part, we present the core steps in our proof. By solving the inequalities of the deviating expected vote share, we convert the problem of finding upper and lower bounds into maximizing the minimum of two upper bounds of ξ, denoted as ξ_h and ξ_ℓ, on the strategies and fractions of type-0 and type-1 agents.

We first introduce the notations. Consider a strategy profile Σ where no agents play weakly dominated strategies. Let β_h^{maj} and β_ℓ^{maj} be the average strategy of the majority agents. The strategy that type-0 agents play (recall that we can assume that they play the same strategy without loss of generality) is $(\beta_\ell^0, \beta_h^0 = 0)$. The strategy that type-1 agents play is $(\beta_\ell^1 = 1, \beta_h^1)$. In this way, the expected vote share of the strategy profile can be written as

$$\varphi_H = \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1 - \alpha - \gamma) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) + \gamma \cdot P_{\ell H} \cdot \beta_\ell^0.$$
$$\varphi_L = \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1 - \alpha - \gamma) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + \gamma \cdot P_{\ell L} \cdot \beta_\ell^0.$$

Σ is an ε-Bayesian ξn-strong equilibrium if and only if no deviation succeed. As said, we can consider only two extremity deviations that minimizes/maximizes the deviating expected vote share respectively. In the first deviation that minimizes φ_L, a ξ fraction of type-1 agents switch to always vote for $\mathbf{R}$. The expected vote share under H state after deviation φ'_H is required to be larger than $\frac{1}{2}$ to successfully defend the deviation. In the second deviation that maximizes φ_H, a ξ fraction of type-0 agents switch to always vote for $\mathbf{A}$. The expected vote share under L state after deviation φ'_L is required to be smaller than $\frac{1}{2}$. Therefore, we have the following constraints.

$$\varphi'_H := \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1-\alpha-\gamma-\xi) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) + \gamma \cdot P_{\ell H} \cdot \beta_\ell^0 > \frac{1}{2} \tag{5.7}$$

$$\varphi'_L := \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1-\alpha-\gamma) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + (\gamma-\xi) \cdot P_{\ell L} \cdot \beta_\ell^0 + \xi < \frac{1}{2} \tag{5.8}$$

We claim that, for any given β_h^1, β_ℓ^0, γ, and ξ, there is a pair of $(\beta_\ell^{maj}, \beta_h^{maj})$ satisfying this condition if and only if the solution to the equation set $\{\varphi'_H = \frac{1}{2}, \varphi'_L = \frac{1}{2}\}$ (denoted as $(\beta_\ell^*, \beta_h^*)$) is in $(0,1)$. The proof of the claim shares the same idea with the proof of Proposition 3 and 4 as we illustrated.

Specifically, the explicit form of the solution is

$$\beta_\ell^* = \frac{1}{\alpha}(\frac{1}{2} - (1-\alpha) + \gamma \cdot (1-\beta_\ell^0)) - \frac{\xi}{\Delta \cdot \alpha}(P_{hH} \cdot P_{hL} \cdot \beta_h^1 + P_{hH} \cdot P_{\ell L} \cdot (1-\beta_\ell^0) + P_{hL}).$$
$$\beta_h^* = \frac{1}{\alpha}(\frac{1}{2} - (1-\alpha-\gamma) \cdot \beta_h^1) + \frac{\xi}{\Delta \cdot \alpha}(P_{hH} \cdot P_{\ell L} \cdot \beta_h^1 + P_{\ell H} \cdot P_{\ell L} \cdot (1-\beta_\ell^0) + P_{\ell H}).$$

Solving the constraint $\beta_h^* < 1$ and $\beta_\ell^* > 0$ on ξ respectively, we get two thresholds on ξ, denoted as ξ_ℓ and ξ_h.

$$\xi < \xi_h := \frac{\Delta(\alpha - 1/2 + (1-\alpha-\gamma) \cdot \beta_h^1)}{P_{hH} \cdot P_{\ell L} \cdot \beta_h^1 + P_{\ell H} \cdot P_{\ell L} \cdot (1-\beta_\ell^0) + P_{\ell H}}. \tag{5.9}$$

$$\xi < \xi_\ell := \frac{\Delta(\alpha - 1/2 + \gamma \cdot (1-\beta_\ell^0))}{P_{hH} \cdot P_{hL} \cdot \beta_h^1 + P_{hH} \cdot P_{\ell L} \cdot (1-\beta_\ell^0) + P_{hL}}. \tag{5.10}$$

Specifically, $\xi_h(0,0) = \frac{\Delta(\alpha-1/2)}{P_{\ell H}}$ and $\xi_\ell(0,0) = \frac{\Delta(\alpha-1/2)}{P_{hL}}$ do not depends on γ.

The following two lemmas guarantee the relationship between bounding ξ and maximizing $\min(\xi_h, \xi_\ell)$. Lemma 9 guarantees a strategy profile sequence satisfying all constraints for every $\min(\xi_h, \xi_\ell)$ that can be reached by a set of β_h^1, β_ℓ^0, and γ. Lemma 10 guarantees that the upper bound of ξ^* is (partially) determined by the maximum of $\min(\xi_h, \xi_\ell)$.

Lemma 9. *Given a set of β_h^1, β_ℓ^0, and γ satisfying $\beta_h^1 \in [0,1]$, $\beta_\ell^0 \in [0,1]$, and let $\xi' = min(\xi_h, \xi_\ell)$ corresponding to β_h^1, β_ℓ^0, and γ. If $\gamma > \xi'$ and $1-\alpha-\gamma > \xi'$ holds, then for all $\xi < \xi'$ there exists a sequence of profiles such that the constraints in Theorem 12 hold for ξ.*

Lemma 10. *Let $\xi' = \max_{\beta_h^1, \beta_\ell^0, \gamma} \min(\xi_h, \xi_\ell)$. No sequence of strategy profiles can satisfy all three constraints in Theorem 12 with $\xi = \max(\xi', \frac{\Delta(\alpha-1/2)}{P_{hL}})$.*

The proof of Lemma 9 resembles that of Proposition 3 (except for the more complicated constraints), and the proof of Lemma 10 resembles that of Proposition 4. Both proofs share the idea that the constraints $\{\varphi'_H > \frac{1}{2}, \varphi'_L < \frac{1}{2}\}$ holds if and only if the equation set $\{\varphi'_h = \frac{1}{2}, \varphi'_L = \frac{1}{2}\}$ has a solution in $(0,1)^2$. The reason that Lemma 10 takes a maximum on ξ' and $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ is that ξ_h and ξ_ℓ only characterize the scenario where $\gamma > \xi'$ and $1-\alpha-\gamma > \xi'$. When $\gamma \leq \xi'$ or $1-\alpha-\gamma \leq \xi'$, Proposition 5 gives the upper bound $\frac{\Delta(\alpha-1/2)}{P_{hL}}$.

5.5.3.3 Maximizing $\min(\xi_h, \xi_\ell)$.

From now on, we start to maximize $\min(\xi_h, \xi_\ell)$ on β_h^1, $1-\beta_\ell^0$, and γ. We find that $\min(\xi_h, \xi_\ell)$ exceeds $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ only if one of the following conditions holds. (1) all the type-1 agents always vote for **A** ($\beta_h^1 = 1$); or (2) all the type-0 agents always vote for **R** ($1-\beta_\ell^0 = 1$).

Proposition 7. *At least one of the following holds. (1) $\max_{\beta_h^1, \beta_\ell^0, \gamma} \min(\xi_h, \xi_\ell) \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$; (2) $\min(\xi_h, \xi_\ell)$ is maximized when $\beta_h^1 = 1$, and (3) $\min(\xi_h, \xi_\ell)$ is maximized when $1-\beta_\ell^0 = 1$.*

Proposition 7 largely decreases the space we need to search for the maximum of $\min(\xi_h, \xi_\ell)$. We only need examine two corner cases $\beta_h^1 = 1$ and $1-\beta_\ell^0 = 1$ and compare the maximum of these cases with $\frac{\Delta(\alpha-1/2)}{P_{hL}}$.

The key idea to prove Proposition 7 is to analyze the derivatives of ξ_h and ξ_ℓ for a spectrum of lines that passes ($\beta_h^1 = 0, 1-\beta_\ell^0 = 0$). That is, for every $t \geq 0$, we consider the derivatives under a linear constraint of $1-\beta_\ell^0 = t \cdot \beta_h^1$. (Specifically, $t = +\infty$ for line $\beta_h^1 = 0$.) If ξ_h (or ξ_ℓ) decreases towards such a line characterized by t, then for every $1-\beta_\ell^0 = t \cdot \beta_h^1$ and every γ, ξ_h (ξ_ℓ, respectively) will not exceed that when $\beta_h = 1-\beta_\ell = 0$, which is no more than $\frac{\Delta(\alpha-1/2)}{P_{hL}}$. On the other hand, if ξ_h and ξ_ℓ both increases towards line t, they will keep increasing until it reaches $\beta_h^1 = 1$ or $1 - \beta_\ell^0 = 1$. Therefore, $\min(\xi_h, \xi_\ell)$ is maximized at $\beta_h^1 = 1$ or $1 - \beta_\ell^0 = 1$ (if it exceeds $\frac{\Delta(\alpha-1/2)}{P_{hL}}$).

It remains to show that ξ_h and ξ_ℓ are always increasing or always non-increasing on every line t. While directly solving the derivatives is difficult, we only care about whether they are positive or negative. Recall the fact that for a function $f = \frac{u}{v}$, the derivative of f is $f' = \frac{u'v-uv'}{v^2}$. Therefore, we follow $\mathrm{sign}(f') = \mathrm{sign}(u'v - uv')$ and have the following observation on the derivatives.

$$\mathrm{sign}(\frac{\partial \xi_h(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1}) = \mathrm{sign}((1-\alpha-\gamma)\cdot P_{\ell H} - (\alpha - \frac{1}{2})\cdot P_{hH}\cdot P_{\ell L} - (\alpha-\frac{1}{2})\cdot P_{\ell H}\cdot P_{\ell L}\cdot t)).$$

$$\mathrm{sign}(\frac{\partial \xi_\ell(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1}) = \mathrm{sign}(t\cdot(\gamma\cdot P_{hL} - (\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L}) - P_{hH}\cdot P_{\ell L}\cdot(\alpha-\frac{1}{2})).$$

An important observation is that both (signs of the) derivatives are linear on t and independent of β_h^1 and $1-\beta_\ell^0$. Therefore, both derivatives are either always positive or always non-positive. This implies that both ξ_h and ξ_ℓ are either always increasing or always non-increasing on each line t. This finishes the proof of Proposition 7.

In the following sections of the proof sketch, we maximize $\min(\xi_h, \xi_\ell)$ under $\beta_h^1 = 1$ and under $1 - \beta_\ell^0 = 1$, respectively. As we assumed that $P_{\ell L} \geq P_{hH}$, these two cases are not symmetric. In the $1-\beta_\ell^0 = 1$ case, $\min(\xi_h, \xi_\ell)$ is upper bounded by linear constraint related to "Steep" Segment 2 and "Tail" Segment 4, while in the $\beta_h^1 = 1$ case, the "Non-Linear" Segment 3 occurs.

5.5.3.4 Case $1 - \beta_\ell^0 = 1$.

When $1-\beta_\ell^0 = 1$, $\min(\xi_h, \xi_\ell)$ is upper bounded by the maximum of the extensions of "Steep" Segment 2 and of "Tail" Segment 4 together. When $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{1+P_{\ell L}}$, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$, which relates to Segment 2; and when $\frac{1}{1+P_{\ell L}} \geq \alpha > \frac{1}{2}$, $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$, which relates to

Segment 4. As $\frac{\Delta(\alpha-1/2)}{P_{hL}} \geq \frac{1}{2}\Delta\alpha$ if and only if $\alpha \geq \frac{1}{1+P_{\ell L}}$, this upper bound is the maximum of the line $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ and line $\frac{1}{2}\Delta\alpha$.

Throughout the case $1-\beta_\ell^0 = 1$, we assume that ξ_h is positively monotone on β_h^1, which holds if and only if an inequality between γ and α holds. The case where the inequality does not hold is covered by Proposition 7, and $\xi_h \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$.

Proposition 8. *When* $\frac{\Delta}{1-\theta} \geq \alpha \geq \frac{1}{1+P_{\ell L}}$*, if* $1-\beta_\ell^0 = 1$ *holds, either* ξ_h *or* ξ_ℓ *cannot exceed* $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$.

Proposition 9. *When* $\frac{1}{1+P_{\ell L}} \geq \alpha \geq \frac{1}{2}$*, if* $1-\beta_\ell^0 = 1$ *holds, either* ξ_h *or* ξ_ℓ *cannot exceed* $\frac{1}{2}\Delta\alpha$.

The proof of Proposition 8 and 9 resemble each other. The cutting point is a specific set of parameters: $\beta_h^1 = 1-\beta_\ell^0 = 1$, and $\gamma = \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. In this case, $\xi_h = \xi_\ell = \frac{1}{2}\Delta\alpha$. (Note that this also serves as a lower bound by Lemma 9, which will eventually become the lower-bound proof for "Tail" Segment 4.) If we decrease β_h^1 or increase γ, ξ_h will decrease and be smaller than $\frac{1}{2}\Delta\alpha$. If we decrease γ, $\xi_h \geq \frac{1}{2}\Delta\alpha$ only when β_h^1 is large, while $\xi_\ell \geq \frac{1}{2}\Delta\alpha$ only when β_h^1 is small. By solving an inequality, we shall figure out that no $\beta_h^1 \in [0,1]$ can satisfy both inequalities. Therefore, $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$. As $\frac{1}{2}\Delta\alpha \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$ when $\alpha \geq \frac{1}{1+P_{\ell L}}$, this also implies Proposition 8.

5.5.3.5 Case $\beta_h^1 = 1$.

Unlike the $1-\beta_\ell^0 = 1$ case, $\min(\xi_h, \xi_\ell)$ contains the "Non-Linear" Segment 3, and the method for Proposition 8 and 9 does not apply. Therefore, we take a more straightforward method by maximizing the $\min(\xi_h, \xi_\ell)$ one parameter at a time. Firstly, for every fixed $1-\beta_\ell^0$, we maximize $\min(\xi_h, \xi_\ell)$ on γ. We denote γ^* as the γ that maximizes $\min(\xi_h, \xi_\ell)$. Then we maximize $\min(\xi_h, \xi_\ell)$ where $\gamma = \gamma^*$ on $1-\beta_\ell^0$. The maximum naturally becomes an upper bound. Finally, the parameters leading to this maximum also serve as proof for a tight lower bound.

Given that ξ_h decreases and ξ_ℓ increases when γ increases, the γ that maximizes $\min(\xi_h, \xi_\ell)$ satisfies $\xi_h = \xi_\ell$. Therefore, solving this equation, we get γ^* and the corresponding ξ (denoted as $\hat{\xi}$) being an upper bound of $\min(\xi_h, \xi_\ell)$

Lemma 11. *When* $\frac{1}{2P_{\ell L}} \geq \alpha \geq \frac{1}{2}$*, for* $\beta_h^1 = 1$ *and any fixed* $1 - \beta_\ell^0$*, let*

$$\hat{\xi} = \frac{\Delta \cdot (\alpha - 1/2 + 1/2 \cdot (1 - \beta_\ell^0))}{(P_{\ell L} \cdot (1 - \beta_\ell^0) + P_{hL}) \cdot (P_{\ell H} \cdot (1 - \beta_\ell^0) + P_{hH} + 1)}.$$

Then, $\min(\xi_h, \xi_\ell) \leq \hat{\xi}$ *holds for all feasible* γ*.*

The next step is the maximize $\hat{\xi}$ on $1 - \beta_\ell^0$. Here we again apply the trick $\text{sign}(f') = \text{sign}(u'v - uv')$ and examine the sign of the derivative $\frac{\partial \hat{\xi}}{\partial (1-\beta_\ell^0)}$. It turns out that the $1 - \beta_\ell^0$ that maximizes $\hat{\xi}$ increases as α decreases, and that three cases are occurring in order as α decreases. (1) Firstly, when α is large, $\hat{\xi}$ is maximized at $1 - \beta_\ell^0 = 0$, in which $\hat{\xi} \leq \frac{\Delta(\alpha - 1/2)}{P_{hL}}$. (2) Secondly, when α becomes smaller, $\hat{\xi}$ is maximized at $1 - \beta_\ell^0 \in (0, 1)$, in which $\hat{\xi} = \xi_{NL}$. (3) Finally, when α further decreases, $\hat{\xi}$ is maximized at $1 - \beta_\ell^0 = 1$, and $\hat{\xi} = \frac{1}{2}\Delta\alpha$. Note this is also considered as a case $1 - \beta_\ell^0 = 1$, for which Section 5.5.3.4 gives an upper and lower bound of $\frac{1}{2}\Delta\alpha$.

By Lemma 11, case (2) is exactly the upper bound for the “Non-Linear” Segment 3, and case (3) is the upper bound for the “Tail” Segment 4. Additionally, the changing point between case (2) and (3) is $\alpha = \frac{1}{1 + P_{\ell L} + (P_{\ell L} - P_{hH})}$, which is also the cut-off point of Segment 3 and 4. Specifically, if this cut-off point is smaller than $\frac{1}{2}$, Segment 4 disappears. On the other hand, $\xi_{NL} \geq \frac{\Delta(\alpha - 1/2)}{P_{hL}}$ if and only if $\alpha \leq \alpha_{NL}$. This becomes the cut-off point of Segment 2 and 3.

As a conclusion, the upper bound given in the case $\beta_h^1 = 1$ has three Segments.

1. For $\alpha \geq \alpha_{NL}$, the upper-bound is at most $\frac{\Delta(\alpha - 1/2)}{P_{hL}}$.

2. For $\alpha_{NL} > \alpha > \frac{1}{1 + P_{\ell L} + (P_{\ell L} - P_{hH})}$, the upper bound is ξ_{NL}.

3. For $\alpha > \frac{1}{1 + P_{\ell L} + (P_{\ell L} - P_{hH})}$, the upper bound is $\frac{1}{2}\Delta\alpha$.

This contributes to the upper bound for Segment 3 and Segment 4 in $\xi^*(\alpha)$.

Finally, we wish to derive a (tight) lower bound from the parameters γ^*, $\beta_h^1 = 1$, and $1-\beta_\ell^0$ that maximizes $\min(\xi_h, \xi_\ell)$ (or, equivalently, $\hat{\xi}$) for "Non-Linear" Segment 3 and "Tail" Segment 4, aligning with the upper bounds, by applying Lemma 9. We need to show that the parameters satisfy the conditions for Lemma 9. β_h^1 and β_ℓ^0 are guaranteed to be $[0,1]$. Therefore, it remains to show that $\gamma^* \geq \hat{\xi}$ and $1-\alpha-\gamma^* \geq \hat{\xi}$. We don't do Segment 2 because it cannot do better than the case where γ is biased, where a lower bound $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ is given.

While the proof for this constraint is mostly calculation and solving inequalities, we would like to give a brief high-level idea here. Recall that as α decreases, the $1-\beta_\ell^0$ that maximizes $\hat{\xi}$ increases. Given that $\beta_h^1 = 1$ always holds, this implies that minority agents are playing a more and more balanced strategy. Therefore, γ^* which guarantees that $\xi_h = \xi_\ell$, will also be close to half of the minority agents, which is $\frac{1-\alpha}{2}$. As a consequence, α decreases, both γ^* and $1-\alpha-\gamma^*$ are increasing, while $\hat{\xi}$ is decreasing. Therefore, it suffices to show that at the beginning of the "Non-Linear" Segment 3, where $\alpha = \alpha_{NL}$, both $\gamma^* \geq \hat{\xi}$ and $1-\alpha-\gamma^* \geq \gamma^*$ (which is easier than directly proving that $1-\alpha-\gamma^* \geq \hat{\xi}$) hold. This is achieved by directly solving the inequalities.

5.5.3.6 Wrap Up the Bounds.

So far, we have shown that the upper and the lower bounds of ξ only derive from one of the four cases. (1) When γ is biased (either type-0 agents or type-1 agents dominates the minority group). (2) When $1-\beta_\ell^0 = 1$ (all type-0 agents always vote for $\mathbf{R}$). (3) When $\beta_h^1 = 1$ (all type-1 agents always vote for $\mathbf{A}$). (4) All other cases, upper-bounded with $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 7. Each case has its own upper bound, and cases (1), (2), and (3) also have lower bounds.

Therefore, the last step is to integrate all the cases and reach a global upper and lower bound for each Segment. For the upper bound, we take the maximum of the upper bound for each case. In each case, there does not exist a strategy profile satisfying all three constraints with ξ equal to their corresponding upper bound. Therefore, overall, no strategy profiles can satisfy all three constraints with ξ equal to the maximum of all upper bounds. For the lower bound, we show that the upper bound in each segment binds a tight lower bound. As a consequence, the bounds consist of the threshold curve $\xi^*(\alpha)$, which finishes the proof.

Now we start to integrate the upper bound of each case.

1. When γ is biased, the upper bound is $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 5 along Segment 2, 3, and 4.

2. When $1-\beta_\ell^0 = 1$, the upper bound is given by the maximum of $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ and $\frac{1}{2}\Delta\alpha$ by Proposition 8 and 9.

3. When $\beta_h^1 = 1$, the upper bound is no more than $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ in Segment 2, ξ_{NL} in Segment 3, and $\frac{1}{2}\Delta\alpha$ in Segment 4, as analyzed in Section 5.5.3.5.

4. In all other cases, the upper bound is at most $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 7.

The ordering of these upper bounds varies in different segments, of which the full proof is in Appendix C.5.1. In Segment 2, $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ is the largest. In Segment 3, ξ_{NL} is the largest. Finally, in Segment 4, $\frac{1}{2}\Delta\alpha \geq \frac{\Delta(\alpha-1/2)}{P_{hL}}$. Therefore, taking the maximum on each segment, the upper bound for ξ is $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ in Segment 2, ξ_{NL} in Segment 3, and $\frac{1}{2}\Delta\alpha$ in Segment 4 (if exists). This is exactly the upper bound given in $\xi^*(\alpha)$.

Finally, we show the tight lower bound corresponding to the upper bound in each segment.

1. In "Steep" Segment 2, Proposition 6 in the biased case gives the lower bound $\frac{\Delta(\alpha-1/2)}{P_{hL}}$, where all minority agents always vote for **A**.

2. In "Non-Linear" Segment 3, we show that the corresponding ξ^* and $1-\beta_\ell^0$ that maximizes $\hat{\xi}$ serves as a valid set of parameters in Section 5.5.3.5. By Lemma 9, ξ_{NL} is also a lower bound.

3. For "Tail" Segment 4, the parameters $\beta_h^1 = 1 - \beta_\ell^0 = 1$, and $\gamma = \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$ (as discussed in the Section 5.5.3.4 and 5.5.3.5) gives the lower bound $\frac{1}{2}\Delta\alpha$ by Lemma 9.

Therefore, the lower bound on ξ is exactly given in $\xi^*(\alpha)$. We finish the proof that $\xi^*(\alpha)$ is exactly the threshold curve for which a strategy profile satisfying all three constraints exists.

5.5.4 Equilibria in Each Segment

In this section, we display the "optimal" equilibria achieving the maximized ξ^* in each segment as an extension of the proof sketch. In each segment, the equilibrium exhibits a different structure and leads to different features of the segment Throughout this section, we assume that $P_{\ell L} \geq P_{hH}$ for a clearer illustration. In each equilibrium described, all majority agents play the same strategy, and all the minority agents in the same group play the same strategy.

We start from $\alpha > \theta$ where an n-strong equilibrium exists. In the equilibrium, majority agents play a modified informative voting such that $\beta_h = 1$ and $\beta_\ell \geq 0$. Deng et al. [10] show that this majority strategy guarantees an n-strong equilibrium regardless of minority behavior. When $P_{hH} = P_{\ell L}$, $\beta_\ell = 0$, and the strategy becomes informative voting that maximizes the margins. When $P_{\ell L} > P_{hH}$, the H side margin $\varphi_H - \frac{1}{2}$ is more fragile than the L side margin $\frac{1}{2} - \varphi_L$ under informative voting. Therefore, β_ℓ is raised to strengthen the L side at the expense of the H side.

In the "Flat" Segment 1, the majority's equilibrium strategy keeps the same pattern of $\beta_h = 1$ and $\beta_\ell \geq 0$. Moreover, β_ℓ decreases as α decreases, reaching 0 at the transition to Segment 2. The minority agents always vote for **A** in the equilibrium strategy. Specifically, when $P_{hH} = P_{\ell L}$, Segment 1 reduces to a point. The flat threshold occurs because the minority **A** votes offset the margin imbalance caused by signal asymmetry ($P_{\ell L} > P_{hH}$). As these **A** votes strengthen the H side margin, the majority can adopt strategies closer to informative voting despite a shrinking population. Strategies closer to informative voting have larger $\beta_h - \beta_\ell$, which yields a larger sum of two margins $\varphi_H - \varphi_L$. Consequently, the margins remain constant with the majority population and keep the segment flat. At the transition to Segment 2, minority A votes exactly balance signal asymmetry, allowing majority agents to adopt pure informative voting.

In the "Steep" Segment 2, the majority of agents play a modified informative voting, but in another direction: $\beta_h < 1$ and $\beta_\ell = 0$. Moreover, β_h decreases as α decreases. The minority agents still always vote for **A**. In this segment, minority **A** votes overcompensate for signal asymmetry, and the L side margin becomes weaker under informative voting. Therefore, the majority agents must sacrifice total margin to balance both sides, causing the threshold decline.

The "Non-Linear" Segment 3 exhibits the most complex equilibrium structure. The majority of agents follow $\beta_h < 1$ and $\beta_\ell = 0$, and β_h increases toward 1 as α decreases. Minority agents split into two groups. In the larger group (Group 1), all the agents always vote for **A**. In the smaller group (Group 0), agents play a strategy of $\beta_h = 0$ and $\beta_\ell \in (0, 1)$. As α decreases, the strategy of Group 0 gets closer to always voting for **R**, reaching it at the transition to Segment 4. In this equilibrium, opposing votes from the two groups offset each other's deviation power, maximizing tolerance to deviations. The non-linear relationship arises because both Group 0's strategy and fraction change simultaneously as α varies, creating multiplicative effects on margin changes.

Finally, in the “Tail” Segment 4, all the majority agents vote informatively. Minority agents maintain their two-group structure: the larger Group 1 always votes for **A**, while the smaller Group 0 always votes for **R**. Opposing votes between groups offset their collective deviation power, while the size difference helps balance signal distribution asymmetry. Unlike Segment 3, strategies remain fixed as α changes, restoring the linear threshold relationship.

CHAPTER 6
CONCLUSION AND FUTURE WORK

This dissertation addresses whether strategic agents can reach informed collective decisions when agents possess different subjective preferences contingent on an objective ground truth. We give an affirmative answer: strategic behavior indeed reaches informed decisions under simple mechanisms such as majority vote and its variants under different preference structures. This sharply contrasts with the classical non-strategic behaviors, which require additional assumptions to reach good outcomes. Consequently, we claim that strategic behavior is a bliss in making informed decisions.

We propose a unified utility framework that holds both subjective and objective preferences: each agent's utility is the sum of a subjective component, which depends only on the chosen alternative, and an objective component, which depends on the unknown ground truth. We adopt the informed majority decision as the criterion of a "good", which is the alternative a majority would favor had the ground truth been a common knowledge.

Chapter 3 proves the foundational result. Under the solution concepts of approximated strong Bayes-Nash equilibrium, every equilibrium of strategic majority voting reaches the informed majority decision. The next two chapters answer questions this result leaves open. Chapter 4 asks whether the informed majority decision can be reached by a mechanism simpler than the equilibria of Chapter 3, which are not easy for voters to locate. It gives a two-round procedure whose natural equilibrium aggregates information through a first-round poll and decides on the aggregate in the second round. Chapter 5 investigates a different preference structure where a strong equilibrium may not exist. The key question is whether strategic behavior induces a good outcome where agents have limited capability to coordinate. It introduces k-strong equilibrium, which caps the size of any deviating coalition at k, and shows that the informed majority decision is reached even when coordination is bounded.

6.1 Limitations

So far, my research is limited to binary voting, where agents select a winner from two alternatives. One of the most important extensions in the future is to study informed decision-making in a wider range of social choice scenarios, including single-winner voting with more alternatives, multi-winner voting, and participatory budgeting. A challenge of this extension lies in the definition of a "good" decision, as we do not have a universally acknowledged criterion like "the majority's wish" in binary voting. One idea is to extend the informed majority decision to the *informed-X-decision*, where X can be a voting rule or an axiom (a desiderata proposed to evaluate, compare, and design voting rules). Given a voting rule r, informed-r requires that the r winner under agents' fully-informed preferences be chosen. Given a social choice desideratum X, informed-X requires that the group decision to satisfy X w.r.t. agents' fully-informed preferences. Informed-X-decision leverages the rich collection of voting rules and axioms in the existing social choice literature to establish comprehensive benchmarks and design targets for mechanisms in our preference-uncertain decision-making scenarios. Another challenge is that the landscape of strategic behaviors and their outcome may change drastically. As indicated by Goertz and Maniquet [23], informative voting may be an equilibrium but leads to the wrong alternative in a model of three alternatives. A more thorough analysis will be a future direction.

We also have limitations on the assumptions about the settings. For example, we assume that agents receive i.i.d. signals, and the preferences of agents follow a certain structure of population (aligned or opposite), and that agents commit to their strategies before observing signals. Accordingly, future directions include investigating agents with correlated signals, agents with different competencies, and a wider range of solution concepts.

6.2 Looking Forward

My research addresses challenges at the intersection of several contemporary movements. The integration of AI systems into decision-making (e.g., content moderation and medical diagnosis) demands frameworks that coordinate human and machine intelligence across different information sources. The rise of online deliberation platforms (Polis, Remesh), participatory governance, and decentralized autonomous organizations reflects growing institutional interest in scalable collective decision-making mechanisms. Meanwhile, the epistemic challenges of our information environment, like algorithmic filter bubbles [88] and echo chambers [89], make information aggregation not merely theoretical but essential for democratic legitimacy. My framework for understanding how strategic agents with heterogeneous information and preferences reach informed decisions provides theoretical foundations for these emerging applications in digital democracy, human-AI collaboration, and decentralized governance.

Moving forward, my research will deepen the analysis of the three key challenges. A promising direction for **preference modeling** is developing richer characterizations spanning the full spectrum between aligned and opposing structures, incorporating multiple social and information groups with overlapping or conflicting networks, and analyzing multi-dimensional or continuous information structures. For **aggregation mechanisms**, a key direction is systematically investigating the informed decisions across comprehensive collections of voting rules and axioms to build a taxonomy suited to preference-uncertain environments. Regarding **strategic behavior and incentives**, extending analysis beyond binary decisions to multi-alternative voting, multi-winner elections [90], and participatory budgeting [91] will reveal how the epistemic perspective of simultaneous preference expression and information aggregation applies in richer decision spaces.

Bridging Theory and Practice. To ensure theoretical insights translate to real-world impact, I will continue conducting experiments with both large language model agents and human subjects, validating theoretical predictions and uncovering behavioral patterns that inform refined models.

Concluding Remarks. As our world becomes increasingly interconnected yet fragmented by information silos, the need for robust, fair, and informed group decision-making mechanisms has never been greater. My research aims to provide the theoretical foundations and practical tools to meet this challengeenabling diverse groups of humans and AI systems to harness their collective knowledge and navigate their differences to reach decisions that reflect both individual values and shared understanding of reality. By bridging game theory, social choice, and behavioral analysis, this agenda strives toward a future where strategic complexity becomes an asset rather than an obstacle to collective wisdom.

REFERENCES

[1] M. d. Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris: L'Imprimerie Royale, 1785.

[2] H. Nurmi, *Voting Paradoxes and How to Deal with Them.* Springer-Verlag Berlin Heidelberg, 1999.

[3] S. Barberà, H. Sonnenschein, and L. Zhou, "Voting by committees," *Econometrica*, vol. 59, no. 3, pp. 595–609, 1991.

[4] D. Austen-Smith and J. S. Banks, "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *The American Political Science Review*, vol. 90, no. 1, pp. 34–45, 1996.

[5] J. Wit, "Rational Choice and the Condorcet Jury Theorem," *Games and Economic Behavior*, vol. 22, no. 2, pp. 364–376, 1998.

[6] R. B. Myerson, "Extended Poisson Games and the Condorcet Jury Theorem," *Games and Economic Behavior*, vol. 25, no. 1, pp. 111–131, 1998.

[7] T. Feddersen and W. Pesendorfer, "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting," *American Political Science Review*, vol. 92, no. 1, pp. 23–35, 1998.

[8] T. Feddersen and W. Pesendorfer, "Voting Behavior and Information Aggregation in Elections With Private Information," *Econometrica*, vol. 65, no. 5, pp. 1029–1058, 1997.

[9] G. Schoenebeck and B. Tao, "Wisdom of the crowd voting: Truthful aggregation of voter information and preferences," in *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 2021, pp. 1872–1883.

[10] X. Deng, B. Tao, and Y. Wang, "Aggregation of antagonistic contingent preferences: When is it possible?" In *International Conference on Web and Internet Economics*, Springer, 2024, Forthcoming. See also https://arxiv.org/abs/2410.08127.

[11] S. Nitzan and J. Paroush, "Collective decision making and jury theorems," in *The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts*, F. Parisi, Ed., Oxford University Press, 2017.

[12] N. R. Miller, "Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem," in *Information Pooling and Group Decision Making*, G. B. and O. G., Eds., JAI Press, 1986, pp. 173–192.

[13] B. Grofman, G. Owen, and S. L. Feld, "Thirteen theorems in search of the truth," *Theory and Decision*, vol. 15, no. 3, pp. 261–278, 1983.

[14] G. Owen, B. Grofman, and S. L. Feld, "Proving a distribution-free generalization of the Condorcet Jury Theorem," *Mathematical Social Sciences*, vol. 17, no. 1, pp. 1–16, 1989.

[15] P. J. Boland, F. Proschan, and Y. L. Tong, "Modelling Dependence in Simple and Indirect Majority Systems," *Journal of Applied Probability*, vol. 26, no. 1, pp. 81–88, 1989.

[16] S. Nitzan and J. Paroush, "The significance of independent decisions in uncertain dichotomous choice situations," *Theory and Decision*, vol. 17, no. 1, pp. 47–60, 1984.

[17] L. Shapley and B. Grofman, "Optimizing group judgmental accuracy in the presence of interdependencies," *Public Choice*, vol. 43, no. 3, pp. 329–343, 1984.

[18] S. Kaniovski, "Aggregation of correlated votes and Condorcet's Jury Theorem," *Theory and Decision*, vol. 69, no. 3, pp. 453–468, 2010.

[19] S. Nitzan and J. Paroush, "Investment in Human Capital and Social Self Protection under Uncertainty," *International Economic Review*, vol. 21, no. 3, pp. 547–557, 1980.

[20] M. Gradstein and S. Nitzan, "Organizational decision-making quality and the severity of the free-riding problem," *Economics Letters*, vol. 23, no. 4, pp. 335–339, 1987.

[21] R. Ben-Yashar and M. Zahavi, "The Condorcet jury theorem and extension of the franchise with rationally ignorant voters," *Public Choice*, vol. 148, no. 3, pp. 435–443, 2011.

[22] H. P. Young, "Condorcet's theory of voting," *American Political Science Review*, vol. 82, pp. 1231–1244, 1988.

[23] J. M. Goertz and F. Maniquet, "Condorcet Jury Theorem: An example in which informative voting is rational but leads to inefficient information aggregation," *Economics Letters*, vol. 125, no. 1, pp. 25–28, 2014.

[24] I. Esponda and E. Vespa, "Hypothetical thinking and information extraction in the laboratory," *American Economic Journal: Microeconomics*, vol. 6, no. 4, pp. 180–202, Nov. 2014.

[25] M. Battaglini, R. B. Morton, and T. R. Palfrey, "The swing voter's curse in the laboratory," *The Review of Economic Studies*, vol. 77, no. 1, pp. 61–89, Jan. 2010.

[26] J. K. Goeree and L. Yariv, "An experimental study of collective deliberation," *Econometrica*, vol. 79, no. 3, pp. 893–921, 2011.

[27] J. Duggan and C. Martinelli, "A Bayesian Model of Voting in Juries," *Games and Economic Behavior*, vol. 37, no. 2, pp. 259–294, 2001.

[28] A. Meirowitz, "Informative voting and condorcet jury theorems with a continuum of types," *Social Choice and Welfare*, vol. 19, no. 1, pp. 219–236, 2002.

[29] A. Meirowitz, "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values," *Journal of Theoretical Politics*, vol. 19, no. 3, pp. 301–327, 2007.

[30] C. Martinelli, "Convergence results for unanimous voting," *Journal of Economic Theory*, vol. 105, no. 2, pp. 278–297, 2002.

[31] A. Chakraborty and P. Ghosh, "Efficient equilibria and information aggregation in common interest voting games," Mimeo, Tech. Rep., 2003.

[32] J. Morgan and P. C. Stocken, "Information aggregation in polls," *American Economic Review*, vol. 98, no. 3, pp. 864–96, Jun. 2008.

[33] J. Kim and M. Fey, "The swing voter's curse with adversarial preferences," *Journal of Economic Theory*, vol. 135, no. 1, pp. 236–252, 2007.

[34] S. N. Ali, M. Mihm, and L. Siga, "Adverse selection in distributive politics," *Available at SSRN 3579095*, 2018.

[35] J. McMurray, "Ideology as opinion: A spatial model of common-value elections," *American Economic Journal: Microeconomics*, vol. 9, no. 4, pp. 108–140, 2017.

[36] P. Barelli, S. Bhattacharya, and L. Siga, "Full information equivalence in large elections," *Econometrica*, vol. 90, no. 5, pp. 2161–2185, 2022.

[37] S. Bhattacharya, "Preference monotonicity and information aggregation in elections," *Econometrica*, vol. 81, no. 3, pp. 1229–1247, 2013.

[38] S. Bhattacharya, "Condorcet jury theorem in a spatial model of elections," Ph.D. dissertation, Indian Institute of Management Calcutta, 2023.

[39] Q. Han, G. Schoenebeck, B. Tao, and L. Xia, "The wisdom of strategic voting," in *Proceedings of the 24th ACM Conference on Economics and Computation*, London, United Kingdom: Association for Computing Machinery, 2023, pp. 885–905, ISBN: 9798400701047.

[40] P. J. Coughlan, " In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting," *The American Political Science Review*, vol. 94, no. 2, pp. 375–393, 2000.

[41] E. Schulte, "Information aggregation and preference heterogeneity in committees," *Theory and Decision*, vol. 69, no. 1, pp. 97–118, 2010.

[42] D. Gerardi, "Jury Verdicts and Preference Diversity," *The American Political Science Review*, vol. 94, no. 2, pp. 395–406, 2000.

[43] D. Gerardi and L. Yariv, "Deliberative voting," *Journal of Economic Theory*, vol. 134, no. 1, pp. 317–338, 2007.

[44] E. Dekel and M. Piccione, "Sequential voting procedures in symmetric binary elections," *Journal of Political Economy*, vol. 108, pp. 34–55, 1 2000.

[45] S. Chopra, E. Pacuit, and R. Parikh, "Knowledge-theoretic properties of strategic voting," in *Logics in Artificial Intelligence*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 18–30.

[46] D. Prelec, H. S. Seung, and J. McCoy, "A solution to the single-question crowd wisdom problem," *Nature*, vol. 541, no. 7638, pp. 532–535, 2017.

[47] H. Hosseini, D. Mandal, N. Shah, and K. Shi, "Surprisingly popular voting recovers rankings, surprisingly!" In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 245–251.

[48] M. Monroe and B. Waggoner, *Public projects with preferences and predictions*, 2025. arXiv: 2403.01042 [cs.GT].

[49] R. J. Aumann, "Acceptable points in general cooperative n-person games," in *Annals of Mathematics Study 40*, ser. Contributions to the Theory of Games, vol. IV, Princeton University Press, 1959, pp. 287–324.

[50] Y. Desmedt and E. Elkind, "Equilibria of plurality voting with abstentions," in *Proceedings of the 11th ACM conference on Electronic commerce*, 2010, pp. 347–356.

[51] S. Barbera, M. Maschler, and J. Shalev, "Voting for voters: A model of electoral evolution," *Games and Economic Behavior*, vol. 37, no. 1, pp. 40–78, 2001.

[52] Z. Rabinovich, S. Obraztsova, O. Lev, E. Markakis, and J. Rosenschein, "Analysis of equilibria in iterative voting schemes," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, 2015.

[53] R. Holzman and N. Law-Yone, “Strong equilibrium in congestion games,” *Games and economic behavior*, vol. 21, no. 1-2, pp. 85–101, 1997.

[54] H. Yin, U. V. Shanbhag, and P. G. Mehta, “Nash equilibrium problems with scaled congestion costs and shared constraints,” *IEEE transactions on automatic control*, vol. 56, no. 7, pp. 1702–1708, 2011.

[55] T. Harks and M. Klimm, “On the existence of pure nash equilibria in weighted congestion games,” *Mathematics of Operations Research*, vol. 37, no. 3, pp. 419–436, 2012.

[56] T. Ichiishi and A. Yamazaki, “Cooperative extensions of the bayesian game,” in *Advances in Mathematical Economics*. Tokyo: Springer Japan, 2006, pp. 273–296.

[57] T. Ichiishi and A. Idzik, “Bayesian cooperative choice of strategies,” *International Journal of Game Theory*, vol. 25, pp. 455–473, 1996.

[58] G. Hahn and N. C. Yannelis, “Coalitional bayesian nash implementation in differential information economies,” *Economic Theory*, vol. 18, pp. 485–509, 2001.

[59] M. Safronov, “Coalition-proof full efficient implementation,” *Journal of Economic Theory*, vol. 177, pp. 659–677, 2018.

[60] X. A. Gao, J. R. Wright, and K. Leyton-Brown, “Incentivizing evaluation with peer prediction and limited access to ground truth,” *Artificial Intelligence*, vol. 275, pp. 618–638, 2019.

[61] H. Guo and N. C. Yannelis, “Robust coalitional implementation,” *Games and economic behavior*, vol. 132, pp. 553–575, 2022.

[62] Q. Han, G. Schoenebeck, B. Tao, and L. Xia, “Strong equilibria in bayesian games with bounded group size,” in *Proceedings of the ACM on Web Conference 2025*, 2025, pp. 4572–4581.

[63] I. Abraham, D. Dolev, and J. Y. Halpern, “Lower bounds on implementing robust and resilient mediators,” in *Theory of Cryptography*, R. Canetti, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 302–319.

[64] I. Abraham, D. Dolev, R. Gonen, and J. Halpern, “Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation,” in *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing*, Denver, Colorado, USA: Association for Computing Machinery, 2006, pp. 53–62.

[65] J. E. Bickel, “Some comparisons among quadratic, spherical, and logarithmic scoring rules,” *Decision Analysis*, vol. 4, no. 2, pp. 49–65, 2007.

[66] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[67] N. Miller, P. Resnick, and R. Zeckhauser, “Eliciting Informative Feedback: The Peer-Prediction Method,” *Management Science*, vol. 51, no. 9, pp. 1359–1373, 2005.

[68] G. Schoenebeck and F.-Y. Yu, “Learning and Strongly Truthful Multi-Task Peer Prediction: A Variational Approach,” in *12th Innovations in Theoretical Computer Science Conference*, ser. Leibniz International Proceedings in Informatics, vol. 185, Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 78:1–78:20.

[69] G. Schoenebeck, F.-Y. Yu, and Y. Zhang, “Information elicitation from rowdy crowds,” in *Proceedings of the Web Conference 2021*, New York, NY, USA: Association for Computing Machinery, 2021, pp. 3974–3986, ISBN: 9781450383127.

[70] D. Prelec, “A bayesian truth serum for subjective data,” *Science*, vol. 306, no. 5695, pp. 462–466, 2004.

[71] J. Witkowski and D. C. Parkes, “Peer Prediction Without a Common Prior,” in *Proceedings of the 13th ACM Conference on Electronic Commerce*, Valencia, Spain, 2012, pp. 964–981.

[72] M. S. Miller and K. E. Drexler, “Markets and computation: Agoric open systems,” in *The Ecology of Computation*, B. A. Huberman, Ed., North-Holland, 1988.

[73] J. Wolfers and E. Zitzewitz, “Prediction markets,” *Journal of Economic Perspectives*, vol. 18, no. 2, pp. 107–126, Jun. 2004.

[74] V. P. Crawford and J. Sobel, “Strategic information transmission,” *Econometrica: Journal of the Econometric Society*, pp. 1431–1451, 1982.

[75] M. Battaglini, “Policy advice with imperfectly informed experts,” *The BE Journal of Theoretical Economics*, vol. 4, no. 1, 2004.

[76] D. Austen-Smith, “Interested experts and policy advice: Multiple referrals under open rule,” *Games and Economic Behavior*, vol. 5, no. 1, pp. 3–43, 1993.

[77] M. Battaglini, “Multiple referrals and multidimensional cheap talk,” *Econometrica*, vol. 70, no. 4, pp. 1379–1401, 2002.

[78] E. Baharad, J. Goldberger, M. Koppel, and S. Nitzan, “Beyond condorcet: Optimal aggregation rules using voting records,” *Theory and decision*, vol. 72, no. 1, pp. 113–130, 2012.

[79] A. C. Berry, "The accuracy of the gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941.

[80] C.-G. Esseen, "On the liapunoff limit of error in the theory of probability," *Arkiv för Matematik, Astronomi och Fysik*, vol. A28, no. 9, pp. 1–19, 1942.

[81] D. Black, "On the rationale of group decision-making," *Journal of Political Economy*, vol. 56, no. 1, pp. 23–34, 1948.

[82] R. I. Dunbar, "Neocortex size as a constraint on group size in primates," *Journal of human evolution*, vol. 22, no. 6, pp. 469–493, 1992.

[83] F. P. Brooks Jr, *The mythical man-month: essays on software engineering*. Addison-Wesley, 1975.

[84] A. G. Ingham, G. Levinger, J. Graves, and V. Peckham, "The ringelmann effect: Studies of group size and group performance," *Journal of experimental social psychology*, vol. 10, no. 4, pp. 371–384, 1974.

[85] E. Koutsoupias and C. H. Papadimitriou, "Worst-case equilibria," in *Symposium on Theoretical Aspects in Computer Science*, 1999, pp. 404–413.

[86] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden, "The price of stability for network design with fair cost allocation," in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004, pp. 295–304.

[87] T. Feddersen and W. Pesendorfer, "The swing voter's curse," *American Economic Review*, vol. 86, pp. 408–424, 1996.

[88] E. Pariser, *The filter bubble: What the internet is hiding from you*. Penguin Books Limited, 2011.

[89] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the national academy of sciences*, vol. 118, no. 9, e2023301118, 2021.

[90] T. Allouche, J. Lang, and F. Yger, "Multi-winner approval voting goes epistemic," in *Uncertainty in Artificial Intelligence*, PMLR, 2022, pp. 75–84.

[91] S. Rey and U. Endriss, "Epistemic selection of costly alternatives: The case of participatory budgeting," *Autonomous Agents and Multi-Agent Systems*, vol. 39, no. 1, p. 1, 2025.

[92] I. Shevtsova, "An improvement of convergence rate estimates in the lyapunov theorem," *Doklady Mathematics*, vol. 82, pp. 862–864, Dec. 2010.

Appendices

APPENDIX A

TECHNICAL APPENDIX FOR CHAPTER 3

A.1 Proof of Theorem 1

Theorem 1. *For any $N_0 \in \mathbb{N}$, there exists an instance of $N > N_0$ agents, which does not exist a strong Bayes-Nash equilibrium.*

Proof. For any N_0, we construct an instance of $N = 2N_0 + 3$ agents. The agents consist of three parts: F is a set of $N_0 + 1$ friendly agents. C is a set of two contingent agents. And U is a set of N_0 unfriendly agents. The valuation function of each set is shown in the table. The threshold $\mu = 0.5$. The prior distribution $P_L = P_H = 0.5$. The signal distribution $P_{hH} = P_{lL} = 0.8$, and $P_{lH} = P_{hL} = 0.2$.

agents	$v(H, \mathbf{A})$	$v(L, \mathbf{A})$	$v(L, \mathbf{R})$	$v(H, \mathbf{R})$
F	100	99	1	0
C	90	0	100	0
U	1	0	100	99

Still, consider three strategy profiles:

- Σ_1: In F, N_0 agents always vote for $\mathbf{A}$, and one agent votes informatively. Two agents of C vote informatively. N_0 agents of U always vote for $\mathbf{R}$.

- Σ_2: $N_0 + 1$ agents of F always vote for $\mathbf{A}$. Two agents of C vote informatively. N_0 agents of U always vote for $\mathbf{R}$.

- Σ_3: $N_0 + 1$ agents of F always vote for $\mathbf{A}$. One C agent votes informatively, and the other agent always votes for $\mathbf{R}$. N_0 agents of U always vote for $\mathbf{R}$.

The expected utility of each type of agents in three profiles is in the table as follows. The three profiles form a deviating cycle of $\Sigma_1 \to \Sigma_2 \to \Sigma_3 \to \Sigma_1$. In Σ_1, the F agent who votes

agents	Σ_1	Σ_2	Σ_3
F	50.396	66.14	50.3
C	85.12	75.2	76
U	50.396	34.46	50.3

informatively has the incentive to always vote for **A**, and the profile becomes Σ_2. In Σ_2, a C agent has the incentive to always vote for **R**, and the profile becomes Σ_3. And in Σ_3, the group of a F agent and two C agents have the incentives to turn to informative voting, and the profile becomes Σ_1.

Now we show that for any other strategy profile Σ, there exists a group of agents with incentives to deviate.

Case 1: **A** is always the winner. In this case, at least $N_0 + 2$ agents always vote for **A**, the expected utility of U agents is 0.5, and the expected utility of C agents is 45. If there exists a group of U agents such that **A** is not always the winner after they deviate to always vote for **R**, their expected utility will increase. Otherwise, there are still at least $N_0 + 2$ agents who always vote for **A** even if all U agents always vote for **R**. Therefore, at least one C agent always votes for **A**. Now consider the deviating group $C \cup U$, where all agents in the group turn to always voting for **R**. Then **R** will always be the winner. The expected utility of U agents increases to 99.5, and the expected utility of C agents increases to 50. Therefore, in this case, Σ is not a strong BNE.

Case 2: **R** is always the winner. In this case, at least $N_0 + 2$ agents always vote for **R**; the expected utility of F agents is 0.5, and the expected utility of C agents is 50. Similarly, if all agents in F voting for **A** can reverse the decision, their expected utility will increase. Otherwise, all agents in U and C always vote for **R**. Then, the group $F \cup C$ has incentives to deviate, where F agents always vote for **A**, and C agents vote informatively. The profile after the deviation is Σ_2, in which both F and C agents have high expected utilities. Therefore, in this case, Σ is not a strong BNE.

Case 3: Neither of the alternatives is always the winner. In this case, any F agent not always voting for **A** and U agent not always voting for **R** has incentives to deviate to get a higher probability that their preferred alternative to be selected. Now suppose all F agents always vote for **A**, and U agents always vote for **R**. In this case, we show that the expected utility of C agents is uniquely maximized by one agent voting informatively and the other agent always voting for **R**, just as Σ_3. Therefore, for any $\Sigma \neq \Sigma_3$, C agents have incentives to deviate to Σ_3. And Σ_3 itself is dominated by Σ_1.

Now without loss of generality, we index the two agents in C as agent 1 and agent 2. Suppose the strategy of agent 1 is $\sigma_1 = (\beta_l^1, \beta_h^1)$, and the strategy of agent 2 is $\sigma_2 = (\beta_l^2, \beta_h^2)$. Since there are $N_0 + 1$ votes for **A** and N_0 votes for **R**, **R** is the winner only when both two agents vote for **R**. Let X_1 and X_2 be the random variable that denotes the vote of agent 1 and agent 2 respectively. $X_i = 1$ stands for "agent i votes for **A**", and $X_i = 0$ stands for "agent i votes for **R**". Therefore, the probability of **A** being winner under H state is

$$\begin{aligned}
\lambda_H^{\mathbf{A}}(\Sigma) =& 1 - \Pr[X_1 = 0 \wedge X_2 = 0 \mid W = H] \\
=& 1 - \Pr[X_1 = 0 \mid W = H] \cdot \Pr[X_2 = 0 \mid W = H] \\
=& 1 - (0.8 \cdot (1 - \beta_h^1) + 0.2 \cdot (1 - \beta_l^1))(0.8 \cdot (1 - \beta_h^2) + 0.2 \cdot (1 - \beta_l^2)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\lambda_L^{\mathbf{R}}(\Sigma) =& \Pr[X_1 = 0 \wedge X_2 = 0 \mid W = L] \\
=& (0.8 \cdot (1 - \beta_l^1) + 0.2 \cdot (1 - \beta_h^1))(0.8 \cdot (1 - \beta_l^2) + 0.2 \cdot (1 - \beta_h^2)).
\end{aligned}$$

And

$$u_1(\Sigma) = u_2(\Sigma) = 0.5 \cdot (90 \cdot \lambda_H^{\mathbf{A}}(\Sigma) + 100 \cdot \lambda_L^{\mathbf{R}}(\Sigma)).$$

Let $x_1 = 1 - \beta_h^1, y_1 = 1 - \beta_l^1, x_2 = 1 - \beta_h^2$, and $y_2 = 1 - \beta_l^2$, and expand the expected utility, we get

$$u_1(\Sigma) = -26.8x_1x_2 + 0.8x_1y_2 + 0.8x_2y_1 + 30.2y_1y_2 + 45.$$

We consider the variables to maximize the expected utility. Note that given x_1 and y_1, if at least one of x_1 and y_1 does not equal to zero, $y_2 = 1$ is a necessary condition for $u_2(\Sigma)$ to be maximized, and vice versa. In the special case where four variables are all 0, Σ stands for both C agents always vote for **A**. In this case, **A** is always the winner, and C agents have incentives to deviate to always vote for **R** as shown in Case 1. Now, given that at least one of the four values is non-zero, we must have $y_1 = y_2 = 1$ to maximize the expected utility. Then the expected utility turns to

$$u_1(\Sigma) = -26.8x_1x_2 + 0.8x_1 + 0.8x_2 + 75.2.$$

Then it's not hard to see that $u_1(\Sigma)$ is uniquely maximized when one of x_1 and x_2 takes 0 and the other takes 1. Then Σ is exactly Σ_3. Therefore, the expected utility of C agents is maximized on Σ_3.

Consequently, we show that in every case, a strategy profile is not a strong BNE. Therefore, a strong BNE does not exist in the instance. $\square$

A.2 Non-binary Setting

A.2.1 Additional Setting

In this section, we propose notions in non-binary settings that are used in technical proofs and not mentioned in the main paper.

Fraction of Agents Given a specific n, we use n_F, n_C, and n_U to denote the number of each type of agents, and use $\hat{\alpha}_F = \frac{n_F}{n}, \hat{\alpha}_C = \frac{n_C}{n}$ and $\hat{\alpha}_U = \frac{n_U}{n}$ to denote the fraction of each type of agents. Note that $\hat{\alpha}_F, \hat{\alpha}_C$, and $\hat{\alpha}_U$ is dependent to n.

In the non-binary setting, we have $\hat{\alpha}_F < \mu$ and $\hat{\alpha}_U < \mu$ for all sufficiently large n, which means that neither friendly nor unfriendly agents can dominate the vote. Note that this is guaranteed by the assumption that both $\mathcal{L}$ and $\mathcal{H}$ are non-empty.

Proposition 10. *There exists a constant $n_\mu > 0$, s.t. for all $n > n_\mu$, $\hat{\alpha}_F < \mu$ and $\hat{\alpha}_U < \mu$.*

Proof. First, we show that $\hat{\alpha}_F < \mu$. For a candidate-friendly agent, we have $\mathcal{L}_i \subsetneq \mathcal{L}$. Therefore, $L_i < L$, and i prefer $\mathbf{A}$ if the world state is L. On the other hand, an agent vote for $\mathbf{A}$ in L is a friendly agent according to the definition. Therefore, the set of friendly agents is exactly the set of agents who will vote $\mathbf{A}$ in world state L. Therefore we have

$$\begin{aligned} n_F =& |N(L, \mathbf{A})| \\ =& n - \lfloor \alpha_L^{\mathbf{R}} \cdot T \rfloor \\ \leq& (1 - \alpha_L^{\mathbf{R}}) \cdot n + 1 \\ =& \alpha_L^{\mathbf{A}} n + 1 \\ =& n(\alpha_L^{\mathbf{A}} + \frac{1}{T}). \end{aligned}$$

According to the definition of L, we know that $\alpha_L^{\mathbf{A}} < \mu$ and $\alpha_L^{\mathbf{A}}$ is independent from n. Therefore, there must exist a $n_\mu > 0$ s.t. for all $n > n_\mu$, $\alpha_L^{\mathbf{A}} + \frac{1}{n} < \mu$, and therefore $\hat{\alpha}_F < \mu$.

Then we show that $\hat{\alpha}_U < 1 - \mu$. Similarly, the set of unfriendly agents is exactly the set of agents who will vote $\mathbf{R}$ in world state H. Therefore,

$$\begin{aligned} n_U =& |N(H, \mathbf{R})| \\ =& \lfloor \alpha_H^{\mathbf{R}} \cdot n \rfloor \\ \leq& \alpha_H^{\mathbf{R}} \cdot n \\ <& (1 - \mu) \cdot n. \end{aligned}$$

The last inequality is due to the definition of H. Therefore, we have $\hat{\alpha}_U < 1 - \mu$. □

For the non-binary results we all assume that $n > n_\mu$, therefore $\hat{\alpha}_F < \mu$ and $\hat{\alpha}_U < \mu$.

Then we define the approximate fraction of each type of agents α_F, α_C, and α_U that are independent of n. As we have shown in Proposition 10, the set of friendly agents is exactly the set of agents who will vote **A** in L, and the set of unfriendly agents is exactly the set of agents who will vote **R** in H. Therefore, we define α_F, α_C, and α_U as follows:

1. $\alpha_F = \alpha_L^{\mathbf{A}}$.

2. $\alpha_U = \alpha_H^{\mathbf{R}}$.

3. $\alpha_C = 1 - \alpha_F - \alpha_U$.

According to the definition we have $\alpha_F < \mu$ and $\alpha_U < 1 - \mu$. And it's not hard to verify that $n_F = n - \lfloor (1 - \alpha_F) \cdot n \rfloor$, $n_U = \lfloor \alpha_U \cdot n \rfloor$, and $n_C = \lfloor (1 - \alpha_F) \cdot n \rfloor - \lfloor \alpha_U \cdot n \rfloor$.

Error Rate Error rate is complementary of the fidelity. Given a strategy profile Σ, let $\lambda_k^{\mathbf{A}}(\Sigma)$ ($\lambda_k^{\mathbf{R}}(\Sigma)$, respectively) be the probability that **A** (**R**, respectively) becomes the winner when world state is k. We can define the error rate I of the mechanism:

$$I(\Sigma) = \sum_{k \in \mathcal{L}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma) + \sum_{k \in \mathcal{H}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma). \tag{A.1}$$

Note that $A(\Sigma) + I(\Sigma) = 1$.

A.2.2 Berry-Esseen Theorem

In this section, we recall the technical theorem which will be used in the proof of our result. Berry-Esseen Theorem [79, 80] bounds the difference between the distribution of the sum of independent variables and the normal distribution.

Definition 11 (Berry-Esseen Theorem)**.** Let $X_1, X_2, \cdots, X_n$ be n independent variables with $E[X_i] = 0$, $Var(X_i) = \sigma_i^2 > 0$, and $E[|X_i|^3] = \rho_i < \infty$. Let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}}.$$

Let F_n denote the CDF of S_n, and Φ to be the CDF of the standard normal distribution. Then there exists an absolute constant C_0 s.t. for all n,

$$\sup_{s\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq C_0 \cdot \frac{\rho_1 + \rho_2 + \cdots + \rho_n}{(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{3/2}}.$$

The upper bound of C_0 is estimated to be 0.5600 by Reference [92].

A.3 Non-binary Results.

In this section, we present our theoretical results in the non-binary setting. All the theorems are natural extensions from the binary setting to the non-binary setting, and the proofs also hold for the binary setting. We'll give remarks to show how to convert the proofs into the binary setting.

A.3.1 Lemma 12 (Lemma 1 for binary setting)

First, we show that a strategy profile with high fidelity will lead to an ε-strong BNE with a sufficiently small ε.

Lemma 12. *Let $e(n)$ be a function of n. For any $n > n_\mu$ and any regular strategy profile Σ^* with n agents, if Σ^* satisfies $A(\Sigma^*) \geq 1 - e(n)$ (equivalently, $I(\Sigma^*) \leq e(n)$), then Σ^* is an ε-strong Bayes-Nash where $\varepsilon = KB((K-1)B+1) \cdot e(n)$.*

Lemma 12 is a natural extension of Theorem B.4 in Schoenebeck and Tao [9].

Proof. Since we have $\hat{\alpha}_F < \mu$ and $\hat{\alpha}_U < 1-\mu$, we do not need to consider the friendly/unfriendly agent dominating case. Consider a deviating strategy profile Σ', there are two possible cases:

1. $I(\Sigma') < ((K-1)B+1) \cdot e(n)$
2. $I(\Sigma') \geq ((K-1)B+1) \cdot e(n)$

Case 1. In the first case, the error rate of Σ' is low. Since both strategy profiles have high fidelity, we show that agents cannot gain large utility from deviating.

Claim 1. *If $I(\Sigma') < ((K-1)B+1) \cdot e(n)$, then $u_i(\Sigma') - u_i(\Sigma^*) \le \varepsilon$ for all agents.*

We know that $I(\Sigma) = \sum_{k \in \mathcal{L}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma) + \sum_{k \in \mathcal{H}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma)$. Therefore, we have $0 \le \lambda_k^{\mathbf{A}}(\Sigma') \le \frac{((K-1)B+1) \cdot e(n)}{P_k}$ for all $k \in \mathcal{L}$ and $0 \le \lambda_k^{\mathbf{R}}(\Sigma') \le \frac{((K-1)B+1) \cdot e(n)}{P_k}$ for all $k \in \mathcal{H}$. At the same time, since $I(\Sigma^*) \le e(n)$, we have $0 \le \lambda_k^{\mathbf{A}}(\Sigma^*) \le \frac{e(n)}{P_k}$ for all $k \in \mathcal{L}$ and $0 \le \lambda_k^{\mathbf{R}}(\Sigma^*) \le \frac{e(n)}{P_k}$ for all $k \in \mathcal{H}$. Therefore, for every $k \in \mathcal{L}$, we have

$$|\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*)| = |\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)| \le \max(\lambda_k^{\mathbf{A}}(\Sigma'), \lambda_k^{\mathbf{A}}(\Sigma^*)) \le \frac{((K-1)B+1) \cdot e(n)}{P_k}.$$

And for every $k \in \mathcal{H}$, we have

$$|\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)| = |\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*)| \le \max(\lambda_k^{\mathbf{R}}(\Sigma'), \lambda_k^{\mathbf{R}}(\Sigma^*)) \le \frac{((K-1)B+1) \cdot e(n)}{P_k}.$$

Therefore, for any $k \in \mathcal{W}$, $|\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)| \le \frac{((K-1)B+1) \cdot e(n)}{P_k}$. Now for an arbitrary agent i, we consider her gain in deviation:

$$\begin{aligned}
& u_i(\Sigma') - u_i(\Sigma') \\
= & \sum_{k=1}^{K} P_k((\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \cdot v_i(k, \mathbf{A}) + (\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*)) \cdot v_i(k, \mathbf{R})) \\
= & \sum_{k \in \mathcal{L}_i} P_k(v_i(k, \mathbf{R}) - v_i(k, \mathbf{A}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
& + \sum_{k \in \mathcal{H}_i} P_k(v_i(k, \mathbf{A}) - v_i(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*))
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{k\in\mathcal{L}_i} P_k(v_i(k,\mathbf{R})-v_i(k,\mathbf{A}))\frac{((K-1)B+1)\cdot e(n)}{P_k}\\
&\quad+\sum_{k\in\mathcal{H}_i} P_k(v_i(k,\mathbf{A})-v_i(k,\mathbf{R}))\frac{((K-1)B+1)\cdot e(n)}{P_k}\\
&\leq \sum_{k\in\mathcal{L}_i} P_k\cdot B\cdot\frac{((K-1)B+1)\cdot e(n)}{P_k}+\sum_{k\in\mathcal{H}_i} P_k\cdot B\cdot\frac{((K-1)B+1)\cdot e(n)}{P_k}\\
&=KB((K-1)B+1)\cdot e(n)\\
&=\varepsilon.
\end{aligned}
$$

Case 2. In the second case, we show that Σ' cannot be a successful deviating strategy in the following steps:

1. Contingent agents get strictly less utility in Σ' than in Σ^* and thus have no incentive to deviate.

2. For a friendly agent t_1 and an unfriendly agent t_2, if one of them gain more than $\Delta = B\cdot e(n)$ from deviation, the other will get strictly less utility. Therefore, D contains either only friendly agents or only unfriendly agents.

3. D with only one type of pre-determined agents cannot gain utility more than ε.

Claim 2. *If $I(\Sigma')\geq((K-1)B+1)\cdot e(n)$, then $u_i(\Sigma')-u_i(\Sigma^*)<0$ for contingent agents.*

Firstly, we have $A(\Sigma^*)\geq A(\Sigma')+(K-1)B\cdot e(n)$. This is because $I(\Sigma')\geq((K-1)B+1)\cdot e(n)$ and $I(\Sigma^*)\leq e(n)$. Then we consider the utility difference of an arbitrary contingent agent i:

$$
\begin{aligned}
u_i(\Sigma')-u_i(\Sigma^*)=&\sum_{k\in\mathcal{L}} P_k(v_i(k,\mathbf{R})-v_i(k,\mathbf{A}))(\lambda_k^{\mathbf{R}}(\Sigma')-\lambda_k^{\mathbf{R}}(\Sigma^*))\\
&+\sum_{k\in\mathcal{H}} P_k(v_i(k,\mathbf{A})-v_i(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*))
\end{aligned}
$$

Recall that $v_i(k,\mathbf{R})-v_i(k,\mathbf{A})>0$ for all $k\in\mathcal{L}$ and $v_i(k,\mathbf{A})-v_i(k,\mathbf{R})>0$ for all $k\in\mathcal{H}$. We consider two cases:

(1) For all world state k, Σ^* leads to the informed majority decision with a higher probability than Σ'. That is, for all $k \in \mathcal{L}$, $\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*) \leq 0$, and for all $k \in \mathcal{H}$, $\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*) \leq 0$. In this case, there is at least one k such that the inequality is strict. Therefore we have $u_i(\Sigma') - u_i(\Sigma') < 0$.

(2) There exists a world state k in which Σ' leads to the informed majority decision with a higher probability than Σ^*. That is, there exists $\mathcal{L}' \subseteq \mathcal{L}$ and $\mathcal{H}' \subseteq \mathcal{H}$ s.t. $\mathcal{L}' \cup \mathcal{H}' \neq \emptyset$, and for $k \in \mathcal{L}'$, $\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*) > 0$; for $k \in \mathcal{H}'$, $\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*) > 0$. Note that $\mathcal{L}' \cup \mathcal{H}' \neq \mathcal{W}$ always holds, otherwise we will have $A(\Sigma') > A(\Sigma^*)$ which is a contradiction. Then we have

$$
\begin{aligned}
&\sum_{k \in \mathcal{L} \setminus \mathcal{L}'} P_k(\lambda_k^{\mathbf{R}}(\Sigma^*) - \lambda_k^{\mathbf{R}}(\Sigma')) + \sum_{k \in \mathcal{H} \setminus \mathcal{H}'} P_k(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
=& A(\Sigma^*) - A(\Sigma') - \sum_{k \in \mathcal{L}'} P_k(\lambda_k^{\mathbf{R}}(\Sigma^*) - \lambda_k^{\mathbf{R}}(\Sigma')) - \sum_{k \in \mathcal{H}'} P_k(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') \\
>& A(\Sigma^*) - A(\Sigma') \\
\geq& (K-1)B \cdot e(n).
\end{aligned}
$$

At the same time, for all $k \in \mathcal{L}$, we have $\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*) \leq \frac{e(n)}{P_k}$, and for all $k \in \mathcal{H}'$, $\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*) \leq \frac{e(n)}{P_k}$. Therefore, the difference of utility will have

$$
\begin{aligned}
& u_i(\Sigma') - u_i(\Sigma') \\
<& -(K-1)B \cdot e(n) + \sum_{k \in \mathcal{L}' \cup \mathcal{H}'} P_k \cdot \frac{e(n)}{P_k} \cdot B \\
\leq& -(K-1)B \cdot e(n) + (K-1)B \cdot e(n) \\
=& 0.
\end{aligned}
$$

In the second line, the first term comes from the sum of all terms for $k \in \mathcal{W} \setminus (\mathcal{L}' \cup \mathcal{H}')$:

$$
\begin{aligned}
&\sum_{k\in\mathcal{L}\setminus\mathcal{L}'} P_k(\lambda_k^{\mathbf{R}}(\Sigma') - \lambda_k^{\mathbf{R}}(\Sigma^*))(v_i(k,\mathbf{R}) - v_i(k,\mathbf{A})) \\
&\quad + \sum_{k\in\mathcal{H}\setminus\mathcal{H}'} P_k(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*))(v_i(k,\mathbf{A}) - v_i(k,\mathbf{R})) \\
=& - \sum_{k\in\mathcal{L}\setminus\mathcal{L}'} P_k(\lambda_k^{\mathbf{R}}(\Sigma^*) - \lambda_k^{\mathbf{R}}(\Sigma'))(v_i(k,\mathbf{R}) - v_i(k,\mathbf{A})) \\
&\quad - \sum_{k\in\mathcal{H}\setminus\mathcal{H}'} P_k(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma'))(v_i(k,\mathbf{A}) - v_i(k,\mathbf{R})) \\
\leq& - \sum_{k\in\mathcal{L}\setminus\mathcal{L}'} P_k(\lambda_k^{\mathbf{R}}(\Sigma^*) - \lambda_k^{\mathbf{R}}(\Sigma')) \cdot 1 - \sum_{k\in\mathcal{H}\setminus\mathcal{H}'} P_k(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \cdot 1 \\
<& - (K-1)B \cdot e(n).
\end{aligned}
$$

And the second term comes from the sum of all terms for $k \in \mathcal{L}' \cup \mathcal{H}'$. Note that since $\mathcal{L}' \cup \mathcal{H}' \subsetneq \mathcal{W}$, we have $|\mathcal{L}' \cup \mathcal{H}'| \leq K - 1$.

After excluding contingent agents from the deviating group, we show that any deviating group D cannot contain both friendly and unfriendly agents.

Claim 3. *Suppose i_1 is an arbitrary friendly agent, and i_2 is an arbitrary unfriendly agent. For $\Delta = K \cdot B(B+1) \cdot e(n)$, we have*

1. *If $u_{i_1}(\Sigma') - u_{i_1}(\Sigma^*) > \Delta$, $u_{i_2}(\Sigma') - u_{i_2}(\Sigma^*) < 0$.*

2. *If $u_{i_2}(\Sigma') - u_{i_2}(\Sigma^*) > \Delta$, $u_{i_1}(\Sigma') - u_{i_1}(\Sigma^*) < 0$.*

We will only prove (1) since the reasoning of (1) and (2) are similar. Without loss of generality, let $i_1 = 1$ and $i_2 = 2$. Suppose $u_1(\Sigma') - u_1(\Sigma^*) > \Delta$, we'll show that $u_2(\Sigma') - u_2(\Sigma^*) < 0$. According to the definition of friendly and unfriendly agents, we have $\mathcal{L}_1 \subsetneq \mathcal{L} \subsetneq \mathcal{L}_2$, and $\mathcal{H}_2 \subsetneq \mathcal{H} \subsetneq \mathcal{H}_1$.

We'll mainly use three facts in the proof:

1. for all $k \in \mathcal{L}$, $\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') \leq \frac{e(n)}{P_k} - \lambda_k^{\mathbf{A}}(\Sigma') \leq \frac{e(n)}{P_k}$; for all $k \in \mathcal{H}$, $\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*) \leq \lambda_k^{\mathbf{A}}(\Sigma') - (1 - \frac{e(n)}{P_k}) \leq \frac{e(n)}{P_k}$.

2. for both $i = 1$ and $i = 2$, $v_i(k, \mathbf{A}) - v_i(k, \mathbf{R})$ is negative for all $k \in \mathcal{L}_1$, and positive for all $k \in \mathcal{H}_2$.

3. for both $i = 1$ and $i = 2$, $v_i(k, \mathbf{A}) - v_i(k, \mathbf{R})$ is increasing in k.

We first consider the difference of 1's expected utility:

$$\Delta < u_1(\Sigma') - u_1(\Sigma^*) = \sum_{k=1}^{K} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*))$$

We can categorize k into three parts:

- $k \in \mathcal{L}_1$. For these k, $v_i(k, \mathbf{A}) - v_i(k, \mathbf{R}) < 0$ and $\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') \leq \frac{e(n)}{P_k}$. Therefore, we have

$$\begin{aligned}
&\sum_{k \in \mathcal{L}_1} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
=&\sum_{k \in \mathcal{L}_1} P_k(v_1(k, \mathbf{R}) - v_1(k, \mathbf{A}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
\leq&\sum_{k \in \mathcal{L}_1} P_k \cdot B \cdot \frac{e(n)}{P_k} \\
=&\sum_{k \in \mathcal{L}_1} B \cdot e(n).
\end{aligned}$$

- $k \in \mathcal{H}_2$. For these k, $v_i(k, \mathbf{A}) - v_i(k, \mathbf{R}) > 0$ and $\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*) \leq \frac{e(n)}{P_k}$. Therefore, we have

$$\begin{aligned}
&\sum_{k \in \mathcal{H}_2} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
\leq&\sum_{k \in \mathcal{H}_2} P_k \cdot B \cdot \frac{e(n)}{P_k} \\
=&\sum_{k \in \mathcal{H}_2} B \cdot e(n).
\end{aligned}$$

- $k \in \mathcal{W} \setminus (\mathcal{L}_1 \cup \mathcal{H}_2)$. Note that this part can be divided into two parts: $k \in \mathcal{L} \cap \mathcal{H}_1$ and $\mathcal{L}_2 \cap \mathcal{H}$. We'll deal with this part later.

Therefore,

$$\begin{aligned}
\Delta <& u_1(\Sigma^*) - u_1(\Sigma') \\
=& \sum_{k=1}^{K} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
\leq& \sum_{k \in \mathcal{W} \setminus (\mathcal{L}_1 \cup \mathcal{H}_2)} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) + \sum_{k \in \mathcal{L}_1 \cup \mathcal{H}_2} B \cdot e(n) \\
=& \sum_{k \in \mathcal{L} \cap \mathcal{H}_1} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + \sum_{k \in \mathcal{L}_1 \cup \mathcal{H}_2} e(n) \cdot B.
\end{aligned}$$

Now we start deal with the first two terms with fact 1 and 3.

1. For the first term $\sum_{k \in \mathcal{L} \cap \mathcal{H}_1} P_k(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*))$, we have $(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R})) > 0$ and $(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \geq -\frac{e(n)}{P_k}$. we want to replace the k in v_1 into L, which will make $(v_1(k, \mathbf{A}) - v_1(k, \mathbf{R}))$ larger. At the same time, we want the whole term to become larger as well. Therefore, we add some term into $(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*))$

as follows:

$$
\begin{aligned}
&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(k,\mathbf{A})-v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*))\\
=&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(k,\mathbf{A})-v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*)+\frac{e(n)}{P_k})\\
&-\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(k,\mathbf{A})-v_1(k,\mathbf{R}))\cdot\frac{e(n)}{P_k}\\
\le&\sum_{k\in\mathcal{L}\cup\mathcal{H}_1} P_k(v_1(L,\mathbf{A})-v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*)+\frac{e(n)}{P_k})\\
&-\sum_{k\in\mathcal{L}\cap\mathcal{H}_1}(v_1(k,\mathbf{A})-v_1(k,\mathbf{R}))\cdot e(n)\\
=&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(L,\mathbf{A})-v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*))\\
&+\sum_{k\in\mathcal{L}\cap\mathcal{H}_1}(((v_1(L,\mathbf{A})-v_1(L,\mathbf{R}))-(v_1(k,\mathbf{A})-v_1(k,\mathbf{R})))\cdot e(n)\\
\le&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(L,\mathbf{A})-v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma')-\lambda_k^{\mathbf{A}}(\Sigma^*))+\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} B\cdot e(n).
\end{aligned}
$$

2. For the second term $-\sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_1(k,\mathbf{A})-v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*)-\lambda_k^{\mathbf{A}}(\Sigma'))$, we use sim-

ilar technique to replace k into L:

$$
\begin{aligned}
&- \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(k,\mathbf{A}) - v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
=& - \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(k,\mathbf{A}) - v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') + \frac{e(n)}{P_k}) \\
&+ \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} (v_1(k,\mathbf{A}) - v_1(k,\mathbf{R})) \cdot e(n) \\
\leq& - \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') + \frac{e(n)}{P_k}) \\
&+ \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} (v_1(k,\mathbf{A}) - v_1(k,\mathbf{R})) \cdot e(n) \\
=& - \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
&+ \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} (((v_1(k,\mathbf{A}) - v_1(k,\mathbf{R})) - (v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))) \cdot e(n) \\
\leq& - \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + \sum_{k \in \mathcal{L}_2 \cap \mathcal{H}} B \cdot e(n).
\end{aligned}
$$

Note that the first inequality comes from fact 3 that $0 < v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}) \leq v_1(k,\mathbf{A}) - v_1(k,\mathbf{R})$ for all $k \in \mathcal{L}_2 \cap \mathcal{H}$.

Now we put everything together and get

$$
\begin{aligned}
\Delta <& u_1(\Sigma') - u_1(\Sigma^*) \\
\leq& \sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(k,\mathbf{A}) - v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_1(k,\mathbf{A}) - v_1(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + \sum_{k\in\mathcal{L}_1\cup\mathcal{H}_2} B\cdot e(n) \\
\leq& \sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
&+ \sum_{k\in\mathcal{L}_1\cup\mathcal{H}_2} B\cdot e(n) + \sum_{k\in\mathcal{L}\cup\mathcal{H}_1} B\cdot e(n) + \sum_{k\in\mathcal{L}_2\cup\mathcal{H}} B\cdot e(n) \\
=& \sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + K\cdot B\cdot e(n).
\end{aligned}
$$

Note that $B \geq v_1(L,\mathbf{A}) - v_1(L,\mathbf{R}) > 0$. Therefore, we have

$$
\begin{aligned}
&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) - \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) \\
>& \frac{\Delta - K\cdot B\cdot e(n)}{B} \\
=& \frac{K\cdot B(B+1)\cdot e(n) - K\cdot B\cdot e(n)}{B} \\
=& K\cdot B\cdot e(n)
\end{aligned}
$$

Now we consider agent 2's side. We show that $u_2(\Sigma') - u_2(\Sigma^*) < 0$. Most calculations will be the same as those of agent 1. Note that one main difference is for agent 2, $v_2(k,\mathbf{A}) - v_2(k,\mathbf{R}) < 0$ for all $k \in \mathcal{W} \setminus (\mathcal{L}_1 \cup \mathcal{H}_2)$.

$$
\begin{aligned}
&u_2(\Sigma') - u_2(\Sigma^*) \\
=&\sum_{k=1}^{K} P_k(v_2(k,\mathbf{A}) - v_2(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
\leq&\sum_{k\in\mathcal{W}\setminus(\mathcal{L}_1\cup\mathcal{H}_2)} P_k(v_2(k,\mathbf{A}) - v_2(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) + \sum_{k\in\mathcal{L}_1\cup\mathcal{H}_2} B\cdot e(n) \\
=&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_2(k,\mathbf{A}) - v_2(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_2(k,\mathbf{A}) - v_2(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + \sum_{k\in\mathcal{L}_1\cup\mathcal{H}_2} B\cdot e(n) \\
\leq&\sum_{k\in\mathcal{L}\cap\mathcal{H}_1} P_k(v_2(L,\mathbf{A}) - v_2(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)) \\
&- \sum_{k\in\mathcal{L}_2\cap\mathcal{H}} P_k(v_2(L,\mathbf{A}) - v_2(L,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma')) + K\cdot B\cdot e(n) \\
<&(v_2(L,\mathbf{A}) - v_2(L,\mathbf{R}))\cdot K\cdot B\cdot e(n) + K\cdot B\cdot e(n) \\
\leq& -1\cdot K\cdot B\cdot e(n) + K\cdot B\cdot e(n) \\
=&0.
\end{aligned}
$$

Therefore, we show that $u_2(\Sigma') - u_2(\Sigma^*) < 0$, which implies the claim.

Remark 2. *Note that in the binary setting, this part of proof can be largely simplified. Note that $\mathcal{L}_1 = \mathcal{H}_2 = \emptyset$, $\mathcal{L}\cup\mathcal{H}_1 = \{L\}$, and $\mathcal{H}\cup\mathcal{L}_2 = \{H\}$.*

Now we come to the final step of Case 2. We'll show that a deviating group D that contains only friendly or only unfriendly agents cannot gain utility more than ε.

Claim 4. *If D contains only friendly agents (or only unfriendly agents), then for every $i \in D$, $u_i(\Sigma') - u_i(\Sigma^*) < \varepsilon$.*

Without loss of generality, suppose D contains only friendly agents. The calculation for D containing only unfriendly agents will be the same. Note that in Σ^* every friendly agent always votes for $\mathbf{A}$. Therefore, when friendly agents in D deviate, the probability of $\mathbf{A}$ being the winner will decrease. Formally, for any $k = 1, 2, \cdots, K$, we have $\lambda_k^{\mathbf{A}}(\Sigma') \leq \lambda_k^{\mathbf{A}}(\Sigma^*)$. Now we consider the difference in the expected utility of an arbitrary friendly agent in D:

$$\begin{aligned}u_i(\Sigma') - u_i(\Sigma^*) = &\sum_{k\in\mathcal{L}_i} P_k(v_i(k,\mathbf{R}) - v_i(k,\mathbf{A}))(\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma'))\\ &+ \sum_{k\in\mathcal{H}_i} P_k(v_i(k,\mathbf{A}) - v_i(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma') - \lambda_k^{\mathbf{A}}(\Sigma^*)).\end{aligned}$$

Note that for all $k \in \mathcal{H}_i$, we have $v_i(k,\mathbf{A}) - v_i(k,\mathbf{R}) > 0$. Therefore, the second sum is non-positive. At the same time, since $\mathcal{L}_i \subsetneq \mathcal{L}$, we have $\lambda_k^{\mathbf{A}}(\Sigma^*) - \lambda_k^{\mathbf{A}}(\Sigma') \leq \lambda_k^{\mathbf{A}}(\Sigma^*) \leq \frac{e(n)}{P_k}$ and $v_i(k,\mathbf{R}) - v_i(k,\mathbf{A}) > 0$. Therefore,

$$u_i(\Sigma') - u_i(\Sigma^*) \leq \sum_{k\in\mathcal{L}_i} P_k \cdot B \cdot \frac{e(n)}{P_k} < K \cdot B \cdot e(n) < \varepsilon.$$

Therefore, we show that Σ^* is an ε-strong BNE where $\varepsilon = KB((K-1)B+1)\cdot e(n)$. □

A.3.2 Theorem 6 (Theorem 3 for binary setting)

Then we show that there always exists a sequence of regular strategy profiles with high fidelity.

Theorem 6. *For an arbitrary sequence of instances, there exists a series of regular strategy profiles $\{\Sigma'_n\}_{n=1}^{\infty}$ and constants $n_0 > 0$, $\phi > 0$ such that for all $n > n_0$, $A(\Sigma'_n) \geq 1 - 2\exp(-2\phi^2 n)$.*

Proof. Let $B(n,p)$ denote a random variable of binomial distribution with n experiments and probability p. Given n and a regular strategy profile Σ_n where all contingent agents play strategy $\sigma = (\beta_1, \beta_2, \cdots, \beta_m)$, we can write $\lambda_k^{\mathbf{A}}$ and $\lambda_k^{\mathbf{R}}$ for each k.

$$\begin{aligned}
\lambda_k^{\mathbf{A}}(\Sigma_n) =& \Pr[\#\text{number of votes for } \mathbf{A} \geq \mu n \mid W = k] \\
=& \Pr[n_F + B(n_C, \sum_{m=1}^{M} P_{mk} \cdot \beta_m) \geq \mu n] \\
=& \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot \beta_m) - n_C \sum_{m=1}^{M} P_{mk} \cdot \beta_m \geq -f_{k\mathbf{A}}^n(\sigma) \cdot n],
\end{aligned}$$

where we define

$$\begin{aligned}
f_{k\mathbf{A}}^n(\sigma) =& \frac{n_F}{n} + \frac{n_C}{n} \sum_{m=1}^{M} P_{mk} \cdot \beta_m - \mu \\
=& \hat{\alpha}_F + \hat{\alpha}_C \sum_{m=1}^{M} P_{mk} \cdot \beta_m - \mu
\end{aligned}$$

Similarly,

$$\lambda_k^{\mathbf{R}}(\Sigma_n) = \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m)) - n_C \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m) \geq -f_{k\mathbf{R}}^n(\sigma) \cdot n],$$

where

$$\begin{aligned}
f_{k\mathbf{R}}^n(\sigma) =& \frac{n_U}{n} + \frac{n_C}{n} \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m) - (1 - \mu) \\
=& \hat{\alpha}_U + \hat{\alpha}_C \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m) - (1 - \mu)
\end{aligned}$$

If we can find a σ', and constant $n_0 > 0$, $\phi > 0$ s.t. for every $n > n_0$, for every $k \in \mathcal{H}$ we have $f_{k\mathbf{A}}^n(\sigma') > \phi$ and for every $k \in \mathcal{L}$ we have $f_{k\mathbf{R}}^n(\sigma') > \phi$, then we can directly apply the Hoeffding Inequality and show that for every $n > n_0$, Σ_n' has high fidelity.

Now we start constructing strategy σ'. First we define the "approximated version" of $f_{k\mathbf{A}}^n$

and $f^n_{k\mathbf{R}}$ which are independent from n. Given a strategy σ, let

$$\hat{f}_{k\mathbf{A}}(\sigma) = \alpha_F + \alpha_C \sum_{m=1}^{M} P_{mk} \cdot \beta_m - \mu \tag{A.2}$$

$$\hat{f}_{k\mathbf{R}}(\sigma) = \alpha_U + \alpha_C \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m) - (1 - \mu). \tag{A.3}$$

The rest of the proof will go as follows:

1. First, we construct a strategy σ' s.t. for every $k \in \mathcal{H}$ we have $\hat{f}_{k\mathbf{A}}(\sigma') > 0$ and for every $k \in \mathcal{L}$ we have $\hat{f}_{k\mathbf{R}}(\sigma') > 0$.

2. Then we find out n_0 and show that for every $n > n_0$, $f^n_{k\mathbf{A}}(\sigma')$ will not deviate from $\hat{f}_{k\mathbf{A}}(\sigma')$ too much, thus will larger than some constant ϕ. Similarly we show that for every $n > n_0$, $f^n_{k\mathbf{A}}(\sigma') > \phi$.

3. Finally, we show that for all $n > n_0$, $A(\Sigma'_n) \geq 1 - 2\exp(-2\phi^2 T)$ by applying the Hoeffding inequality.

Part 1. Constructing σ'. The construction of σ has three steps. We start from a simple strategy and then modify it to the target step by step.

Step 1. Consider a simple strategy $\sigma_0 = (\beta^*, \beta^*, \cdots, \beta^*)$. That is, the agent always votes for $\mathbf{A}$ with probability β^* and votes for $\mathbf{R}$ with probability $1 - \beta^*$ no matter what signal she receives. β^* satisfies $\alpha_F + \alpha_C \cdot \beta^* - \mu = 0$. Since $\alpha_F < \mu$ and $\alpha_U < 1 - \mu$, we have $\beta^* \in (0, 1)$. Moreover, we have $\hat{f}_{k\mathbf{A}}(\sigma_0) = 0$ and $\hat{f}_{k\mathbf{R}}(\sigma_0) = 0$ for all $k \in \mathcal{W}$.

Step 2. Let $m^* = \lfloor \frac{M}{2} \rfloor$. m^* will serve as a threshold, as we regard signals not larger than m^* as "low signals" and those exceeding m^* as "high signals". Specifically, for a world state k, define $P_{lk} = \sum_{m=1}^{m^*} P_{mk}$, and $P_{hk} = \sum_{m=m^*+1}^{M} P_{mk}$. According to the stochastic dominance assumption, we have $0 \leq P_{lk_1} < P_{lk_2}$ and $P_{hk_1} > P_{hk_2} \geq 0$ for all $k_1 > k_2$. Let $\beta_l = \beta^* - \delta_l$, and $\beta_h = \beta^* + \delta_h$, where $\delta_l > 0$ and $\delta_h > 0$ are constant satisfying $\delta_h = \delta_l \cdot \frac{P_{lH}}{P_{hH}}$. We can carefully select and fix δ_h and δ_l to make β_h and β_l inside $(0, 1)$. Then we construct strategy σ_1: if the agent receives a low signal $m \leq m^*$, she votes for $\mathbf{A}$ with probability β_l and $\mathbf{R}$

with probability $1-\beta_l$; if she receives a high signal $m > m^*$, she votes for **A** with probability β_h and **R** with probability $1-\beta_h$. Then we show that

- for all $k \in \mathcal{H}$, $\hat{f}_{k\mathbf{A}}(\sigma_1) \geq 0$, and
- for all $k \in \mathcal{L}$, $\hat{f}_{k\mathbf{R}}(\sigma_1) > 0$.

For the $\mathcal{H}$ side, we'll first show that $\hat{f}_{H\mathbf{A}}(\sigma_1) = 0$. Then we show that for all $k \in \mathcal{H}$, $\hat{f}_{k\mathbf{A}}(\sigma_1) \geq \hat{f}_{H\mathbf{A}}(\sigma_1)$. For $k = H$, we have

$$\begin{aligned}
\hat{f}_{H\mathbf{A}}(\sigma_1) =& \alpha_F + \alpha_C(P_{lH} \cdot \beta_l + P_{hH} \cdot \beta_h) - \mu \\
=& \alpha_F + \alpha_C(P_{lH} \cdot (\beta^* - \delta_l) + P_{hH} \cdot (\beta^* + \delta_h)) - \mu \\
=& \alpha_C \cdot (-P_{lH} \cdot \delta_l + P_{hH} \cdot \delta_l \cdot \frac{P_{lH}}{P_{hH}}) \\
=& 0.
\end{aligned}$$

For other $k \in \mathcal{H}$, because $\beta_l < \beta_h$, $P_{lk} \leq P_{lH}$,and $P_{hk} \geq P_{hH}$, we have $\hat{f}_{k\mathbf{A}}(\sigma_1) \geq \hat{f}_{H\mathbf{A}}(\sigma_1) \geq 0$.

For the $\mathcal{L}$ side, similarly, we'll first show that $\hat{f}_{L\mathbf{R}}(\sigma_1) > 0$. Then we show that for all $k \in \mathcal{L}$, $\hat{f}_{k\mathbf{R}}(\sigma_1) > \hat{f}_{L\mathbf{R}}(\sigma_1)$. For $k = L$ we have

$$\begin{aligned}
\hat{f}_{L\mathbf{R}}(\sigma_1) =& \alpha_U + \alpha_C(P_{lL} \cdot (1 - \beta_l) + P_{hL} \cdot (1 - \beta_h)) - (1 - \mu) \\
=& \alpha_U + \alpha_C(P_{lL} \cdot (1 - \beta^* + \delta_l) + P_{hL} \cdot (1 - \beta^* - \delta_h)) - (1 - \mu) \\
=& \alpha_C \cdot (P_{lL} \cdot \delta_l - P_{hL} \cdot \delta_l \cdot \frac{P_{lH}}{P_{hH}}) \\
=& \frac{\alpha_C \cdot \delta_l}{P_{hH}}(P_{lL} \cdot P_{hH} - P_{hL} \cdot P_{lH}) \\
>& 0.
\end{aligned}$$

The final inequality comes that $P_{lH} < P_{lL}$,and $P_{hH} > P_{hL}$ since $L < H$. For other $k \in \mathcal{L}$, because $(1 - \beta_l) > (1 - \beta_h)$, $P_{lk} \geq P_{lL}$,and $P_{hk} \leq P_{hL}$, we have $\hat{f}_{k\mathbf{A}}(\sigma_1) \geq \hat{f}_{L\mathbf{A}}(\sigma_1) > 0$.

Remark 3. *In the binary setting, we just let* $\sigma_1 = (\beta_l, \beta_h)$*. And we will find that* $\hat{f}_{H\mathbf{A}}(\sigma_1) = 0$ *and* $\hat{f}_{L\mathbf{R}}(\sigma_1) > 0$*.*

Step 3. In step 3, we modify β_h to make both sides strictly positive. Let $\beta'_l = \beta_l$, and $\beta'_h = \beta_h + \delta'_h$, where $\delta'_h > 0$ is a constant satisfying $\alpha_C \cdot \delta'_h \leq \frac{\hat{f}_{L\mathbf{R}}(\sigma_1)}{2}$ and $\beta_h + \delta'_h \leq 1$. Therefore, $0 \leq \beta'_l < \beta'_h \leq 1$. Now we are ready to construct our final strategy σ': if the agent receives a low signal $m \leq m^*$, she votes for **A** with probability $\beta'_m = \beta'_l$ and **R** with probability $1 - \beta'_l$; if she receives a high signal $m > m^*$, she votes for **A** with probability $\beta'_m = \beta'_h$ and **R** with probability $1 - \beta'_h$. Then we show that

- for all $k \in \mathcal{H}$, $\hat{f}_{k\mathbf{A}}(\sigma') > 0$, and
- for all $k \in \mathcal{L}$, $\hat{f}_{k\mathbf{R}}(\sigma') > 0$.

For the $\mathcal{H}$ side, we still consider H first:

$$
\begin{aligned}
\hat{f}_{H\mathbf{A}}(\sigma') =& \alpha_F + \alpha_C(P_{lH} \cdot \beta'_l + P_{hH} \cdot \beta'_h) - \mu \\
=& \alpha_F + \alpha_C(P_{lH} \cdot \beta_l + P_{hH} \cdot (\beta_h + \delta'_h)) - \mu \\
=& \hat{f}_{H\mathbf{A}}(\sigma_1) + \alpha_C \cdot P_{hH} \cdot \delta'_h \\
=& \alpha_C \cdot P_{hH} \cdot \delta'_h \\
>& 0.
\end{aligned}
$$

For other $k \in \mathcal{H}$, because $\beta'_l < \beta'_h$, $P_{lk} \leq P_{lH}$,and $P_{hk} \geq P_{hH}$, we have $\hat{f}_{k\mathbf{A}}(\sigma') \geq \hat{f}_{H\mathbf{A}}(\sigma') > 0$.

For the $\mathcal{L}$ side, we consider L first:

$$
\begin{aligned}
\hat{f}_{L\mathbf{R}}(\sigma') =& \alpha_U + \alpha_C(P_{lL} \cdot (1-\beta'_l) + P_{hL} \cdot (1-\beta'_h)) - (1-\mu) \\
& \alpha_U + \alpha_C(P_{lL} \cdot (1-\beta_l) + P_{hL} \cdot (1-\beta_h - \delta'_h)) - (1-\mu) \\
=& \hat{f}_{L\mathbf{R}}(\sigma') - \alpha_C \cdot P_{hL} \cdot \delta'_h \\
\geq& \frac{\hat{f}_{L\mathbf{R}}(\sigma')}{2} \\
>& 0.
\end{aligned}
$$

For other $k \in \mathcal{L}$, because $(1-\beta'_l) > (1-\beta'_h)$, $P_{lk} \geq P_{lL}$,and $P_{hk} \leq P_{hL}$, we have $\hat{f}_{k\mathbf{A}}(\sigma') \geq \hat{f}_{L\mathbf{A}}(\sigma') > 0$.

Remark 4. *In the binary setting, similarly, let* $\sigma' = (\beta'_l, \beta'_h)$*. We will find that* $\hat{f}_{H\mathbf{A}}(\sigma') > 0$ *and* $\hat{f}_{L\mathbf{R}}(\sigma') > 0$

Part 2 Determine n_0 and ϕ. In this part, we'll show that for all $k \in \mathcal{H}$, ($k \in \mathcal{L}$, respectively), $f^n_{k\mathbf{A}}(\sigma')$ ($f^n_{k\mathbf{R}}(\sigma')$, respectively) will not deviate from $\hat{f}_{k\mathbf{A}}(\sigma')$ ($\hat{f}_{k\mathbf{R}}(\sigma')$, respectively) too much. Since $\hat{f}_{k\mathbf{A}}(\sigma')$ and $\hat{f}_{k\mathbf{R}}(\sigma')$ do not depend on n, we can determine constant n_0 and ϕ, s.t. for all $n > n_0$, $f^n_{k\mathbf{A}}(\sigma') > \phi$, and $f^n_{k\mathbf{R}}(\sigma') > \phi$. We start from the $\mathcal{H}$ side. Recall that

$$
f^n_{k\mathbf{A}}(\sigma) = \frac{n_F}{n} + \frac{n_C}{n} \sum_{m=1}^{M} P_{mk} \cdot \beta_m - \mu.
$$

And recall that $n_F = n - \lfloor (1-\alpha_F) \cdot n \rfloor$, $n_U = \lfloor \alpha_U \cdot n \rfloor$, and $n_C = \lfloor (1-\alpha_F) \cdot n \rfloor - \lfloor \alpha_U \cdot n \rfloor$.

Therefore, for all $k \in \mathcal{H}$, we have

$$\begin{aligned}
f^n_{k\mathbf{A}}(\sigma') =& 1 - \frac{\lfloor (1-\alpha_F)\cdot n \rfloor}{n} + \frac{\lfloor (1-\alpha_F)\cdot n \rfloor - \lfloor \alpha_U \cdot n \rfloor}{n} \sum_{m=1}^{M} P_{mk} \cdot \beta'_m - \mu \\
\geq & 1 - (1-\alpha_F) + ((1-\alpha_F) - \frac{1}{n} - \alpha_U) \sum_{m=1}^{M} P_{mk} \cdot \beta'_m - \mu \\
=& \alpha_F + \alpha_C \cdot \sum_{m=1}^{M} P_{mk} \cdot \beta'_m - \mu - \frac{1}{n} \sum_{m=1}^{M} P_{mk} \cdot \beta'_m \\
\geq & \hat{f}_{k\mathbf{A}}(\sigma') - \frac{1}{n} \\
\geq & \hat{f}_{H\mathbf{A}}(\sigma') - \frac{1}{n}
\end{aligned}$$

Since $\hat{f}_{H\mathbf{A}}(\sigma') > 0$ is independent from n, there exists a $n_{\mathcal{H}} > 0$, s.t. for all $n > n_{\mathcal{H}}$ and all $k \in \mathcal{H}$, $f^n_{k\mathbf{A}}(\sigma') \geq \frac{\hat{f}_{H\mathbf{A}}(\sigma')}{2}$.

For the $\mathcal{L}$ side, similarly for all $k \in \mathcal{L}$, we have

$$\begin{aligned}
f^n_{k\mathbf{R}}(\sigma') =& \frac{n_U}{n} + \frac{n_C}{n} \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) - (1-\mu) \\
=& \frac{\lfloor \alpha_U \cdot n \rfloor}{n} + \frac{\lfloor (1-\alpha_F)\cdot n \rfloor - \lfloor \alpha_U \cdot n \rfloor}{n} \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) - (1-\mu) \\
\geq & \alpha_U - \frac{1}{n} + ((1-\alpha_F) - \frac{1}{n} - \alpha_U) \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) - (1-\mu) \\
=& \alpha_U + \alpha_C \cdot \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) - (1-\mu) - \frac{1}{n}(1 + \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m)) \\
\geq & \hat{f}_{k\mathbf{R}}(\sigma') - \frac{2}{n} \\
\geq & \hat{f}_{L\mathbf{R}}(\sigma') - \frac{2}{n}.
\end{aligned}$$

Since $\hat{f}_{L\mathbf{R}}(\sigma') > 0$ is independent from n, there exists a $n_{\mathcal{L}} > 0$, s.t. for all $n > n_{\mathcal{L}}$ and all $k \in \mathcal{L}$, $f^n_{k\mathbf{R}}(\sigma') \geq \frac{\hat{f}_{L\mathbf{R}}(\sigma')}{2}$.

Remark 5. *This part is different in the binary setting due to the different rounding method. However, we can still show that $f^n_{H\mathbf{A}}(\sigma') \geq \hat{f}_{H\mathbf{A}}(\sigma') - \frac{2}{n}$ and $f^n_{L\mathbf{R}}(\sigma') \geq \hat{f}_{L\mathbf{R}}(\sigma') - \frac{2}{n}$ by the same reasoning.*

Therefore, putting $\mathcal{H}$ and $\mathcal{L}$ sides together, let $\phi = \min(\frac{\hat{f}_{H\mathbf{A}}(\sigma')}{2}, \frac{\hat{f}_{L\mathbf{R}}(\sigma')}{2})$, and $n_0 = \max(n_{\mathcal{H}}, n_{\mathcal{L}})$, we have for all $n > n_0$, for all $k \in \mathcal{H}$, $f^n_{k\mathbf{A}}(\sigma') \geq \phi$, and all $k \in \mathcal{L}$, $f^n_{k\mathbf{R}}(\sigma') \geq \phi$.

Part 3. Bound the fidelity. Now we come back to the fidelity and bound it by the Hoeffding Inequality. For all $n > n_0$, for the $\mathcal{H}$ side we have

$$\begin{aligned}
\lambda_k^{\mathbf{A}}(\Sigma'_n) =& \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot \beta'_m) - n_C \sum_{m=1}^{M} P_{mk} \cdot \beta'_m \geq -f^n_{k\mathbf{A}}(\sigma) \cdot n] \\
\geq& \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot \beta'_m) - n_C \sum_{m=1}^{M} P_{mk} \cdot \beta'_m \geq -\phi \cdot n] \\
\geq& 1 - 2\exp(-2\phi^2 n).
\end{aligned}$$

Similarly, for the $\mathcal{L}$ side we have

$$\begin{aligned}
\lambda_k^{\mathbf{R}}(\Sigma'_n) =& \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m)) - n_C \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) \geq -f^n_{k\mathbf{R}}(\sigma) \cdot n] \\
\geq& \Pr[B(n_C, \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m)) - n_C \sum_{m=1}^{M} P_{mk} \cdot (1-\beta'_m) \geq -\phi \cdot n] \\
\geq& 1 - 2\exp(-2\phi^2 n).
\end{aligned}$$

Therefore for all $n > n_0$,

$$\begin{aligned}
A(\Sigma'_n) =& \sum_{k \in \mathcal{L}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma'_n) + \sum_{k \in \mathcal{H}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma'_n) \\
\geq& \sum_{k \in \mathcal{L}} P_k(1 - 2\exp(-2\phi^2 n)) + \sum_{k \in \mathcal{H}} P_k(1 - 2\exp(-2\phi^2 n)) \\
=& 1 - 2\exp(-2\phi^2 n).
\end{aligned}$$

□

A.3.3 Theorem 5 (Theorem 2 for binary setting)

Then we give the theorem of the equivalence between fidelity and strong equilibrium.

Theorem 5. *Given an arbitrary sequence of instances and an arbitrary sequence of regular strategy profiles* $\{\Sigma_n\}$ *, let* $\{A(\Sigma_n)\}_{n=1}^{\infty}$ *be the sequence of fidelity of* Σ*.*

- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$*, for every* $n > n_\mu$*,* Σ_n *is an* ε*-strong BNE with* $\varepsilon = o(1)$*.*

- *If* $\lim_{n\to\infty} A(\Sigma_n) = 1$ *does NOT hold, there exist infinitely many* n *such that* Σ_n *is not an* ε*-strong BNE with constant* ε

Proof. Let $e(n) = (1 - A(\Sigma_n))$, and we can directly apply Lemma 12, and get that for every n, Σ_n is an ε-strong BNE where $\varepsilon = KB((K-1)B+1)\cdot(1-A(\Sigma_n))$.

Case 1: $\lim_{n\to\infty} A(\Sigma_n) = 1$. In this case, since $A(\Sigma_n)$ converges to 1, ε, which is positive proportion to $1 - A(\Sigma_n)$, will converge to 0 as $n \to \infty$.

Case 2: $\lim_{n\to\infty} A(\Sigma_n) = 1$ does not hold. Then there exists a constant $\delta > 0$ and an infinite set $\mathcal{N} \subseteq \mathbb{N}$, s.t. for all $n \in \mathcal{N}$, $A(\Sigma_n) \le 1-\delta$. From Theorem 6 we know that we can construct a series of symmetric deviating strategies $\{\Sigma'_n\}$, and find constants $n_0 > 0, \phi > 0$, s.t. for all $n > n_0$, $A(\Sigma'_n) \le 1 - 2\exp(-2\phi^2 n)$. Note that in both $\{\Sigma_n\}$ and $\{\Sigma'_n\}$ every friendly (unfriendly, respectively) agent always votes for $\mathbf{A}$ ($\mathbf{R}$, respectively). Therefore, the deviating group contains only contingent agents, and we only need to consider the expected utility of contingent agents. Given an arbitrary $(n > n_0) \wedge (n \in \mathcal{N})$, recall the difference of expected utility for an agent between Σ_n and Σ'_n:

$$\begin{aligned} u_i(\Sigma'_n) - u_i(\Sigma_n) = &\sum_{k\in\mathcal{L}} P_k(v_i(k,\mathbf{R}) - v_i(k,\mathbf{A}))(\lambda_k^{\mathbf{R}}(\Sigma'_n) - \lambda_k^{\mathbf{R}}(\Sigma_n)) \\ &+ \sum_{k\in\mathcal{H}} P_k(v_i(k,\mathbf{A}) - v_i(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma'_n) - \lambda_k^{\mathbf{A}}(\Sigma_n)). \end{aligned}$$

And recall the definition of fidelity:

$$A(\Sigma) = \sum_{k\in\mathcal{L}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma) + \sum_{k\in\mathcal{H}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma).$$

For Σ'_n we have $A(\Sigma'_n) \ge 1 - 2\exp(-2\phi^2 n)$. Therefore, for every $k \in \mathcal{H}$ we have $\lambda_k^{\mathbf{A}}(\Sigma'_n) \ge 1 - \frac{2\exp(-2\phi^2 n)}{P_k}$, and for every $k \in \mathcal{L}$, $\lambda_k^{\mathbf{R}}(\Sigma'_n) \ge 1 - \frac{2\exp(-2\phi^2 n)}{P_k}$. For Σ_n, on the other hand, we have $A(\Sigma_n) \leq 1-\delta$ for all $n \in \mathcal{N}$. Therefore, there exists some $k^* \in \mathcal{W}$, s.t.

1. if $k^* \in \mathcal{H}$, $\lambda_{k^*}^{\mathbf{A}}(\Sigma_n) \leq 1-\delta$, or

2. if $k^* \in \mathcal{L}$, $\lambda_{k^*}^{\mathbf{R}}(\Sigma_n) \leq 1-\delta$.

Without loss of generality, we assume $k^* \in \mathcal{H}$. The reasoning for $k^* \in \mathcal{L}$ will be almost the same. Then in the difference of expected utility, we have

$$\begin{aligned}
u_i(\Sigma_n') - u_i(\Sigma_n) =& P_{k^*}(v_i(k^*,\mathbf{A}) - v_i(k^*,\mathbf{R}))(\lambda_{k^*}^{\mathbf{A}}(\Sigma_n') - \lambda_{k^*}^{\mathbf{A}}(\Sigma_n)) \\
&+ \sum_{k\in\mathcal{L}} P_k(v_i(k,\mathbf{R}) - v_i(k,\mathbf{A}))(\lambda_k^{\mathbf{R}}(\Sigma_n') - \lambda_k^{\mathbf{R}}(\Sigma_n)) \\
&+ \sum_{k\in\mathcal{H}\setminus\{n^*\}} P_k(v_i(k,\mathbf{A}) - v_i(k,\mathbf{R}))(\lambda_k^{\mathbf{A}}(\Sigma_n') - \lambda_k^{\mathbf{A}}(\Sigma_n)) \\
\geq& P_{k^*}\cdot 1\cdot(1 - \frac{2\exp(-2\phi^2 n)}{P_{k^*}} - (1-\delta)) - \sum_{k\in\mathcal{W}\setminus\{k^*\}} P_k\cdot B\cdot\frac{2\exp(-2\phi^2 n)}{P_k} \\
=& \delta P_{k^*} - ((K-1)B+1)\cdot 2\exp(-2\phi^2 n). \\
\geq& \delta\cdot\min_{k\in\mathcal{W}}(P_k) - ((K-1)B+1)\cdot 2\exp(-2\phi^2 n).
\end{aligned}$$

Note that the first term $\delta\cdot\min_{k\in\mathcal{W}}(P_k)$ is a constant, while the second term $((K-1)B+1)\cdot 2\exp(-2\phi^2 n)$ converge to 0 as $n\to\infty$. Therefore, there exists a $n_\delta \geq n_0$ s.t. for all $(n > n_\delta)\wedge(n\in\mathcal{N})$, $u_i(\Sigma_n') - u_i(\Sigma_n) > \frac{1}{2}\delta\cdot\min_{k\in\mathcal{W}}(P_k)$. Therefore, for every such n, Σ_n is NOT an ε-strong BNE for all $\varepsilon \leq \frac{1}{2}\delta\cdot\min_{k\in\mathcal{W}}(P_k)$. □

A.3.4 Theorem 13 (Theorem 4 for binary setting)

Then we give the theorem based on the probability analysis of the fidelity and provide a criterion for judging whether a profile sequence is of high fidelity.

Consider a strategy profile sequence $\{\Sigma_n\}$. For every n, we define random variable X_i^n as "agent i votes for $\mathbf{A}$". That is, in the instance of n agents, $X_i^n = 1$ if agent i votes for $\mathbf{A}$, $X_i^n = 0$ if i votes for $\mathbf{R}$. Then we can write $\lambda_k^{\mathbf{A}}(\Sigma_n)$ and $\lambda_k^{\mathbf{R}}(\Sigma_n)$:

$$
\begin{aligned}
\lambda_k^{\mathbf{A}}(\Sigma_n) =& \Pr[\sum_{i=1}^{n} X_i^n \geq \mu \cdot n \mid k] \\
=& \Pr[\sum_{i=1}^{n} X_i^n - \sum_{i=1}^{n} E[X_i^n \mid k] \geq \mu \cdot n - \sum_{i=1}^{n} E[X_i^n \mid k] \mid k] \\
\lambda_k^{\mathbf{R}}(\Sigma_n) =& \Pr[\sum_{i=1}^{n}(1 - X_i^n) > (1-\mu)n \mid k] \\
=& \Pr[\sum_{i=1}^{n}(1 - X_i^n) - \sum_{i=1}^{n} E[1 - X_i^n \mid k] > (1-\mu)n - \sum_{i=1}^{n} E[1 - X_i^n \mid k] \mid k]
\end{aligned}
$$

Let the excess expected vote share be

$$
\begin{aligned}
f_{k\mathbf{A}}^n =& \frac{1}{n}\sum_{i=1}^{n} E[X_i^n \mid k] - \mu. \\
f_{k\mathbf{R}}^n =& \frac{1}{n}\sum_{i=1}^{n} E[1 - X_i^n \mid k] - (1-\mu).
\end{aligned}
$$

Then, for every n, let

$$
f^n = \min\left(\min_{k\in\mathcal{H}}(f_{k\mathbf{A}}^n), \min_{k\in\mathcal{L}}(f_{k\mathbf{R}}^n)\right).
$$

Theorem 13. *Given an arbitrary sequence of instances and an arbitrary sequence of strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$, let f^n be defined for every Σ^n.*

- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n = +\infty$, $\lim_{n\to\infty} A(\Sigma_n) = 1$.
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n < 0$ *(including $-\infty$), i.e., there exists a constant $\eta < 0$ and an infinite set $\mathcal{N}$, s.t. for every $n \in \mathcal{N}$, $\sqrt{n} \cdot f^n \leq \eta$, then there exists a constant $n_\eta > 0$ such that for all $(n \in \mathcal{N}) \wedge (n > n_\eta)$, $A(\Sigma_n)$ has a constant distance from 1.*
- *If* $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$ *(not including $+\infty$), i.e., there exists a constant $\eta \geq 0$ and an infinite set $\mathcal{N}$, s.t. for every $n \in \mathcal{N}$, $\sqrt{n} \cdot f^n \leq \eta$, and there exists a constant ψ s.t. for every $n \in \mathcal{N}$ and every $k \in \mathcal{W}$, $Var(\sum_{i=1}^{n} X_i^n \mid k) \geq \psi \cdot n$, then there exists a*

constant $n_\eta > 0$ such that for all $(n \in \mathcal{N}) \wedge (n > n_\eta)$, $A(\Sigma_n)$ has a constant distance from 1.

Proof. **Case 1:** $\liminf_{n\to\infty} \sqrt{n} \cdot f^n = +\infty$. In this case, we use the Hoeffding Inequality to give each $\lambda_k^{\mathbf{A}}(\Sigma_n)$ (or $\lambda_k^{\mathbf{R}}(\Sigma_n)$) a lower bound. For all $k \in \mathcal{H}$, according to the definition of f^n:

$$\begin{aligned}\lambda_k^{\mathbf{A}}(\Sigma_n) =& \Pr[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E[X_i^n \mid k] \geq -f_{k\mathbf{A}}^n \cdot n \mid k] \\ \geq& \Pr[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E[X_i^n \mid k] \geq -f^n \cdot n \mid k].\end{aligned}$$

Let n_0 be such that for all $n > n_0$, $\sqrt{n} \cdot f^n > 0$. Then for all $n > n_0$, we can directly apply the Hoeffding Inequality:

$$\begin{aligned}\lambda_k^{\mathbf{A}}(\Sigma_n) \geq& \Pr[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E[X_i^n \mid k] \geq -f^n \cdot n \mid k] \\ \geq& 1 - 2\exp(-2(f^n)^2 n).\end{aligned}$$

Similarly, for all $k \in \mathcal{L}$, we have

$$\begin{aligned}\lambda_k^{\mathbf{R}}(\Sigma_n) =& \Pr[\sum_{i=1}^n (1 - X_i^n) - \sum_{i=1}^n E[1 - X_i^n \mid k] \geq -f_{k\mathbf{R}}^n \cdot n \mid k] \\ \geq& \Pr[\sum_{i=1}^n (1 - X_i^n) - \sum_{i=1}^n E[1 - X_i^n \mid k] \geq -f^n \cdot n \mid k] \\ \geq& 1 - 2\exp(-2(f^n)^2 n).\end{aligned}$$

Therefore, the fidelity of Σ_n satisfies

$$A(\Sigma_n) = \sum_{k\in\mathcal{L}} P_k \cdot \lambda_k^{\mathbf{R}}(\Sigma_n) + \sum_{k\in\mathcal{H}} P_k \cdot \lambda_k^{\mathbf{A}}(\Sigma_n) \tag{A.4}$$

$$\geq \sum_{k\in\mathcal{W}} P_k \cdot (1 - 2\exp(-2(f^n)^2 n) \tag{A.5}$$

$$=1 - 2\exp(-2(f^n)^2 n). \tag{A.6}$$

Case 2: $\liminf_{n\to\infty} \sqrt{n} \cdot f^n < 0$. In this case, there exists a constant $\eta < 0$ and an infinite set $\mathcal{N}$, s.t. for every $n \in \mathcal{N}$, $\sqrt{n} \cdot f^n \leq \eta$. For every n, we define k_n as follows. For each n, define

$$g_k^n = \begin{cases} f_{k\mathbf{A}}^n, & \text{if } k \in \mathcal{H}, \\ f_{k\mathbf{R}}^n, & \text{if } k \in \mathcal{L}. \end{cases}$$

Choose an arbitrary minimizer

$$k_n \in \arg\min_{k\in\mathcal{W}} g_k^n.$$

That is, k_n is the world state whose $f_{k_n,\mathbf{A}}^n$ (if $k_n \in \mathcal{H}$) or $f_{k_n,\mathbf{R}}^n$ (if $k_n \in \mathcal{L}$) reach the minimum f^n. Note that for different n, k_n may be different, and some of them will be in $\mathcal{H}$ while others may be in $\mathcal{L}$. However, the reasoning for different k_n will be the same. Consider the fidelity of Σ_n for some $n \in \mathcal{N}$ when world state $W = k_n$, and without loss of generality, suppose $k_n \in \mathcal{H}$:

$$\begin{aligned} \lambda_{k_n}^{\mathbf{A}}(\Sigma_n) &= \Pr[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E[X_i^n \mid k_n] \geq -f_{k_n\mathbf{A}}^n \cdot n \mid k_n] \\ &\leq \Pr[\sum_{i=1}^n X_i^n - \sum_{i=1}^n E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n]. \end{aligned}$$

In this case, we have $\eta < 0$, thus $-\eta\sqrt{n} > 0$. Therefore, we can directly apply the Hoeffding

Inequality and get

$$\begin{aligned}\lambda^{\mathbf{A}}_{k_n}(\Sigma_n) \leq &\Pr[\sum_{i=1}^{n} X_i^n - \sum_{i=1}^{n} E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n] \\ \leq &\exp(-2\eta^2).\end{aligned}$$

For every $n \in \mathcal{N}$, we can prove $\lambda^{\mathbf{A}}_{k_n}(\Sigma_n) \leq \exp(-2\eta^2)$ if $k_n \in \mathcal{H}$, or $\lambda^{\mathbf{R}}_{k_n}(\Sigma_n) \leq \exp(-2\eta^2)$ if $k_n \in \mathcal{L}$. Therefore, for every $n \in \mathcal{N}$, the fidelity

$$\begin{aligned}A(\Sigma_n) \leq &1 - P_{k_n}(1 - \exp(-2\eta^2)) \\ \leq &1 - (1 - \exp(-2\eta^2)) \min_{k \in \mathcal{W}}(P_k).\end{aligned}$$

Case 3: $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$. In this case, we apply the Berry-Esseen Theorem, which bounds the difference between the sum of random variables and a normal distribution. We define $\mathcal{N}$ and k_n similarly as in Case 2. Suppose $k_n \in \mathcal{H}$. First we rewrite the form of $\lambda^{\mathbf{A}}_{k_n}(\Sigma_n)$. Let $Y_i^n = X_i^n - E[X_i^n \mid k_n]$, and $s_N = \sqrt{\sum_{i=1}^{n} Var(Y_i^n \mid k_n)} = \sqrt{\sum_{i=1}^{n} Var(X_i^n \mid k_n)}$. We have

$$\lambda^{\mathbf{A}}_{k_n}(\Sigma_n) \leq \Pr[\sum_{i=1}^{n} X_i^n - \sum_{i=1}^{n} E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n]$$

Let n' be the number of agents whose $Var(X_i^n \mid k_n) > 0$. Without loss of generality, let $Var(X_i^n \mid k_n) > 0$ for $i = 1, 2, \cdots, n'$, and $Var(X_i^n \mid k_n) = 0$ for $i = n'+1, n'+2, \cdots, n$. From the assumption $\sum_{i=1}^{n} Var(X_i^n \mid k_n) \geq \psi \cdot n$ we know that $n' \geq \psi \cdot n$ (because $Var(X_i^n \mid k_n) \leq 1$). Since $Var(X_i^n \mid k_n) = 0$ for $i = n'+1, n'+2, \cdots, n$, we have $X_i^n = E[X_i^n \mid k_n]$ conditioned on k_n for these i. Therefore, we have

$$\begin{aligned}\lambda^{\mathbf{A}}_{k_n}(\Sigma_n) \leq &\Pr[\sum_{i=1}^{n} X_i^n - \sum_{i=1}^{n} E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n] \\ = &\Pr[\sum_{i=1}^{n'} X_i^n - \sum_{i=1}^{n'} E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n]\end{aligned}$$

Then we rewrite the formula in Y_i^n and s_N:

$$\begin{aligned}\lambda_{k_n}^{\mathbf{A}}(\Sigma_n) &\leq \Pr[\sum_{i=1}^{n'} X_i^n - \sum_{i=1}^{n'} E[X_i^n \mid k_n] \geq -\eta\sqrt{n} \mid k_n] \\ &= \Pr\left[\frac{\sum_{i=1}^{n'} Y_i^n}{s_N} \geq -\frac{\eta\sqrt{n}}{s_N}\middle| k_n\right].\end{aligned}$$

Then we apply the Berry-Esseen Theorem and get

$$\lambda_{k_n}^{\mathbf{A}}(\Sigma_n) \leq 1 - \left(\Phi\left(-\frac{\eta\sqrt{n}}{s_N}\right) - C_0 \cdot \frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}}\right),$$

where Φ is the CDF of standard normal distribution, and $C_0 < 1$ is a constant from the theorem. We deal with this inequality term by term to show that $\lambda_{k_n}^{\mathbf{A}}(\Sigma_n)$ has a constant difference with 1. First, we show

$$\Phi\left(\frac{\eta\sqrt{n}}{s_N}\right) \geq \Phi\left(-\frac{\eta\sqrt{n}}{\sqrt{\psi n}}\right) = \Phi\left(-\frac{\eta}{\sqrt{\psi}}\right).$$

This is because $s_N^2 = \sum_{i=1}^{n} Var(Y_i^n \mid k_n) \geq \psi \cdot n$. The following table reveals how things works.

s_N	$\geq$	$\sqrt{\psi \cdot n}$
$\frac{\eta\sqrt{n}}{s_N}$	$\leq$	$\frac{\eta\sqrt{n}}{\sqrt{\psi n}}$
$-\frac{\eta\sqrt{n}}{s_N}$	$\geq$	$-\frac{\eta\sqrt{n}}{\sqrt{\psi n}}$
$\Phi\left(-\frac{\eta\sqrt{n}}{s_N}\right)$	$\geq$	$\Phi\left(-\frac{\eta\sqrt{n}}{\sqrt{\psi n}}\right)$

Table A.1: Comparison on the left and the right side.

Secondly, we start to deal with

$$\frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}}.$$

Firstly, since $-1 \leq Y_i \leq 1$, we have $E[|Y_i^n|^3 \mid k_n] \leq 1$. Therefore,

$$\frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}} \leq \frac{n'}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}}.$$

Then, notice that since $\sigma_i^2 = 0$ for $i = n'+1, n'+2, \cdots, n$, we have $\sum_{i=1}^{n'} Var(Y_i^n \mid k_n) = \sum_{i=1}^{n} Var(Y_i^n \mid k_n) \geq \psi \cdot n$. Therefore,

$$\frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}} \leq \frac{n'}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}} \leq \frac{n'}{(\psi \cdot n)^{3/2}}.$$

Finally, since $n' \leq n$, and $n_C \geq \alpha_C n - 1$,

$$\frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'_C} Var(Y_i^n \mid k_n)\right)^{3/2}} \leq \frac{n'}{(\psi \cdot n)^{3/2}} \leq \frac{1}{\sqrt{\psi^3 \cdot n}}$$

Now we are ready to wrap things up:

$$\begin{aligned}
\lambda_{k_n}^{\mathbf{A}}(\Sigma_n) \leq & 1 - \left(\Phi\left(-\frac{\eta\sqrt{n}}{s_N}\right) - C_0 \cdot \frac{\sum_{i=1}^{n'} E[|Y_i^n|^3 \mid k_n]}{\left(\sum_{i=1}^{n'} Var(Y_i^n \mid k_n)\right)^{3/2}}\right) \\
\leq & 1 - \left(\Phi\left(-\frac{\eta}{\sqrt{\psi}}\right) - C_0 \cdot \frac{1}{\sqrt{\psi^3 \cdot n}}\right) \\
\leq & 1 - \left(\Phi\left(-\frac{\eta}{\sqrt{\psi}}\right) - \frac{1}{\sqrt{\psi^3 \cdot n}}\right)
\end{aligned}$$

The last inequality comes from that $C_0 \leq 0.5600$ [92].

The same reasoning also works for other $n \in \mathcal{N}$. Therefore, let $\phi = \Phi\left(-\frac{\eta}{\sqrt{\psi}}\right)$, then there must exist a n_ϕ, s.t. for all $(n > n_\phi) \wedge (n \in \mathcal{N})$, $\lambda_{k_n}^{\mathbf{A}}(\Sigma_n) \leq 1 - \frac{1}{2}\phi$. Therefore, $A(\Sigma_n) \leq 1 - \frac{1}{2}P_{k_n} \cdot \phi \leq 1 - \frac{1}{2}\min_{k\in\mathcal{W}}(P_k) \cdot \phi$. □

Similarly to the binary setting, if we apply Theorem 5 to each case of Theorem 13, we get a criterion for judging whether a profile sequence is an equilibrium.

Corollary 5. *Given an arbitrary sequence of instances and arbitrary sequence of regular strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$ let f^n is defined for every Σ^n.*

- *If $\liminf_{n\to\infty} \sqrt{n} \cdot f^n = +\infty$, then for all sufficiently large n, Σ_n is an ε-strong BNE with $\varepsilon = o(1)$.*
- *If $\liminf_{n\to\infty} \sqrt{n} \cdot f^n < 0$ (including $-\infty$), there exist infinitely many n such that Σ_n is not an ε-strong BNE with some constant ε.*
- *If $\liminf_{n\to\infty} \sqrt{n} \cdot f^n \geq 0$ (not including $+\infty$), and there exists a constant ψ s.t. for every $n \in \mathcal{N}$ and every $k \in \mathcal{W}$, $Var(\sum_{i=1}^{n} X_i^n \mid k) \geq \psi \cdot n$, there exist infinitely many n such that Σ_n is not an ε-strong BNE with some constant ε.*

A.3.5 Corollary 6 (Corollary 2 for binary setting)

Finally, we can directly get a dichotomy for symmetric strategy profiles from Theorem 13, by showing that every case of symmetric profiles falls into some case.

Given arbitrary n and a symmetric strategy profile induced by $\sigma = (\beta_1, \beta_2. \cdots, \beta_M)$ the excess expected vote share is

$$\begin{aligned}
f_{k\mathbf{A}}^n =& \frac{1}{N}\sum_{i=1}^{n}(\sum_{m=1}^{M} P_{mk} \cdot \beta_m) - \mu = \sum_{m=1}^{M} P_{mk} \cdot \beta_m - \mu, \\
f_{k\mathbf{R}}^n =& \frac{1}{N}\sum_{i=1}^{n}(\sum_{m=1}^{M} P_{mk} \cdot (1-\beta_m)) - (1-\mu) = \sum_{m=1}^{M} P_{mk} \cdot (1-\beta_m) - (1-\mu), \\
f^n =& \min\left(\min_{k\in\mathcal{H}} f_{k\mathbf{A}}^n, \min_{k\in\mathcal{L}} f_{k\mathbf{R}}^n\right).
\end{aligned}$$

An interesting observation is that the excess expected vote share of symmetric profiles is independent of N. Therefore, for simplicity, we use $f_{k\mathbf{A}}, f_{k\mathbf{R}}$, and f to denote the excess expected vote share for symmetric profiles.

Corollary 6. *For an arbitrary strategy $\sigma = (\beta_1, \beta_2. \cdots, \beta_M)$ and an arbitrary sequence of instances, let $\{\Sigma_n\}$ be the sequence of strategy profile Σ_n induced by σ, and f be the excess expected vote share of the strategy profiles.*

- *If $f > 0$, then there exists a constant $n_0 > 0$, s.t. for all $n > n_0$, $A(\Sigma_n) \geq 1 - 2\exp(-\frac{1}{2}f^2 n)$.*

- *If $f \leq 0$, then there exist constants $n_0 > 0$ and $\eta' > 0$, s.t. for every $n > n_0$. $A(\Sigma_n) \leq 1 - \eta'$.*

Proof. For $f > 0$, we have $\liminf n \to \infty \sqrt{n} \cdot f = +\infty$. Therefore, we can apply Case 1 of Theorem 5 (Inequality A.6), and have

$$A(\Sigma_n) \geq 1 - 2\exp\left(-2f^2 n\right).$$

For $f < 0$, we have $\liminf n \to \infty \sqrt{n} \cdot f = -\infty$. Therefore, we can apply Case 2 of Theorem 5 with any constant $\eta < 0$. Then there exists a n_η such that for all $n > n_\eta$, we have $A(\Sigma_n) \leq 1 - (1 - \exp(-2\eta^2)) \min_{k \in \mathcal{W}}(P_k)$. Therefore, let $\eta' = (1 - \exp(-2\eta^2)) \min_{k \in \mathcal{W}}(P_k)$, we have $A(\Sigma_n) \leq 1 - \eta'$.

For $f = 0$, we have $\liminf n \to \infty \sqrt{n} \cdot f = 0$. Therefore, we need to consider the variance. $\sum_{n=1}^{n} Var(X_i^n \mid k)$. For a single contingent agent i, given $W = k$, we have

$$\Pr[X_i^n = 0 \mid k] = \sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m)$$
$$\Pr[X_i^n = 1 \mid k] = \sum_{m=1}^{M} P_{mk} \cdot \beta_m.$$

Therefore, we have

$$Var(X_i^n \mid k) = \left(\sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m)\right)\left(\sum_{m=1}^{M} P_{mk} \cdot \beta_m\right)$$

is a constant. Let $\psi = \min_{k \in \mathcal{W}} \left(\sum_{m=1}^{M} P_{mk} \cdot (1 - \beta_m)\right)\left(\sum_{m=1}^{M} P_{mk} \cdot \beta_m\right)$, and we get for every $k \in \mathcal{W}$, $\sum_{n=1}^{n} Var(X_i^n \mid k) \geq \psi \cdot n$.

We show by contradiction that $\psi = 0$ will not happen. Suppose it is not the case and

$\psi = 0$. Let $k_\psi = \arg\min_{k \in \mathcal{W}} \left(\sum_{m=1}^M P_{mk} \cdot (1-\beta_m)\right) \left(\sum_{m=1}^M P_{mk} \cdot \beta_m\right)$. Then we have

$$\left(\sum_{m=1}^M P_{mk_\psi} \cdot (1-\beta_m)\right) \left(\sum_{m=1}^M P_{mk_\psi} \cdot \beta_m\right) = 0.$$

Without loss of generality, assume $\left(\sum_{m=1}^M P_{mk_\psi} \cdot \beta_m\right) = 0$. Then we consider $f_{k_\psi \mathbf{A}}$, and have

$$f_{k_\psi \mathbf{A}} = \alpha_C \sum_{m=1}^M P_{mk_\psi} \cdot \beta_m - \mu = -\mu < 0,$$

which contradicts $f = 0$. Therefore, $\psi = 0$ will not happen.

Since we have guaranteed that $\psi > 0$, we can apply Theorem 13. Let $\eta > 0$ be any positive constant. We can apply Case 3 of Theorem 13 and get that there exists a n_η, s.t. for all $n > n_\eta$, $A(\Sigma_n) \leq 1 - \frac{1}{2} \min_{k \in \mathcal{W}} (P_k)^2 \cdot \Phi\left(-\frac{\eta}{\sqrt{\psi}}\right)$. Therefore, let $\eta' = \frac{1}{2} \min_{k \in \mathcal{W}} (P_k)^2 \cdot \Phi\left(-\frac{\eta}{\sqrt{\psi}}\right)$, we have $A(\Sigma_n) \leq 1 - \eta'$ for all $n > n_\eta$. $\square$

APPENDIX B

TECHNICAL APPENDIX FOR CHAPTER 4

B.1 Full Proofs.

We abuse the notation for deterministic first/second round strategy such that $\sigma_i^1(m) = \mathbf{A}$ means that i always votes for $\mathbf{A}$ (with probability 1) in the first round when her signal is m. Other deterministic votes follow a similar notation.

B.1.1 Proof of Lemma 5

Lemma 5 (Formal)**.** *Let Σ^* be an arbitrary profile, and let $\varepsilon = 2B(B+2)(1 - A(\Sigma^*))$. If there exists a group of agents D and another strategy profile Σ' such that agents in D have incentives to deviate to Σ' so that (1) $\forall\ i \in D$, $u_i(\Sigma') \geq u_i(\Sigma^*)$, and (2) there exists $i \in D$ such that $u_i(\Sigma') > u_i(\Sigma^*) + \varepsilon$, then either $D \subseteq F$ or $D \subseteq U$.*

In this proof, we use Σ^* to denote Σ_n^* for simplicity. We show that for any other strategy profile Σ', there does not exist a group of agents wishing to deviate. Let $e = 1 - A(\Sigma^*)$. We know that e converges to 0 as n goes to infinity. We take $\varepsilon = 2B(B+2)e$.

Proof of Lemma 5. We consider two different cases of the fidelity of Σ'.

Case 1: $(1 - A(\Sigma')) < (2B+2)e$**.** In this case, since the fidelity of Σ' is also close to one, two profiles make no significant difference, and no agent can gain more than $\varepsilon = 2B(B+2)e$.

Consider the difference of expected utility in two profiles.

$$\begin{aligned} u_i(\Sigma') - u_i(\Sigma^*) = {} & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*))(v_i(H,\mathbf{A}) - v_i(H,\mathbf{R})) \\ & + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*))(v_i(L,\mathbf{R}) - v_i(L,\mathbf{A})). \end{aligned}$$

Since we know that $A(\Sigma^*) = 1 - e$, we have $\lambda_H^{\mathbf{A}}(\Sigma^*) \geq 1 - \frac{e}{P_H}$ and $\lambda_L^{\mathbf{R}}(\Sigma^*) \geq 1 - \frac{e}{P_L}$. Similarly, for Σ', we have $\lambda_H^{\mathbf{A}}(\Sigma') \geq 1 - \frac{(2B+2)e}{P_H}$ and $\lambda_L^{\mathbf{R}}(\Sigma') \geq 1 - \frac{(2B+2)e}{P_L}$. Then we can bound the difference between these probabilities.

$$\begin{aligned}
\lambda_L^{\mathbf{R}}(\Sigma^*) - \lambda_L^{\mathbf{R}}(\Sigma') &\leq \frac{(2B+2)e}{P_L}, \\
\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*) &\leq \frac{e}{P_L}, \\
\lambda_H^{\mathbf{A}}(\Sigma^*) - \lambda_H^{\mathbf{A}}(\Sigma') &\leq \frac{(2B+2)e}{P_H}, \\
\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*) &\leq \frac{e}{P_H}.
\end{aligned}$$

Then we can bound the expected utilities.

- For the friendly agents, $v_i(H, \mathbf{A}) > v_i(H, \mathbf{R})$ and $v_i(L, \mathbf{A}) > v_i(L, \mathbf{R})$. Then

$$\begin{aligned}
u_i(\Sigma') - u_i(\Sigma^*) &\leq P_L \cdot \frac{(2B+2)e}{P_L} \cdot B + P_H \cdot \frac{e}{P_H} \cdot B \\
&= 2B(B+2)e.
\end{aligned}$$

- For contingent agents, $v_i(H, \mathbf{A}) > v_i(H, \mathbf{R})$ and $v_i(L, \mathbf{A}) < v_i(L, \mathbf{R})$. Then

$$\begin{aligned}
u_i(\Sigma') - u_i(\Sigma^*) &\leq P_L \cdot \frac{e}{P_L} \cdot B + P_H \cdot \frac{e}{P_H} \cdot B \\
&= 2Be.
\end{aligned}$$

- For the unfriendly agents, we have $v_i(H, \mathbf{A}) < v_i(H, \mathbf{R})$ and $v_i(L, \mathbf{A}) < v_i(L, \mathbf{R})$. Then

$$\begin{aligned}
u_i(\Sigma') - u_i(\Sigma^*) &\leq P_L \cdot \frac{e}{P_L} \cdot B + P_H \cdot \frac{(2B+2)e}{P_H} \cdot B \\
&= 2B(B+2)e.
\end{aligned}$$

Therefore, no agents can gain more than $\varepsilon = 2B(B+2)e$.

Case 2: $(1 - A(\Sigma')) \geq (2B+2)e$**.** In this case, we proceed with our proof in two steps.

1. Firstly, contingent agents will not deviate, because their expected utility will strictly decrease.

2. Secondly, a deviating group cannot contain both friendly and unfriendly agents.

Claim 5. *If* $(1 - A(\Sigma')) \geq (2B+2)e$*, then for any contingent agent* i*,* $u_i(\Sigma') - u_i(\Sigma^*) < 0$*.*

Still, recall the difference between the two expected utilities.

$$\begin{aligned} u_i(\Sigma') - u_i(\Sigma^*) = {} & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*))(v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})) \\ & + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})). \end{aligned}$$

From $A(\Sigma^*) = e$ and $A(\Sigma') \leq 1 - (2B+2)e$, we have

$$P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma^*) + P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma^*) - P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma') + P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma') \geq (2B+1)e.$$

Then we consider different cases between the λ.

- If $\lambda_H^{\mathbf{A}}(\Sigma') \leq \lambda_H^{\mathbf{A}}(\Sigma^*)$ and $\lambda_L^{\mathbf{R}}(\Sigma') \leq \lambda_L^{\mathbf{R}}(\Sigma^*)$, at least one of them will be strict. Then $u_i(\Sigma') - u_i(\Sigma^*) < 0$.

- If $\lambda_H^{\mathbf{A}}(\Sigma') > \lambda_H^{\mathbf{A}}(\Sigma^*)$, then we have $\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*) < \frac{e}{P_H}$. On the other hand, $\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*) \leq -\frac{(2B+1)e}{P_L}$. Then,

$$u_i(\Sigma') - u_i(\Sigma^*) \leq P_H \cdot \frac{e}{P_H} \cdot B + P_L \cdot \left(-\frac{(2B+1)e}{P_L}\right) \cdot 1 < 0.$$

- If $\lambda_L^{\mathbf{R}}(\Sigma') > \lambda_L^{\mathbf{R}}(\Sigma^*)$, with similar reasoning, we can show that $u_i(\Sigma') - u_i(\Sigma^*) < 0$.

Claim 6. *Let* i_1 *be any friendly agent and* i_2 *be any unfriendly agent. Then*

1. If $u_{i_1}(\Sigma') - u_{i_1}(\Sigma^*) > \varepsilon$*, then* $u_{i_2}(\Sigma') - u_{i_2}(\Sigma^*) < 0$*.*

2. If $u_{i_2}(\Sigma') - u_{i_2}(\Sigma^*) > \varepsilon$, then $u_{i_1}(\Sigma') - u_{i_1}(\Sigma^*) < 0$.

We only prove 1, and 2 will follow the same reasoning. Without loss of generality, let $i_1 = 1$ and $i_2 = 2$.

Still, recall the difference between the expected utility.

$$\begin{aligned} u_i(\Sigma') - u_i(\Sigma^*) = {} & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*))(v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})) \\ & + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})). \end{aligned}$$

We consider two cases.

Case 1: $\lambda_H^{\mathbf{A}}(\Sigma') \geq \lambda_H^{\mathbf{A}}(\Sigma^*)$. In this case, since $\lambda_H^{\mathbf{A}}(\Sigma^*)$ is already close to 1, the gain in agent 1's expected utility on H's side will not exceed $P_H \cdot \frac{e}{P_H} \cdot B = eB < \varepsilon$. Therefore, $\lambda_L^{\mathbf{R}}(\Sigma') < \lambda_L^{\mathbf{R}}(\Sigma^*)$. In this case, the expected utility of agent 2 will strictly decrease.

Case 2: $\lambda_H^{\mathbf{A}}(\Sigma') < \lambda_H^{\mathbf{A}}(\Sigma^*)$. In this case, the P_L term is positive for agent 1, while the P_H term is positive for agent 2. For agent 1, we have $v_1(H, \mathbf{A}) \geq v_1(L, \mathbf{A}) > v_1(L, \mathbf{R}) \geq v_1(H, \mathbf{R})$. Then since $u_1(\Sigma') - u_1(\Sigma^*)$ is strictly positive, we must have

$$P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma^*)) > P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma^*)).$$

For agent 2, on the other hand, we have $v_2(L, \mathbf{R}) \geq v_2(H, \mathbf{R}) > v_2(H, \mathbf{A}) \geq v_2(L, \mathbf{A})$. Then there must be $u_2(\Sigma') - u_2(\Sigma^*) < 0$. □

B.1.2 Proof of Theorem 9

The proof of Theorem 9 resembles that of Theorem 8. Lemma 3 guarantees the fidelity. For the equilibrium parts, applying Lemma 5, it suffices to consider a deviating group with only friendly agents or only unfriendly agents. Finally, Claim 7 shows that such deviation cannot succeed.

Lemma 3. *Under any environment sequence $\{\mathcal{I}_n\}$, every constantly separable sequence of* *profiles* $\{\Sigma^*\}$ *satisfies that (1)* $\lim_{n\to\infty} A(\Sigma^*_n) = 1$*, and (2) for every* n*,* Σ^*_n *is an* ε*-strong BNE where* ε *converges to 0 as* n *goes to* ∞*.*

Proof. The high-level idea of Σ^* is that contingent agents determine the world state with high probability via the first-round voting and vote for the correct alternative in the second round. Since predetermined agents behave regularly, and both α_F and α_U do not exceed 0.5, the alternative that all contingent agents vote for will be the winner. Now we show that both $\lim_{n\to\infty} \lambda^{\mathbf{A}}_H(\Sigma^*) = 1$ and $\lim_{n\to\infty} \lambda^{\mathbf{R}}_L(\Sigma^*) = 1$.

For now we fix an arbitrary n. For an agent i, let X_i be the random variable indicating i's vote in the first round. $X_i = 0$ means i votes for $\mathbf{R}$, and $X_i = 1$ means i votes for $\mathbf{A}$. Then the result of the first round voting x can be represented as $x = \sum_i X_i$. If the first round outcome exceeds $x \geq n_F + n_C \cdot P_{hH} - \delta$, then all the agents i will vote for $\mathbf{A}$, and $\mathbf{A}$ becomes the winner. Therefore, the probability that $\mathbf{A}$ becomes the winner conditioned on the world state H satisfies

$$\lambda^{\mathbf{A}}_H(\Sigma^*) \geq \Pr\left[\sum_i X_i \geq n_F + n_C \cdot P_{hH} - \delta \cdot n \;\middle|\; H\right].$$

Since contingent agent votes informatively in the first round, the expectation of the first round outcome $E[\sum_i X_i \mid H] = n_F + n_C \cdot P_{hH}$. Then for all $n \geq n_0$, we can apply the Hoeffding inequality to the probability:

$$\begin{aligned}
\lambda^{\mathbf{A}}_H(\Sigma^*) \geq\ & \Pr\left[\sum_i X_i \geq n_F + n_C \cdot P_{hH} - \delta \cdot n \;\middle|\; H\right] \\
=\ & \Pr\left[\sum_i X_i - E[\sum_i X_i \mid H] \geq n_F + n_C \cdot P_{hH} - \delta \cdot n - E[\sum_i X_i \mid H] \;\middle|\; H\right] \\
=\ & \Pr\left[\sum_i X_i - E[\sum_i X_i \mid H] \geq -\delta \cdot n \;\middle|\; H\right] \\
\geq\ & 1 - \exp(-2\delta^2 \cdot n).
\end{aligned}$$

Similarly, when the world state is L, the expectation of the first-round outcome is

$E[\sum_i X_i \mid L] = n_F + n_C \cdot P_{hL}$.

$$
\begin{aligned}
\lambda_L^{\mathbf{R}}(\Sigma^*) \geq\ & \Pr\left[\sum_i X_i \leq n_F + n_C \cdot P_{hL} + \delta \cdot n \;\middle|\; L\right] \\
=\ & \Pr\left[\sum_i X_i - E[\sum_i X_i \mid L] \leq n_F + n_C \cdot P_{hL} + \delta \cdot n - E[\sum_i X_i \mid L] \;\middle|\; L\right] \\
=\ & \Pr\left[\sum_i X_i - E[\sum_i X_i \mid L] \leq \delta \cdot n \;\middle|\; L\right] \\
\geq\ & 1 - \exp(-2\delta^2 \cdot n).
\end{aligned}
$$

Therefore, both $\lambda_H^{\mathbf{A}}(\Sigma_n^*)$ and $\lambda_L^{\mathbf{R}}(\Sigma_n^*)$ converge to 1 as n goes to infinity. Since fidelity $A(\Sigma^*) = P_H \cdot \lambda_H^{\mathbf{A}}(\Sigma_n^*) + P_L \cdot \lambda_L^{\mathbf{R}}(\Sigma_n^*)$, $\lim_{n\to\infty} A(\Sigma^*) = 1$. $\square$

Lemma 4. *Σ_n^* is an ε-strong BNE with $\varepsilon = o(1)$.*

By applying Lemma 5, we only need to show that the deviating group containing only one type of predetermined agents cannot succeed. Therefore, "informative + threshold" strategy profile Σ^* is an ε-strong BNE with ε converging to 0.

Claim 7. *Suppose D is the deviating group that contains only friendly or only unfriendly agents, then $u_i(\Sigma') - u_i(\Sigma^*) < \varepsilon$ for all $i \in D$.*

We prove the friendly agent side, and the unfriendly agent side will follow the same reasoning. The expected utility of a friendly agent increases only when the likelihood that **A** becomes the winner increases. However, in Σ_n^* , all the friendly agents always vote for **A** in both rounds. If friendly agents deviate in the second round, there will be strictly fewer agents voting for **A** in expectation, and **A** will be less likely to be the winner. If friendly agents deviate in the first round, the result of the first-round voting x will have a higher probability of being smaller, and contingent agents playing the threshold strategy will tend to vote **R**. Therefore, no matter how D deviates, the likelihood of **A** being the winner will not increase, and friendly agents cannot get a gain of more than ε.

B.1.3 Proof for Lemma 6

Lemma 6. *The sequence of "informative + sincere" strategy profiles is constantly separable.*

Proof. Recall that

$$x^* = n_F + \frac{\log \frac{P_L \cdot (v_i(L,\mathbf{R}) - v_i(L,\mathbf{A}))}{P_H \cdot (v_i(H,\mathbf{A}) - v_i(H,\mathbf{R}))} + n_C \cdot \log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH} P_{\ell L}}{P_{\ell H} P_{hL}}}.$$

As n_F and n_C increases with n, while $\log \frac{P_L \cdot (v_i(L,\mathbf{R}) - v_i(L,\mathbf{A}))}{P_H \cdot (v_i(H,\mathbf{A}) - v_i(H,\mathbf{R}))}$ remains constant, it suffices to show that

$$n_F + n_C \cdot P_{hL} + \delta \cdot n \le n_F + \frac{n_C \cdot \log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH} P_{\ell L}}{P_{\ell H} P_{hL}}} \le n_F + n_C \cdot P_{hH} - \delta \cdot n,$$

which is equivalent to

$$P_{hL} + \delta \cdot \frac{n}{n_C} \le \frac{\log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH} P_{\ell L}}{P_{\ell H} P_{hL}}} \le P_{hH} - \delta \cdot \frac{n}{n_C}.$$

□

We first deal with the left inequality.

$$\begin{aligned}
& P_{hL} + \delta \cdot \frac{n}{n_C} \le \frac{\log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH} P_{\ell L}}{P_{\ell H} P_{hL}}} \\
\Leftrightarrow\ & P_{hL} + \delta \cdot \frac{n}{n_C} \le \frac{\log \frac{P_{\ell L}}{P_{\ell H}}}{\log \frac{P_{hH}}{P_{hL}} + \log \frac{P_{\ell L}}{P_{\ell H}}} \\
\Leftrightarrow\ & \delta \cdot \frac{n}{n_C} \cdot (\log \frac{P_{hH}}{P_{hL}} + \log \frac{P_{\ell L}}{P_{\ell H}}) \le (P_{hL} \cdot \log \frac{P_{hL}}{P_{hH}} + \log P_{\ell L} \cdot \frac{P_{\ell L}}{P_{\ell H}}).
\end{aligned}$$

Note that the RHS takes a form of the KL-divergence between distribution $(P_{hL}, P_{\ell L})$ and $P_{hH}, P_{\ell H})$. Given that $P_{hH} > P_{hL}$, RHS is a strictly positive constant independent from n.

On the other hand, $n_C \approx \alpha_C \cdot n$. Therefore, by taking

$$\delta \leq \frac{1}{2}\alpha_C \cdot \frac{(P_{hL} \cdot \log \frac{P_{hL}}{P_{hH}} + \log P_{\ell L} \cdot \frac{P_{\ell L}}{P_{\ell H}})}{(\log \frac{P_{hH}}{P_{hL}} + \log \frac{P_{\ell L}}{P_{\ell H}})},$$

the left inequality is satisfied.

Likewise, the right inequality is equivalent to

$$\delta \cdot \frac{n}{n_C} \cdot (\log \frac{P_{hH}}{P_{hL}} + \log \frac{P_{\ell L}}{P_{\ell H}}) \leq (P_{hH} \cdot \log \frac{P_{hH}}{P_{hL}} + \log P_{\ell H} \cdot \frac{P_{\ell H}}{P_{\ell L}}),$$

and by taking

$$\delta \leq \frac{1}{2}\alpha_C \cdot \frac{(P_{hH} \cdot \log \frac{P_{hH}}{P_{hL}} + \log P_{\ell H} \cdot \frac{P_{\ell H}}{P_{\ell L}})}{(\log \frac{P_{hH}}{P_{hL}} + \log \frac{P_{\ell L}}{P_{\ell H}})},$$

the right inequality is satisfied. Therefore, by a sufficiently small δ and all sufficiently large n, the constraint for constant separability is satisfied.

B.1.4 Proof of Theorem 7

Before starting the proof, we extend the definition of the first-round voting share. Let $\phi_H^F(\Sigma), \phi_H^U(\Sigma)$, and $\phi_H^C(\Sigma)$ be the expectation on the number of votes for $\mathbf{A}$ from three types of agents conditioned on the world state being H and the strategy profile being Σ. We have $\phi_H(\Sigma) = \phi_H^F(\Sigma) + \phi_H^U(\Sigma) + \phi_H^C(\Sigma)$. The L side is defined similarly.

Proof. We first give the proof for regular profiles. Suppose $\{\Sigma_n\}$ is a regular profile sequence such that $\lim_{n\to\infty} A(\Sigma_n) = 1$ does not hold. This means that there exists a constant δ and an infinite set $\mathcal{N}$ such that for all $n \in \mathcal{N}$, $A(\Sigma_n) \leq 1 - \delta$. We show that there exists a constant $\varepsilon = \frac{\delta \cdot P_H P_L}{2}$ and an infinite set $\mathcal{N}'$ such that for all $n \in \mathcal{N}'$, Σ_n is NOT an ε-strong BNE.

We will first construct a different profile sequence $\{\Sigma_n'\}$ whose fidelity converges to 1. Then for all sufficiently large $n \in \mathcal{N}$, both $\lambda_H^{\mathbf{A}}$ and $\lambda_L^{\mathbf{R}}$ are close to one. On the other hand, Since $A(\Sigma_n) \leq 1 - \delta$, we have at least one of $\lambda_H^{\mathbf{A}}(\Sigma_n) \leq 1 - \delta$ and $\lambda_L^{\mathbf{R}}(\Sigma_n) \leq 1 - \delta$ holds.

Then we can see that all the contingent agents have incentives to deviate to Σ'_n to get a constant gain in the expected utility.

For simplicity, we use Σ and Σ' to denote Σ_n and Σ'_n. Let Σ'_n be the following strategy profile. The deviating group D contains all the contingent agents.

- All predetermined agents play the same strategy as in Σ.
- If $\phi_H^F(\Sigma) + \phi_H^U(\Sigma) \geq \phi_L^F(\Sigma) + \phi_L^U(\Sigma)$, all contingent agents vote informatively in the first round. For the second round, all contingent agents vote for **A** if $x \geq \frac{\phi_H(\Sigma')+\phi_L(\Sigma')}{2}$ and vote for **R** otherwise.
- If $\phi_H^F(\Sigma) + \phi_H^U(\Sigma) < \phi_L^F(\Sigma) + \phi_L^U(\Sigma)$, contingent agents "reverse" their strategy. In the first round, a contingent agent votes for **A** if receiving l and for **R** if receiving h. And in the second round, all contingent agents vote for **A** if $x < \frac{\phi_H(\Sigma')+\phi_L(\Sigma')}{2}$ and vote for **R** otherwise.

The high-level idea of Σ' is similar to Σ^* in Theorem 9: contingent agents determine the world state with high probability via the first-round voting and vote for the correct alternative in the second round. Since predetermined agents behave regularly, and both α_F and α_U do not exceed 0.5, the alternative that all contingent agents vote for will be the winner. Now we show that both $\lim_{n\to\infty} \lambda_H^{\mathbf{A}}(\Sigma') = 1$ and $\lim_{n\to\infty} \lambda_L^{\mathbf{R}}(\Sigma') = 1$.

Without loss of generality, suppose $\phi_H^F(\Sigma) + \phi_H^U(\Sigma) \geq \phi_L^F(\Sigma) + \phi_L^U(\Sigma)$. The reasoning for the other case will be similar. For an agent i, let X_i be the random variable indicating i's vote in the first round. $X_i = 0$ means i votes for **R**, and $X_i = 1$ means i votes for **A**. Then the result of the first round voting x can be represented as $x = \sum_i X_i$. Then the probability such that **A** becomes the winner (which is the probability that $x \geq \frac{\phi_H(\Sigma')+\phi_L(\Sigma')}{2}$) conditioned on the world state H can be written as

$$\lambda_H^{\mathbf{A}}(\Sigma') = \Pr[\sum_i X_i \geq \frac{\phi_H(\Sigma') + \phi_L(\Sigma')}{2} \mid H].$$

Note that $E[\sum_i X_i \mid H] = \phi_H(\Sigma')$.

Since contingent agent votes informatively in the first round, $\phi_H^C(\Sigma') = n_C \cdot P_{hH}$ and $\phi_L^C(\Sigma') = n_C \cdot P_{hL}$. Since $P_{hH} > P_{hL}$, we have $\phi_H^C(\Sigma') - \phi_L^C(\Sigma') = n_C \cdot (P_{hH} - P_{hL}) \geq n \cdot (\alpha_C - \frac{1}{n}) \cdot (P_{hH} - P_{hL})$. Then there exists a n_0 such that for all $n > n_0$, $\phi_H^C(\Sigma') - \phi_L^C(\Sigma') \geq \frac{1}{2} n \cdot \alpha_C \cdot (P_{hH} - P_{hL})$. Then since $\phi_H^F(\Sigma) + \phi_H^U(\Sigma) \geq \phi_L^F(\Sigma) + \phi_L^U(\Sigma)$, we have $n > n_0$, $\phi_H(\Sigma') - \phi_L(\Sigma') \geq \frac{1}{2} n \cdot \alpha_C \cdot (P_{hH} - P_{hL})$. Then we can apply the Hoeffding inequality to the probability:

$$\begin{aligned}
\lambda_H^{\mathbf{A}}(\Sigma') = & \Pr\left[\sum_i X_i \geq \frac{\phi_H(\Sigma') + \phi_L(\Sigma')}{2} \;\middle|\; H\right] \\
= & \Pr\left[\sum_i X_i - E[\sum_i X_i \mid H] \geq \frac{\phi_H(\Sigma') + \phi_L(\Sigma')}{2} - E[\sum_i X_i \mid H] \;\middle|\; H\right] \\
= & \Pr\left[\sum_i X_i - \phi_H(\Sigma') \geq \frac{\phi_L(\Sigma') - \phi_H(\Sigma')}{2} \;\middle|\; H\right] \\
\geq & \Pr[\sum_i X_i - \phi_H(\Sigma') \geq -\frac{1}{4} n \cdot \alpha_C \cdot (P_{hH} - P_{hL}) \mid H] \\
\geq & 1 - \exp(-\frac{1}{8}(\alpha_C \cdot (P_{hH} - P_{hL})^2 \cdot n)).
\end{aligned}$$

The L side works similarly.

Therefore, both $\lambda_H^{\mathbf{A}}(\Sigma_n')$ and $\lambda_L^{\mathbf{R}}(\Sigma_n')$ converge to 1 as n goes to infinity.

Next, we show that contingent agents have incentives to deviate from Σ_n to Σ_n'. Let $\mathcal{N}'$ be the set of all $n \in \mathcal{N}$ such that both $\lambda_H^{\mathbf{A}}(\Sigma_n') \geq 1 - \frac{\delta \cdot P_H P_L}{2B}$ and $\lambda_L^{\mathbf{R}}(\Sigma_n') \geq 1 - \frac{\delta \cdot P_H P_L}{2B}$ holds. We know that $\mathcal{N}'$ is also an infinite set. We fix an arbitrary $n \in \mathcal{N}'$. Without loss of generality, suppose $\lambda_H^{\mathbf{A}}(\Sigma_n) \leq 1 - \delta$. Then we consider the difference in the expected utility of a contingent agent.

$$\begin{aligned}
u_i(\Sigma_n') - u_i(\Sigma_n) = & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma_n') - \lambda_H^{\mathbf{A}}(\Sigma_n))(v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})) \\
& + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma_n') - \lambda_L^{\mathbf{R}}(\Sigma_n))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})).
\end{aligned}$$

By assigning the bounds above, we have

$$\begin{aligned} u_i(\Sigma'_n) - u_i(\Sigma_n) &\geq P_H \cdot (1 - \frac{\delta \cdot P_H P_L}{2B} - 1 + \delta) \cdot 1 + P_L \cdot (-\frac{\delta \cdot P_H P_L}{2B}) \cdot B \\ &\geq P_H \cdot \delta - \frac{\delta \cdot P_H P_L}{2} \\ &> \frac{\delta \cdot P_H P_L}{2}. \end{aligned}$$

The similar reasoning can be achieved for $\lambda_L^{\mathbf{R}}(\Sigma_n) \leq 1 - \delta$.

Therefore, all the contingent agents have incentives to deviate from Σ_n to Σ'_n to get a constant gain. This means that Σ_n is NOT an ε-strong BNE for some constant ε.

B.1.4.1 Proof for non-regular profiles.

We adopt the same idea to construct a strategy profile where the deviating group forces the state to be revealed by the first-round vote and votes for the preferred alternative in the second round. Then we show that as the fidelity converges to 1, the group has incentives to do so.

For non-regular profiles, a deviating group with only contingent agents may not work, as there is no assumption on the behavior of predetermined agents in the second round. (For example, if $\alpha_C < 0.5$, and all the predetermined agents always vote for **A**, then the winner will always be **A** regardless of the votes from contingent agents.) Fortunately, the following lemma shows that if the fidelity of the deviating strategy profile is close to 1, then either all the friendly agents get an expected utility increase or all the unfriendly agents get an expected utility increase. Moreover, which type gets utility gain does not depend on the deviating strategy Σ'_n.

Lemma 2. *Let $\{\Sigma_n\}$ be a sequence of profiles such that there exists a constant δ and an infinite set $\mathcal{N}$ such that for all $n \in \mathcal{N}$, $A(\Sigma_n) \leq 1 - \delta$. Then there exists an infinite set $\mathcal{N}' \subseteq \mathcal{N}$ such that for each $n \in \mathcal{N}'$, at least one of the following holds for all sequences of profiles $\{\Sigma'_n\}$ with $\lim_{n\to\infty} A(\Sigma') = 1$:*

1. For any friendly agent i_1, $u_{i_1}(\Sigma'_n) - u_{i_1}(\Sigma_n) \geq 0$.

2. For any unfriendly agent i_2, $u_{i_2}(\Sigma'_n) - u_{i_2}(\Sigma_n) \geq 0$.

Moreover, for a fixed n and Σ_n, either 1 holds for all Σ'_n or 2 holds for all Σ'_n.

The proof of Lemma 2 comes from the fact that a predetermined agent has different sensitivities to utility changes in different states of the world. For a friendly agent i_1, $v_{i_1}(H, \mathbf{A}) > v_{i_1}(L, \mathbf{A}) > v_{i_1}(L, \mathbf{R}) > v_{i_1}(H, \mathbf{R})$. Therefore, i_1 is more sensitive to utility change in world state H than that in world state L. On the other hand, an unfriendly agent i_2 is more sensitive in world state L. This difference in sensitivity guarantees that when the fidelity increases, the expected utility of both types of agents cannot decrease at the same time.

Proof of Lemma 2. Without loss of generality, suppose $u_{i_1}(\Sigma'_n) - u_{i_1}(\Sigma_n) \leq 0$ holds for i_1 (which means (1) does not hold). We show that (2) $u_{i_2}(\Sigma'_n) - u_{i_2}(\Sigma_n) > 0$ must hold for i_2. Recall that the expected utility of an agent i is defined as follows.

$$\begin{aligned} u_i(\Sigma) = & P_L(\lambda_L^{\mathbf{A}}(\Sigma) \cdot v_i(L, \mathbf{A}) + \lambda_L^{\mathbf{R}}(\Sigma) \cdot v_i(L, \mathbf{R})) \\ & + P_H(\lambda_H^{\mathbf{A}}(\Sigma) \cdot v_i(H, \mathbf{A}) + \lambda_H^{\mathbf{R}}(\Sigma) \cdot v_i(H, \mathbf{R})). \end{aligned}$$

Therefore, the difference between $u_i(\Sigma'_n)$ and $u_i(\Sigma_n)$ can be represented as follows.

$$\begin{aligned} & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma') - \lambda_H^{\mathbf{A}}(\Sigma))(v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})) \\ & + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma') - \lambda_L^{\mathbf{R}}(\Sigma))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})). \end{aligned}$$

Since $A(\Sigma_n) \leq 1 - \delta$, at least one of $\lambda_H^{\mathbf{A}}(\Sigma_n) \leq 1 - \delta$ and $\lambda_L^{\mathbf{R}}(\Sigma_n) \leq 1 - \delta$ holds. We deal with two cases separately.

Case 1: $\lambda_H^{\mathbf{A}}(\Sigma_n) \leq 1 - \delta$. In this case, we have $\lambda_H^{\mathbf{A}}(\Sigma'_n) - \lambda_H^{\mathbf{A}}(\Sigma_n) > 0$. Note that for the friendly agent i_1, $v_{i_1}(H, \mathbf{A}) - v_{i_1}(H, \mathbf{R}) \geq -(v_{i_1}(L, \mathbf{R}) - v_{i_1}(L, \mathbf{A}))$. Therefore, $u_{i_1}(\Sigma') -$

$u_{i_1}(\Sigma) \leq 0$ implies

$$P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma_n') - \lambda_H^{\mathbf{A}}(\Sigma_n)) \leq P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma_n') - \lambda_L^{\mathbf{R}}(\Sigma_n)).$$

On the other hand, for any unfriendly agent i_2, $-v_{i_2}(H, \mathbf{A}) - v_{i_2}(H, \mathbf{R}) \leq v_{i_2}(L, \mathbf{R}) - v_{i_2}(L, \mathbf{A})$. Therefore, $u_{i_2}(\Sigma_n') - u_{i_2}(\Sigma_n) \geq 0$.

Case 2: $\lambda_L^{\mathbf{R}}(\Sigma_n) \leq 1-\delta$**.** In this case, we have $\lambda_L^{\mathbf{R}}(\Sigma_n') - \lambda_L^{\mathbf{R}}(\Sigma_n) > 0$. If $\lambda_H^{\mathbf{A}}(\Sigma_n') - \lambda_H^{\mathbf{A}}(\Sigma_n) \leq 0$, then

$$u_{i_2}(\Sigma_n') - u_{i_2}(\Sigma_n) \geq P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma_n') - \lambda_L^{\mathbf{R}}(\Sigma_n))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})) > 0.$$

If $\lambda_H^{\mathbf{A}}(\Sigma_n') - \lambda_H^{\mathbf{A}}(\Sigma_n) > 0$, we apply the same reasoning as in Case 1. Therefore, $u_{i_2}(\Sigma_n') - u_{i_2}(\Sigma_n) \geq 0$ always holds. $\square$

Now we construct the deviating strategy profile sequence $\{\Sigma_n'\}$, which is a combination of two strategy profile sequences $\{\Sigma_F\}$ and $\{\Sigma_U\}$. In both Σ_F and Σ_U contingent agents collaborate with one type of predetermined agents and deviate together.

In Σ_F, all the friendly and contingent agents vote informatively in the first round and play a threshold strategy in the second round.

- All *unfriendly* agents play the same strategy as in Σ.
- If $\phi_H^U(\Sigma_n) \geq \phi_L^U(\Sigma_n)$, all friendly agents and contingent agents vote informatively in the first round. For the second round, all friendly agents and contingent agents vote for $\mathbf{A}$ if $x \geq \frac{\phi_H(\Sigma_F)+\phi_L(\Sigma_F)}{2}$ and vote for $\mathbf{R}$ otherwise.
- If $\phi_H^U(\Sigma_n) < \phi_L^U(\Sigma_n)$, friendly agents and contingent agents "reverse" their strategy. In the first round, a friendly (or contingent) agent votes for $\mathbf{A}$ if receiving l and for $\mathbf{R}$ if receiving h. In the second round, all friendly agents and contingent agents vote for $\mathbf{A}$ if $x \geq \frac{\phi_H(\Sigma_F)+\phi_L(\Sigma_F)}{2}$

Σ_U follows the same "informative and threshold" pattern, except that the deviators are all the unfriendly agents and all the contingent agents.

Since $\alpha_F < 0.5$ and $\alpha_U < 0.5$, if contingent agents and one type of predetermined agents vote for the same alternative, that alternative will be the winner. Therefore, with similar reasoning to the regular case, both Σ_F and Σ_U have fidelity converging to 1, and contingent agents have incentives to deviate to either of them.

Applying Lemma 2, we know that for each n where the fidelity of Σ_n does not converge to 1, there is one type of predetermined agents whose expected utilities will increase if the deviators deviate to either Σ_F or Σ_U. Moreover, the type of agents that get higher expected utilities does not depend on which deviating strategy they play. Therefore, when friendly agents get higher expected utilities, the deviators are all the friendly agents and contingent agents with $\Sigma'_n = \Sigma_F$; otherwise, the deviators are all the unfriendly agents and contingent agents with $\Sigma'_n = \Sigma_U$. Then, the expected utilities of contingent increase by at least a constant, and those of the deviating predetermined agents increase. Therefore, for each Σ_n whose fidelity is not close to 1, the deviators have incentives to deviate to either Σ_F or Σ_U. Therefore, Σ_n is NOT an ε-strong BNE with constant ε.

□

B.1.5 Formal Statement and Proof of Theorem 10

Theorem 10. *Let $\{\hat{\Sigma}_n\}$ be a one-round regular strategy profile sequence in an environment sequence $\{\mathcal{I}_n\}$. If for all n, $\hat{\Sigma}$ is an $\bar{\varepsilon}$-strong BNE in one-round voting with $\bar{\varepsilon}$ converges to 0, then for any two-round strategy profile sequence $\{\Sigma_n\}$ in environment $\{\mathcal{I}_n\}$, where for any n, Σ_n is a second-round consistent strategy profile of $\hat{\Sigma}_n$, Σ_n is an ε-strong BNE in the two-round voting for all n with ε converging to 0.*

Proof. The proof proceeds in two steps. Firstly, we show that the fidelity of $\{\Sigma_n\}$ converges to 1. Secondly, by applying Lemma 5, we know that a deviating group either contains only friendly agents or only unfriendly agents. Finally, we show that such a deviating group cannot succeed.

Step 1: $A(\Sigma_n)$ converges to 1. In Σ_n, the behavior and result in the first-round voting have no effect to the second-round voting that determines the winner. Agents in the second round behave exactly as they are playing a one-round voting game. Therefore, $A(\Sigma_n) = A(\hat{\Sigma}_n)$. Then since $\hat{\Sigma}_n$ is an $\bar{\varepsilon}$-strong BNE with $\bar{\varepsilon}$ converges to 0, we have $A(\hat{\Sigma}_n)$ converges to 1 according to the if and only if characterization in Han et al. [39, Theorem 2]. Therefore, $A(\Sigma_n)$ also converges to 1.

Step 2: equilibrium. For each n, let $\varepsilon = 2B(B+2)(1 - A(\Sigma_n))$. Fix an arbitrary n. Suppose there exists a strategy profile Σ_n' such that a deviating group D of agents have incentives to deviate to Σ_n', then by Lemma 5, D either contains only friendly agents or only unfriendly agents.

Then we show that such D cannot succeed. This follows the last part of Lemma 4. Without loss of generality, suppose D contains only friendly agents. Since $\hat{\Sigma}_n$ is a regular profile, we know that in Σ_n, friendly agents always vote for **A** in the second round. Then friendly agents cannot make themselves better off. The changes in the first-round voting have no effect on the second round, and changes in the second-round voting will only decrease the votes for **A** and cannot increase the probability that **A** becomes the winner. Therefore, for any friendly agents, the expected utility cannot increase by ε.

Therefore, we have shown that for each n, Σ_n is an ε-strong BNE with $\varepsilon = 2B(B+2)(1 - A(\Sigma_n))$. Since the fidelity converges to 1, ε will converge to 0. □

B.2 More Details on the experiments

B.2.1 Configuration

The experiment is run on a Lenovo ThinkBook 15 G3 AGL with AMD Ryzen 5 5600U with Radeon Graphics 2.30 GHz. The program is written in Python 3.9.

We adopt online APIs to create generative AI agents. The API for Deepseek R1 is provided by Together AI, and the API for GPT-4o is provided by OpenAI's official API platform.

B.2.2 Examples for prompts and answers of agents.

Here we give two examples of the prompts and the answers of the agents. More data can be found in the supplementary materials.

B.2.2.1 Deepseek R1, biased signal, accurate information, and one-round voting.

Signal. Bad Impression.

Prompt. "Consider the following hypothetical situation. Suppose you are a recruiter in a company. Your company is hiring for a new position by voting. There are in total 40 voters. You believe that the candidate should be hired only if he is qualified. Most of the company agrees with you, but a small minority <about 12%> seem to support hiring every candidate regardless of qualifications, and another small minority <about 7%> does not support hiring any candidate regardless of qualifications. Before meeting with the candidate, you thought she has <about 50% chance> to be qualified. If the candidate is qualified, you will likely have a good impression after meeting with her <about 80% chance>. If the candidate is not qualified, you will be less likely to have a good impression <about 60% chance>. It seems that everyone is on the same page about both the candidate's likelihood of being qualified and how informative an interview is, but the interview impression is private and independent among all the voters. You got a BAD impression after meeting the candidate. After interviewing, but before any discussions, the company holds a vote. There will be only one round, where each voter votes YES or NO on whether to hire this candidate. The outcome is determined by whether the candidate gets a majority of yes votes. Analyze what is your best vote. The voting time is limited, so you cannot think everlastingly. Summarize and make an immediate decision in two minutes. Your answer should start with the analysis and end with a single character Y or N in a separate line without any other characters or symbols, indicating your vote. ",

Vote. N.

Answer. The candidate has a 33.3% posterior probability of being qualified after a bad impression. With 5 automatic YES and 3 NO votes, the remaining 32 voters' expected YES votes (weighted by qualification likelihood) sum to $21.33 + 5 = 26.33$, likely securing a majority. However, given the higher probability the candidate is unqualified (66.7%), voting NO aligns with preventing unqualified hires despite the likely majority.

B.2.2.2 Deepseek R1, biased signal, vague information, and two-round voting.

Here we give an example where an agent in two-round voting switches its vote in the second round.

Signal. Good Impression.

First-round Prompt. Consider the following hypothetical situation. Suppose you are a recruiter in a company. Your company is hiring for a new position by voting. There are in total 40 voters. You believe that the candidate should be hired only if he is qualified. Most of the company agrees with you, but a small minority seem to support hiring every candidate regardless of qualifications, and another small minority does not support hiring any candidate regardless of qualifications. Before meeting with the candidate, you thought her likelihood of being qualified was roughly balanced. If the candidate is qualified, you would almost certainly form a favorable impression after meeting her. If the candidate is not qualified, you would be far less inclined to have a positive impression, with odds comparable to a rare occurrence. It seems that everyone is on the same page about both the candidate's likelihood of being qualified and how informative an interview is, but the interview impression is private and independent among all the voters. You got a GOOD impression after meeting the candidate. After interviewing, but before any discussions, the company holds a vote. There will be two rounds. In each round, a voter votes yes or no. In the first round, the individual votes are private, but the number of votes will be announced to all the voters. This round does not determine the outcome. In the second round, the voting is also private; the outcome is determined by whether the candidate gets a majority of YES votes. Analyze what your best vote is in the first round. The voting time is limited, so you cannot think everlastingly. Summarize and make an immediate decision in two minutes. Your answer should start with the analysis and end with a single character Y or N in a separate line without any other characters or symbols, indicating your vote.

First-round Vote. Y.

First-round Answer. To determine the optimal first-round vote, consider Bayesian updating: a GOOD impression strongly indicates the candidate is qualified (91% probability). The first-round vote count informs the second-round voters. Voting Y truthfully contributes to an accurate signal, aiding rational voters in updating their beliefs. Extremists' votes are fixed, but their impact is diluted if rational voters coordinate via the first-round tally. Truthful voting maximizes collective accuracy.

Second-round Prompt. Consider the following hypothetical situation. Suppose you are a recruiter in a company. Your company is hiring for a new position by voting. There are in total 40 voters. You believe that the candidate should be hired only if he is qualified. Most of the company agrees with you, but a small minority seem to support hiring every candidate regardless of qualifications, and another small minority does not support hiring any candidate regardless of qualifications. Before meeting with the candidate, you thought her likelihood of being qualified was roughly balanced. If the candidate is qualified, you would almost certainly form a favorable impression after meeting her. If the candidate is not qualified, you would be far less inclined to have a positive impression, with odds comparable to a rare occurrence. It seems that everyone is on the same page about both the candidate's likelihood of being qualified and how informative an interview is, but the interview impression is private and independent among all the voters. You got a GOOD impression after meeting the candidate. After interviewing, but before any discussions, the company holds a vote. There will be two rounds. In each round, a voter votes yes or no. In the first round, the individual votes are private, but the number of votes will be announced to all the voters. This round does not determine the outcome. In the second round, the voting is also private; the outcome is determined by whether the candidate gets a majority of YES votes. After the first round, 14 of 40 voters voted yes. Analyze what is your best vote in the second round. The voting time is limited, so you cannot think everlastingly. Summarize and make an immediate decision in two minutes. Your answer should start with the analysis and end with a single character Y or N in a separate line without any other characters or symbols, indicating your vote.

Second-round Vote. N.

Second-round Answer. After observing 14/40 first-round YES votes, Bayesian updating suggests the candidate is likely unqualified. Given rational voters' independent impressions, a low YES count aligns more with an unqualified candidate (expected 8-12 YES) than a qualified (2̃8-32). Despite my GOOD impression, the aggregate signal outweighs individual likelihood due to extreme unlikelihood of 14 YES under qualification. Vote NO to align with probable unqualification.

APPENDIX C

TECHNICAL APPENDIX FOR CHAPTER 5

C.1 Proof of Lemma 7

Proof. Suppose the strategies other agents play are Σ_{-i}. The only scenario where i's vote can change the outcome is the pivotal case, where both alternatives need exactly one vote to pass their threshold. Therefore, different strategies i plays only affect his/her expected utility conditioned on i being pivotal. The ex-ante expected utility of i with strategy profile $\Sigma = (\sigma, \Sigma_{-i})$ conditioned on him/her being pivotal is as follows. Note that the posterior belief on the world state conditioned on i being pivotal does not depend on i's strategy.

$$\begin{aligned} u_i(\Sigma \mid piv) = &\Pr[L \mid piv, \Sigma_{-i}] \cdot ((P_{hL} \cdot \beta_h + P_{\ell L} \cdot \beta_\ell) \cdot v_i(L, \mathbf{A}) \\ &+ (P_{hL} \cdot (1 - \beta_h) + P_{\ell L} \cdot (1 - \beta_\ell)) \cdot v_i(L, \mathbf{R})) \\ &+ \Pr[H \mid piv, \Sigma_{-i}] \cdot ((P_{hH} \cdot \beta_h + P_{\ell H} \cdot \beta_\ell) \cdot v_i(H, \mathbf{A}) \\ &+ (P_{hH} \cdot (1 - \beta_h) + P_{\ell H} \cdot (1 - \beta_\ell)) \cdot v_i(H, \mathbf{R})). \end{aligned}$$

When agent i plays a different strategy $\sigma' = (\beta'_\ell, \beta'_h)$, the expected utility conditioned on pivotal can be written in the similar form. Let $\Sigma' = (\sigma', \Sigma_{-i})$. Then, the difference between the expected utility when i plays σ and σ' is exactly the expected utility difference conditioned on i being pivotal.

$$\begin{aligned} u_i(\Sigma') - u_i(\Sigma) &= u_i(\Sigma' \mid piv) - u_i(\Sigma \mid piv) \\ &= \Pr[L \mid piv, \Sigma_{-i}] \cdot ((P_{hL} \cdot (\beta'_h - \beta_h) + P_{\ell L} \cdot (\beta'_\ell - \beta_\ell)) \cdot (v_i(L, \mathbf{A}) - v_i(L, \mathbf{R})) \\ &+ \Pr[H \mid piv, \Sigma_{-i}] \cdot ((P_{hH} \cdot (\beta'_h - \beta_h) + P_{\ell H} \cdot (\beta'_\ell - \beta_\ell)) \cdot (v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})). \end{aligned}$$

For majority agents, $v_i(L, \mathbf{A}) - v_i(L, \mathbf{R}) < 0$ and $v_i(H, \mathbf{A}) - v_i(H, \mathbf{R}) > 0$. When $\beta_h < 1$ and $\beta_\ell > 0$, let $\beta'_h = \beta_h + \varepsilon$ and $\beta'_\ell = \beta_\ell - \varepsilon \cdot \frac{P_{hL}+P_{hH}}{P_{\ell L}+P_{\ell H}}$, where $\varepsilon > 0$ is a small constant that guarantees $\beta_h \leq 1$ and $\beta_\ell \geq 0$. Then we have $P_{hL} \cdot (\beta'_h - \beta_h) + P_{\ell L} \cdot (\beta'_\ell - \beta_\ell) < 0$ and $P_{hH} \cdot (\beta'_h - \beta_h) + P_{\ell H} \cdot (\beta'_\ell - \beta_\ell) > 0$, which implies that $u_i(\Sigma') - u_i(\Sigma) > 0$ and that σ is dominated by σ'. Now, without loss of generality, suppose $\beta_h = 1$. In this case, for all $\sigma' \neq \sigma$, either $P_{hL} \cdot (\beta'_h - \beta_h) + P_{\ell L} \cdot (\beta'_\ell - \beta_\ell) > 0$ or $P_{hH} \cdot (\beta'_h - \beta_h) + P_{\ell H} \cdot (\beta'_\ell - \beta_\ell) < 0$ hold. (If $P_{hH} \cdot (\beta'_h - \beta_h) + P_{\ell H} \cdot (\beta'_\ell - \beta_\ell) < 0$ does not hold and $\sigma' \neq \sigma$, then $\beta'_\ell > \beta_\ell$ must hold. Combined with the fact that $P_{hH} > P_{hL}$ and $P_{\ell H} < P_{\ell L}$, $P_{hL} \cdot (\beta'_h - \beta_h) + P_{\ell L} \cdot (\beta'_\ell - \beta_\ell) > 0$ holds.) Therefore, by constructing Σ_{-i} so that $\Pr[L \mid piv, \Sigma_{-i}]$ is sufficiently close to 0 or 1, we can find a scenario where $u_i(\Sigma') - u_i(\Sigma) < 0$, which implies that Σ is not weakly dominated by Σ'. The reasoning for $\beta_\ell = 0$ and that for the minority agents resembles the reasoning above. □

C.2 Expected Vote share and fidelity.

[39] provides useful tool to determine the fidelity of a profile sequence by the *expected vote share* of alternative **A** under different world states.

Definition 12 (Expected vote share). Given an instance of n agents, and a strategy profile Σ, let random variable X_i^n be "agent i votes for **A**": $X_i^n = 1$ if agent i votes for **A**, and $X_i^n = 0$ if i votes for **R**. Then the expected vote share (of **A**) is defined as follows:

$$\varphi_H = \frac{1}{n} \sum_{i=1}^{n} E[X_i^n \mid H]. \tag{C.1}$$

$$\varphi_L = \frac{1}{n} \sum_{i=1}^{n} E[X_i^n \mid L]. \tag{C.2}$$

Specifically, φ_H (φ_L, respectively) is the expected vote share of **A** conditioned on world state H (world state L, respectively).

Note that when the majority vote winning threshold is $\frac{1}{2}$, $\varphi_H = f_H + \frac{1}{2}$,and $\varphi_L = f_L + \frac{1}{2}$. For the convenience of presentation, we rewrite Theorem 4 into expected vote share.

Theorem 4. *Given an arbitrary sequence of instances and arbitrary sequence of strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$, let φ_H^n and φ_L^n be the expected vote share for each Σ_n. Let*

$$F = \liminf_{n\to\infty} \sqrt{n} \cdot \min\left(\varphi_H^n - \frac{1}{2}, \frac{1}{2} - \varphi_L^n\right)$$

- *If $F = +\infty$, the fidelity of Σ_n converges to 1 , i.e., $\lim_{n\to\infty} A(\Sigma_n) = 1$.*
- *If $F < 0$ (including $-\infty$), $A(\Sigma_n)$ does NOT converge to 1.*
- *If $F \geq 0$ (not including $+\infty$), and the variance of $\sum_{i=1}^{n} X_i^n$ is $\Theta(n)$ under every world state, $A(\Sigma_n)$ does NOT converge to 1.*

C.3 Proof of Theorem 11

Proof. When $\xi \leq \frac{1}{2}$, for each n, let each Σ_n be all the agents unanimously vote for **A**. According to Lemma 7, no agents play a weakly dominated strategy. When the world state is L, the winner is still A, so $\lambda_L^{\mathbf{R}}(\Sigma_n) = 0$, which implies that $A(\Sigma_n) \leq P_H < 1$. Finally, since $\xi \leq \frac{1}{2}$, any deviation group contains at most $\frac{n}{2}$ agents. Consequently, any possible deviating strategy still contains at least $\frac{n}{2}$ votes for **A**, which guarantees that **A** is the winner. Since the outcome of the vote cannot be changed, no agent can get an expected utility gain exceeding any $\varepsilon > 0$. Therefore, Σ_n is a $\frac{n}{2}$-strong equilibrium for any n.

When $\xi > \frac{1}{2}$, we show that a group of majority agents can always successfully deviate to a deviating strategy profile that asymptotically reaches the informed majority decision with probability converging to 1.

Suppose $\{\Sigma_n\}$ is a strategy profile sequence such that $A(\Sigma_n)$ does not converge to 1. This implies that there exists a constant $\delta > 0$ and an infinite set $N_1 \subseteq N_{\geq 0}$ such that for all $n \in N_1$, $1 - A(\Sigma_n) \geq \delta$. Now we construct the deviating strategy sequence $\{\Sigma_n'\}$. We fix an arbitrary n. Let D be a set of majority agents with $D \geq \xi \cdot n - 1$. Given that $\xi > \frac{1}{2}$, $|D| > \frac{1}{2}$ for all sufficiently large n. For all agents not in D, let $\bar{\beta_h} = \frac{1}{n-|D|}\sum_{i\notin D}\beta_h^i$ and $\bar{\beta_\ell} = \frac{1}{n-|D|}\sum_{i\notin D}\beta_\ell^i$ be the average strategy of all such agents. Now we define the strategy of agents in D. All agents play the same strategy with $\beta_h = \frac{n}{2|D|} + \bar{\beta_h}\cdot(1-\frac{n}{|D|}) + \psi$, and $\beta_\ell = \frac{n}{2|D|} + \bar{\beta_h}\cdot(1-\frac{n}{|D|}) - \psi$. $\psi > 0$ is a small constant that still keeps β_h and β_ℓ between $[0,1]$. This construction averages the two groups of agents described in the sketch and creates a symmetric deviating strategy. With the condition that $|D| > \frac{n}{2}$, it is not hard to show that such ψ exists. In this way, we figure out the expected vote share of Σ_n.

$$\varphi_H = \frac{1}{2} + \frac{|D|}{n}\cdot\psi\cdot(\frac{P_{hH}}{P_{hH}+P_{hL}} - \frac{P_{\ell H}}{P_{\ell L}+P_{\ell H}}) > \frac{1}{2}.$$
$$\varphi_L = \frac{1}{2} + \frac{|D|}{n}\cdot\psi\cdot(\frac{P_{hL}}{P_{hH}+P_{hL}} - \frac{P_{\ell L}}{P_{\ell L}+P_{\ell H}}) < \frac{1}{2}.$$

Both inequalities hold because $\Delta > 0$. Therefore, by applying the first case of Theorem 4, the fidelity of Σ'_n converges to 1. Therefore, the majority agents in D get a constant utility increase under infinitely many n, and $\{\Sigma_n\}$ fails to be an equilibrium sequence with ε converging to 0. $\square$

C.4 Technical Lemma on reaching informed majority decision.

Here we give some useful results that will be applied in the full proof.

Below are the lemmas that we apply to deal with the case where the variance of a strategy profile becomes $o(n)$.

Lemma 13. *For any instance of n agents and an arbitrary strategy profile Σ, let $\delta_H = \frac{1}{n}Var(\sum_{i=1}^n X_i^n \mid H)$ and $\delta_L = \frac{1}{n}Var(\sum_{i=1}^n X_i^n \mid L)$, then $\varphi_H - \varphi_L \leq \frac{2\cdot\Delta\cdot\delta_H}{\min(P_{hH},P_{\ell H})}$ and $\varphi_H - \varphi_L \leq \frac{2\cdot\Delta\cdot\delta_L}{\min(P_{hL},P_{\ell L})}$.*

Proof. For each i, we denote agent i' strategy as $(\beta_\ell^i, \beta_h^i)$. In world state H, $X_i^n = 1$ with probability $P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i$. Therefore,

$$\begin{aligned}
\delta_H &= \frac{1}{n} Var(\sum_{i=1}^{n} X_i^n \mid H) \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} Var(X_i^n \mid H) \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} (P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i) \cdot (1 - P_{hH} \cdot \beta_h^i - P_{\ell H} \cdot \beta_\ell^i).
\end{aligned}$$

Similarly, when the world state is L,

$$\delta_L = \frac{1}{n} \cdot \sum_{i=1}^{n} (P_{hL} \cdot \beta_h^i + P_{\ell L} \cdot \beta_\ell^i) \cdot (1 - P_{hL} \cdot \beta_h^i - P_{\ell L} \cdot \beta_\ell^i).$$

On the other hand,

$$\begin{aligned}
\varphi_H - \varphi_L &= \frac{1}{n} \sum_{i=1}^{n} (E[X_i^n \mid H] - E[X_i^n \mid L]) \\
&= \frac{1}{n} \sum_{i=1}^{n} \cdot \Delta \cdot (\beta_h^i - \beta_\ell^i).
\end{aligned}$$

Let $\delta_H^i = (P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i) \cdot (1 - P_{hH} \cdot \beta_h^i - P_{\ell H} \cdot \beta_\ell^i)$. If $P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i \leq \frac{1}{2}$, $\delta_H^i \geq \frac{1}{2}(P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i)$. Therefore, $\beta_h^i - \beta_\ell^i \leq \beta_h^i \leq \frac{2\delta_H^i}{P_{hH}}$. If $P_{hH} \cdot \beta_h^i + P_{\ell H} \cdot \beta_\ell^i > \frac{1}{2}$, $\delta_H^i > \frac{1}{2}(1 - P_{hH} \cdot \beta_h^i - P_{\ell H} \cdot \beta_\ell^i)$, then

$$\begin{aligned}
\beta_h^i - \beta_\ell^i =& (1 - \beta_\ell^i) - (1 - \beta_h^i) \\
&\leq 1 - \beta_\ell^i \\
&\leq \frac{2\delta_H^i}{P_{\ell H}}.
\end{aligned}$$

Similarly, for world state L, we also have $\beta_h^i - \beta_\ell^i \leq \frac{2\delta_L}{P_{\ell L}}$ or $\beta_h^i - \beta_\ell^i \leq \frac{2\delta_L}{P_{hL}}$. Note that Therefore, by summing up all the i, we have $\varphi_H - \varphi_L \leq \frac{2 \cdot \Delta \cdot \delta_H}{\min(P_{hH}, P_{\ell H})}$ and $\varphi_H - \varphi_L \leq \frac{2 \cdot \Delta \cdot \delta_L}{\min(P_{hL}, P_{\ell L})}$. □

Lemma 14. *For any instance of n agents and an arbitrary strategy profile Σ where all* *agents play non-dominated strategies, let maj be the set of all majority agents,* $\delta_H^{maj} = \frac{1}{n}Var(\sum_{i\in maj} X_i^n \mid H)$ *and* $\delta_L^{maj} = \frac{1}{n}Var(\sum_{i\in maj} X_i^n \mid L)$*. Then* $\delta_H^{maj} \geq \frac{\min(P_{\ell H}, P_{hH})}{2\cdot\Delta}\cdot(\varphi_H - \varphi_L)$*, and* $\delta_L^{maj} \geq \frac{\min(P_{\ell L}, P_{hL})}{2\cdot\Delta}\cdot(\varphi_H - \varphi_L)$*.*

Proof. The proof resembles that of Lemma 13. For each i, we denote agent i' strategy as $(\beta_\ell^i, \beta_h^i)$. In world state H, $X_i^n = 1$ with probability $P_{hH}\cdot\beta_h^i + P_{\ell H}\cdot\beta_\ell^i$. Therefore,

$$Var(X_i^n \mid H) = (P_{hH}\cdot\beta_h^i + P_{\ell H}\cdot\beta_\ell^i)(1 - P_{hH}\cdot\beta_h^i - P_{\ell H}\cdot\beta_\ell^i).$$

Likewise,

$$Var(X_i^n \mid L) = (P_{hL}\cdot\beta_h^i + P_{\ell L}\cdot\beta_\ell^i)(1 - P_{hL}\cdot\beta_h^i - P_{\ell L}\cdot\beta_\ell^i).$$

By the proof of Lemma 13, we know that $\beta_h^i - \beta_\ell^i \leq \frac{2Var(X_i^n|H)}{\min(P_{\ell H}, P_{hH})}$ and $\beta_h^i - \beta_\ell^i \leq \frac{2Var(X_i^n|L)}{\min(P_{\ell L}, P_{hL})}$. Now we sum up the inequalities for all $i \in maj$. This will be

$$\frac{2\delta_H^{maj}\cdot n}{\min(P_{\ell H}, P_{hH})} \geq \sum_{i\in maj}\beta_h^i - \beta_\ell^i,$$
$$\frac{2\delta_L^{maj}\cdot n}{\min(P_{\ell L}, P_{hL})} \geq \sum_{i\in maj}\beta_h^i - \beta_\ell^i.$$

On the other hand, $\varphi_H - \varphi_L = \frac{1}{n}\cdot\sum_{i=1}^n \Delta\cdot(\beta_h^i - \beta_\ell^i)$. For all the minority playing non-dominated strategies, $\beta_h^i = 0$ or $\beta_\ell^i = 1$ holds, which implies that $\beta_h - \beta_\ell \leq 0$. Therefore, $\varphi_H - \varphi_L \leq \frac{1}{n}\cdot\sum_{i\in maj}\Delta\cdot(\beta_h^i - \beta_\ell^i)$. Therefore, we have

$$\begin{aligned}\delta_H^{maj} &\geq \frac{\min(P_{\ell H}, P_{hH})}{2\cdot n}\cdot\sum_{i\in maj}\beta_h^i - \beta_\ell^i \\ &\geq \frac{\min(P_{\ell H}, P_{hH})}{2\cdot n}\cdot\frac{n}{\Delta}\cdot(\varphi_H - \varphi_L)\end{aligned}$$

Likewise, $\delta_L^{maj} \geq \frac{\min(P_{\ell L}, P_{hL})}{2\cdot\Delta}\cdot(\varphi_H - \varphi_L)$. □

Lemma 14 implies that, if the majority agents' strategies in the strategy profile sequence have a sub-linear variance, the difference between the expected vote shares is also sub-linear.

Corollary 7. *Let maj be the set of all majority agents. For an arbitrary sequence of instances and arbitrary sequence of strategy profiles $\{\Sigma_n\}_{n=1}^{\infty}$ where all majority agents play non-dominated strategies, if for any $\delta > 0$ there exist infinitely many n such that in some world state k, $Var(\sum_{i\in maj} X_i^n \mid k) < \delta \cdot n$, then for any $\varepsilon > 0$, there exist infinitely many n such that in some world state k, $\varphi_H - \varphi_L < \varepsilon$.*

C.5 Proof of Theorem 12

C.5.1 Lemmas

In this section, we give the technical lemmas to prove Theorem 12. They either form crucial prerequisites for some intermediate results or guarantee some parameters are in the correct range.

Lemma 15 restricts deviations to groups involving only minority agents.

Lemma 15. *Given a strategy profile Σ, if there exists $e > 0$ such that $\lambda_H^{\mathbf{A}}(\Sigma) \geq 1 - e$ and $\lambda_L^{\mathbf{R}}(\Sigma) \geq 1 - e$ holds, then for any deviating strategy profile Σ', at least one of the following holds. (1) No agent has a utility increase more than $\frac{2B(B+1)e}{P_H \cdot P_L}$. (2) Every majority agent gets a strictly negative utility change.*

Proof. The difference on the ex-ante expected utility of agent i under two strategy profiles is

$$\begin{aligned} u_i(\Sigma_n') - u_i(\Sigma_n) = {} & P_H \cdot (\lambda_H^{\mathbf{A}}(\Sigma_n') - \lambda_H^{\mathbf{A}}(\Sigma_n))(v_i(H, \mathbf{A}) - v_i(H, \mathbf{R})) \\ & + P_L \cdot (\lambda_L^{\mathbf{R}}(\Sigma_n') - \lambda_L^{\mathbf{R}}(\Sigma_n))(v_i(L, \mathbf{R}) - v_i(L, \mathbf{A})). \end{aligned}$$

We discuss on different cases on the fidelity of Σ_n'.

Suppose $A(\Sigma_n') \geq 1 - \frac{(B+1)e}{P_H \cdot P_L}$. For a majority agent i, $u_i(\Sigma_n') - u_i(\Sigma_n) \leq P_H \cdot e \cdot B + P_L \cdot e \cdot B = Be$. For a minority agent i, $u_i(\Sigma_n') - u_i(\Sigma_n) \leq P_H \cdot \frac{(B+1)e}{P_H^2 \cdot P_L} \cdot B + P_L \cdot \frac{(B+1)e}{P_H \cdot P_L^2} \cdot B = \frac{2B(B+1)e}{P_H \cdot P_L}$. Therefore, no agent can get expected utility gain more than $\frac{2B(B+1)e}{P_H \cdot P_L}$.

Suppose $A(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$. We first show that majority agents will not join the deviating group. Then we show that at most $\xi \cdot n$ minority agents cannot successfully deviate. $A(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$, at least one of $\lambda_H^{\mathbf{A}}(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$ and $\lambda_L^{\mathbf{R}}(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$ holds. When $\lambda_H^{\mathbf{A}}(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$, the utility difference of a majority agent i will be $u_i(\Sigma'_n) - u_i(\Sigma_n) < P_H \cdot (e - \frac{(B+1)e}{P_H \cdot P_L}) \cdot 1 + P_L \cdot e \cdot B < (P_H - 1) \cdot e + (P_L - 1) \cdot B \cdot e < 0$. When $\lambda_L^{\mathbf{R}}(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$ holds, the utility difference of a majority agent i is also strictly negative according to similar reasoning. Therefore, a majority agent has no incentives to join the deviation group. $\square$

The following lemma shows that the deviation succeeds if and only if the fidelity of the deviating strategy does not converge to 1.

Lemma 16. *Let $\{\Sigma_n\}$ be an arbitrary sequence of strategy profile such that $A(\Sigma_n)$ converges to 1 and $\{\Sigma'_n\}$ be a sequence of deviating strategy profile from $\{\Sigma_n\}$ where for each n, the deviating group D contains only minority agents.*

1. *If $\lim_{n\to\infty} A(\Sigma'_n) = 1$, the expected utility increase of the minority agents converges to 0.*

2. *If $\lim_{n\to\infty} A(\Sigma'_n) = 1$ does not hold, there exists a constant ε and infinitely many n such that every minority agent gains at least ε in the expected utility.*

Proof. Consider the utility difference of an arbitrary minority agent between Σ_n and Σ'_n. $A(\Sigma_n)$ converges to 1 implies that $\lambda_H^{\mathbf{A}}(\Sigma_n)$ and $\lambda_L^{\mathbf{R}}(\Sigma_n)$ converges to 1.

When $\lim_{n\to\infty} A(\Sigma'_n) = 1$, $\lambda_H^{\mathbf{A}}(\Sigma'_n)$ and $\lambda_L^{\mathbf{R}}(\Sigma'_n)$ converges to 1. Therefore, for any constant $\varepsilon > 0$, we can find a $N > 0$ such that for all $n > N$, all $\lambda_H^{\mathbf{A}}(\Sigma_n)$, $\lambda_L^{\mathbf{R}}(\Sigma_n)$, $\lambda_H^{\mathbf{A}}(\Sigma'_n)$, and $\lambda_L^{\mathbf{R}}(\Sigma'_n)$ is at least $1 - \varepsilon$

$$u_i(\Sigma'_n) - u_i(\Sigma_n) \leq P_H \cdot \varepsilon \cdot B + P_L \cdot \varepsilon \cdot B = \varepsilon \cdot B.$$

Therefore, the expected utility gain of any minority agents converge to 0.

On the other hand, when $\lim_{n\to\infty} A(\Sigma'_n) = 1$ does not hold, there are a constant $\delta > 0$ and infinitely many n (denote the set as N') in which either $\lambda_H^{\mathbf{A}}(\Sigma'_n) \le 1-\delta$ or $\lambda_L^{\mathbf{R}}(\Sigma'_n) \le 1-\delta$ holds. Without loss of generality, suppose $\lambda_H^{\mathbf{A}}(\Sigma'_n) \le 1-\delta$ hold. On the other hand, for all sufficiently large n, both $\lambda_H^{\mathbf{A}}(\Sigma_n)$ and $\lambda_L^{\mathbf{R}}(\Sigma_n)$ are at least $1 - P_H P_L \cdot \delta/(2B)$. Therefore,

$$\begin{aligned} u_i(\Sigma'_n) - u_i(\Sigma_n) &\ge P_H \cdot (\delta - P_H P_L \cdot \delta/(2B)) \cdot 1 - P_L \cdot P_H P_L \delta/(2B) \cdot B \\ &\ge P_H \cdot \delta \cdot (\frac{2B-1}{2B} - \frac{P_L^2}{2}) \\ &> 0. \end{aligned}$$

Therefore, there exists a constant $\varepsilon = P_H \cdot \delta \cdot (\frac{2B-1}{2B} - \frac{P_L^2}{2})$ and infinitely many n such that every minority agent gain at least ε in the expected utility. □

The following two lemmas guarantee that α_{NL} is between $\frac{1}{2P_{\ell L}}$ and $\frac{1}{1+P_{\ell L}}$.

Lemma 17. $\alpha_{NL} < \frac{1}{2P_{\ell L}}$.

Proof. By manipulating the inequality, $\alpha_{NL} < \frac{1}{2P_{\ell L}}$ is equivalent to

$$(1-P_{hH})^2 + P_{hH}P_{\ell L}^2 + 3P_{hH}P_{\ell L} + P_{\ell L}^2 + 2P_{hL}P_{\ell L}\sqrt{P_{hH}P_{\ell L}P_{hL}P_{\ell H}} - 2P_{hH}P_{\ell L}^3 - P_{hH}^2 P_{\ell L} - 2P_{\ell L} > 0.$$

Given that $P_{hL} < P_{hH} \le P_{\ell L}$, $\sqrt{P_{hH}P_{\ell L}P_{hL}P_{\ell H}} \ge P_{hL}P_{\ell L}$. Therefore, for fixed P_{hL} and $P_{\ell L}$, we consider the following function of P_{hH}

$$\begin{aligned} f(P_{hH}) &= (1-P_{hH})^2 + P_{hH}P_{\ell L}^2 + 3P_{hH}P_{\ell L} + P_{\ell L}^2 + 2P_{hL}^2 P_{\ell L}^2 - 2P_{hH}P_{\ell L}^3 - P_{hH}^2 P_{\ell L} - 2P_{\ell L} \\ &= P_{hL} \cdot P_{hH}^2 - (2P_{\ell L}^3 - P_{\ell L}^2 - 3P_{\ell L} + 2) \cdot P_{hH} + P_{\ell L}^2 + 2P_{hL}^2 P_{\ell L}^2 - 2P_{\ell L} + 1. \end{aligned}$$

It is not hard to verify that $f(P_{hH}) \le LHS$. Therefore, it suffices to show that $f(P_{hH}) > 0$ for all $P_{hL} \le P_{hH} \le P_{\ell L}$. We prove this by showing that $f(P_{hL}) > 0$ and that f increases on P_{hH}.

First, we have

$$
\begin{aligned}
f(P_{hL}) &= P_{\ell L}^2 + P_{hL}P_{\ell L}^2 + 3P_{hL}P_{\ell L} + P_{\ell L}^2 + 2P_{hL}^2P_{\ell L}^2 - 2P_{hL}P_{\ell L}^3 - P_{hL}^2P_{\ell L} - 2P_{\ell L} \\
&= 4P_{\ell L}^2P_{hL}^2 \\
&> 0.
\end{aligned}
$$

Secondly,

$$
\begin{aligned}
\frac{df}{dP_{hH}} &= 2P_{hL} \cdot P_{hH} - (2P_{\ell L}^3 - P_{\ell L}^2 - 3P_{\ell L} + 2) \\
&> 2(1 - P_{\ell L}^2) - (2P_{\ell L}^3 - P_{\ell L}^2 - 3P_{\ell L} + 2) \\
&= P_{\ell L} \cdot (1 - P_{\ell L}) \cdot (2P_{\ell L} - 1) \\
&> 0.
\end{aligned}
$$

Therefore, $LHS \geq f(P_{hH}) > f(P_{hL}) > 0$. This implies $\alpha_{NL} < \frac{1}{2P_{\ell L}}$. $\square$

Lemma 18. $\alpha_{NL} \geq \frac{1}{1+P_{\ell L}}$.

Proof. By manipulating the inequality, $\alpha_{NL} \geq \frac{1}{1+P_{\ell L}}$ is equivalent to

$$
P_{hH}^2 - P_{hH} + 3P_{hH}P_{\ell L}^2 + P_{\ell L}^2 + 2P_{hL}(1 + P_{\ell L})\sqrt{P_{hH}P_{\ell L}P_{hL}P_{\ell H}} - 2P_{hH}P_{\ell L}^3 - P_{hH}^2P_{\ell L} - P_{\ell L} \leq 0.
$$

Likewise, we view $P_{\ell L}$ as fixed and define $f(P_{hH}) = LHS$ be a function of P_{hH}. We show that $f(P_{\ell L}) = 0$ and f is increasing on P_{hH}.

Firstly,

$$
\begin{aligned}
f(P_{\ell L}) &= P_{\ell L}^2 - P_{\ell L} + 3P_{\ell L}^3 + P_{\ell L}^2 + 2(1 - P_{\ell L}^2) \cdot P_{\ell L} \cdot (1 - P_{\ell L}) - 2P_{\ell L}^4 - P_{\ell L}^3 - P_{\ell L} \\
&= 0.
\end{aligned}
$$

Then we consider the derivative of f.

$$\frac{df}{dP_{hH}} = 2(1-P_{\ell L}) \cdot P_{hH} + 3P_{\ell L}^2 - 2P_{\ell L}^3 - 1 + \frac{P_{hL}(1+P_{\ell L}) \cdot \sqrt{P_{hL}P_{\ell L}}}{\sqrt{P_{hH}P_{\ell H}}} \cdot (1-2P_{hH}).$$

We discuss the derivative on $P_{hH} < \frac{1}{2}$ and $P_{hH} \geq \frac{1}{2}$ separately.

When $P_{hL} < P_{hH} < \frac{1}{2}$, $(1-2P_{hH}) \geq 0$. Therefore, given that $\sqrt{P_{hH}P_{\ell H}} \leq \frac{1}{2}$, we have

$$\frac{df}{dP_{hH}} \geq 2(1-P_{\ell L}) \cdot P_{hH} + 3P_{\ell L}^2 - 2P_{\ell L}^3 - 1 + 2P_{hL}(1+P_{\ell L}) \cdot \sqrt{P_{hL}P_{\ell L}} \cdot (1-2P_{hH}).$$

Note that RHS (denoted by $f_1'(P_{hH})$) is linear to P_{hH}. As we aim to show that it suffices to show that $f_1'(P_{hH}) \geq 0$ for all $P_{hL} \leq P_{hH} \leq \frac{1}{2}$, it suffices to show that $f_1'(P_{hH}) \geq 0$ when $P_{hH} = P_{hL}$ and when $P_{hH} = \frac{1}{2}$.

$$\begin{aligned}
f_1'(P_{hL}) &= 2(1-P_{\ell L})^2 - (1-P_{\ell L})^2 \cdot (2P_{\ell L}+1) + 2(1-P_{\ell L})(1+P_{\ell L}) \cdot (2P_{\ell L}-1) \cdot \sqrt{P_{\ell L}(1-P_{\ell L})} \\
&= (2P_{\ell L}-1) \cdot (1-P_{\ell L}) \cdot \sqrt{1-P_{\ell L}} \cdot (2 \cdot (1+P_{\ell L}) \cdot \sqrt{P_{\ell L}} - \sqrt{1-P_{\ell L}}) \\
&> 0. \\
f_1'(\frac{1}{2}) &= 1 - P_{\ell L} + 3P_{\ell L}^2 - 2P_{\ell L}^3 - 1 \\
&= P_{\ell L} \cdot (1-P_{\ell})(2P_{\ell L}-1) \\
&> 0.
\end{aligned}$$

When $\frac{1}{2} \leq P_{hH} \leq P_{\ell L}$, $(1-2P_{hH}) \leq 0$. Therefore, by $P_{hL} < P_{hH} \leq P_{\ell L}$, $\sqrt{P_{hH}P_{\ell H}} \geq \sqrt{P_{hL}P_{\ell L}}$, we have

$$\frac{df}{dP_{hH}} \geq 2(1-P_{\ell L}) \cdot P_{hH} + 3P_{\ell L}^2 - 2P_{\ell L}^3 - 1 + P_{hL}(1+P_{\ell L}) \cdot (1-2P_{hH}).$$

Similarly, the RHS (denoted by $f'2$) is linear to P_{hH}. Therefore, it suffices to show that

$f_2'(\frac{1}{2}) \geq 0$ and $f_2'(P_{\ell L}) \geq 0$.

$$\begin{aligned} f_2'(\frac{1}{2}) &= f_1'(\frac{1}{2}) > 0. \\ f_2'(P_{\ell L}) &= 2(1-P_{\ell L}) \cdot P_{\ell L} + 3P_{\ell L}^2 - 2P_{\ell L}^3 - 1 + (1-P_{\ell L})(1+P_{\ell L})(1-2P_{\ell L}) \\ &= 0. \end{aligned}$$

Therefore, for all $P_{hH} \in (P_{hL}, P_{\ell L}]$, f is increasing on P_{hH}. Therefore, $\alpha_{NL} \geq \frac{1}{1+P_{\ell L}}$. $\square$

The following lemma guarantees the switch between the "Steep" Segment 2 and "Non-Linear" Segment 3 is numerically correct. That is, $\xi_{NL} \geq \frac{\Delta(\alpha-1/2)}{P_{hL}}$ if and only if $\alpha \leq \alpha_{NL}$.

Lemma 19. *For $\frac{1}{2P_{\ell L}} > \alpha > \alpha_{NL}$, $\xi_{NL} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. For $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $\xi_{NL} \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$*

Proof. Rewrite $\xi_{NL} \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ as the following formula.

$$(2\alpha-1)\cdot 2\sqrt{2(1-\alpha P_{\ell H})(1-2\alpha P_{\ell L})P_{\ell H}P_{\ell L}} \leq P_{hL} + (2\alpha-1)\cdot(4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L})$$

The proof proceeds as follows. First, we show that when $P_{hL} + (2\alpha-1)\cdot(4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$ and $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$, the inequality holds at $\alpha_{NL} \geq \alpha > \frac{1}{2}$. Secondly, we show that for all $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $P_{hL} + (2\alpha-1)\cdot(4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$ holds. Lastly, we show that when $\frac{1}{2P_{\ell L}} \geq \alpha > \alpha_{NL}$, the inequality does not hold.

When $P_{hL} + (2\alpha-1)\cdot(4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$, it is equivalent to solve the inequality by squaring both side, i.e.,

$$8(2\alpha-1)^2\cdot((1-\alpha P_{\ell H})(1-2\alpha P_{\ell L})P_{\ell H}P_{\ell L}) \leq (P_{hL} + (2\alpha-1)\cdot(4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}))^2.$$

Refactoring on α, we get the following quadratic inequality.

$$\begin{aligned}
&\alpha^2 \cdot 4(P_{\ell H}^2 + 4P_{hH}P_{\ell L^2}) \\
&- \alpha \cdot 4(2P_{hH}P_{\ell L}^2 + 3P_{hH}P_{\ell L} + P_{hH}^2 - 2P_{hH} - P_{\ell L} + 2) \\
&+ (P_{hL}^2 - 4P_{\ell H}P_{\ell L} + 4P_{\ell L}^2 + P_{\ell H}^2 + 4P_{hL}P_{\ell L} + 2P_{\ell H}P_{hL}) \\
&\geq 0.
\end{aligned}$$

Using the root formula of the quadratic equation, the solution of this quadratic inequality is

$$\alpha \leq \frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 - 2P_{hL} \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2} = \alpha_{NL}$$
$$\alpha \geq \frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 + 2P_{hL} \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2}$$

By the similar reasoning of Lemma 17, we can show that

$$\begin{aligned}
&\frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 + 2P_{hL} \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2} \\
\geq\ &\frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 + 2P_{hL}^2 \cdot P_{\ell L}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2} \\
\geq\ &\frac{1}{2P_{\ell L}}.
\end{aligned}$$

Therefore, when $P_{hL} + (2\alpha - 1) \cdot (4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$ and $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$, the inequality holds at $\alpha_{NL} \geq \alpha > \frac{1}{2}$.

Now we are going to show that when $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $P_{hL} + (2\alpha - 1) \cdot (4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$ holds. Note that this constraint is a quadratic inequality on α. Let $g(\alpha) = P_{hL} + (2\alpha - 1) \cdot (4\alpha P_{\ell H}P_{\ell L} - P_{\ell H} - 2P_{\ell L})$. If we can show that $g(\alpha)$ is monotonically decreasing on α for $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$, and find a $\frac{1}{2P_{\ell L}} \geq \alpha' \geq \alpha_{NL}$ such that $g(\alpha') \geq 0$, we show that $g(\alpha) \geq 0$ for all $\alpha_{NL} \geq \alpha > \frac{1}{2}$. Refactoring g, we have

$$g(\alpha) = \alpha^2 \cdot 8P_{\ell L}P_{\ell L} - \alpha \cdot 2(2P_{\ell L}P_{\ell H} + 2P_{\ell L} + P_{\ell H}) + 2P_{\ell L} + P_{\ell H} + P_{hL}.$$

As a quadratic function, g is decreasing on $\alpha \leq \frac{1}{4} + \frac{1}{4P_{\ell H}} + \frac{1}{8P_{\ell L}}$. Note that

$$\begin{aligned}
\frac{1}{4} + \frac{1}{4P_{\ell H}} + \frac{1}{8P_{\ell L}} &> \frac{1}{4} + \frac{1}{4P_{\ell L}} + \frac{1}{8P_{\ell L}} \\
&\geq \frac{1}{8P_{\ell L}} + \frac{1}{4P_{\ell L}} + \frac{1}{8P_{\ell L}} \\
&= \frac{1}{2P_{\ell L}}.
\end{aligned}$$

Therefore, $g(\alpha)$ is monotonically decreasing on α for $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$.

Now we set

$$\alpha' = \frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 - 2P_{hL}^2 \cdot P_{\ell L}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2}.$$

Note that $\alpha \geq \alpha_{NL}$. And note that Lemma 17 implies $\alpha' \leq \frac{1}{2P_{\ell L}}$ (by showing that $f(P_{hH})$ in Lemma 17 is positive).

$$\begin{aligned}
g(\alpha') = \frac{4P_{\ell L}P_{hL}^2}{(P_{\ell H}^2 + 4P_{hH}P_{\ell L}^2)^2} &\cdot ((1 - P_{hH})^4 + (1 - P_{hH})^2 \cdot (3P_{hH} - 2)) \cdot P_{\ell L} \\
&+ 2(P_{hH}^3 + P_{hH}^2 - 4P_{hH} + 2) \cdot P_{\ell L}^2 + 4(2P_{hH} - 1) \cdot P_{\ell L}^3 + 2(1 - P_{hH}) \cdot P_{\ell L}^4).
\end{aligned}$$

Therefore it suffices to show that for all $\frac{1}{2} < P_{\ell L} < 1$ and $P_{hL} < P_{hH} \leq P_{\ell L}$, $f(P_{hH}, P_{\ell L}) =$

$$(1-P_{hH})^4+(1-P_{hH})^2\cdot(3P_{hH}-2))\cdot P_{\ell L}+2(P_{hH}^3+P_{hH}^2-4P_{hH}+2)\cdot P_{\ell L}^2+4(2P_{hH}-1)\cdot P_{\ell L}^3+2(1-P_{hH})\cdot P_{\ell L}^4 \geq 0.$$

To solve this problem, we calculate f's higher derivatives on $P_{\ell L}$ and show that all of them are strictly positive from higher ones to lower ones. Specifically, we have

$$\frac{\partial f}{\partial P_{\ell L}} = (1-P_{hH})^2 \cdot (3P_{hH}-2)) + 4(P_{hH}^3 + P_{hH}^2 - 4P_{hH} + 2) \cdot P_{\ell L} + 12(2P_{hH}-1) \cdot P_{\ell L}^2 + 8(1-P_{hH}) \cdot P_{\ell L}^3$$
$$\frac{\partial^2 f}{\partial P_{\ell L}^2} = 4(P_{hH}^3 + P_{hH}^2 - 4P_{hH} + 2) + 24(2P_{hH}-1) \cdot P_{\ell L} + 24(1-P_{hH}) \cdot P_{\ell L}^2.$$
$$\frac{\partial^3 f}{\partial P_{\ell L}^3} = 24(2P_{hH}-1) + 48(1-P_{hH}) \cdot P_{\ell L}.$$

Firstly, note that $\frac{\partial^3 f}{\partial P_{\ell L}^3} = 24(1-P_{hH}) \cdot (2P_{\ell L}-1) \geq 0$. If we want to show that $\frac{\partial^2 f}{\partial P_{\ell L}^2}$ is positive, under constraint $P_{\ell L} > P_{hH}$ and $P_{hH} > (1-P_{\ell L})$, it suffices to show that $\frac{\partial^2 f}{\partial P_{\ell L}^2}$ is positive when $P_{\ell L} = P_{hH}$ and when $P_{\ell L} = 1 - P_{hH}$.

$\frac{\partial^2 f}{\partial P_{\ell L}^2}(P_{hH}, P_{\ell L} = 1-P_{hH}) = 4(1-P_{hH}) \cdot (5(1-P_{hH}) - 8(1-P_{hH}) + 5) = 4(1-P_{hH}) \cdot (5P_{hH}^2 + 2(1-P_{hH})) > 0.$

$$\begin{aligned}\frac{\partial^2 f}{\partial P_{\ell L}^2}(P_{hH}, P_{\ell L} = P_{hH}) &= 4 \cdot (-5P_{hH}^3 + 19P_{hH}^2 - 10P_{hH} + 2) \\ &= 4 \cdot (2(2P_{hH}-1)^2 - 2P_{hH} \cdot (2P_{hH}-1)^2 + 3P_{hH}^2 + 3P_{hH}^3) \\ &> 0.\end{aligned}$$

Therefore, $\frac{\partial^2 f}{\partial P_{\ell L}^2} \geq 0$ for all feasible P_{hH} and $P_{\ell L}$. Now we show that $\frac{\partial f}{\partial P_{\ell L}} \geq 0$ for all feasible P_{hH} and $P_{\ell L}$ by showing that $\frac{\partial f}{\partial P_{\ell L}}(P_{hH}, P_{\ell L} = 1-P_{hH}) \geq 0$ and $\frac{\partial f}{\partial P_{\ell L}}(P_{hH}, P_{\ell L} = P_{hH}) \geq 0$.

$$\begin{aligned}\frac{\partial f}{\partial P_{\ell L}}(P_{hH}, P_{\ell L} = 1-P_{hH}) &= (1-P_{hH})^2 \cdot (4(1-P_{hH})^2 - 11(1-P_{hH}) + 9) \\ &= P_{\ell L}^2 \cdot ((2(1-P_{hH}) - 3)^2 + 1 - P_{hH}) \\ &> 0.\end{aligned}$$

$$\frac{\partial f}{\partial P_{\ell L}}(P_{hH}, P_{\ell L} = P_{hH}) = (9P_{hH} - P_{hH}^2 - 2) \cdot (4P_{hH}^2 - 3P_{hH} + 1)$$
$$\geq (9/2 - 3) \cdot ((2P_{hH} - 1)^2 + P_{hH})$$
$$\geq 0.$$

Therefore, $\frac{\partial f}{\partial P_{\ell L}} \geq 0$ for all feasible P_{hH} and $P_{\ell L}$. Now we show that $f(P_{hH}, P_{\ell L}) \geq 0$ for all feasible P_{hH} and $P_{\ell L}$ by showing that $f(P_{hH}, P_{\ell L} = 1 - P_{hH}) \geq 0$ and $f(P_{hH}, P_{\ell L} = P_{hH}) \geq 0$.

$$f(P_{hH}, P_{\ell L} = 1 - P_{hH}) = (2 + P_{hH}) \cdot 1 - P_{hH}^3 > 0$$
$$f(P_{hH}, P_{\ell L} = P_{hH}) = (4P_{hH}^2 - 3P_{hH} + 1)^2 > 0.$$

Therefore, for all $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $P_{hL} + (2\alpha - 1) \cdot (4\alpha P_{\ell H} P_{\ell L} - P_{\ell H} - 2P_{\ell L}) \geq 0$ holds. This implies that $\xi_{NL} \geq \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$ holds for all $\alpha_{NL} \geq \alpha > \frac{1}{2}$.

Finally, we show that for all $\frac{1}{2P_{\ell L}} \geq \alpha > \alpha_{NL}$, $\xi_{NL} < \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$. Note that $g(\alpha)$ is monotone on $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$, $g(\alpha_{NL}) > 0$, and

$$g(\frac{1}{2P_{\ell L}}) = -\frac{(1 - P_{\ell L}) \cdot (P_{\ell L} + P_{hH} - 1)}{P_{\ell L}} < 0.$$

By the intermediate value theorem, there exists a unique solution $\alpha_{NL} < \alpha'' \leq \frac{1}{2P_{\ell L}}$ such that $g(\alpha'') = 0$. For all $\alpha'' < \alpha \leq \frac{1}{2P_{\ell L}}$, $g(\alpha) < 0$. This directly implies that $\xi_{NL} < \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$. For all $\alpha_{NL} < \alpha < \alpha''$, The first part of this proof implies that $\xi_{NL} < \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$. This finishes the proof. $\square$

Deng et al. [10] show that when the fraction of the majority type agents is above the threshold, a good equilibrium always exists. In the equilibrium, the majority agents play the following "optimal strategy". A majority agent with h signal votes for **A** with probability $\beta_h = \frac{1}{2(P_{\ell L} + P_{\ell H})}$, and a majority agent with ℓ signal votes for **A** with probability $\beta_\ell = 0$.

Corollary 8. *[10] When the fraction of the majority $\alpha > \theta$, for any $1 \leq k \leq n$, any* *sequence of strategy profiles where all the majority type agents play the optimal strategy profile forms an ε-ex-ante Bayesian k-strong equilibrium with ε converging to 0. Such a sequence of equilibria leads to the informed majority decision with probability converging to 1.*

By Lemma 7, the optimal strategy majority agents play is not weakly dominated. Therefore, when minority agents play any non-weakly-dominated strategy, such an equilibrium satisfies the constraint where no agents play weakly dominated strategies.

C.5.2 Proof for "Flat" Segment 1.

In this section, it is sufficient to show the following two statements. Proposition 4 shows that when $\alpha \leq \theta$ and $k = (1-\theta)\cdot n$, even constraint (2) and (3) are not compatible. Proposition 3 shows that when $\alpha \geq \frac{\Delta}{1-\theta}$ and $k < (1-\theta)\cdot n$, a strategy profile satisfying all three constraints exists. (Note that given that $\theta = \frac{1}{2} + \frac{P_{\ell H}}{2P_{\ell L}}$, $1-\theta = \frac{\Delta}{2P_{\ell L}}$ and $\frac{1}{2P_{\ell L}} = \frac{\Delta}{1-\theta}$.)

Proposition 3. *For all $\alpha \geq \frac{\Delta}{1-\theta}$, for any $\xi < 1-\theta$ and $k = \xi \cdot n$, there exists a strategy profile sequence that satisfies all three constraints.*

Proof. For each n, let Σ_n be the following. All the minority agents always vote for **A**. The majority agents play the following strategy.

$$\beta_\ell^{maj} = \frac{1}{\alpha}(\alpha - \frac{1}{2} - \frac{P_{hL}}{\Delta}\cdot\xi) - \delta\cdot(P_{hH}+P_{hL}).$$
$$\beta_h^{maj} = \frac{1}{\alpha}(\alpha - \frac{1}{2} + \frac{P_{\ell L}}{\Delta}\cdot\xi) + \delta\cdot(P_{\ell H}+P_{\ell L}).$$

Here $\delta > 0$ is a constant such that $1 \geq \beta_\ell^{maj} \geq 0$ and $1 \geq \beta_h^{maj} \geq 0$, and at least one of $\beta_\ell^{maj} = 0$ and $\beta_h^{maj} = 1$ hold. We first show that such δ exists.

For β_ℓ^{maj}, $\frac{1}{\alpha}(\alpha - \frac{1}{2} - \frac{P_{hL}}{\Delta}\cdot\xi) < \frac{1}{\alpha}\cdot\alpha < 1$. Since $\alpha \geq \frac{\Delta}{1-\theta} = \frac{1}{2} + \frac{P_{hL}}{2P_{\ell L}}$, there is $\frac{\alpha-1/2}{P_{hL}} \geq \frac{1}{2P_{\ell L}}$.

Moreover, given that $\xi < 1-\theta = \frac{1}{2} - \frac{P_{\ell H}}{2P_{\ell L}} = \frac{\Delta}{2P_{\ell L}}$,

$$\begin{aligned}\alpha - \frac{1}{2} - \frac{P_{hL}}{\Delta}\cdot\xi &> \alpha - \frac{1}{2} - \frac{P_{hL}}{\Delta}\cdot\frac{\Delta}{2P_{\ell L}}\\ &= \alpha - \frac{1}{2} - \frac{P_{hL}}{2P_{\ell L}}\\ &\geq 0.\end{aligned}$$

For β_h^{maj}, $\alpha - \frac{1}{2} + \frac{P_{\ell L}}{\Delta}\cdot\xi \geq \alpha - \frac{1}{2} > 0$. Given that $\alpha \geq \frac{1}{2} + \frac{P_{hL}}{2P_{\ell L}}$ and that $\xi < \frac{\Delta}{2P_{\ell L}}$,

$$\frac{1}{\alpha}(\alpha - \frac{1}{2} + \frac{P_{\ell L}}{\Delta}\cdot\xi) < \frac{1}{\alpha}(\alpha - \frac{1}{2} + \frac{P_{\ell L}}{2P_{\ell L}}) = \frac{1}{\alpha}\cdot\alpha = 1.$$

Therefore, such $\delta > 0$ exists.

Now we are ready to show that the sequence $\{\Sigma_n\}$ satisfies all three constraints. According to Lemma 7, it is not hard to verify that all agents are playing a non-weakly dominated strategy. For the fidelity, we compute the expected vote share for **A** of the profile, denoted by φ_H and φ_L.

$$\begin{aligned}\varphi_H &= \alpha\cdot\delta\cdot\Delta + \xi + \frac{1}{2} > \frac{1}{2} + \xi.\\ \varphi_L &= \frac{1}{2} - \alpha\cdot\delta\cdot\Delta < \frac{1}{2}.\end{aligned}$$

Moreover, the difference $\alpha\cdot\delta$ does not depend on n. Therefore, $\liminf \sqrt{n}\cdot(\varphi_H - \frac{1}{2}) = +\infty$ and $\liminf \sqrt{n}\cdot(\frac{1}{2} - \varphi_L) = +\infty$. By applying the first case of Theorem 4, the fidelity of Σ_n converges to 1.

Now we show that Σ_n is an equilibrium. Let $e = 2\exp(-(\alpha\cdot\delta\cdot\Delta)^2\cdot n)$ and Σ'_n be a different strategy profile such that at most $\xi\cdot n$ agents deviate. Applying Lemma 15, we know that either no agents receive more than $\frac{2B(B+1)e}{P_H P_L}$ utility increase or minority agents will not join the deviating group.

Now suppose the deviation group D contains only minority types of agents. Let φ'_L and

φ'_H be the expected vote share of Σ'_n. Since in Σ_n all minorities always vote for **A**, any deviation will increase the vote share of **R**. Therefore, $\varphi'_L \geq \varphi^n_L = \alpha \cdot \delta \cdot \Delta$. Moreover, since there are at most $\xi \cdot n$ deviators, $\varphi'_H \geq \varphi^n_H - \xi = \alpha \cdot \delta \cdot \Delta$. By the Hoeffding Inequality, $\lambda^{\mathbf{A}}_H(\Sigma'_n) \geq 1 - e$ and $\lambda^{\mathbf{R}}_L(\Sigma'_n) \geq 1 - e$. Therefore, $A(\Sigma'_n) \geq 1 - e$, which contradicts the assumption that $A(\Sigma'_n) < 1 - \frac{(B+1)e}{P_H \cdot P_L}$. Therefore, such deviation cannot succeed, and Σ_n is an $\frac{2B(B+1)e}{P_H P_L}$-ex-ante Bayesian k-strong equilibrium. □

Proposition 4. *For all $\alpha \leq \theta$, for any profile sequence $\{\Sigma_n\}$ such that $A(\Sigma_n)$ converges to 1 and no agents play weakly dominated strategies, there is a constant $\varepsilon > 0$ and infinitely many n such that Σ_n is NOT an ε-ex-ante Bayesian k-strong equilibrium, where $k = (1-\theta) \cdot n$.*

Proof. In this case, we show that a group of $(1-\theta)\cdot n$ of minority type agents can successfully deviate by always voting for **A** or always voting for **R** simultaneously. Suppose $\{\Sigma\}$ is a strategy profile sequence such that $A(\Sigma)$ converges to 1.

For every n, we construct two deviating strategy profile sequences $\{\Sigma^1\}$ and $\{\Sigma^2\}$. Let D be the deviating group containing $(1-\theta)\cdot n$ minority type agents. Let $\bar{\sigma} = \sum_{i \notin D} \sigma_i = (\bar{\beta}_\ell, \bar{\beta}_h)$ be the average of strategies of non-deviators, which includes all the majority agents and the $(1-\alpha-\theta)\cdot n$ minority agents. Note that $(\bar{\beta}_\ell, \bar{\beta}_h)$ may or may not be identical in different n in the sequence.

In $\{\Sigma^1\}$, all the deviators vote for **R**. Then the expected vote share for **A** of Σ^1 is

$$\varphi^1_H = \theta \cdot (P_{hH} \cdot \bar{\beta}_h + P_{\ell H} \cdot \bar{\beta}_\ell),$$
$$\varphi^1_L = \theta \cdot (P_{hL} \cdot \bar{\beta}_h + P_{\ell L} \cdot \bar{\beta}_\ell).$$

In $\{\Sigma^2\}$, all the deviators vote for **A**. Then the expected vote share for **A** of Σ^2 is

$$\varphi^2_H = \theta \cdot (P_{hH} \cdot \bar{\beta}_h + P_{\ell H} \cdot \bar{\beta}_\ell) + (1-\theta),$$
$$\varphi^2_L = \theta \cdot (P_{hL} \cdot \bar{\beta}_h + P_{\ell L} \cdot \bar{\beta}_\ell) + (1-\theta).$$

We first consider the case where there are infinitely many n such that in the sequence $\{\Sigma\}$, the majority agent strategies in Σ have sub-linear variance. Suppose for any $\delta > 0$ there exist infinitely many n such that in some world state k, $Var(\sum_{i\in maj} X_i^n \mid k) < \delta \cdot n$. Firstly, $\{\Sigma^1\}$ and $\{\Sigma^2\}$ also have infinitely many Σ where majority agent strategies have sub-linear variance. This is because the majority agents do not change their strategies in these two deviations. Then by Corollary 7, for any $\varepsilon > 0$, there exist infinitely many n such that in some world state k, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$. (Note that for every such n, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$ holds simultaneously.)

Then we show that in this case, at least one of $\{\Sigma^1\}$ and $\{\Sigma^2\}$ falls into the second case in Theorem 4, which indicates that its fidelity does not converge to 1. If $\liminf_n \sqrt{n}\cdot(\varphi_H^1 - \frac{1}{2}) < 0$, $\{\Sigma^1\}$ falls into the second case in Theorem 4. Suppose $\liminf_n \sqrt{n}\cdot(\varphi_H^1 - \frac{1}{2}) \geq 0$, we show that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^2) < 0$. $\liminf_n \sqrt{n}\cdot(\varphi_H^1 - \frac{1}{2}) \geq 0$ implies that, for any constant $\delta' > 0$ for all sufficiently large n, $\varphi_H^1 - \frac{1}{2} > -\delta'$. Then by Corollary 7, there exist infinitely many n such that in some world state k, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$. Note that $\varphi_H^1 = \varphi_H^2 - (1-\theta)$. Combining these three facts, and setting the constant such that $\delta' + \varepsilon < \frac{1-\theta}{2}$. we know that there exist infinitely many n such that

$$\varphi_L^2 > \varphi_H^2 - \varepsilon = \varphi_H^1 - \varepsilon + (1-\theta) > \frac{1}{2} - \delta' - \varepsilon + (1-\theta) > \frac{1}{2} + \frac{1-\theta}{2}.$$

This directly implies that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^2) < 0$. Therefore, the fidelity of either $\{\Sigma^1\}$ or $\{\Sigma^2\}$ does not converge to 1. Precisely, there exists a $\delta > 0$ and infinitely many n such that $A(\Sigma^1) \leq 1 - \delta$ or $A(\Sigma^2) \leq 1 - \delta$ holds. As the fidelity decreases, the utility of the minority agents increases, and the deviation is successful. This implies that $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

For now we suppose that the majority agent strategies Σ have linear variance for all sufficiently large n. Likewise, the majority agent strategies in Σ^1 and Σ^2 also have linear variances. In this case, if we can show that for all n, either $\varphi_H^1 \leq \frac{1}{2}$ or $\varphi_L^2 \geq \frac{1}{2}$ holds, then at least one of them holds for infinitely many n. By the second or the third case (as the variance is linear) of Theorem 4, either the fidelity of Σ^1 or the fidelity of Σ^2 does not converge to 1. This implies that the deviation is successful, and $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

It remains to show that $\varphi_H^1 \leq \frac{1}{2}$ or $\varphi_L^2 \geq \frac{1}{2}$ holds for an arbitrary n. Let $\varphi'_H = \varphi_H^1$ and $\varphi'_L = \varphi_L^2$. We fix an arbitrary n. Suppose $\varphi'_H > \frac{1}{2}$, we show that $\varphi'_L > \frac{1}{2}$ so that the deviation can succeed by all the deviators always voting for **A**. Given that $\theta = \frac{P_{\ell L}+P_{\ell H}}{2P_{\ell L}}$ this is equivalent to show that

$$P_{hL} \cdot \bar{\beta}_h + P_{\ell L} \cdot \bar{\beta}_\ell > \frac{P_{\ell H}}{P_{\ell L} + P_{\ell H}}$$

under the condition that

$$P_{hH} \cdot \bar{\beta}_h + P_{\ell H} \cdot \bar{\beta}_\ell > \frac{P_{\ell L}}{P_{\ell L} + P_{\ell H}}.$$

Denote $\hat{f}'_H = P_{hH} \cdot \bar{\beta}_h + P_{\ell H} \cdot \bar{\beta}_\ell$ and $\hat{f}'_L = P_{hL} \cdot \bar{\beta}_h + P_{\ell L} \cdot \bar{\beta}_\ell$. Now, let function $g(\beta_h) = P_{hL} \cdot \beta_h + P_{\ell L} \cdot \frac{\hat{f}'_H - P_{hH} \cdot \beta_h}{P_{\ell H}} - \frac{P_{\ell H}}{P_{\ell L}+P_{\ell H}}$. Then it is sufficient to show that $g(\bar{\beta}_h) > 0$.

Firstly, $\frac{dg}{d\beta_h} = \frac{P_{hL} \cdot P_{\ell H} - P_{hH} \cdot P_{\ell L}}{P_{\ell H}} < 0$. Therefore, $g(\beta_h)$ is minimized when β_h is maximized. Therefore,

$$\begin{aligned} g(\bar{\beta}_h) &\geq g(1) \\ &= P_{hL} + P_{\ell L} \cdot \frac{\hat{f}'_H - P_{hH} \cdot 1}{P_{\ell H}} - \frac{P_{\ell H}}{P_{\ell L} + P_{\ell H}}. \end{aligned}$$

Recall that $\hat{f}'_H > \frac{P_{\ell L}}{P_{\ell L}+P_{\ell H}}$. Therefore,

$$\begin{aligned}
\hat{f}'_H - P_{hH} &> \frac{P_{\ell L}}{P_{\ell L}+P_{\ell H}} - P_{hH} \\
&= \frac{(P_{\ell L} - P_{hH}\cdot P_{\ell L}) - P_{hH}\cdot P_{\ell H}}{P_{\ell L}+P_{\ell H}} \\
&= \frac{P_{\ell L}\cdot P_{\ell H} - P_{hH}\cdot P_{\ell H}}{P_{\ell L}+P_{\ell H}} \\
&= \frac{(P_{\ell L} - P_{hH})\cdot P_{\ell H}}{P_{\ell L}+P_{\ell H}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
g(\bar{\beta}_h) &> \frac{P_{\ell L}}{P_{\ell H}} \cdot \frac{(P_{\ell L} - P_{hH})\cdot P_{\ell H}}{P_{\ell L}+P_{\ell H}} + (P_{hL} - \frac{P_{\ell H}}{P_{\ell L}+P_{\ell H}}) && \text{(C.3)} \\
&= \frac{(P_{\ell L} - P_{hH})\cdot P_{\ell L}}{P_{\ell L}+P_{\ell H}} + \frac{(P_{hL} - P_{\ell H})\cdot P_{\ell L}}{P_{\ell L}+P_{\ell H}} && \text{(C.4)} \\
&= 0. && \text{(C.5)}
\end{aligned}$$

Therefore, $\hat{f}'_L > \frac{P_{\ell H}}{P_{\ell L}+P_{\ell H}}$. Therefore, by applying the second or the third case of Theorem 4, either the fidelity of Σ_n^1 or that of Σ_n^2 does not converge to 1. By Lemma 16, minority can gain a constant utility increase from at least one of such deviation under infinitely many n. Therefore, $\{\Sigma_n\}$ fail to be an equilibrium sequence with ε converging to 0. □

C.5.3 Segment 2 to 4: Overview

The rest of the proof will proceed in the order of the proof sketch. Here is the outline.

1. We solve the case when γ is biased (either γ or $1-\alpha-\gamma$ is close to 0) and restrict the case when γ is not biased to be "double-symmetric": all type-0 agents play the same strategy, and all type-1 agents play the same strategy.

2. We show that a sequence of strategy profiles satisfies all constraints on ξ if and only if $\xi < \xi_h$ and $\xi < \xi_\ell$. Here ξ_h and ξ_ℓ are two constraints derived by the constraints on the expected vote share ($\varphi'_H > \frac{1}{2}$ and $\varphi_L < \frac{1}{2}$, like that in the "Flat" Segment

1). Therefore, the problem of finding the upper and lower bound is converted into maximizing $\min(\xi_h, \xi_\ell)$.

3. We reduce the search space of maximizing $\min(\xi_h, \xi_\ell)$ into two corner cases. In the first case, all type-0 agents always vote for **R**. In the second case, all type-1 agents always vote for **A**. The threshold ξ^* take the maximum of these two cases and $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$, which derives from the biased case.

4. We tackle the two extreme cases respectively. In the case where all type-0 agents always vote for **R** (the first case), the extension of "Steep" Segment 2 and "Tail" Segment 4 serves as an upper bound. On the other hand, the case where all type-1 agents always vote for **A** (the second case) derives the "Non-Linear" Segment 3 and the "Tail" Segment 4.

5. We take the maximum of all the cases and reach the upper and the lower bounds.

We first introduce notations. Consider a strategy profile Σ where no agents play weakly dominated strategies. For the majority type agents, we only care about their contribution to the expected vote share, so it suffices to know the average of their strategies. Let β_h^{maj} and β_ℓ^{maj} be the average strategy of the majority agents. We partition minority agents into two subgroups. The type-0 minority agents always vote for **R** when they receive h signals, i.e. $\beta_h^0 = 0$. The type-1 minority agents always vote for **A** when they receive ℓ signals, i.e. $\beta_\ell^1 = 1$. Each minority agent will only be categorized into exactly one type. Specifically, those who vote anti-informatively will either be a type-1 or type-0 agent, but will not be both. Let β_ℓ^0 and β_h^1 be the average probability of voting for **A** for a type-0 agent receiving signal ℓ and a type-1 agent receiving signal h, respectively. Let $0 \leq \gamma \leq 1-\alpha$ be the fraction of type-0 agents.

In this way, the expected vote share of the strategy profile can be written as

$$\varphi_H = \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1-\alpha-\gamma) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) + \gamma \cdot P_{\ell H} \cdot \beta_\ell^0.$$
$$\varphi_L = \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1-\alpha-\gamma) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + \gamma \cdot P_{\ell L} \cdot \beta_\ell^0.$$

C.5.4 Segment 2 to 4: Preliminaries

We first show that, when γ is very biased, no strategy profile can tolerate a deviation group of $\xi \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$.

Proposition 5. *For $\alpha \leq \frac{\Delta}{1-\theta}$, for any $\xi' \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$, no sequence with infinitely many strategy profiles where $\gamma \leq \xi'$ or $1-\alpha-\gamma \leq \xi'$ can satisfy all three constraints in Theorem 12 with $\xi = \xi'$.*

Proof. The high-level idea is that the deviator group exhausts one type of minority agent. Suppose Σ is a strategy profile with $\gamma \leq \xi'$. (The $1-\alpha-\gamma \leq \xi'$ side follows similar reasoning given $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}} \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{\ell H}}$). Let the deviation group contain all the type-0 agents and $(\xi-\gamma)\cdot n$ of type-1 agents. We consider two possible deviations.

In the first deviation, denoted as Σ^1, the type-1 agents in the deviation group are those with the minimum probability to vote for **A** when receiving the h signal. All the deviators switch to always vote for **A**. We denote the expected share for Σ^1 as φ_H^1 and φ_L^1.

In the second deviation, denoted as Σ^2, the type-1 agents in the deviation group are those with the maximum probability to vote for **A** when receiving the h signal. All the deviators switch to always vote for **R**. We denote the expected share for Σ^2 as φ_H^2 and φ_L^2.

Note that $\varphi_H^1 \geq \varphi_H^2 + \xi'$ and $\varphi_L^1 \geq \varphi_L^2 + \xi'$. Firstly, all the deviators (with a fraction of ξ') in Σ^1 always vote for **A**, while all the deviators in Σ^2 always vote for **R**. Secondly, the majority agents play the same strategy in two deviations. Finally, the non-deviator minority agents in Σ^1 are more likely to vote for **A** than the non-deviator minority agents.

Similar to the proof of Proposition 4, we first consider the case where there are infinitely many n such that in $\{\Sigma\}$, majority agent strategies have sub-linear variance. Suppose for any $\delta > 0$ there exist infinitely many n such that in some world state k, $Var(\sum_{i\in maj} X_i^n \mid k) < \delta \cdot n$. Firstly, $\{\Sigma^1\}$ and $\{\Sigma^2\}$ also have infinitely many Σ where majority agent strategies have sub-linear variance. Then by Corollary 7, for any $\varepsilon > 0$, there exist infinitely many n such that in some world state k, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$. (Note that for every such n, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$ hold simultaneously.)

Then we show that in this case, at least one of $\{\Sigma^1\}$ and $\{\Sigma^2\}$ falls into the second case in Theorem 4, which indicates that its fidelity does not converge to 1. If $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) < 0$, $\{\Sigma^2\}$ falls into the second case in Theorem 4. Suppose $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) \geq 0$, we show that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^1) < 0$. Note that $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) \geq 0$ implies that, for any constant $\delta' > 0$ for all sufficiently large n, $\varphi_H^2 - \frac{1}{2} > -\delta'$. Then by Corollary 7, there exist infinitely many n such that in some world state k, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$. Note that $\varphi_H^1 \geq \varphi_H^2 + \xi'$. Combining these three facts, and setting the constant such that $\delta' + \varepsilon < \xi'$, we know that there exist infinitely many n such that

$$\varphi_L^1 > \varphi_H^1 - \varepsilon \geq \varphi_H^2 - \varepsilon + \xi' > \frac{1}{2} - \delta' - \varepsilon + \xi' > \frac{1}{2} + \frac{\xi'}{2}.$$

This directly implies that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^1) < 0$. Therefore, the fidelity of either $\{\Sigma^1\}$ or $\{\Sigma^2\}$ does not converge to 1. Precisely, there exists a $\delta > 0$ and infinitely many n such that $A(\Sigma^1) \leq 1 - \delta$ or $A(\Sigma^2) \leq 1 - \delta$ holds. As the fidelity decreases, the utility of the minority agents increases, and the deviation is successful. This implies that $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

For now we suppose that the majority agent strategies in Σ (and consequently Σ^1 and Σ^2) have linear variance for all sufficiently large n. In this case, if we can show that for all n, either $\varphi_H^2 \leq \frac{1}{2}$ or $\varphi_L^1 \geq \frac{1}{2}$ holds, then at least one of them holds for infinitely many n. By the second or the third case of Theorem 4, either the fidelity of Σ^1 or the fidelity of Σ^2 does not converge to 1. By Lemma 16, this implies that the deviation is successful, and $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

By the definition of Σ^1 and Σ^2 ,we have

$$\varphi_L^1 \geq \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1 - \alpha - \xi) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + \xi.$$

$$\varphi_H^2 \leq \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1 - \alpha - \xi) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}).$$

We show that either $\varphi_H^2 \leq \frac{1}{2}$ or $\varphi_L^1 \geq \frac{1}{2}$, which implies that the deviating succeed, and the strategy profile is not an equilibrium.

Note that the solution to equation

$$\alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1 - \alpha - \xi) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + \xi = \frac{1}{2}$$
$$\alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1 - \alpha - \xi) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) = \frac{1}{2}$$

is

$$\beta_h^{maj} = \frac{1}{\alpha}\left(\frac{1}{2} + \frac{P_{\ell H} \cdot \xi}{\Delta} - (1 - \alpha - \xi) \cdot \beta_h^1\right).$$
$$\beta_\ell^{maj} = \frac{1}{\alpha}\left(\frac{1}{2} - \frac{P_{hH} \cdot \xi}{\Delta} - (1 - \alpha - \xi)\right).$$

Denote this solution as β_h^* and β_ℓ^*. Note that when $\xi \geq \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$, $\beta_\ell^* \leq 0$

Then we rewrite the inequality of φ_L^1 and φ_H^2 as

$$\varphi_H^2 \leq \frac{1}{2} + \alpha \cdot (P_{hH} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell H} \cdot (\beta_\ell^{maj} - \beta_\ell^*)).$$
$$\varphi_L^1 \geq \frac{1}{2} + \alpha \cdot (P_{hL} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell L} \cdot (\beta_\ell^{maj} - \beta_\ell^*)).$$

If $\varphi_H^2 > \frac{1}{2}$, there must be $P_{hH} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell H} \cdot (\beta_\ell^{maj} - \beta_\ell^*) > 0$. By $P_{hH} > P_{hL}$, $P_{\ell L} > P_{\ell H}$, and $(\beta_\ell^{maj} - \beta_\ell^*) \geq 0$, there must be $(P_{hL} \cdot (\beta_h^{maj} - \beta_h^*) + P_\ell \cdot (\beta_\ell^{maj} - \beta_\ell^*) > 0$. Therefore, $\varphi_L^1 \geq \frac{1}{2}$.

Therefore, there exists a constant ε such that for every strategy Σ in the sequence $\{\Sigma\}$ such that $\gamma \leq \xi'$ or $1-\alpha-\gamma \leq \xi'$, Σ is NOT an ε-ex-ante Bayesian $\xi'\cdot n$ strong equilibrium. $\square$

At the same time, the biased case also contributes a tight lower bound, which eventually becomes the lower bound for Segment 2.

Proposition 6. *When $\frac{\Delta}{1-\theta} \geq \alpha$, for any $\xi < \frac{\Delta(\alpha-1/2)}{P_{hL}}$, there exists a sequence of strategy profiles $\{\Sigma\}$ that satisfies the constraints in Theorem 12 with $\xi = \frac{\Delta(\alpha-1/2)}{P_{hL}}$*

Proof. For a fixed $\xi < \frac{\Delta(\alpha-1/2)}{P_{hL}}$ and for every n, the strategy profile is as follows. All the minority agents always vote for **A** regardless of their signals. All the majority agents play $\beta_h^{maj} = 1+\frac{1}{\alpha}\cdot(-\frac{1}{2}+\frac{\xi\cdot P_{\ell L}}{\Delta})+\delta$ and $\beta_\ell^{maj} = 0$, where $\delta \in (\frac{\xi+(\alpha-1/2)\cdot P_{\ell H}}{\alpha\cdot P_{hH}} - \frac{\xi\cdot P_{\ell L}}{\alpha\cdot\Delta}, \frac{(\alpha-1/2)\cdot P_{\ell L}}{\alpha\cdot P_{hL}} - \frac{\xi\cdot P_{\ell L}}{\alpha\cdot\Delta})$. Given that $\xi < \frac{\Delta(\alpha-1/2)}{P_{hL}}$, we can verify that the feasible set of δ is not empty.

$$
\begin{aligned}
\frac{(\alpha-1/2)\cdot P_{\ell L}}{\alpha\cdot P_{hL}} - \frac{\xi+(\alpha-1/2)\cdot P_{\ell H}}{\alpha\cdot P_{hH}} &> \frac{(\alpha-1/2)\cdot P_{\ell L}}{\alpha\cdot P_{hL}} - \frac{(\alpha-1/2)\cdot P_{\ell H}}{\alpha\cdot P_{hH}} - \frac{\Delta(\alpha-1/2)\cdot P_{\ell L}}{\alpha\cdot P_{hL}\cdot P_{hH}} \\
&= \frac{(\alpha-1/2)}{\alpha}\cdot\frac{P_{\ell L}\cdot P_{hH} - \Delta - P_{\ell H}\cdot P_{hL}}{P_{hL}\cdot P_{hH}} \\
&= 0.
\end{aligned}
$$

It is not hard to verify that $\beta_h^{maj} \geq 0$. Given that $\alpha \leq \frac{\Delta}{1-\theta} = \frac{1}{2P_{\ell L}}$ and $\xi < \frac{\Delta(\alpha-1/2)}{P_{hL}}$, we can verify that $\beta_h^{maj} \leq 1$.

$$
\beta_h^{maj} < 1 + \frac{1}{\alpha}\cdot(-\frac{1}{2} + \frac{(\alpha-1/2)\cdot P_{\ell L}}{P_{hL}}) = 1 + \frac{1}{\alpha}\cdot\frac{2\alpha\cdot P_{\ell L}-1}{2P_{hL}} \leq 1.
$$

By Lemma 1, no agents play weakly dominated strategies. For constraint 2, we calculated the expected vote share for **A** under two world states respectively.

$$
\begin{aligned}
\varphi_H &= \alpha\cdot P_{hH}\cdot(1+\frac{1}{\alpha}\cdot(-\frac{1}{2}+\frac{\xi\cdot P_{\ell L}}{\Delta})+\delta) + (1-\alpha). \\
\varphi_L &= \alpha\cdot P_{hL}\cdot(1+\frac{1}{\alpha}\cdot(-\frac{1}{2}+\frac{\xi\cdot P_{\ell L}}{\Delta})+\delta) + (1-\alpha).
\end{aligned}
$$

For φ_H we have

$$
\begin{aligned}
\varphi_H &> \alpha \cdot P_{hH} \cdot (1 + \frac{1}{\alpha} \cdot (-\frac{1}{2} + \frac{\xi \cdot P_{\ell L}}{\Delta}) + \frac{\xi + (\alpha - 1/2) \cdot P_{\ell H}}{\alpha \cdot P_{hH}} - \frac{\xi \cdot P_{\ell L}}{\alpha \cdot \Delta}) + (1-\alpha) \\
&= P_{hH} \cdot (\alpha - \frac{1}{2} + \frac{\xi + (\alpha - 1/2) \cdot P_{\ell H}}{P_{hH}}) + (1-\alpha) \\
&= P_{hH} \cdot (\alpha - \frac{1}{2}) + (\alpha - 1/2) \cdot P_{\ell H} + \xi + (1-\alpha) \\
&= \frac{1}{2} + \xi.
\end{aligned}
$$

And for φ_L, we have

$$
\begin{aligned}
\varphi_L &< \alpha \cdot P_{hL} \cdot (1 + \frac{1}{\alpha} \cdot (-\frac{1}{2} + \frac{\xi \cdot P_{\ell L}}{\Delta}) + \frac{(\alpha - 1/2) \cdot P_{\ell L}}{\alpha \cdot P_{hL}} - \frac{\xi \cdot P_{\ell L}}{\alpha \cdot \Delta}) + (1-\alpha) \\
&= P_{hL} \cdot (\alpha - 1/2 + \frac{(\alpha - 1/2) \cdot P_{\ell L}}{P_{hL}}) + (1-\alpha) \\
&= \frac{1}{2}.
\end{aligned}
$$

Moreover, φ_H and φ_L do not depend on n. Therefore, $\liminf \sqrt{n} \cdot (\varphi_H - \frac{1}{2}) = +\infty$ and $\liminf \sqrt{n} \cdot (\frac{1}{2} - \varphi_L) = +\infty$. By applying the first case of Theorem 4, the fidelity of Σ_n converges to 1.

Now we show that Σ is an equilibrium with $\xi^* = \frac{\Delta \cdot (\alpha - 1/2)}{P_{hL}}$. Note that $\xi^* \leq 1 - \alpha$ when $\alpha \leq \theta$. By Lemma 15, it suffices to show that a fraction $\xi < \xi^*$ of minority agents cannot successfully deviate. Suppose a $\xi < \xi^*$ fraction of minority agents change their strategies. Consider the expected vote share after deviation, denoted as φ'_H and φ'_L. Recall that all minority agents always vote for **A** in Σ. Therefore, $\varphi'_L \leq \varphi_L < \frac{1}{2}$. On the other hand, the deviator can decrease φ_H by no more than its fraction ξ. Therefore, $\varphi'_H \geq \varphi_H - \xi > \frac{1}{2}$. Therefore, applying the first case of Theorem 4 and Lemma 16, the fidelity still converges to 1 after any deviation, and the expected utility gain of the minority agents converges to 0. Therefore, the strategy profile proposed is an ε-equilibrium with ε converging to 0. $\square$

Now we come to solve the case where $\gamma > \xi$ and $1 - \alpha - \gamma > \xi$, i.e., the deviation group cannot exhaust one type of minority agents. The following lemma shows that in the strategy profile maximizing ξ, each type of minority agents should play the same strategy.

Lemma 8. *Let $\{\Sigma_n\}$ be a sequence of strategy profiles that satisfy the constraints in Theorem 12 with ξ. Consider the following profiles $\{\Sigma'_n\}$, where each n, Σ_n and Σ'_n has the following relationship.*

1. *Each majority agent in Σ'_n play the same strategy as in Σ_n.*

2. *All the type-1 minority agents in Σ'_n play the average strategy of type-1 minority agents in Σ_n, denoted by $\beta^1_\ell = 1$ and β^1_h.*

3. *All the type-0 minority agents in Σ'_n play the average strategy of type-0 minority agents in Σ_n, denoted by $\beta^0_h = 0$ and β^0_ℓ.*

Then $\{\Sigma'_n\}$ also satisfies the constraints on ξ.

Proof. The first constraint is straightforward.

For the second and the third constraint, we consider the expected vote share. Let φ_H and φ_L be the expected vote share of Σ_n and φ'_H and φ'_L be the expected vote share of Σ'_n. For simplicity, we use Σ and Σ' to denote Σ_n and Σ'_n.

For an arbitrary n, the extremities of expected vote share in Σ. For the H side, φ_H is minimized by selecting $\xi \cdot n$ agents in type-1 agents with maximum probability to vote for **A** with signal h and switching them to always vote for **R**. The change bringing to φ_H is $-\frac{1}{n}\sum_{i\in D}(P_{hH}\cdot\beta^i_h + P_{\ell H}) \le -\xi\cdot(P_{hH}\cdot\beta^1_h + P_{\ell H})$. For the L side, φ_L is maximized by selecting $\xi{\cdot}n$ agents in type-0 agents with minimum probability to vote for **A** with signal ℓ and switching them to always vote for **A**. The change bringing to φ_L is $\xi - \frac{1}{n}\sum_{i\in D}(P_{hL}+P_{\ell L}\cdot\beta^i_\ell) \ge \xi - \xi\cdot(P_{hL} + P_{\ell L}\cdot\beta^0_\ell)$.

Given that Σ satisfies constraint 3, any deviation from Σ still have fidelity converging to 1 (otherwise the deviation succeed). By Theorem 4, for any constant $\varepsilon > 0$ and all sufficiently large n, there must be

$$\varphi_H - \frac{1}{n}\sum_{i\in D}(P_{hH}\cdot\beta_h^i + P_{\ell H}) \geq \frac{1}{2} - \varepsilon$$
$$\varphi_L + \xi - \frac{1}{n}\sum_{i\in D}(P_{hL} + P_{\ell L}\cdot\beta_\ell^i) \leq \frac{1}{2} + \varepsilon.$$

By taking $\varepsilon < \frac{\xi}{4}$, this implies that $\varphi_H - \varphi_L \geq \frac{\xi}{2}$ for all sufficiently large n. Then Lemma 14, the variance of the majority agent's strategy is already $\Theta(n)$ for all sufficiently large n. Therefore, for any deviation from Σ_n, the variance is $\Theta(n)$. Given that Σ_n satisfies the third constraint (equilibrium), these deviation fails, and their fidelities converges to 1. This rules out the third case in Theorem 4 and further show that

$$\liminf \sqrt{n}\cdot(\varphi_H - \frac{1}{n}\sum_{i\in D}(P_{hH}\cdot\beta_h^i + P_{\ell H}) - \frac{1}{2}) = +\infty,$$
$$\liminf \sqrt{n}\cdot(\frac{1}{2} - (\varphi_L + \xi - \frac{1}{n}\sum_{i\in D}(P_{hL} + P_{\ell L}\cdot\beta_\ell^i))) = +\infty.$$

This also implies that $\liminf \sqrt{n}\cdot(\varphi_H - \frac{1}{2}) = +\infty$ and $\liminf \sqrt{n}\cdot(\frac{1}{2} - \varphi_L) = +\infty$.

Then we consider the Σ_n'. We start from the second constraint (fidelity converging to 1). Note that $\varphi_H' = \varphi_H$ and $\varphi_L' = \varphi_L$ for all n. Therefore, $\liminf \sqrt{n}\cdot(\varphi_H' - \frac{1}{2}) = +\infty$ and $\liminf \sqrt{n}\cdot(\frac{1}{2} - \varphi_L') = +\infty$ also hold. Therefore, by the first case in Theorem 4, the fidelity of Σ_n' converges to 1.

Now we prove the third constraint holds by showing that no deviation in Σ_n' can succeed. Consider the extremities for Σ_n'. For the H side, φ_H' is minimized by selecting $\xi\cdot n$ agents in type-1 (where they play the same strategy). Therefore, the change bringing to φ_H' is at most $-\xi\cdot(P_{hH}\cdot\beta_h^1 + P_{\ell H})$. Likewise, the change bringing to φ_L' is at most $\xi - \xi\cdot(P_{hL} + P_{\ell L}\cdot\beta_\ell^0)$.

Given that $\varphi_H = \varphi'_H$ and $\varphi_L = \varphi'_L$, there must be

$$\varphi'_H - \xi \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) \geq \varphi_H - \frac{1}{n}\sum_{i \in D}(P_{hH} \cdot \beta_h^i + P_{\ell H})$$
$$\varphi'_L + \xi - \xi \cdot (P_{hL} + P_{\ell L} \cdot \beta_\ell^0) \leq \varphi_L + \xi - \frac{1}{n}\sum_{i \in D}(P_{hL} + P_{\ell L} \cdot \beta_\ell^i).$$

Therefore, it also holds that

$$\liminf \sqrt{n} \cdot (\varphi'_H - \xi \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) - \frac{1}{2}) = +\infty,$$
$$\liminf \sqrt{n} \cdot (\frac{1}{2} - (\varphi'_L + \xi - \xi \cdot (P_{hL} + P_{\ell L} \cdot \beta_\ell^0))) = +\infty.$$

Therefore, for any deviating strategy profile Σ''_n from Σ'_n, whose expected vote share is φ''_H and φ''_L, there must be $\liminf \sqrt{n} \cdot \min(\varphi''_H - \frac{1}{2}, \frac{1}{2} - \varphi''_L) = +\infty$. By the first case of Theorem 4 and Lemma 16, the fidelity of Σ''_n converges to 1, and the deviation cannot succeed. Therefore, $\{\Sigma'_n\}$ satisfies the third constraint. □

Therefore, if we show that any strategy profile where each type of minority agents plays the same strategy does not satisfy the constraints, then no strategy profile can satisfy the constraints. In the rest of the proof, we focus on strategy profiles where each type of minority agents plays the same strategy.

C.5.5 Segment 2 to 4: Transition to Maximization Problem.

Consider a strategy profile Σ where the average strategy of majority agents is $(\beta_\ell^{maj}, \beta_h^{maj})$, every type-1 minority agents play $(1, \beta_h^1)$, every type-0 minority agents play $(\beta_\ell^0, 0)$, and the fraction of type-0 minority agents is γ.

Consider the condition for Σ to be an equilibrium under ξ. First, after a ξ fraction of type-0 agents switch to always vote for **A**, the deviating expected vote share under world state L (denoted as φ'_L) should be strictly less than $\frac{1}{2}$. Second, after a ξ fraction of type-1 agents switch to always vote for **R**, the deviating expected vote share under world state

H (denoted as φ'_H) should be strictly greater than $\frac{1}{2}$. Therefore, we have the following conditions.

$$\varphi'_H = \alpha \cdot (P_{hH} \cdot \beta_h^{maj} + P_{\ell H} \cdot \beta_\ell^{maj}) + (1 - \alpha - \gamma - \xi) \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) + \gamma \cdot P_{\ell H} \cdot \beta_\ell^0 > \frac{1}{2}$$
$$\varphi'_L = \alpha \cdot (P_{hL} \cdot \beta_h^{maj} + P_{\ell L} \cdot \beta_\ell^{maj}) + (1 - \alpha - \gamma) \cdot (P_{hL} \cdot \beta_h^1 + P_{\ell L}) + (\gamma - \xi) \cdot P_{\ell L} \cdot \beta_\ell^0 + \xi < \frac{1}{2}$$

We claim that, for any given β_h^1, β_ℓ^0, γ, and ξ, there is a pair of $(\beta_\ell^{maj}, \beta_h^{maj})$ satisfying this condition if and only if the solution of $\varphi'_H = \frac{1}{2}$ and $\varphi'_L = \frac{1}{2}$ (denoted as $(\beta_\ell^*, \beta_h^*)$) is in $(0,1)^2$. Specifically, the explicit form of the solution is

$$\beta_\ell^* = \frac{1}{\alpha}(\frac{1}{2} - (1-\alpha) + \gamma \cdot (1 - \beta_\ell^0)) - \frac{\xi}{\Delta \cdot \alpha}(P_{hH} \cdot P_{hL} \cdot \beta_h^1 + P_{hH} \cdot P_{\ell L} \cdot (1 - \beta_\ell^0) + P_{hL}).$$
$$\beta_h^* = \frac{1}{\alpha}(\frac{1}{2} - (1-\alpha-\gamma) \cdot \beta_h^1) + \frac{\xi}{\Delta \cdot \alpha}(P_{hH} \cdot P_{\ell L} \cdot \beta_h^1 + P_{\ell H} \cdot P_{\ell L} \cdot (1 - \beta_\ell^0) + P_{\ell H}).$$

Combining with the constraint that $(\beta_\ell^*, \beta_h^*) \in (0,1)^2$, we get two constraints on ξ, denoted as ξ_ℓ and ξ_h.

$$\xi < \xi_h = \frac{\Delta(\alpha - 1/2 + (1-\alpha-\gamma) \cdot \beta_h^1)}{P_{hH} \cdot P_{\ell L} \cdot \beta_h^1 + P_{\ell H} \cdot P_{\ell L} \cdot (1 - \beta_\ell^0) + P_{\ell H}}.$$
$$\xi < \xi_\ell = \frac{\Delta(\alpha - 1/2 + \gamma \cdot (1 - \beta_\ell^0))}{P_{hH} \cdot P_{hL} \cdot \beta_h^1 + P_{hH} \cdot P_{\ell L} \cdot (1 - \beta_\ell^0) + P_{hL}}.$$

Specifically, when $\beta_h^1 = (1 - \beta_\ell^0) = 0$, $\xi_h = \frac{\Delta(\alpha - \frac{1}{2})}{P_{\ell H}}$ and $\xi_\ell = \frac{\Delta(\alpha - \frac{1}{2})}{P_{hL}}$. When $\gamma = \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, and $\beta_h^1 = (1 - \beta_\ell^0) = 1$, $\xi_h = \xi_\ell = \frac{1}{2}\Delta\alpha$.

Lemma 9. *Given a set of β_h^1, β_ℓ^0, and γ satisfying $\beta_h^1 \in [0,1]$, $\beta_\ell^0 \in [0,1]$, and let $\xi' = min(\xi_h, \xi_\ell)$ corresponding to β_h^1, β_ℓ^0, and γ. If $\gamma \geq \xi'$ and $1 - \alpha - \gamma \geq \xi'$ holds, then for all $\xi < \xi'$ there exists a sequence of profiles such that the constraints in Theorem 12 holds for ξ.*

Proof. We fix an arbitrary $\xi < \xi'$, and let β_h^* and β_ℓ^* defined as above. By the definition of ξ_h and ξ_ℓ, β_h^* and β_ℓ^* are in $(0,1)$. Now we set $\beta_h^{maj} = \beta_h^* + \delta \cdot (P_{\ell H} + P_{\ell L})$ and $\beta_\ell^{maj} = \beta_\ell^* - \delta(P_{hH} + P_{hL})$, where $\delta > 0$ guarantees that β_ℓ^{maj} and β_h^{maj} are in $[0,1]$, and either

$\beta_h^{maj} = 1$ or $\beta_\ell^{maj} = 0$ holds.

Now we consider a sequence of strategy profiles Σ. For each n, Σ is as follows. All majority agents play β_h^{maj} and β_ℓ^{maj}. A fraction γ of minority agents play $\beta_h^0 = 0$ and β_ℓ^0. The rest of the minority agents play $\beta_\ell^1 = 1$ and β_h^1. We show such strategy satisfies all three constraints.

First, it is not hard to verify that all agents play non-weakly dominated strategy. Secondly, we consider the fidelity of this strategy. In this proof, φ'_H and φ'_L refers two deviating expected vote share defined as above. Note that

$$\varphi_H = \varphi'_H + \xi \cdot (P_{hH} \cdot \beta_h^1 + P_{\ell H}) > \varphi'_H.$$

$$\varphi_L = \varphi'_L - \xi \cdot (1 - P_{\ell L} \cdot \beta_\ell^0) < \varphi'_L.$$

Therefore, it is sufficient to show that $\varphi'_H > \frac{1}{2}$ and $\varphi'_L < \frac{1}{2}$. Recall that β_h^* and β_ℓ^* is the solution of the set of equation ($\varphi'_H = \frac{1}{2}$, $\varphi'_L = \frac{1}{2}$). Therefore, assigning β_h^{maj} and β_ℓ^{maj},

$$\varphi'_H = \frac{1}{2} + \alpha \cdot \delta \cdot (P_{hH} \cdot (P_{\ell H} + P_{\ell L}) - P_{\ell H} \cdot (P_{hH} + P_{hL})) = \frac{1}{2} + \alpha \cdot \delta \cdot \Delta > \frac{1}{2}$$
$$\varphi'_L = \frac{1}{2} + \alpha \cdot \delta \cdot (P_{hL} \cdot (P_{\ell H} + P_{\ell L}) - P_{\ell L} \cdot (P_{hH} + P_{hL})) = \frac{1}{2} - \alpha \cdot \delta \cdot \Delta < \frac{1}{2}$$

Therefore, for the second constraint, we show that $\varphi_H > \frac{1}{2}$ and $\varphi_L < \frac{1}{2}$, and that they do not depend on n. By the first case of Theorem 4, the fidelity of Σ converges to 1.

For the third constraint, by Lemma 15, it suffices to consider a deviating group with only minority agents. For any deviating strategy with deviating expected vote share φ''_H and φ''_L, there must be $\varphi''_H \geq \varphi'_H$ and $\varphi''_L \leq \varphi'_L$. Given that $\varphi'_H > \frac{1}{2}$ and $\varphi'_L < \frac{1}{2}$ and do not depend on n, this directly implies $\liminf \sqrt{n} \cdot \min(\varphi''_H - \frac{1}{2}, \frac{1}{2} - \varphi''_L) = +\infty$. By the first case of Theorem 4 and Lemma 16, the fidelity of Σ''_n converges to 1, and the deviation cannot succeed. Therefore, $\{\Sigma_n\}$ satisfies the third constraint. $\square$

Lemma 10. *Let* $\xi' = \max_{\beta_h^1, \beta_\ell^0, \gamma} \min(\xi_h, \xi_\ell)$. *No sequence strategy profile can satisfy all three* *constraints in Theorem 12 with* $\xi = \max(\xi', \frac{\Delta(\alpha-1/2)}{P_{hL}})$.

Proof. Let $\{\Sigma\}$ be a sequence of strategy profiles satisfying constraints 1 and 2. We show that constraint 3 does not hold, i.e., there exists a constant ε such that infinitely many strategy profiles in $\{\Sigma\}$ are not ε-ex-ante Bayesian $\xi \cdot n$ strong equilibrium.

When $\{\Sigma\}$ contains infinitely many strategy profiles where $\gamma \leq \xi'$ or $1 - \alpha - \gamma \leq \xi'$, we apply Proposition 5. If $\xi' < \frac{\Delta(\alpha-1/2)}{P_{hL}}$, $\{\Sigma\}$ also contains infinitely many strategy profiles where $\gamma \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$ or $1 - \alpha - \gamma \leq \frac{\Delta(\alpha-1/2)}{P_{hL}}$, and Proposition 5 implies that the constraint is violated with $\xi = \frac{\Delta(\alpha-1/2)}{P_{hL}}$; if $\xi' \geq \frac{\Delta(\alpha-1/2)}{P_{hL}}$, Proposition 5 implies that the constraint is violated with $\xi = \xi'$. Therefore, $\{\Sigma\}$ cannot satisfy all three constraints with $\xi = \max(\xi', \frac{\Delta(\alpha-1/2)}{P_{hL}})$.

Now we consider the case where $\gamma > \xi'$ and $1 - \alpha - \gamma > \xi'$ holds for all sufficiently large n. By Lemma 8, without loss of generality, we can assume that all type-0 agents play the same strategy and all type-1 agents play the same strategy. For an arbitrary n, let φ_H and φ_L be the expected vote share of Σ. We consider two deviations. In the first deviation Σ^1, a ξ' fraction of type-0 agents switch to always vote for **A**. We denote the expected vote share after this deviation as φ_H^1 and φ_L^1. In the second deviation Σ^2, a ξ' fraction of type-1 agents switch to always vote for **R**. We denote the expected vote share after this deviation as φ_H^2 and φ_L^2. Note that φ_H' and φ_L' defined in Equation 5.7 and 5.8 is exactly $\varphi_H' = \varphi_H^2$ and $\varphi_L' = \varphi_L^1$. We will show that for infinitely many n, either Σ^1 or Σ^2 is a successful deviation.

Now we do a case study on the variance of Σ. We first consider the case where $\{\Sigma\}$ has infinitely many Σ where the majority agent strategies have sub-linear variance. This part of proof resembles that of Proposition 4. Suppose for any $\delta > 0$ there exist infinitely many n such that in some world state k, $Var(\sum_{i \in maj} X_i^n \mid k) < \delta \cdot n$. Firstly, $\{\Sigma^1\}$ and $\{\Sigma^2\}$ also has infinitely many Σ where the majority agent strategies have sub-linear variance. Then by Corollary 7, for any $\varepsilon > 0$, there exist infinitely many n such that in some world state k, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$. (Note that for every such n, $\varphi_H^1 - \varphi_L^1 < \varepsilon$ and $\varphi_H^2 - \varphi_L^2 < \varepsilon$ holds simultaneously.)

Then we show that in this case, at least one of $\{\Sigma^1\}$ and $\{\Sigma^2\}$ falls into the second case in Theorem 4, which indicates that its fidelity does not converge to 1. If $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) < 0$, $\{\Sigma^2\}$ falls into the second case in Theorem 4. Suppose $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) \geq 0$, we show that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^1) < 0$. $\liminf_n \sqrt{n}\cdot(\varphi_H^2 - \frac{1}{2}) \geq 0$ implies that, for any constant $\delta' > 0$ for all sufficiently large n, $\varphi_H^2 - \frac{1}{2} > -\delta'$. Then by Corollary 7, there exist infinitely many n such that in some world state k, $\varphi_H^2 - \varphi_L^2 < \varepsilon$ and $\varphi_H^1 - \varphi_L^1 < \varepsilon$. On the other hand, note that

$$\begin{aligned}
\varphi_H^1 &= \varphi_H^2 + \xi'\cdot(P_{hH}\cdot\beta_h^1 + P_{\ell H}) + \xi'\cdot(1 - \beta_\ell^0\cdot P_{\ell H}) \\
&= \varphi_H^2 + \xi'\cdot(1 + P_{\ell H} + P_{hH}\cdot\beta_h^1 - \beta_\ell^0\cdot P_{\ell H}) \\
&\geq \varphi_H^2 + \xi'\cdot(P_{hL} + P_{\ell H}).
\end{aligned}$$

Combining these three inequalities, and setting the constant such that $\delta' + \varepsilon < \frac{\xi'\cdot(P_{hL}+P_{\ell H})}{2}$. we know that there exist infinitely many n such that

$$\varphi_L^1 > \varphi_H^1 - \varepsilon \geq \varphi_H^2 - \varepsilon + \xi'\cdot(P_{hL} + P_{\ell H}) > \frac{1}{2} - \delta' - \varepsilon + \xi'\cdot(P_{hL} + P_{\ell H}) > \frac{1}{2} + \frac{\xi'\cdot(P_{hL} + P_{\ell H})}{2}.$$

This directly implies that $\liminf_n \sqrt{n}\cdot(\frac{1}{2} - \varphi_L^1) < 0$. Therefore, the fidelity of either $\{\Sigma^1\}$ or $\{\Sigma^2\}$ does not converge to 1. Precisely, there exists a $\delta > 0$ and infinitely many n such that $A(\Sigma^1) \leq 1 - \delta$ or $A(\Sigma^2) \leq 1 - \delta$ holds. By Lemma 8, the utility of the minority agents increases as the fidelity decreases, and at least one of the deviations is successful. This implies that $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

Now we consider the case where Σ (and consequently Σ^1 and Σ^2) have majority agent strategies with $\Theta(n)$ variance of all sufficiently large n. It suffices to show that at least one of $\varphi'_L \geq \frac{1}{2}$ and $\varphi'_H \leq \frac{1}{2}$ holds for infinitely many n. By the second or the third case (as the variance is linear) of Theorem 4, either the fidelity of Σ^1 or the fidelity of Σ^2 does not converge to 1. This implies that the deviation is successful, and $\{\Sigma\}$ is not a sequence of ε-ex-ante k-strong BNE with ε converging to 0.

Recall that β_h^* and β_ℓ^* are the solution to $\varphi'_H = \frac{1}{2}$ and $\varphi'_L = \frac{1}{2}$. By the definition of ξ_h and ξ_ℓ, either $\beta_h^* \geq 1$ or $\beta_\ell^* \leq 0$ holds. Without loss of generality, let $\beta_h^* \geq 1$. Now we consider β_h^{maj} and β_ℓ^{maj}. Note that $\beta_h^{maj} \leq \beta_h^*$ always holds.

$$\varphi'_H = \frac{1}{2} + \alpha \cdot (P_{hH} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell H} \cdot (\beta_\ell^{maj} - \beta_\ell^*)).$$
$$\varphi'_L = \frac{1}{2} + \alpha \cdot (P_{hL} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell L} \cdot (\beta_\ell^{maj} - \beta_\ell^*)).$$

We show that $\varphi'_H > \frac{1}{2}$ implies $\varphi'_L > \frac{1}{2}$. By $\beta_h^{maj} \leq \beta_h^*$, $\varphi'_H > \frac{1}{2}$ implies $\beta_\ell^{maj} > \beta_\ell^*$ and $P_{hH} \cdot (\beta_h^{maj} - \beta_h^*) + P_{\ell H} \cdot (\beta_\ell^{maj} - \beta_\ell^*) > 0$. By $P_{hH} > P_{hL}$ and $P_{\ell H} < P_{\ell L}$, $\varphi'_L > \varphi'_H > \frac{1}{2}$. $\square$

C.5.6 Segment 2 to 4: Maximizing $\min(\xi_h, \xi_\ell)$.

Now we try to find the condition that maximizes $\min(\xi_h, \xi_\ell)$. We view ξ_h and ξ_ℓ as functions of β_h^1 and $(1-\beta_\ell^0)$ and consider their partial derivative on β_h^1 and $(1-\beta_\ell^0)$. It is not hard to verify that $\frac{\partial \xi_h}{\partial(1-\beta_\ell^0)} < 0$ and $\frac{\partial \xi_\ell}{\partial \beta_h^1} < 0$. On the other two, recall the fact that for a function $f = \frac{u}{v}$, the derivative of f is $f' = \frac{u'v - uv'}{v^2}$. Therefore, we follow $\text{sign}(f') = \text{sign}(u'v - uv')$ and have the following observation on the derivatives.

$$\text{sign}(\frac{\partial \xi_h}{\partial \beta_h^1}) = \text{sign}((1-\alpha-\gamma) \cdot P_{\ell H} \cdot (1 + P_{\ell L} \cdot (1-\beta_\ell^0)) - P_{hH} \cdot P_{\ell L} \cdot (\alpha - \frac{1}{2})),$$
$$\text{sign}(\frac{\partial \xi_\ell}{\partial(1-\beta_\ell^0)}) = \text{sign}(\gamma \cdot P_{hL} \cdot (1 + P_{hH} \cdot \beta_h^1) - P_{hH} \cdot P_{\ell L} \cdot (\alpha - \frac{1}{2})).$$

We first discuss on ξ_h. When $(1-\alpha-\gamma) \leq \frac{P_{hH} \cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$, $\frac{\partial \xi_h}{\partial \beta_h^1} \leq 0$ at $(1-\beta_\ell^0) = 0$. Therefore, for any $(\beta_\ell^0, \beta_h^1)$, $\xi_h(\beta_h, 1-\beta_\ell^0) < \xi_h(\beta_h^1, 0) \leq \xi_h(0,0) = \frac{\Delta(\alpha - \frac{1}{2})}{P_{\ell H}}$. On the other hand, when $(1-\alpha-\gamma) > \frac{P_{hH} \cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$, $\frac{\partial \xi_h}{\partial \beta_h^1} > 0$ for all $\beta_\ell^0 \in [0,1]$. Similarly, when $\gamma \leq \frac{P_{hH} \cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$, for any $(\beta_\ell^0, \beta_h^1)$, $\xi_\ell(\beta_h, 1-\beta_\ell^0) < \xi_\ell(0,0) = \frac{\Delta(\alpha - \frac{1}{2})}{P_{hL}}$, and when $\gamma > \frac{P_{hH} \cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$, $\frac{\partial \xi_\ell}{\partial(1-\beta_\ell^0)} > 0$ holds. Therefore, the following proposition holds.

Proposition 11. *When either* $(1-\alpha-\gamma) \leq \frac{P_{hH} \cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$ *or* $\gamma \leq \frac{P_{hH} \cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$ *holds, either* ξ_h *or* ξ_ℓ *cannot exceed* $\frac{\Delta(\alpha - \frac{1}{2})}{P_{hL}}$, *and* $\min(\xi_h, \xi_\ell)$ *is maximized at* $\frac{\Delta(\alpha - \frac{1}{2})}{P_{hL}}$.

Now we consider the case when $(1-\alpha-\gamma) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ and $\gamma > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$. We deal with this case by analyzing the derivatives under a linear constraint of β_h^1 and $(1-\beta_\ell^0)$, i.e. $1-\beta_\ell^0 = t\cdot\beta_h^1$ for every $t>0$. Then we have

$$\mathrm{sign}(\frac{\partial \xi_h(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1}) = \mathrm{sign}((1-\alpha-\gamma)\cdot P_{\ell H} - (\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L} - (\alpha-\frac{1}{2})\cdot P_{\ell H}\cdot P_{\ell L}\cdot t)).$$
$$\mathrm{sign}(\frac{\partial \xi_\ell(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1}) = \mathrm{sign}(t\cdot(\gamma\cdot P_{hL} - (\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L}) - P_{hH}\cdot P_{\ell L}\cdot(\alpha-\frac{1}{2})).$$

Therefore, $\frac{\partial \xi_h(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} \geq 0$ if and only if $t \leq \frac{1-\alpha-\gamma}{P_{\ell L}\cdot(\alpha-1/2)} - \frac{P_{hH}}{P_{\ell H}}$, and $\frac{\partial \xi_\ell(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} \geq 0$ if and only if $t \geq \frac{P_{hH}\cdot P_{\ell L}\cdot(\alpha-\frac{1}{2})}{(\gamma\cdot P_{hL}-(\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L})}$. With this observation, we have the following lemma.

Lemma 20. *When $(1-\alpha-\gamma) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ and $\gamma > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, at least one of the three holds. (1) $\max_{\beta_h^1,\beta_\ell^0,\gamma}\min(\xi_h,\xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$; (2) $\min(\xi_h,\xi_\ell)$ is maximized when $\beta_h^1 = 1$, and (3) $\min(\xi_h,\xi_\ell)$ is maximized when $1-\beta_\ell^0 = 1$.*

Proof. We consider a pair of β_h^1 and $1-\beta_\ell^0$ other than (0,0). Let $t = \frac{1-\beta_\ell^0}{\beta_h^1}$.

1. If $t \geq \frac{1-\alpha-\gamma}{P_{\ell L}\cdot(\alpha-1/2)} - \frac{P_{hH}}{P_{\ell H}}$, there is $\frac{\partial \xi_h(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} \leq 0$ which implies $\xi_h(\beta_h^1, 1-\beta_\ell^0) < \xi_h(0,0)$.
2. If $t \leq \frac{P_{hH}\cdot P_{\ell L}\cdot(\alpha-\frac{1}{2})}{(\gamma\cdot P_{hL}-(\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L})}$, there is $\frac{\partial \xi_\ell(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} \leq 0$ which implies $\xi_\ell(\beta_h^1, 1-\beta_\ell^0) < \xi_\ell(0,0)$.
3. Finally, suppose $\frac{P_{hH}\cdot P_{\ell L}\cdot(\alpha-\frac{1}{2})}{(\gamma\cdot P_{hL}-(\alpha-\frac{1}{2})\cdot P_{hH}\cdot P_{\ell L})} < t < \frac{1-\alpha-\gamma}{P_{\ell L}\cdot(\alpha-1/2)} - \frac{P_{hH}}{P_{\ell H}}$. In this case, $\frac{\partial \xi_h(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} > 0$ and $\frac{\partial \xi_\ell(\beta_h^1, t\cdot\beta_h^1)}{\partial \beta_h^1} > 0$. Therefore, if $\beta_h^1 < 1$ and $(1-\beta_\ell^0) < 1$, We consider the pair $\beta_h^1 + \delta \leq 1$ and $(1-\beta_\ell^0) + t\cdot\delta \leq 1$, where $\delta > 0$ is a small constant. Then we have $\xi_h(\beta_h^1, 1-\beta_\ell^0) < \xi_h(\beta_h^1+\delta, 1-\beta_\ell^0+t\cdot\delta)$ and $\xi_\ell(\beta_h^1, 1-\beta_\ell^0) < \xi_\ell(\beta_h^1+\delta, 1-\beta_\ell^0+t\cdot\delta)$.

If the third case exists, $\min(\xi_h,\xi_\ell)$ is maximized on either $\beta_h^1 = 1$ or $1-\beta_\ell^0 = 1$. Otherwise, $\min(\xi_h,\xi_\ell)$ cannot exceed $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. □

Combining Proposition 11 and Lemma 20, we have the following result.

Proposition 7. *At least one of the following holds. (1) $\max_{\beta_h^1,\beta_\ell^0,\gamma}\min(\xi_h,\xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$; (2) $\min(\xi_h,\xi_\ell)$ is maximized when $\beta_h^1 = 1$, and (3) $\min(\xi_h,\xi_\ell)$ is maximized when $1-\beta_\ell^0 = 1$.*

In the rest of the proof, we will deal the three cases in Proposition 7. Case (1) always lead to $\frac{\Delta\cdot(\alpha-1/2)}{P_{\ell H}}$. Next, we show that in case (3), ξ will upper bounded by either $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ (Proposition 8) or $\frac{1}{2}\Delta\alpha$ (Proposition 9). Finally, we deal with the most difficult case (2) and discover the non-linear segment (Lemma 11 and Lemma 21).

C.5.7 Segment 2 to 4: Case $1-\beta_\ell^0=1$.

Now we are ready to show that, for case (3) $1-\beta_\ell^0=1$, when $\frac{\Delta}{1-\theta}\geq\alpha\geq\frac{1}{1+P_{\ell L}}$, $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ is the upper bound of ξ.

Proposition 8. *When* $(1-\alpha-\gamma)>\frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$, $\gamma>\frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, *and* $\frac{\Delta}{1-\theta}\geq\alpha\geq\frac{1}{1+P_{\ell L}}$, *if* $1-\beta_\ell^0=1$ *holds, either* ξ_h *or* ξ_ℓ *cannot exceed* $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$.

Proof. This proof discusses on different case on γ. Firstly, consider the special case where $\gamma=\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. In this case, and when $\beta_h^1=(1-\beta_\ell^0)=1$, $\xi_h=\xi_\ell=\frac{1}{2}\Delta\alpha$, which is smaller than $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$ when $\alpha\geq\frac{1}{1+P_\ell}$. Now we consider case when $\beta_h^1<1$. We know that when $(1-\alpha-\gamma)>\frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ and $\gamma>\frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, ξ_h is increasing on β_h^1 and decreasing on $(1-\beta_\ell^0)$, and ξ_ℓ is increasing on $(1-\beta_\ell^0)$ and decreasing on β_h^1. Therefore, for all $\beta_h^1<1$, ξ_h decreases and is less than $\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$.

Then we consider the case when $\gamma\neq\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. Note that ξ_h decreases and ξ_ℓ increases when γ increases. Therefore, when $\gamma>\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, ξ_h is even smaller compared to itself when $\gamma=\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, under every same pair of β_h^1 and $1-\beta_\ell^0$. Therefore, in this case, $\xi_h\leq\frac{1}{2}\Delta\alpha<\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$.

Now we consider $\gamma<\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. We will find the β_h^1 such that $\xi_h(\beta_h^1,1)=\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$ and $\xi_\ell(\beta_h^1,1)=\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$ respectively, denoted as β_h^{1h} and $\beta_h^{1\ell}$. Then we will show that $\beta_h^{1h}\geq\beta_h^{1\ell}$. Given that ξ_h is increasing and ξ_ℓ is decreasing on β_h^1, this implies that no β_h^1 satisfies $\min(\xi_h(\beta_h^1,1),\xi_\ell(\beta_h^1,1))>\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$.

Here we give the explicit form of β_h^{1h} and $\beta_h^{1\ell}$.

$$\begin{aligned}\beta_h^{1h} &= \frac{(\alpha-1/2)\cdot(P_{\ell H}P_{\ell L}+P_{\ell H}-P_{hL})}{P_{hL}\cdot(1-\alpha-\gamma)-(\alpha-1/2)P_{hH}P_{\ell L}}.\\ \beta_h^{1\ell} &= \frac{\gamma\cdot P_{hL}-(\alpha-1/2)P_{hH}P_{\ell L}}{(\alpha-1/2)\cdot P_{hH}P_{hL}}.\end{aligned}$$

Note that when $P_{hL}\cdot(1-\alpha-\gamma)-(\alpha-1/2)P_{hH}P_{\ell L}\leq 0$, for any $0<\beta_h^1<1$, $\xi_h<\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$, and the statement holds. Now we consider the case when $P_{hL}\cdot(1-\alpha-\gamma)-(\alpha-1/2)P_{hH}P_{\ell L}>0$. Note that

$$\begin{aligned}&\frac{(\alpha-1/2)\cdot(P_{\ell H}P_{\ell L}+P_{\ell H}-P_{hL})}{P_{hL}\cdot(1-\alpha-\gamma)-(\alpha-1/2)P_{hH}P_{\ell L}}\geq\frac{\gamma\cdot P_{hL}-(\alpha-1/2)P_{hH}P_{\ell L}}{(\alpha-1/2)\cdot P_{hH}P_{hL}}\\ \Leftrightarrow\ & P_{hL}^2\cdot\gamma\cdot(1-\alpha-\gamma)-P_{hL}(\alpha-\frac{1}{2})\cdot P_{hH}P_{\ell L}\cdot(\gamma+1-\alpha-\gamma)+(\alpha-\frac{1}{2})^2P_{hH}^2P_{\ell L}^2\\ &\leq(\alpha-\frac{1}{2})^2P_{hH}P_{\ell L}P_{hL}P_{\ell H}+(\alpha-\frac{1}{2})^2\cdot P_{hH}P_{hL}\cdot(P_{\ell H}-P_{hL})\\ \Leftrightarrow\ & P_{hL}^2\cdot\gamma\cdot(1-\alpha-\gamma)-P_{hL}(\alpha-\frac{1}{2})\cdot P_{hH}P_{\ell L}\cdot(1-\alpha)+(\alpha-\frac{1}{2})^2P_{hH}^2P_{\ell L}^2\\ &\leq(\alpha-\frac{1}{2})^2P_{hH}P_{\ell L}P_{hL}P_{\ell H}+(\alpha-\frac{1}{2})^2\cdot P_{hH}P_{hL}\cdot(P_{\ell H}-P_{hL}).\end{aligned}$$

Note that the first term $P_{hL}^2\cdot\gamma\cdot(1-\alpha-\gamma)$ increases on γ when $\gamma\leq\frac{1}{2}(1-\alpha)$. Given the constraint that $\gamma<\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}\leq\frac{1}{2}(1-\alpha)$, it suffices to show that the inequality holds when $\gamma=\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. Recall that when $\beta_h^1=1$, $\xi_h=\xi_\ell=\frac{1}{2}\Delta\alpha<\frac{\Delta(\alpha-\frac{1}{2})}{P_{hL}}$, ξ_h increases on β_h^1, and ξ_ℓ decreases on β_h^1. Therefore, there must be $\beta_h^{1h}>\beta_h^{1\ell}$, which finishes the proof. □

Proposition 9. *When* $(1-\alpha-\gamma)>\frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$, $\gamma>\frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, *and* $\frac{1}{1+P_{\ell L}}\geq\alpha\geq\frac{1}{2}$, *if* $1-\beta_\ell^0=1$ *holds, either* ξ_h *or* ξ_ℓ *cannot exceed* $\frac{1}{2}\Delta\alpha$, *and* $\min(\xi_h,\xi_\ell)$ *is maximized at* $\frac{1}{2}\Delta\alpha$.

Proof. This proof resembles the proof of Proposition 8. First, when $\gamma=\frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, and $\beta_h^1=(1-\beta_\ell^0)=1$, $\xi_h=\xi_\ell=\frac{1}{2}\Delta\alpha$. We know that when $(1-\alpha-\gamma)>\frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ and $\gamma>\frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, ξ_h is increasing on β_h^1 and decreasing on $(1-\beta_\ell^0)$, and ξ_ℓ is increasing on $(1-\beta_\ell^0)$ and decreasing on β_h^1. Therefore, for all $(1-\beta_\ell^0)=1$ and $\beta_h^1<1$, $\xi_h<\frac{1}{2}\Delta\alpha$. When

$\gamma > \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, ξ_h decreases and is smaller than $\frac{1}{2}\Delta\alpha$.

Now we consider $\gamma < \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$. We will find the β_h^1 such that $\xi_h(\beta_h^1, 1) = \frac{1}{2}\Delta\alpha$ and $\xi_\ell(\beta_h^1, 1) = \frac{1}{2}\Delta\alpha$ respectively, denoted as β_h^{1h} and $\beta_h^{1\ell}$. Then we will show that $\beta_h^{1h} \geq \beta_h^{1\ell}$. Given that ξ_h is increasing and ξ_ℓ is decreasing on β_h^1, this implies that no β_h^1 satisfies $\min(\xi_h(\beta_h^1, 1), \xi_\ell(\beta_h^1, 1)) > \frac{1}{2}\Delta\alpha$.

Here we give the explicit form of β_h^{1h} and $\beta_h^{1\ell}$.

$$\beta_h^{1h} = \frac{\alpha\cdot(P_{\ell H}P_{\ell L} + P_{\ell H} - 2) + 1}{2(1-\alpha-\gamma) - \alpha\cdot P_{hH}P_{\ell L}}$$
$$\beta_h^{1\ell} = \frac{2\gamma - 1 + \alpha\cdot(2 - P_{hL} - P_{hH}P_{\ell L})}{\alpha\cdot P_{hH}P_{hL}} = \frac{2\gamma - 1 + \alpha\cdot(1 + P_{\ell L}P_{\ell H})}{\alpha\cdot P_{hH}P_{hL}}$$

Note that given $\alpha \leq \frac{1}{1+P_{\ell L}} < \frac{1}{1+P_{\ell L}P_{hH}}$ and $\gamma < \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2} \leq \frac{1-\alpha}{2}$, $2(1-\alpha-\gamma)-\alpha\cdot P_{hH}P_{\ell L} > 2-(1+P_{hH}P_{\ell L})\cdot\alpha > 0$. Now we show that $\beta_h^{1h} \geq \beta_h^{1\ell}$ holds.

$$\begin{aligned}
&\frac{\alpha\cdot(P_{\ell H}P_{\ell L} + P_{\ell H} - 2) + 1}{2(1-\alpha-\gamma) - \alpha\cdot P_{hH}P_{\ell L}} \geq \frac{2\gamma - 1 + \alpha\cdot(1 + P_{\ell L}P_{\ell H})}{\alpha\cdot P_{hH}P_{hL}} \\
\Leftrightarrow\ & 4\gamma^2 - 2\gamma\cdot(3(1-\alpha) - \alpha\cdot P_{\ell L}) + 2(1-\alpha)^2 - \alpha\cdot(1-\alpha)\cdot P_{\ell L} \\
&\quad + \alpha\cdot(2\alpha-1)\cdot(P_{\ell L}P_{\ell H} - P_{hH}P_{hL}) + \alpha^2(P_{hH}P_{\ell H} - P_{\ell L}P_{hL}) \geq 0 \\
\Leftrightarrow\ & 4((1-\alpha-\gamma) - \frac{1-\alpha}{2})\cdot((1-\alpha-\gamma) - \frac{\alpha\cdot P_{\ell L}}{2}) \\
&\quad + \alpha\cdot(2\alpha-1)\cdot(P_{\ell L}P_{\ell H} - P_{hH}P_{hL}) + \alpha^2(P_{hH}P_{\ell H} - P_{\ell L}P_{hL}) \geq 0.
\end{aligned}$$

We show that every term in the LHS is non-negative.

- Since we assume $\gamma < \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$, there is $1-\alpha-\gamma > \frac{1-(P_{hL}+P_{hH})\cdot\alpha}{2} \geq \frac{1-\alpha}{2}$. Moreover, when $\alpha < \frac{1}{1+P_{\ell L}}$, $\alpha\cdot P_{\ell L} < \frac{P_{\ell L}}{1+P_{\ell L}} < 1-\alpha$. Therefore, the first term is non-negative.

- For the second term, we have $\alpha > \frac{1}{2}$, $P_{\ell L} \geq P_{hH}$, and $P_{\ell H} \geq P_{hL}$. Therefore, the second term is non-negative.

- For the last term, given that $P_{hL} < P_{hH} \leq P_{\ell L}$ and $P_{hH} + P_{\ell H} = P_{hL} + P_{\ell L} = 1$, there must be $P_{hH}P_{\ell H} \geq P_{\ell L}P_{hL}$. Therefore, the last term is non-negative.

Therefore, we show that $\beta_h^{1h} \geq \beta_h^{1\ell}$, which finishes the proof. $\square$

C.5.8 Segment 2 to 4: Case $\beta_h^1 = 1$.

For now we start to deal with case (2) in Proposition 7 where $\beta_h^1 = 1$. This case contains a non-linear upper bound which cannot be easily characterized. The high-level idea is as follows. Firstly, note that ξ_h decreases while ξ_ℓ increases on γ. Therefore, for a fixed $1-\beta_\ell^0$, there exists a unique γ such that $\xi_h = \xi_\ell$. Moreover, if this gamma is valid, it maximizes $\min(\xi_h, \xi_\ell)$ for this fixed β_ℓ^0. Therefore, we first calculate this γ and the corresponding ξ_h and ξ_ℓ for an arbitrary fixed $1-\beta_\ell^0$ (Lemma 11). Then we maximize ξ_h (also equals to ξ_ℓ) on $1-\beta_\ell^0$ under the optimal γ (Lemma 21). This brings us an upper bound for case (2).

Lemma 11. *When* $(1-\alpha-\gamma) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$, $\gamma > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$, *and* $\frac{1}{2P_{\ell L}} \geq \alpha \geq \frac{1}{2}$, *for* $\beta_h^1 = 1$ *and any fixed* $1-\beta_\ell^0$, *let*

$$\hat{\xi} = \frac{\Delta \cdot (\alpha - 1/2 + 1/2 \cdot (1-\beta_\ell^0))}{(P_{\ell L}\cdot(1-\beta_\ell^0) + P_{hL})\cdot(P_{\ell H}\cdot(1-\beta_\ell^0) + P_{hH} + 1)}.$$

Then, $\min(\xi_h, \xi_\ell) \leq \hat{\xi}$ *holds for all feasible* γ.

Proof. First, we claim that $\hat{\xi}$ is reached by fixing $\beta_h^1 = 1$ and $1-\beta_\ell^0$ and modifying γ such that $\xi_h = \xi_\ell$. By solving the equation $\xi_h = \xi_\ell$ for γ, we get the solution

$$\hat{\gamma} = \frac{\frac{1}{2}(P_{\ell L}\cdot(1-\beta_\ell^0) + P_{hL} + 1) - \alpha\cdot(P_{hH}P_{\ell L} + P_{\ell H}P_{\ell L}\cdot(1-\beta_\ell^0) + P_{\ell H})}{(P_{\ell L}\cdot(1-\beta_\ell^0) + P_{hL})\cdot(P_{\ell H}\cdot(1-\beta_\ell^0) + P_{hH} + 1)}.$$

Correspondingly, when $\gamma = \hat{\gamma}$, $\xi_h = \xi_\ell = \hat{\xi}$.

Now we show that $\hat{\xi}$ is an upper bound of $\min(\xi_h, \xi_\ell)$. Recall that ξ_h is decreasing on γ and ξ_ℓ is increasing on γ. Therefore, for all $\gamma \geq \hat{\gamma}$, $\xi_h \leq \hat{\xi}$; and for all $\gamma \leq \hat{\gamma}$, $\xi_\ell \leq \hat{\xi}$. $\square$

Lemma 21. *Let*

$$\hat{\xi} = \frac{\Delta \cdot (\alpha - 1/2 + 1/2 \cdot (1-\beta_\ell^0))}{(P_{\ell L} \cdot (1-\beta_\ell^0) + P_{hL}) \cdot (P_{\ell H} \cdot (1-\beta_\ell^0) + P_{hH} + 1)}.$$

For $(1-\beta_\ell^0) \in [0,1]$,

- *when* $\alpha \geq \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}$, $\max_{(1-\beta_\ell^0)} \hat{\xi} = \frac{\Delta(\alpha-1/2)}{P_{hL}\cdot(P_{hH}+1)}$ *with* $\arg\max_{(1-\beta_\ell^0)} \hat{\xi} = 0$*;*
- *when* $\frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})} > \alpha > \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\max_{(1-\beta_\ell^0)} \hat{\xi} = \xi_{NL}$*, and the maximum* $\arg\max_{(1-\beta_\ell^0)} \hat{\xi} \in (0,1)$*;*
- *when* $\alpha \leq \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\max_{(1-\beta_\ell^0)} \hat{\xi} = \frac{1}{2}\Delta\alpha$ *with* $\arg\max_{(1-\beta_\ell^0)} \hat{\xi} = 1$.

Proof. Consider the derivative of $\hat{\xi}$ on $(1-\beta_\ell^0)$.

$$\begin{aligned}
\operatorname{sign}\left(\frac{\partial\hat{\xi}}{\partial(1-\beta_\ell^0)}\right) &= \operatorname{sign}((P_{\ell L} \cdot (1-\beta_\ell^0) + P_{hL}) \cdot (P_{\ell H} \cdot (1-\beta_\ell^0) + P_{hH} + 1) \\
&\quad - (\alpha - 1/2 + 1/2 \cdot (1-\beta_\ell^0)) \\
&\quad \cdot (P_{\ell L} \cdot (P_{\ell H} \cdot (1-\beta_\ell^0) + P_{hH} + 1) + P_{\ell H} \cdot (P_{\ell L} \cdot (1-\beta_\ell^0) + P_{hL})) \\
&= \operatorname{sign}(-\frac{1}{2}P_{\ell L}P_{\ell H} \cdot (1-\beta_\ell^0)^2 - (2\alpha-1)P_{\ell L}P_{\ell H} \cdot (1-\beta_\ell^0) \\
&\quad - (\alpha - \frac{1}{2}) \cdot (P_{hH}P_{\ell L} + P_{hL}P_{\ell H} + P_{\ell L}) + \frac{1}{2}(P_{hL}P_{hH} + P_{hL})).
\end{aligned}$$

Denote the right-hand side

$$\begin{aligned}
f(1-\beta_\ell^0) = & -\frac{1}{2}P_{\ell L}P_{\ell H} \cdot (1-\beta_\ell^0)^2 - (2\alpha-1)P_{\ell L}P_{\ell H} \cdot (1-\beta_\ell^0) \\
& - (\alpha - \frac{1}{2}) \cdot (P_{hH}P_{\ell L} + P_{hL}P_{\ell H} + P_{\ell L}) + \frac{1}{2}(P_{hL}P_{hH} + P_{hL})
\end{aligned}$$

We have the following observations. Firstly, $f(1-\beta_\ell^0)$ is decreasing on $(1-\beta_\ell^0)$ for

$(1-\beta_\ell^0) \in [0,1]$. Secondly, when $1-\beta_\ell^0 = 0$,

$$\begin{aligned} f(1-\beta_\ell^0) = & -(\alpha - \frac{1}{2}) \cdot (P_{hH}P_{\ell L} + P_{hL}P_{\ell H} + P_{\ell L}) + \frac{1}{2}(P_{hL}P_{hH} + P_{hL}) \\ = & \frac{1}{2} \cdot (1 + P_{hL} + P_{hH}P_{\ell L}) - \alpha \cdot (P_{hH}P_{\ell L} + P_{hL}P_{\ell H} + P_{\ell L}) \end{aligned}$$

Thirdly, when $1-\beta_\ell^0 = 1$

$$\begin{aligned} f(1-\beta_\ell^0) = & -\frac{1}{2}P_{\ell L}P_{\ell H} - (2\alpha-1)P_{\ell L}P_{\ell H} \\ & -(\alpha - \frac{1}{2}) \cdot (P_{hH}P_{\ell L} + P_{hL}P_{\ell H} + P_{\ell L}) + \frac{1}{2}(P_{hL}P_{hH} + P_{hL}) \\ = & 1 - \alpha \cdot (1 + P_{\ell L} + (P_{\ell L} - P_{hH})). \end{aligned}$$

Therefore, when $\alpha \geq \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}$, $\text{sign}(\frac{\partial \hat{\xi}}{\partial(1-\beta_\ell^0)}) \leq 0$ for all $(1-\beta_\ell^0) \in [0,1]$, and $\hat{\xi}$ is maximized at $(1-\beta_\ell^0) = 0$ with $\max_{(1-\beta_\ell^0)} \hat{\xi} = \frac{\Delta(\alpha-1/2)}{P_{hL}\cdot(P_{hH}+1)}$.

When $\alpha \leq \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\text{sign}(\frac{\partial \hat{\xi}}{\partial(1-\beta_\ell^0)}) \geq 0$ for all $(1-\beta_\ell^0) \in [0,1]$, and $\hat{\xi}$ is maximized at $(1-\beta_\ell^0) = 1$ with $\max_{(1-\beta_\ell^0)} \hat{\xi} = \frac{1}{2}\Delta\alpha$.

Finally, when $\frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})} > \alpha > \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\frac{\partial \hat{\xi}}{\partial(1-\beta_\ell^0)}$ is positive when $(1-\beta_\ell^0) = 0$ and negative when $(1-\beta_\ell^0) = 1$. Therefore, $\hat{\xi}$ is maximized when $(1-\beta_\ell^0)$ satisfies $f(1-\beta_\ell^0) = 0$, which is

$$(1-\beta_\ell^0) = (1-2\alpha) + \frac{\sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{P_{\ell L} \cdot P_{\ell H}}.$$

Note that $f(1-\beta_\ell^0) = 0$ is a quadratic equation. By the monotonicity of f and the intermediate value theorem, a real-number solution is guaranteed. On the other hand, the other solution of the equation is out of $[0,1]$. Therefore, this is the only solution in $[0,1]$ and is guaranteed to be in $[0,1]$.

The corresponding $\hat{\xi}$ is exactly ξ_{NL}. $\square$

C.5.9 Segment 2 to 4: Wrap up the Upper Bounds.

So far, we have shown that the upper and the lower bounds of ξ only derive from one of the four cases. (1) When γ is biased (either type-0 agents or type-1 agents dominates the minority group). (2) When $1-\beta_\ell^0 = 1$ (all type-0 agents always vote for **R**). (3) When $\beta_h^1 = 1$ (all type-1 agents always vote for **A**). (4) All other cases, upper-bounded with $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 7. Each case has its own upper bound, and cases (1), (2), and (3) also have lower bounds.

Therefore, the last step is to integrate all the cases and reach a global upper and lower bound for each Segment. For the upper bound, we take the maximum of the upper bound for each case. In each case, there does not exist a strategy profile satisfying all three constraints with ξ equal to their corresponding upper bound. Therefore, overall, no strategy profiles can satisfy all three constraints with ξ equal to the maximum of all upper bounds. For the lower bound, we show that the upper bound in each segment binds a tight lower bound. As a consequence, the bounds consist of the threshold curve $\xi^*(\alpha)$, which finishes the proof.

Now we start to integrate the upper bound of each case.

1. When γ is biased, the upper bound is $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 5 along Segment 2, 3, and 4.

2. When $1-\beta_\ell^0 = 1$, the upper bound is given by the maximum of $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ and $\frac{1}{2}\Delta\alpha$ by Proposition 8 and 9.

3. When $\beta_h^1 = 1$, the upper bound is no more than $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ in Segment 2, ξ_{NL} in Segment 3, and $\frac{1}{2}\Delta\alpha$ in Segment 4 by Proposition 11 and 21.

4. In all other cases, the upper bound is at most $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ by Proposition 7.

The order of value of these upper bounds varies in different segments, of which the full proof is in Appendix C.5.1. In Segment 2, $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ is the largest. In Segment 3, ξ_{NL} is the largest. Finally, in Segment 4, $\frac{1}{2}\Delta\alpha \geq \frac{\Delta(\alpha-1/2)}{P_{hL}}$. Therefore, taking the maximum on each segment, the upper bound for ξ is $\frac{\Delta(\alpha-1/2)}{P_{hL}}$ in Segment 2, ξ_{NL} in Segment 3, and $\frac{1}{2}\Delta\alpha$ in Segment 4 (if it exists). This is exactly the upper bound given in $\xi^*(\alpha)$.

Here we give the formal proof for the upper bounds.

Proposition 12 (Segment 2)**.** *When $\frac{1}{2P_{\ell L}} > \alpha \geq \alpha_{NL}$, no (sequence of) strategy profiles can satisfy all three conditions for $\xi = \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$.*

Proof. By Lemma 10, it suffices to show that $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ for all possible β_h^1, $1-\beta_\ell^0$, and γ. By Proposition 11, when $(1-\alpha-\gamma) \leq \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$ or $\gamma \leq \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$ holds, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. For the rest of the proof, we assume $(1-\alpha-\gamma) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$ and $\gamma > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$. Then it suffices to show that in all three cases in Proposition 7, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$.

For case (1) where $\beta_h^1 = 1 - \beta_\ell^0 = 0$, $\xi_h = \frac{\Delta\cdot(\alpha-1/2)}{P_{\ell H}} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ and $\xi_\ell = \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. The statement holds.

For case (2) where $\beta_h^1 = 1$, we apply Lemma 11 and Lemma 21. Firstly, Lemma 11 and Lemma 21 characterize that, for $\alpha \geq \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}$, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta(\alpha-1/2)}{P_{hL}\cdot(P_{hH}+1)} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$; and when $\frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})} > \alpha > \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\min(\xi_h, \xi_\ell) \leq \xi_{NL}$. Secondly, Lemma 19 implies that when $\frac{1}{2P_{\ell L}} \geq \alpha > \alpha_{NL}$, $\xi_{NL} \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. Finally, Lemma 21 and Lemma 19 together shows that $\alpha_{NL} < \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}$. This is because, when $\alpha = \alpha_{NL}$, $\gamma_4 = \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$, while when $\alpha \geq \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}$, $\xi_{NL} \leq \frac{\Delta(\alpha-1/2)}{P_{hL}\cdot(P_{hH}+1)} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. Combining these three facts, we know that

1. When $\frac{1}{2P_{\ell L}} \geq \alpha \geq \min(\frac{1}{2P_{\ell L}}, \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})})$, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta(\alpha-1/2)}{P_{hL}\cdot(P_{hH}+1)} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$.

2. When $\min(\frac{1}{2P_{\ell L}}, \frac{1+P_{hL}+P_{hH}P_{\ell L}}{2(P_{hH}P_{\ell L}+P_{hL}P_{\ell H}+P_{\ell L})}) > \alpha \geq \alpha_{NL}$, $\min(\xi_h, \xi_\ell) \leq \xi_{NL} \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$.

Together this shows that $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ when $\frac{1}{2P_{\ell L}} > \alpha \geq \alpha_{NL}$.

For case (3) where $1 - \beta_\ell^0 = 1$, Proposition 8 guarantees that $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. Lemma 18 guarantees that the range of α in Proposition 8 covers $\frac{1}{2P_{\ell L}} > \alpha \geq \alpha_{NL}$.

Therefore, in all three cases, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. This finishes the proof. $\square$

Proposition 13 (Segment 4)**.** *When $\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} \geq \frac{1}{2}$ and $\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})} \geq \alpha \geq \frac{1}{2}$, no (sequence of) strategy profiles can satisfy all three conditions for $\xi = \frac{1}{2}\Delta\alpha$.*

Proof. Note that when $\alpha \leq \frac{1}{1+P_{\ell L}}$, $\frac{1}{2}\Delta\alpha \geq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. By Lemma 10, Proposition 11, and Proposition 7, it suffices to show that in all three cases in Proposition 7, $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$.

For case (1) where $\beta_h^1 = 1 - \beta_\ell^0 = 0$, $\xi_h = \frac{\Delta\cdot(\alpha-1/2)}{P_{\ell H}} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ and $\xi_\ell = \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. When $\alpha \leq \frac{1}{1+P_\ell}$, $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}} \leq \frac{1}{2}\Delta\alpha$.

For case (2) where $\beta_h^1 = 1$, by Lemma 11 and Lemma 21, for any $\beta_h^1 = 1$ and $\beta_\ell^0 \in [0, 1]$, $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$.

For case (3) where $1 - \beta_\ell^0 = 1$, Proposition 9 guarantees that $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$.

Therefore, in all three cases, $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$. This finishes the proof. □

Proposition 14 (Segment 3)**.** *When $\alpha_{NL} \geq \alpha \geq \max(\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}, \frac{1}{2})$, no (sequence of) strategy profiles can satisfy all three conditions for $\xi = \xi_{NL}$.*

Proof. Note that when $\alpha < \alpha_{NL}$, $\xi_{NL} > \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. By Lemma 10, Proposition 11, and Proposition 7, it suffices to show that in all three cases in Proposition 7, $\min(\xi_h, \xi_\ell) \leq \xi_{NL}$.

For case (1) where $\beta_h^1 = 1 - \beta_\ell^0 = 0$, $\xi_h = \frac{\Delta\cdot(\alpha-1/2)}{P_{\ell H}} < \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$ and $\xi_\ell = \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$. By Lemma 19, for $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}} \leq \xi_{NL}$.

For case (2) where $\beta_h^1 = 1$, by Lemma 11 and Lemma 21, for any $\beta_h^1 = 1$ and $\beta_\ell^0 \in [0, 1]$, $\min(\xi_h, \xi_\ell) \leq \xi_{NL}$.

For case (3) where $1 - \beta_\ell^0 = 1$, Proposition 8 guarantees that for $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{1+P_{\ell L}}$, $\min(\xi_h, \xi_\ell) \leq \frac{\Delta\cdot(\alpha-1/2)}{P_{hL}}$, and Proposition 9 guarantees that $\min(\xi_h, \xi_\ell) \leq \frac{1}{2}\Delta\alpha$. By Lemma 19, for $\alpha_{NL} \geq \alpha > \frac{1}{2}$, $\frac{\Delta\cdot(\alpha-1/2)}{P_{hL}} \leq \xi_{NL}$. Lemma 21 implies that when $\alpha \geq \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}$, $\xi_{NL} \geq \frac{1}{2}\Delta\alpha$. Therefore, for all $\alpha_{NL} \geq \alpha \geq \max(\frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})}, \frac{1}{2})$, $\min(\xi_h, \xi_\ell) \leq \xi_{NL}$.

Therefore, in all three cases, $\min(\xi_h, \xi_\ell) \leq \xi_{NL}$. This finishes the proof. □

Before we wrap up the lower bounds, it remains to prove the lower bounds for Segment 3 and 4, which consists of the following subsection

C.5.10 Lower bound for Segment 3 and Segment 4.

We will start from the simpler Segment 4.

Proposition 15. *When* $\frac{1}{1+P_{\ell L}} \geq \alpha \geq \frac{1}{2}$*, the following strategy profile* Σ *satisfies the constraints in Theorem 12 with* $\xi^* = \frac{1}{2} \cdot \Delta \cdot \alpha$*: all the majority agents report informatively, a fraction* $\gamma^* = \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$ *of minority agents vote for* $\mathbf{R}$*, and the other minority agents* $(1-\alpha-\gamma^* = \frac{1-(P_{hL}+P_{hH})\cdot\alpha}{2}$ *) always vote for* $\mathbf{A}$*.*

Proof. Given that $\gamma^* \leq (1-\alpha)/2$, it suffices to show that $\gamma^* \geq \xi^*$. Note that this is equivalent to $\alpha \leq \frac{1}{P_{\ell L}+P_{\ell H}+\Delta} = \frac{1}{2P_{\ell L}}$. This holds for all $\frac{1}{1+P_{\ell L}} \geq \alpha \geq \frac{1}{2}$.

Now we show that all three constraints are satisfied. Firstly, by Lemma 1, no agents play weakly dominated strategies. For the second constraint of fidelity, we calculated the expected vote share for $\mathbf{A}$ under two world states, respectively.

$$\varphi_H = \alpha \cdot P_{hH} + \frac{1-(P_{hL}+P_{hH})\cdot\alpha}{2} = \frac{1}{2} + \frac{1}{2}\cdot\Delta\cdot\alpha$$
$$\varphi_L = \alpha \cdot P_{hL} + \frac{1-(P_{hL}+P_{hH})\cdot\alpha}{2} = \frac{1}{2} - \frac{1}{2}\cdot\Delta\cdot\alpha.$$

Therefore, $\varphi_H > 0.5$ and $\varphi_L < 0.5$. Moreover, φ_H and φ_L do not depend on n. Therefore, $\liminf \sqrt{n}\cdot(\varphi_H - \frac{1}{2}) = +\infty$ and $\liminf \sqrt{n}\cdot(\frac{1}{2} - \varphi_L) = +\infty$. By applying the first case of Theorem 4, the fidelity of Σ_n converges to 1.

Now we show that Σ is an equilibrium with $\xi^* = \frac{1}{2}\cdot\Delta\cdot\alpha$. By Lemma 15, it suffices to consider a deviating group with only minority agents. For any $\xi < \frac{1}{2}\cdot\Delta\cdot\alpha$, $\varphi_H > \frac{1}{2}+\xi$ and $\varphi_L < \frac{1}{2}-\xi$. Therefore, for any group with a fraction ξ of deviators and any deviating strategy, $\varphi'_H > \frac{1}{2}$ and $\varphi'_L < \frac{1}{2}$. Likewise, applying the first case of Theorem 4 and Lemma 16, the fidelity still converges to 1 after any deviation, and the expected utility gain of minority agents converges to 0. Therefore, the strategy profile proposed is an ε-equilibrium with ε converging to 0. □

Then we give the lower bound for Segment 3.

Proposition 16. *When* $\alpha_{NL} > \alpha \geq \max(\frac{1}{2}, \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})})$*, there exists a (sequence of) strategy profiles that satisfy the constraints with* $\xi = \xi_{NL}$*.*

By Lemma 9, it suffices to give a valid set of $\beta_h^1, (1-\beta_\ell^0)$, and γ. We specify the following set.

$$
\begin{aligned}
\beta_h^{1*} &= 1 \\
1-\beta_\ell^{0*} &= (1-2\alpha) + \frac{\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}}{P_{\ell L}\cdot P_{\ell H}} \\
\gamma^* &= \frac{\frac{1}{2}(P_{\ell L}\cdot(1-\beta_\ell^{0*})+2-P_{\ell L})-\alpha\cdot(P_{hH}P_{\ell L}+P_{\ell H}P_{\ell L}\cdot(1-\beta_\ell^{0*})+P_{\ell H})}{(P_{\ell L}\cdot(1-\beta_\ell^{0*})+P_{hL})\cdot(P_{\ell H}\cdot(1-\beta_\ell^{0*})+P_{hH}+1)}.
\end{aligned}
$$

In this case, by Lemma 21 and the proof of Lemma 11, $\min(\xi_h, \xi_\ell) = \xi_{NL}$.

We show that these parameters are in a valid scope. Specifically, there are three restrictions to be proved.

1. $1-\beta_\ell^{0*} \in [0,1]$.
2. $(1-\alpha-\gamma^*) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ and $\gamma^* > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$.
3. $\gamma^* > \xi_{NL}$ and $1-\alpha-\gamma^* > \xi_{NL}$.

Lemma 21 guarantees that $1-\beta_\ell^{0*} \in [0,1]$.

Proposition 17. *For all* $\alpha_{NL} > \alpha > \max(\frac{1}{2}, \frac{1}{1+P_{\ell L}+(P_{\ell L}-P_{hH})})$, $(1-\alpha-\gamma^*) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha-\frac{1}{2})$ *and* $\gamma^* > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha-\frac{1}{2})$ *holds.*

Proof. Recall the definition of ξ_h and ξ_ℓ.

$$
\begin{aligned}
\xi_h &= \frac{\Delta(\alpha-1/2+(1-\alpha-\gamma)\cdot\beta_h^1)}{P_{hH}\cdot P_{\ell L}\cdot\beta_h^1+P_{\ell H}\cdot P_{\ell L}\cdot(1-\beta_\ell^0)+P_{\ell H}}. \\
\xi_\ell &= \frac{\Delta(\alpha-1/2+\gamma\cdot(1-\beta_\ell^0))}{P_{hH}\cdot P_{hL}\cdot\beta_h^1+P_{hH}\cdot P_{\ell L}\cdot(1-\beta_\ell^0)+P_{hL}}.
\end{aligned}
$$

Now we view ξ_h and ξ_ℓ as functions of $\gamma \in (-\infty, +\infty)$, denoted as $\xi_h(\gamma)$ and $\xi_\ell(\gamma)$. $\xi_h(\gamma)$ is monotonically decreasing on γ, and $\xi_\ell(\gamma)$ is monotonically increasing on γ. Moreover, in Proposition 11, we show that when $\gamma \le \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$, for any $(\beta_h^1, 1-\beta_\ell^0) \in [0,1]^2$, $\xi_\ell(\gamma) \le \frac{\Delta(\alpha-1/2)}{P_{hL}}$. Similarly, when $\gamma \ge (1-\alpha) - \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$, for any $(\beta_h^1, 1-\beta_\ell^0) \in [0,1]^2$, $\xi_h(\gamma) \le \frac{\Delta(\alpha-1/2)}{P_{\ell H}}$. Therefore, for $\gamma \in R \setminus (\frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2}), (1-\alpha) - \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2}))$, for any $(\beta_h^1, 1-\beta_\ell^0) \in [0,1]^2$, $\min(\xi_h, \xi_\ell) \le \frac{\Delta(\alpha-1/2)}{P_{hL}}$. However, given the $\beta_h^{1*}, 1-\beta_\ell^{0*}$, and γ^* provided, $\hat{\xi} = \xi_{NL} > \frac{\Delta(\alpha-1/2)}{P_{hL}}$. This directly implies that $(1-\alpha-\gamma) > \frac{P_{hH}\cdot P_{\ell L}}{P_{\ell H}}(\alpha - \frac{1}{2})$, $\gamma > \frac{P_{hH}\cdot P_{\ell L}}{P_{hL}}(\alpha - \frac{1}{2})$ holds. □

Finally, we prove that $\gamma^* > \xi_{NL}$ and $1-\alpha-\gamma^* > \xi_{NL}$. We first give two alternative forms of γ^*.

Lemma 22.

$$\begin{aligned}
\gamma^* &= \frac{(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H}) + \frac{1}{2}\cdot(\frac{1}{P_{\ell H}} - 2\alpha)\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}}{4(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H}) + (\frac{1}{P_{\ell L}} + \frac{2}{P_{\ell H}} - 4\alpha)\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}} \\
&= \frac{2P_{\ell L}^2 - (3-2\alpha\cdot P_{\ell H})\cdot P_{\ell H}\cdot P_{\ell L} + P_{\ell H}\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}}{2\cdot(2P_{\ell L} - P_{\ell H})^2}.
\end{aligned}$$

Proof. We assign $1-\beta_\ell^{0*}$ into the numerator and denominator of γ^* respectively to get the first alternative form. Recall that

$$1-\beta_\ell^{0*} = (1-2\alpha) + \frac{\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}}{P_{\ell L}\cdot P_{\ell H}}.$$

For the numerator,

$$
\begin{aligned}
&\frac{1}{2}(P_{\ell L} \cdot (1-\beta_{\ell}^{0*}) + 2 - P_{\ell L}) - \alpha \cdot (P_{hH}P_{\ell L} + P_{\ell H}P_{\ell L} \cdot (1-\beta_{\ell}^{0*}) + P_{\ell H}) \\
&= 1 - \alpha P_{\ell L} + \frac{\sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{2P_{\ell H}} \\
&\quad - \alpha \cdot \left(P_{hH}P_{\ell L} + P_{\ell H} + (1-2\alpha) \cdot P_{\ell L}P_{\ell H} + \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}\right) \\
&= 1 - \alpha P_{\ell L} + \frac{\sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{2P_{\ell H}} \\
&\quad - \alpha \cdot \left(P_{\ell H} + P_{\ell L} - 2\alpha \cdot P_{\ell L}P_{\ell H} + \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}\right) \\
&= \frac{1}{2} \cdot (\frac{1}{P_{\ell H}} - 2\alpha) \cdot \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}} + 1 - 2\alpha P_{\ell L} - \alpha P_{\ell H} + 2\alpha^2 \cdot P_{\ell L}P_{\ell H} \\
&= (1-2\alpha P_{\ell L}) \cdot (1-\alpha P_{\ell H}) + \frac{1}{2} \cdot (\frac{1}{P_{\ell H}} - 2\alpha) \cdot \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}.
\end{aligned}
$$

Similarly, for the denominator,

$$
\begin{aligned}
&(P_{\ell L} \cdot (1-\beta_{\ell}^{0*}) + P_{hL}) \cdot (P_{\ell H} \cdot (1-\beta_{\ell}^{0*}) + P_{hH} + 1) \\
&= \left(\frac{\sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{P_{\ell H}} + 1 - 2\alpha P_{\ell L}\right) \\
&\quad \cdot \left(\frac{\sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{P_{\ell L}} + 2 - 2\alpha P_{\ell H}\right) \\
&= 2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) + 2(1-2\alpha P_{\ell L}) \cdot (1-\alpha P_{\ell H}) \\
&\quad + (\frac{1-2\alpha P_{\ell L}}{P_{\ell L}} + \frac{2-2\alpha P_{\ell H}}{P_{\ell H}}) \cdot \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}} \\
&= 4(1-2\alpha P_{\ell L}) \cdot (1-\alpha P_{\ell H}) + (\frac{1}{P_{\ell L}} + \frac{2}{P_{\ell H}} - 4\alpha) \cdot \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}.
\end{aligned}
$$

The second form is reached by multiplying

$$
4(1-2\alpha P_{\ell L}) \cdot (1-\alpha P_{\ell H}) - (\frac{1}{P_{\ell L}} + \frac{2}{P_{\ell H}} - 4\alpha) \cdot \sqrt{2 \cdot (1-\alpha \cdot P_{\ell H}) \cdot (1-2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}
$$

together to the numerator and denominator to eliminate the square root in the denominator.

For the numerator, it becomes

$$
\begin{aligned}
& -\frac{1}{P_{\ell L}}(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}} \\
& +(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot(4(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})-(1-2\alpha P_{\ell H})\cdot(1+\frac{2P_{\ell L}}{P_{\ell H}}-4\alpha P_{\ell L})) \\
= & -\frac{1}{P_{\ell L}}(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}} \\
& +\frac{1}{P_{\ell H}}(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot((3-2\alpha\cdot P_{\ell H})\cdot P_{\ell H}-2P_{\ell L})
\end{aligned}
$$

For the denominator it becomes

$$
\begin{aligned}
& 16(1-2\alpha P_{\ell L})^2\cdot(1-\alpha P_{\ell H})^2-2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}\cdot(\frac{1}{P_{\ell L}}+\frac{2}{P_{\ell H}}-4\alpha)^2 \\
= & \, 2(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H}) \\
& \cdot(8(1+2P_{\ell L}P_{\ell H}\alpha^2-(2P_{\ell L}+P_{\ell H})\alpha)-(16P_{\ell L}P_{\ell H}\alpha^2+(2P_{\ell L}+P_{\ell H})\alpha+\frac{(2P_{\ell L}+P_{\ell H})^2}{P_{\ell L}\cdot P_{\ell H}})) \\
= & \, 2(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot\frac{8P_{\ell L}P_{\ell H}-(2P_{\ell L}+P_{\ell H})^2}{P_{\ell L}\cdot P_{\ell H}} \\
= & \, \frac{2(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})\cdot(2P_{\ell L}-P_{\ell H})^2}{P_{\ell L}\cdot P_{\ell H}}.
\end{aligned}
$$

Therefore, putting the numerator and the denominator together, we have

$$
\gamma^*=\frac{2P_{\ell L}^2-(3-2\alpha\cdot P_{\ell H})\cdot P_{\ell H}\cdot P_{\ell L}+P_{\ell H}\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}}{2\cdot(2P_{\ell L}-P_{\ell H})^2}.
$$

□

Proposition 18. *For all* $\alpha_{NL}>\alpha>\frac{1}{2}$, $\gamma^*>\xi_{NL}$.

Proof. By Lemma 11, we know that

$$
\xi_{NL}=\frac{\Delta\cdot(\alpha-1/2+1/2\cdot(1-\beta_\ell^{0*}))}{(P_{\ell L}\cdot(1-\beta_\ell^{0*})+P_{hL})\cdot(P_{\ell H}\cdot(1-\beta_\ell^{0*})+P_{hH}+1)}.
$$

The denominator of γ^* and that of ξ_{NL} are the same. Therefore, it suffices to compare the numerator and prove

$$\frac{1}{2}(P_{\ell L}\cdot(1-\beta_\ell^{0*})+2-P_{\ell L})-\alpha\cdot(P_{hH}P_{\ell L}+P_{\ell H}P_{\ell L}\cdot(1-\beta_\ell^{0*})+P_{\ell H}) > \Delta\cdot(\alpha-1/2+1/2\cdot(1-\beta_\ell^{0*})).$$

From Lemma 22 (the first form), we know the numerator of γ^* equals to

$$(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})+\frac{1}{2}\cdot(\frac{1}{P_{\ell H}}-2\alpha)\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}.$$

Similarly, for ξ_{NL}, we have

$$\begin{aligned}&\Delta\cdot(\alpha-1/2+1/2\cdot(1-\beta_\ell^{0*}))\\ =&\frac{\Delta}{2P_{\ell L}\cdot P_{\ell H}}\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}.\end{aligned}$$

By subtracting the numerator of ξ_{NL} from that of γ^*, we have

$$(1-2\alpha P_{\ell L})\cdot(1-\alpha P_{\ell H})+\frac{1}{2}\cdot(\frac{1}{P_{\ell H}}-2\alpha-\frac{\Delta}{P_{\ell L}\cdot P_{\ell H}})\cdot\sqrt{2\cdot(1-\alpha\cdot P_{\ell H})\cdot(1-2\alpha\cdot P_{\ell L})\cdot P_{\ell L}P_{\ell H}}$$

Note that

$$\frac{1}{P_{\ell H}}-2\alpha-\frac{\Delta}{P_{\ell L}\cdot P_{\ell H}}=\frac{P_{\ell L}-(P_{\ell L}-P_{\ell H})}{P_{\ell L}\cdot P_{\ell H}}-2\alpha=\frac{1}{P_{\ell L}}-2\alpha.$$

Given the assumption that $\alpha<\alpha_{NL}<\frac{1}{2P_{\ell L}}$, every term in the subtraction is strictly positive. This implies that $\gamma^*>\xi_{NL}$. □

Finally, we will prove $1-\alpha-\gamma^*>\xi_{NL}$ by showing that $1-\alpha-\gamma^*\geq\gamma^*$. This consists of two steps. First, we characterize the range where $1-\alpha-\gamma^*\geq\gamma^*$ holds. Second, we show that this range covers $\alpha_{NL}\geq\alpha\geq\frac{1}{2}$.

Lemma 23. *When* $(2P_{hH}-P_{\ell L}-P_{\ell L}P_{hH})<0$, *for all* $\frac{1}{2P_{\ell L}}\geq\alpha>\frac{1}{2}$, $1-\alpha-\gamma^*\geq\gamma^*$.

When $(2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH}) \geq 0$, *for all*

$$\frac{1}{2} < \alpha \leq \frac{P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}} - \frac{P_{\ell H} \cdot \sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH})}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}},$$

$1 - \alpha - \gamma^* \geq \gamma^*$.

Proof. We solve the inequality $1 - \alpha - \gamma^* \geq \gamma^*$, using the second alternative form of γ^* in Lemma 22.

$$\frac{2P_{\ell L}^2 - (3 - 2\alpha \cdot P_{\ell H}) \cdot P_{\ell H} \cdot P_{\ell L} + P_{\ell H} \cdot \sqrt{2 \cdot (1 - \alpha \cdot P_{\ell H}) \cdot (1 - 2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}}}{2 \cdot (2P_{\ell L} - P_{\ell H})^2} \geq \frac{1 - \alpha}{2}$$

$$\begin{aligned}\Leftrightarrow\ & P_{\ell H} \cdot \sqrt{2 \cdot (1 - \alpha \cdot P_{\ell H}) \cdot (1 - 2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H}} \\ & \geq (2P_{\ell L}^2 + P_{\ell H}^2 - P_{\ell L}P_{\ell H}) - \alpha \cdot (2P_{\ell H}^2 P_{\ell L} + 4P_{\ell L}^2 - 4P_{\ell H}P_{\ell L} + P_{\ell H})\end{aligned}$$

Given that the left-hand side of the inequality is positive (assuming $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$), it suffices to consider the inequality by squaring both sides, i.e.,

$$\begin{aligned}& P_{\ell H}^2 \cdot 2 \cdot (1 - \alpha \cdot P_{\ell H}) \cdot (1 - 2\alpha \cdot P_{\ell L}) \cdot P_{\ell L}P_{\ell H} \\ & \geq ((2P_{\ell L}^2 + P_{\ell H}^2 - P_{\ell L}P_{\ell H}) - \alpha \cdot (2P_{\ell H}^2 P_{\ell L} + 4P_{\ell L}^2 - 4P_{\ell H}P_{\ell L} + P_{\ell H}))^2.\end{aligned}$$

By refactoring on α, this inequality is equivalent to

$$(2P_{\ell L} - P_{\ell H})^2 \cdot (\alpha^2 \cdot (4P_{\ell L}^2 + P_{\ell H}^2 - 4P_{\ell L}P_{\ell H}P_{hH}) - 2\alpha \cdot (2P_{\ell L}^2 + P_{\ell H}^2 - P_{\ell L}P_{\ell H}P_{hH}) + (P_{\ell L}^2 + P_{\ell H}^2)) \geq 0.$$

Applying the root formula of the quadratic equation on

$$\alpha^2 \cdot (4P_{\ell L}^2 + P_{\ell H}^2 - 4P_{\ell L}P_{\ell H}P_{hH}) - 2\alpha \cdot (2P_{\ell L}^2 + P_{\ell H}^2 - P_{\ell L}P_{\ell H}P_{hH}) + (P_{\ell L}^2 + P_{\ell H}^2) = 0,$$

we get two solutions.

$$\alpha = \frac{P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}} \pm \frac{P_{\ell H} \cdot \sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH})}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}}$$

When $(2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH}) < 0$, the two solutions are complex. This directly implies that

$$\alpha^2 \cdot (4P_{\ell L}^2 + P_{\ell H}^2 - 4P_{\ell L}P_{\ell H}P_{hH}) - 2\alpha \cdot (2P_{\ell L}^2 + P_{\ell H}^2 - P_{\ell L}P_{\ell H}P_{hH}) + (P_{\ell L}^2 + P_{\ell H}^2) > 0$$

always holds. Therefore, for all $\frac{1}{2P_{\ell L}} \geq \alpha > \frac{1}{2}$, $1 - \alpha - \gamma^* \geq \gamma^*$.

When $(2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH}) \geq 0$, the two solutions are real numbers. This implies that when

$$\frac{1}{2} < \alpha \leq \frac{P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}} - \frac{P_{\ell H} \cdot \sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH})}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}},$$

$1 - \alpha - \gamma^* \geq \gamma^*$. □

Lemma 24. *When* $(2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH}) \geq 0$*,*

$$\alpha_{NL} \leq \frac{P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}} - \frac{P_{\ell H} \cdot \sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH})}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}}.$$

Proof. Recall that

$$\alpha_{NL} = \frac{2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2 - 2P_{hL} \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2}.$$

We first use scaling methods to unify the square roots in these formulas. Note that when $(2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH}) \geq 0$

$$\sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (2P_{hH} - P_{\ell L} - P_{\ell L}P_{hH})} \leq \sqrt{P_{\ell H} \cdot P_{\ell L} \cdot (P_{hH} - P_{\ell L}P_{hH})} = \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}.$$

Therefore, it suffices to show that

$$\alpha_{NL} \leq \frac{P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}} - \frac{P_{\ell H} \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}}{P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}}.$$

This inequality is equivalent to

$$\begin{aligned}
&(P_{\ell H}^2 + 2P_{\ell L}^2 - P_{\ell H}P_{hH}P_{\ell L}) \cdot (2P_{\ell H}^2 + 8P_{hH}P_{\ell L}^2) \\
&- (P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}) \cdot (2P_{hH}P_{\ell L}^2 + P_{hH} \cdot (P_{hH} + 3P_{\ell L}) - (3P_{hH} + P_{\ell L}) + 2) \\
&+ 2 \cdot (P_{hL} \cdot (P_{\ell H}^2 + 4P_{\ell L}^2 - 4P_{\ell H}P_{hH}P_{\ell L}) - P_{\ell L} \cdot (P_{\ell H}^2 + 4P_{hH}P_{\ell L}^2)) \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}} \\
\geq\ & 0.
\end{aligned}$$

By reorganizing the formulas, the non-square-root term equals to $(P_{\ell L} - P_{hH}) \cdot (2P_{\ell L} - P_{\ell H})^2 \cdot (P_{\ell H} + 2P_{\ell L}P_{hH})$, and the square-root term equals to $-2(P_{\ell L} - P_{hH}) \cdot (2P_{\ell L} - P_{\ell H})^2 \cdot \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}$. Therefore, the inequality is equivalent to

$$(P_{\ell L} - P_{hH}) \cdot (2P_{\ell L} - P_{\ell H})^2 \cdot (P_{\ell H} + 2P_{\ell L}P_{hH} - 2\sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}) \geq 0.$$

By $P_{hH} > P_{hL}$ and $P_{\ell L} > P_{\ell H}$, $P_{\ell L}P_{hH} > \sqrt{P_{hH}P_{hL}P_{\ell H}P_{\ell L}}$. Together with the assumption $P_{\ell L} \geq P_{hH}$. Therefore, the inequality holds, which finishes the proof. □

Combining Lemma 23 and 24, we have the following proposition.

Proposition 19. *For all* $\alpha_{NL} > \alpha > \frac{1}{2}$, $1 - \alpha - \gamma^* > \xi_{NL}$.

Therefore, combining Lemma 21, Proposition 17,18, and 19, Proposition 16 is proved.

C.5.11 Segment 2 to 4: Wrap up the Lower Bound.

Finally, we show the tight lower bound corresponding to the upper bound in each segment.

1. In "Steep" Segment 2, Proposition 6 in the biased case gives the lower bound $\frac{\Delta(\alpha-1/2)}{P_{hL}}$, where all minority agents always vote for **A**.

2. In "Non-Linear" Segment 3, we show that the corresponding ξ^* and $1-\beta_\ell^0$ that maximizes $\hat{\xi}$ serves as a valid set of parameters in Proposition 16, so ξ_{NL} is also a lower bound.

3. For "Tail" Segment 4, the parameters $\beta_h^1 = 1-\beta_\ell^0 = 1$, and $\gamma = \frac{1-(P_{\ell L}+P_{\ell H})\cdot\alpha}{2}$ gives the lower bound $\frac{1}{2}\Delta\alpha$ by Proposition 15.

Therefore, the lower bound on ξ is exactly given in $\xi^*(\alpha)$. We finish the proof that $\xi^*(\alpha)$ is exactly the threshold curve for which a strategy profile satisfying all three constraints exists.