The London School of Economics and Political Science

Essays on Information Economics

Chia-Hung Yeh

A thesis submitted to the Department of Economics of the London School of Economics
for the degree of Doctor of Philosophy

London, July 2025

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 41,845 words.

Statement of inclusion of previous work

I confirm that Chapter 3 of the thesis is a revised version of the work submitted to the Department of Economics of the London School of Economics for the degree of Master of Research in July 2021.

Abstract

Chapter 1 analyses a two-period model of information selling where a risk-neutral seller offers binary signals of varying precision to a risk-neutral buyer. The seller cannot observe signal realisations, creating information asymmetry affecting pricing strategies. We demonstrate that only high-precision signals are offered in equilibrium, as they generate superior information rent whilst enabling natural market segmentation. The buyer reveals private information through their purchasing behaviour, enabling full rent extraction despite information asymmetry. Our analysis reveals that expanding signal menus does not enhance seller profits, and optimal mechanism design converges to perfect signals. These findings challenge conventional wisdom regarding product variety, demonstrating that quality concentration dominates menu diversification in information markets.

Chapter 2 extends the dynamic information selling framework by introducing seller risk aversion. Whilst buyer behaviour remains unchanged, seller risk aversion fundamentally transforms optimal pricing strategies by creating tension between profit maximisation and revenue smoothing. Risk-averse sellers may abandon high-type only strategies in favour of conservative pricing that guarantees universal participation and predictable revenue. Our analysis identifies threshold levels of risk aversion at which optimal strategies shift, depending on signal quality and prior beliefs. Unlike risk-neutral sellers who prefer perfect signals, risk-averse sellers deliberately choose lower-quality signals to increase trading probability. These findings demonstrate that risk factors significantly influence information market design with implications for real-world providers.

Chapter 3 examines whether deliberative mechanisms enhance collective decision-making when committee members possess opposed preferences regarding outcomes. Using cheap talk communication and majority voting, we analyse three equilibrium configurations and their efficiency properties. The analysis reveals that deliberative mechanisms improve upon decision-making based solely on prior beliefs only under restrictive conditions: signal informativeness must exceed the prior and the likelihood of recruiting well-intentioned agents must be sufficiently high. When either condition fails, principals achieve superior outcomes by foregoing deliberation entirely. These findings suggest that deliberation proves counterproductive rather than beneficial in most realistic environments with significant preference conflicts.

Acknowledgements

I am profoundly grateful to my supervisor, Dr Francesco Nava. His patient guidance, intellectual generosity, and unwavering encouragement have shaped my research in countless ways. His consistent belief in my potential has given me the confidence to move forward. His mentorship has not only enriched this project but has also influenced the way I approach learning and independent thought.

I would also like to express my sincere gratitude to Professor Gilat Levy for her guidance and support. Her insights and advice have been greatly appreciated.

I would like to thank my colleagues at the Department, Lakmini Staskus, and Kelly Lewis, for their friendship, support, and camaraderie. Sharing this journey with them has made the experience more meaningful and less daunting.

I am deeply thankful to my mother, Sherry Lee, and my sibling, Marine Yeh, for their steady patience and steadfast belief in me. Their support has sustained me in ways that words cannot fully express.

Special thanks go to i-dle, whose art and message have given me strength during difficult moments. Their devotion to expression and creativity has been a powerful source of inspiration.

Finally, heartfelt thanks to myself for the resilience, dedication, and quiet strength.

In pursuit of light, I have wandered the contours of a labyrinth,
each turn revealing not answers, but echoes.
Perhaps, within those echoes, something lasting remains,
delicately inscribed in ink, serenely held within.

Contents

List of Figures

Chapter 1

Dynamic Sale of Information

1.1 Introduction

A decision maker often seeks external information to improve the likelihood of making a correct decision, particularly when uncertain about the true state of the world, which determines the ex post payoff. In many cases, information may be acquired sequentially if doing so increases expected value. However, external information is typically costly, with sellers setting prices and offering signals of varying quality. This paper investigates the pricing strategies of an information seller who seeks to maximise expected profit. Four central questions are addressed: How should the seller design a price menu for different levels of information quality to capture as much information rent as possible? Can the seller extract full information rent and achieve the first-best outcome? Will a broader menu with more signal types increase expected profit? If the seller can determine the information structure, what is the optimal design?

Consider the following example. An investor must choose between taking a long or short position on a firm. The choice depends on the investor belief about the future performance of the firm. The investor may consult an industry expert who provides costly advice in the form of signals of varying precision. Suppose the expert offers two exogenous signals: one high precision and one lower precision. Further, suppose the investor can purchase a second signal after observing the first. In this context, how should a profit-maximising expert price signals dynamically over two periods? Can full information rent be extracted? Will adding more signal types improve profit? If the expert chooses signal precision, what structure maximises expected profit?

The model, introduced in Section 1.3, features a two-period setting with binary signals differing in precision. Section 1.4 characterises equilibrium strategies using the Perfect Bayesian Equilibrium concept. The analysis shows that only the high-precision signal is used in equilibrium, as it yields the highest information rent. The seller sets prices to induce the buyer to acquire high-type signals in both periods.

The intuition is as follows. In the second period, high-type signals generate strictly greater information rent than low-type signals, independent of the first-period realisation. Thus, low-type signals are not purchased in the second period. In the first period, if a high-type signal confirms the prior belief, the buyer does not purchase a second signal, as the posterior becomes extreme. If the realisation contradicts the prior, information rent is created, which the seller can extract fully by setting prices that bind the participation constraint. In contrast, low-type signals give the buyer (weakly) more reason to purchase a second signal, but the seller cannot observe the first realisation and thus cannot extract rent effectively. Moreover, low-type signals generate less rent in the first period. Hence, offering them reduces expected profit. In equilibrium, the seller offers only high-precision signals, thereby removing information asymmetry and achieving the first-best outcome.

As shown by Coase (1972), commitment devices may help overcome time inconsistency. However, in this model, commitment does not raise the seller expected payoff, for two reasons: (i) information asymmetry arises only in the final period, eliminating self-competition across time; and (ii) signals are not durable, so buyers are not constrained to a single purchase.

Two corollaries follow, discussed in Section 1.5.2 and Section 1.5.3. First, increasing the number of signal types does not raise expected profit. If an equilibrium involves multiple signals, say (q_1^*, q_2^*), the outcome would be unchanged if only these two signals were available. Thus, only the highest-precision signal is purchased. Second, under symmetric binary signals, the optimal design is to offer only perfect signals (i.e., precision equal to one). Any less precise signal can be replaced by a more precise one that yields a higher expected payoff, leading to the original signal being abandoned. The seller therefore maximises profit by offering only perfect signals.

The paper proceeds as follows. Section 1.2 reviews the relevant literature on information selling and mechanism design. Section 1.3 presents the model framework with binary signals and establishes the game structure. Section 1.4 characterises equilibrium strategies using Perfect Bayesian Equilibrium, demonstrating that only high-precision signals are used in equilibrium. Section 1.5 examines first-best outcomes, commitment mechanisms, and optimal information design. Section 1.6 concludes with policy implications and directions for future research.

1.2 Literature Review

The paper closely aligns with literature regarding selling information, particularly under conditions of uncertainty about a payoff-relevant state and heterogeneous buyer beliefs. Some key contributions in this field are listed as follows. Hörner and Skrzypacz (2016) analyse an information provider's private type and derive a gradual persuasion rule. Additionally, Esö and Szentes (2007) examine how a seller controls the release of payoff-relevant information when contracting on the actions of decision maker. Hörner and Skrzypacz (2016) primarily examine information sellers with private preferences and identify a gradual persuasion rule as the optimal strategy for disclosing private information. Bergemann and Bonatti (2015) explore the sale of cookies for online advertising. Zhong (2018) assumes an information provider with a binary private signal selling statistical experiment and characterises the profit-maximising menu, consisting of a continuum of experiments. Daskalakis et al. (2016) explore joint designs for selling physical goods and private information, showing that such designs resemble optimal multi-item mechanisms. Malenko and Malenko (2019) analyse information sales to voters, characterising when such sales lead to more informative voting outcomes. Bergemann et al. (2018) investigate information sales under budget constraints, restricting mechanism design to menus of statistical experiments.

The most closely related work is Bergemann et al. (2014), which examines the optimal design of menus consisting of Blackwell experiments, each comprising a set of binary signals. The subgame that begins at the start of the second period in this paper can be regarded as a special case of the one-period model in Bergemann et al. (2014). In both settings, the seller cannot directly observe the interim belief of the buyer at that stage. However, the nature of the interim belief creates a key distinction. Bergemann et al. (2014) allow for fully flexible interim beliefs, whereas in this model the interim belief space contains only two possible elements, by the nature of assumptions. This restriction effectively nullifies the incentive compatibility constraints. Intuitively, separating equilibria do not arise in this paper because the seller can always improve their payoff by employing a pooling pricing strategy. This paper also differs in two structural aspects. First, the information asymmetry here arises from the inability of the seller to observe the signal realisation at the end of the first period. However, in some cases, the buyer unbiasedly reveals information to the seller through their signal acquisition behaviour. In contrast, Bergemann et al. (2014) assume that the buyer holds private information about the true state, which the seller cannot observe. Second, this paper considers a dynamic two-period model, whereas Bergemann et al. (2014) analyse a static one-period setting. The dynamic structure gives the seller greater flexibility in extracting surplus, as it becomes possible to extract second-period surplus in advance without violating the participation constraint of the buyer. Whilst Bergemann et al. (2014) show that the optimal solution involves offering two signals, one of which must perfectly reveal the true state, this paper shows that only a single perfect signal is ever offered under any equilibrium.

Che and Mierendorff (2019) study a decision maker who allocates limited attention dynamically over different news sources that are biased towards alternative actions, showing that the decision maker adopts a learning strategy biased towards the current belief when the belief is extreme and against that belief when it is moderate. Whilst they examine the demand side of information acquisition with attention constraints, our work examines the supply side with information sellers optimally pricing different signal qualities.

The paper also contributes to the literature on dynamic information acquisition. Zhong (2022) analyses a dynamic model in which a decision maker acquires information about payoffs with flow costs, showing that the optimal policy involves signals arriving according to a Poisson process. Whilst Zhong (2022) focuses on the buyer in continuous time, this paper studies the pricing problem of the seller in a discrete two-period setting with unobservable signal realisations. Doval and Skreta (2022) develop tools for dynamic mechanism design where only short-term mechanisms can be committed to, making information acquisition part of the design. This paper is related through its focus on commitment limitations in a dynamic environment, though the emphasis here is on information asymmetry from unobservable realisations. Liang et al. (2022) show that optimal dynamic information acquisition from multiple correlated sources becomes myopic after finitely many periods. This paper complements that result by examining the seller's side and showing that optimal pricing strategies also follow a myopic pattern, selecting only high-precision signals for the purpose of extracting information rent.

Note that this paper contributes to the literature which differs from some classic studies, including Admati and Pfleiderer (1990) pioneered the explicit consideration of information sales, examining how monopolists reveal noisy signals to traders in financial markets. On the other hand, this paper also differs from mechanism design studying fully general space of contracts (e.g., Nöldeke and Samuelson, 2018).

1.3 Model

This section presents a dynamic model of information acquisition involving two players: a decision maker, henceforth the *buyer*, and an expert, henceforth the *seller*. The interaction unfolds over two discrete periods. The objective is to determine how the seller can price signals of varying types to maximise expected profit. This pricing is constrained by the behaviour of the buyer, whose decisions depend on how beliefs evolve over time. In particular, the seller must satisfy two constraints: a *participation constraint*, ensuring that the buyer finds it optimal to acquire a signal, and an *incentive compatibility constraint*, ensuring that the buyer prefers the intended signal over any alternative.

The model assumes a binary state of the world, denoted $\theta \in \{0, 1\} \equiv \Theta$, which determines the ex-post payoff of the buyer. At the outset, the buyer holds a prior belief that the state is $\theta = 0$, referred to as the *status quo*, with probability $\pi \in \left(\frac{1}{2}, 1\right)$. At the end of period $t = 2$, the buyer selects an action $y \in \{0, 1\}$ with the aim of matching the true state. The payoff from the action is given by:

$$u_B^y(\theta, y) = \mathbb{1}_{y=\theta}.$$

Before making a decision on y, the buyer may acquire an informative binary signal $q_t \in \{q^H, q^L\} \equiv Q$ at the beginning of each period $t \in \{1, 2\}$. Signals are independently drawn across periods and provide noisy information about the state. Each signal realisation is denoted $s_t \in \{0, 1\} \equiv S$, and the buyer chooses from two signal types that differ in informativeness: a high-type signal q^H and a low-type signal q^L, where $1 > q^H > q^L > \frac{1}{2}$. Given a signal of type $q_t \in Q$, the likelihood of receiving an accurate signal realisation is:

$$\mathbb{P}(s_t = \theta \mid \theta) = q_t, \quad \mathbb{P}(s_t \neq \theta \mid \theta) = 1 - q_t.$$

Alternatively, the buyer may also choose not to acquire any signal in any given period, which is denoted $q_t = \emptyset$. Let $Q \equiv \{q^H, q^L\}$ denote the set of informative signals and $Q^0 \equiv Q \cup \{\emptyset\}$ the full choice set. Likewise, let $S^0 \equiv S \cup \{\emptyset\}$ denote the full set of signal realisation.

The seller determines signal pricing by posting a price menu in each period. Let $\mathcal{P}_t(\cdot)$ denote the pricing function in period $t \in \{1, 2\}$. In period 1, the seller posts a function $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$, which assigns a non-negative price to each potential signal choice $q_1 \in Q^0$. In period 2, the seller posts a function $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$, mapping the first-period signal q_1 and the second-period choice q_2 to a non-negative price. It is assumed that $\mathcal{P}_1(\emptyset) = \mathcal{P}_2(q_1, \emptyset) = 0$ for all $q_1 \in Q^0$, so that no payment is required when no signal is acquired.

The game proceeds as follows. At the beginning of period $t = 1$, the seller posts a price menu, $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$, with $\mathcal{P}_1(\emptyset) = 0$. Upon observing this menu, the buyer selects a signal $q_1 \in Q^0$. If an informative signal $q_1 \in Q$ is chosen, the buyer pays $\mathcal{P}_1(q_1)$ and observes a signal realisation $s_1 \in \{0, 1\}$. The buyer then updates their belief using Bayes' rule:

$$\mu_1(q_1, s_1) \equiv \mathbb{P}(\theta = 0 \mid q_1, s_1).$$

If no signal is acquired, (i.e. $q_1 = \emptyset$), the belief remains at the prior π, and the realisation is defined as $s_1 = \emptyset$.

The seller observes the choice of the the buyer, q_1, but crucially not the realisation s_1, and posts a second-period price menu, $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$, with $\mathcal{P}_2(q_1, \emptyset) = 0$. The buyer then chooses a second signal $q_2 \in Q^0$. If an informative signal is chosen, they pay $\mathcal{P}_2(q_1, q_2)$ and observe a second realisation $s_2 \in \{0, 1\}$. The buyer subsequently forms a final posterior belief:

$$\mu_2\left((q_1, q_2), (s_1, s_2)\right) \equiv \mathbb{P}\left(\theta = 0 \mid q_1, q_2, s_1, s_2\right).$$

Finally, the buyer chooses an action $y \in \{0, 1\}$, and payoffs are realised. The ex-post payoff to the seller, denoted u_S, is the sum of payments received across both periods:

$$u_S(q_1, q_2) = \mathcal{P}_1(q_1) + \mathcal{P}_2(q_1, q_2).$$

The ex-post payoff to the buyer, denoted u_B, equals the outcome payoff from the final decision minus the total cost of information:

$$u_B(\theta, y, q_1, q_2) = u_B^y(\theta, y) - \mathcal{P}_1(q_1) - \mathcal{P}_2(q_1, q_2).$$

Both players are risk neutral and apply no discounting. Beliefs are updated according to Bayes' rule throughout the game.

This model highlights a fundamental tension in dynamic information markets. On the one hand, a high-type signal enables the buyer to make better-informed decisions. On the other hand, acquiring information is costly, and the marginal value of future signals depends on how prior beliefs evolve. The seller's challenge lies in designing a pricing mechanism that extracts surplus without observing the buyer's private signal realisations. This creates a trade-off between front-loading prices and preserving incentives for future trade.

For intuition, consider an investor deciding whether to buy or short-sell a stock. The investor begins with a prior belief and may consult an analyst who offers either a brief opinion (a low-type signal) or a detailed report (a high-type signal). If the initial signal contradicts the prior, the investor may seek further confirmation. The analyst must decide whether to charge a high price up front or offer information cheaply to increase the likelihood of future purchases. This captures the dynamic trade-offs that the model aims to formalise.

1.4 Equilibrium Characterisation

This section formally defines and characterises equilibrium strategies and outcomes. We adopt the equilibrium concept of Perfect Bayesian Equilibrium (PBE), focusing on equilibria in pure strategies. A strategy profile satisfies PBE if it meets three conditions. First, sequential rationality requires each player to choose strategies that maximise expected payoff given their beliefs and the observed history at each decision point. Second, belief consistency ensures that beliefs are updated using Bayes' rule wherever applicable, so that expectations about unobserved actions are rational. Third, subgame perfection requires strategies to constitute a Nash equilibrium in every subgame, thereby eliminating non-credible threats.

The strategy sets follow from the structure of the game, with each player facing distinct decision problems at different information sets. A key difference arises in how the buyer and the seller respond to risk. The strategy of the buyer remains unchanged from the risk-neutral case, since their preferences and decision rules depend only on posterior beliefs and pricing menus. In contrast, the strategy of the seller must account for risk aversion, which shifts the trade-off between expected revenue and its variability across periods and states.

The strategy of the buyer consists of three components: a first-period signal choice function, $\mathcal{Q}_1 : \mathbb{R}^2_+ \to Q^0$, which maps the observed menu prices to a signal type; a second-period signal choice function, $\mathcal{Q}_2 : Q^0 \times S^0 \times \mathbb{R}^2_+ \to Q^0$, which maps the observed menu prices in the second period, the first-period signal choice q_1, and its realisation $s_1 \in S^0$, to a signal type; a final action rule, $\mathcal{Y} : (Q^0)^2 \times (S^0)^2 \to \{0, 1\}$, which determines the decision based on all acquired information.

The strategy of the seller consists of two pricing functions. The first-period pricing function, $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$, maps each signal type to a non-negative price. The second-period pricing function, $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$, maps the first-period signal choice of the buyer and each second-period signal type to a non-negative price. Risk aversion fundamentally alters this pricing problem, as the seller must now weigh the higher expected profits from aggressive strategies against the revenue volatility inherent in approaches that concentrate profits in specific market scenarios.

Remark 1.1 (Full Surplus Extraction Benchmark). Before proceeding with the general equilibrium analysis, we establish a crucial benchmark that guides interpretation of our results. When the seller is restricted to offering only high-precision signals q^H, full surplus extraction becomes achievable through strategic dynamic pricing. The mechanism exploits the binary nature of information revelation. Given prior $\pi > \frac{1}{2}$, buyers receiving confirming signal realisation $s_1 = 0$ have their beliefs confirmed and pushed toward certainty so that they require no additional information. Conversely, buyers receiving signal realisation $s_1 = 1$ face contradicted beliefs, with posterior belief falling closer to $\frac{1}{2}$, creating uncertainty and valuable demand for clarification. This stark separation perfectly reveals their first-period realisation through observable returning behaviour. The seller leverages this revelation by extracting full information rent from returning buyers in period $t = 2$, whilst those with $s_1 = 0$ exit after the first period. Crucially, anticipating complete extraction from future returns, the seller can offer first-period signals at minimal prices, even potentially free, ensuring participation whilst capturing all surplus through the dynamic mechanism. The total extraction equals the theoretical maximum: the complete value of information given the prior and signal precision.

Bridge to Main Analysis The full extraction benchmark raises a provocative question: might damaged goods paradoxically enhance profitability by sustaining dynamic demand? The logic appears compelling. High-type signals create an all-or-nothing revelation, as buyers receiving $s_1 = 0$ achieve sufficient certainty and exit, forgoing second-period purchases entirely. By contrast, lower-type signals preserve valuable uncertainty even after confirming realisations. When $q^L < q^H$, buyers observing confirming realisation $s_1 = 0$ maintain sufficient doubt to warrant additional information acquisition, potentially unlocking revenue streams that high-type signals foreclose. A sophisticated seller might thus deliberately dilute signal quality to cultivate persistent demand across periods. Yet this seemingly clever strategy harbours a fatal contradiction.

Lower-type signals generate less first-period value whilst failing to capture the second-period rents they create. Moreover. the very uncertainty that sustains demand also enables buyers to extract information surplus, which implies a failure of full surplus extraction. The seller confronts an irreducible trade-off: damaged goods must either impose prohibitive prices that eliminate participation, or accommodate demand through concessionary pricing that leaks information rents through the screening mechanism. The subsequent analysis establishes this constraint as binding. The ostensible sophistication of menu diversification conceals fundamental inefficiencies, validating quality concentration as the unique profit-maximising mechanism rather than a naive strategy.

The following subsections characterise equilibrium strategies using backward induction, beginning with the unchanged final decision of the buyer and working backward through the modified information acquisition and pricing choices that reflect the risk preferences of the seller.

1.4.1 Final Decision of the Buyer

We begin our equilibrium analysis at the end of the game, applying backward induction to understand how rational players behave at each stage. At this final decision node, the buyer has gathered all available information throughout the game and must choose an action $y \in \{0, 1\}$ to maximise expected payoff, $\mathbb{E}[u_B(\cdot)]$. By the end of period $t = 2$, the buyer has observed signal realisations from both periods and formed a posterior belief, $\mu_2 = \mathbb{P}(\theta = 0 \mid (q_1, q_2), (s_1, s_2))$, using Bayes' rule. This posterior belief represents the best assessment of the buyer of the probability that the true state is $\theta = 0$ (i.e., the status quo), given all observed information.

A crucial insight is that since all prices for acquired signals have been paid by this point, they become sunk costs which do not affect the final decision of the buyer. Furthermore, given the symmetric payoff structure, where a correct match yields $u_B^y = 1$ and an incorrect match yields $u_B^y = 0$, the buyer faces a straightforward optimisation problem: which action maximises the probability of matching the true state? The expected payoff from choosing action $y = 0$ is simply the probability that $\theta = 0$, which equals μ_2. Similarly, the expected payoff from choosing action $y = 1$ is the probability that $\theta = 1$, which equals $1 - \mu_2$. This creates a natural threshold at probability one-half. That is, if the posterior belief satisfies $\mu_2 \geq \frac{1}{2}$, the buyer chooses $y = 0$; otherwise, the buyer chooses $y = 1$. Accordingly, the expected outcome payoff is given by:

$$\mathbb{E}\left[u_B^y\right] = \max\left\{\mu_2, 1 - \mu_2\right\},$$

which equals μ_2 if $\mu_2 \in \left[\frac{1}{2}, 1\right]$, and $1 - \mu_2$ if $\mu_2 \in \left[0, \frac{1}{2}\right]$. This expression elegantly captures the value of information: better information leads to more accurate posterior beliefs, which in turn leads to higher expected payoffs from decision-making. We formalise the optimal decision rule regarding the final action in the following proposition:

Proposition 1.1. *Given the posterior belief $\mu_2 \in (0, 1)$, the optimal decision rule is:*

$$y^* = 0 \;\; \textit{if} \;\; \mu_2 \in \left[\frac{1}{2}, 1\right]; \, y^* = 1 \;\; \textit{if} \;\; \mu_2 \in \left[0, \frac{1}{2}\right].$$

Accordingly, the expected payoff of the buyer is:

$$\mathbb{E}\left[u_B\right] = \max\left\{\mu_2, 1 - \mu_2\right\} - \mathcal{P}_1(q_1) - \mathcal{P}_2(q_1, q_2).$$

Proof. See Appendix 1.A.1. □

This result establishes a crucial benchmark for the subsequent analysis. The expected payoff of the buyer from the final decision equals $\max\{\mu_2, 1 - \mu_2\}$, which reflects the value generated by having accurate information about the true state. The expected payoff is minimised when $\mu_2 = \frac{1}{2}$, corresponding to maximal uncertainty. In contrast, it is maximised as μ_2 approaches either 0 or 1, reflecting minimal uncertainty and an expected payoff approaching 1. This characterises the conditions under which information is valuable: signals are most useful when they shift beliefs away from the midpoint $\mu_2 = \frac{1}{2}$ and towards more extreme levels of certainty. This insight forms the basis for understanding information acquisition behaviour in earlier periods.

Having established how the buyer selects an optimal final action, the analysis now proceeds backward to examine when and why additional costly information would be acquired in the second period. The central question concerns under what conditions further information has the potential to improve final decision quality.

1.4.2 Signal Acquisition by the Buyer in the Second Period

Having established how the buyer selects an optimal final action, the analysis now proceeds backward to examine the information acquisition strategy in the second period. This stage reveals a fundamental insight: information has value only when it has the potential to alter the decision made by the buyer. This leads to the central concept of *pivotal* signals, which plays a key role in the analysis that follows.

At the beginning of period $t = 2$, the buyer faces a new decision: whether to acquire additional information before making a final choice. At this stage, the buyer has observed the first-period signal realisation s_1, formed a posterior belief μ_1 based on the observed information, and been presented with a second-period price menu, $\mathcal{P}_2(q_1, q_2)$, by the seller. The buyer now chooses amongst the following three options. They can acquire a high-type signal (i.e., $q_2 = q^H$), a low-type signal (i.e., $q_2 = q^L$), or no additional information (i.e., $q_2 = \emptyset$).

Let us begin with the simplest case when the buyer chooses $q_2 = \emptyset$. If the buyer forgoes additional information, no further payment to the seller is involved, and the posterior belief remains unchanged, $\mu_2 = \mu_1$. Following the previous analysis, the expected payoff of buyer is given by:

$$\mathbb{E}\left[u_B\right] = \max\left\{\mu_1, 1-\mu_1\right\} - \mathcal{P}_1(q_1).$$

This option is appealing when the buyer is already confident about the true state. If the posterior belief is already extreme, with μ_1 sufficiently close to either 0 or 1, the implied uncertainty is minimal, and the potential benefit of acquiring additional information is unlikely to justify its cost.

The Information Rent of Second-Period Signal. When the buyer acquires an informative signal $q_2 \in Q$, they understand that the posterior belief will be updated according to Bayes' rule. This update depends on the signal realisation s_2, leading to different posterior values based on the observed outcome: $\mu_2(q_1, s_1, q_2, s_2 = 0)$ or $\mu_2(q_1, s_1, q_2, s_2 = 1)$. The buyer evaluates the expected benefit of the information acquisition before buying the signal. Specifically, the buyer forms a belief about the distribution of possible signal realisations, which maps directly to the corresponding posterior beliefs. These posterior beliefs, in turn, determine the expected payoff before the final decision is made, which is characterised in the following proposition:

Proposition 1.2. *Given any posterior belief $\mu_1 \in (0, 1)$, the expected payoff of the buyer after an informative signal $q_2 \in Q$ is acquired (before s_2 is realised) is:*

$$\begin{aligned}\mathbb{E}\left[u_B|\mu_1, q_2\right] &= \mathbb{E}\left[\max\left\{\mu_2, 1-\mu_2\right\} \mid \mu_1, q_2\right] - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) \\ &= \max\left\{\mu_1, 1-q_2\right\} + \max\left\{\mu_1, q_2\right\} - \mu_1 - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot),\end{aligned}$$

where $\mathcal{P}_t(\cdot)$ are the prices set by the seller.

Proof. See Appendix 1.A.2. □

This seemingly complex expression reflects a simple idea. The terms $\max\{\mu_1, 1 - q_2\}$ $+ \max\{\mu_1, q_2\}$ represent the expected payoff from improved decision-making after receiving the signal. The term μ_1 corresponds to the expected payoff under the current decision rule of the buyer. The difference between the two values, defined as the *information rent*, represents the improvement in the expected payoff of the buyer from optimal decision-making that results from acquiring signal, q_2. It measures the value created by better information, net of the current decision-making capability, which motivates the following formal definition:

Definition 1.1 (Information Rent)**.** Given any posterior belief $\mu_1 \in (0, 1)$ at the beginning of period $t = 2$, the information rent associated with signal $q_2 \in Q$ is defined as:

$$\varphi(q_2, \mu_1) = \mathbb{E}\left[\max\left\{\mu_2, 1 - \mu_2\right\} \mid \mu_1, q_2\right] - \max\left\{\mu_1, 1 - \mu_1\right\},$$

where μ_2 denotes the posterior belief after observing the signal realisation s_2, given the signal choice q_2.

The Pivotal Signal Concept. The analysis above leads to a crucial insight. A signal q_2 is only valuable when they can potentially change the final decision of the buyer. This motivates our formal definition of a *pivotal* signal.

Definition 1.2 (Pivotal Signal)**.** An informative signal $q_2 \in Q$ is *pivotal* with respect to the posterior belief μ_1 if and only if it can induce different optimal actions depending on the signal realisation. Formally, q_2 is pivotal if and only if, $\mu_2(q_1, s_1, q_2, s_2 = 0) > \frac{1}{2} > \mu_2(q_1, s_1, q_2, s_2 = 1)$. That is, the informative signal q_2 induces different decisions given different signal realisations (i.e., $y(s_2 = 0) \neq y(s_2 = 1)$).

A signal is pivotal when it can 'flip' the final decision of the buyer. That is, the buyer prefers the action $y = 0$, if they observe $s_2 = 0$; the buyer prefers the action $y = 1$, if they observe $s_2 = 1$. When a signal cannot change the optimal action regardless of its realisation, it provides no information rent to the buyer. The following lemma characterises the condition under which a signal is pivotal.

Lemma 1.1. *Given any posterior belief* $\mu_1 \in (0,1)$*, the signal* q_2 *is pivotal if and only if,* $\mu_1 \in [1-q_2, q_2]$.

Proof. See Appendix 1.A.2. □

This interval $[1-q_2, q_2]$ represents the 'uncertainty zone', where the signal generates strictly positive information rent. Intuitively, if the current belief is sufficiently extreme, or, the signal is relatively uninformative, acquiring additional information does not affect the optimal decision. Given that $\mu_1 > q_2$, the posterior belief μ_1 is sufficiently close to the end of $\theta = 0$, which implies that the buyer is already very confident that $\theta = 0$ is the true state, where even a contradictory signal cannot create enough doubt to change their mind. Conversely, given that $\mu_1 < 1 - q_2$, the buyer is very confident that $\theta = 1$ must be the true state, and again, no signal can create sufficient doubt. The buyer is sufficiently uncertain that new information could genuinely change their preferred action, if and only if the posterior belief is moderate $\mu_1 \in [1-q_2, q_2]$. This result is consistent with the information rent of the signals, which is characterised through the following proposition.

Proposition 1.3 (Information Rent Characterisation)**.** *Given any informative signal* $q_2 \in Q$ *and any posterior belief* $\mu_1 \in (0,1)$ *such that* $\mu_1 \in [1-q_2, q_2]$*, the information rent can be expressed as:*

$$\varphi(q_2, \mu_1) = q_2 - \max\{\mu_1, 1-\mu_1\}.$$

Given any informative signal $q_2 \in Q$ *and any posterior belief* $\mu_1 \in (0,1)$ *such that* $\mu_1 \notin [1-q_2, q_2]$*, the information rent is zero,* $\varphi(q_2, \mu_1) = 0$.

Proof. See Appendix 1.A.2. □

The above proposition confirms that only pivotal signals generate positive information rent. Non-pivotal signals yield zero information rent and do not assist the buyer in improving decision quality. We now show that the buyer has no incentive to acquire an informative signal $q_2 \in Q$ at a strictly positive price $\mathcal{P}_2(\cdot, q_2) > 0$ if the signal q_2 is not pivotal. Intuitively, a rational and sophisticated buyer recognises when additional information offers no value and therefore refrains from paying for it. More precisely, when the information rent is zero, there is no improvement in expected payoff, and any positive payment would reduce the payoff of the buyer. This creates a natural constraint on information demand and limits the ability of the seller to extract rent from overconfident buyers.

Proposition 1.4. *Under any equilibrium, $\mathcal{P}_2(q_2) > 0$ and $\mu_1 \notin [1 - q_2, q_2]$ implies $q_2^* \neq q_2$, $\forall q_2 \in Q$. The buyer will not acquire a non-pivotal signal q_2 at any strictly positive price.*

Proof. See Appendix 1.A.2. □

Note that the buyer is indifferent between acquiring and not acquiring an informative signal if the signal is offered for free (i.e., $\mathcal{P}_2(\cdot, q_2) = 0$). However, setting a zero price is not optimal for the seller unless the seller believes that no information rent can be extracted in the second period with probability one.

Quality Dominance: High-Type Signals Have Broader Applicability. A crucial asymmetry arises when comparing different signal qualities in terms of their potential to generate information rent for buyers. This asymmetry is fundamental to understanding why information markets tend to favour high-quality signals. The key insight is that high-type signals are valuable across a wider range of buyer beliefs than low-type signals. This follows from a straightforward mathematical relationship. Given that $q^H > q^L > \frac{1}{2}$, the corresponding pivotal regions satisfy the strict containment $[1 - q^L, q^L] \subset [1 - q^H, q^H]$.

Proposition 1.5 (Quality Dominance in Pivotal Regions)**.** *Let $\mu_1 \in (0, 1)$ be the posterior belief before period $t = 2$, and suppose that $q^H > q^L > \frac{1}{2}$. If the low-type signal $q_2 = q^L$ is pivotal, then the high-type signal $q_2 = q^H$ is also pivotal.*

Proof. See Appendix 1.A.2. □

The intuition is as follows. High-type signals, due to their greater precision, lead to greater dispersion in posterior beliefs compared to that of low-type signals. When a buyer acquires a high-type signal, the resulting posterior belief shifts further from the prior belief than it would under a low-type signal with the same signal realisation. This larger informational impact allows high-type signals to affect decisions over a broader range of prior beliefs. For example, consider a buyer who holds a moderate belief about the true state. A low-type signal may be too weak to induce a change in decision, whereas a high-type signal with the same realisation might generate enough uncertainty to prompt reconsideration. This gives an advantage to sellers who offer more precise signals.

More importantly, the converse of Proposition 1.5 does not hold. A high-type signal may be pivotal even when a low-type signal is not. This asymmetry creates a fundamental competitive advantage for high-type signals, as they can serve buyers whilst the low-type signals cannot.

Corollary 1.5.1 (Complete Characterisation of Pivotal Cases)**.** *Given any posterior belief $\mu_1 \in (0,1)$ and signal types $(q^H, q^L) \in \left(\frac{1}{2}, 1\right)^2$, exactly one of the following three conditions must hold:*

(i) Both signal types (q^H, q^L) are pivotal if $\mu_1 \in [1-q^L, q^L]$;

(ii) Only the high-type signal q^H is pivotal if $\mu_1 \in [1-q^H, q^H] \setminus [1-q^L, q^L]$;

(iii) Neither signal is pivotal if $\mu_1 \notin [1-q^H, q^H]$.

Proposition 1.5 provides the foundation for our central finding that low-type signals are never offered in equilibrium. When both signal types are pivotal, the seller prefers to offer only high-type signals because they generate higher extractable information rent. When only high-type signals are pivotal, low-type signals have no value. This creates a systematic bias towards high-quality information provision in dynamic information markets.

The Buyer's Optimisation Problem. Having established when information has value and how different signal types compete, we can now characterise the complete decision-making framework of the buyer in the second period. The buyer faces a classic consumer choice problem. Given the available options and their prices, which signal (if any) maximises net surplus?

By Lemma 1.1, given any signal $q_2 \in Q$, it is pivotal if and only if the prior belief satisfies $\mu_1 \in [1-q_2, q_2]$. By Proposition 1.2, if the buyer acquires q_2, their expected payoff is:

$$\mathbb{E}\left[u_B | q_2 \in Q\right] = q_2 - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot),$$

where the posterior belief is assumed to be centred at q_2 in expectation under pivotality. This elegant result captures the essence of information value: the expected payoff of the buyer equals the signal precision minus the total cost of information acquisition. By contrast, if the buyer does not acquire a second-period signal (i.e., $q_2 = \varnothing$), their expected payoff is:

$$\mathbb{E}\left[u_B\right] = \max\left\{\mu_1, 1-\mu_1\right\} - \mathcal{P}_1(\cdot).$$

This represents the outside option of the buyer, which is the expected payoff from making a decision based on current information alone. Hence, the buyer has an incentive to acquire the informative signal $q_2 \in Q$ if and only if doing so yields a weakly higher expected payoff, that is:

$$q_2 - \mathcal{P}_2(\cdot, q_2) \geq \max\{\mu_1, 1 - \mu_1\},$$

or, $\mathcal{P}_2(q_2) = 0$ in which the seller offers the signal at a zero price. This inequality defines the participation constraint, where the outside option corresponds to the expected payoff without acquiring a signal. This participation constraint has a natural economic interpretation. The result can also be interpreted as requiring that the price of the signal not exceed the improvement in expected outcome payoff:

$$\mathcal{P}_2(\cdot, q_2) \leq q_2 - \max\{\mu_1, 1 - \mu_1\} = \varphi(q_2, \mu_1),$$

which is equivalent to the information rent associated with the signal $q_2 \in Q$ by Definition 1.1. If multiple signal types satisfy the participation constraint, the buyer faces a quality choice decision. The risk-neutral buyer will choose the signal that maximises their surplus, defined as the difference between expected payoff improvement and price paid, which creates the incentive compatibility constraint:

$$q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq q_2' - \mathcal{P}_2(\cdot, q_2') \text{ for all } q_2' \in Q.$$

In addition, our analysis of pivotal signals imposes an additional constraint. The buyer will not pay a strictly positive price for a non-pivotal signal, which creates a natural boundary on when information has value, regardless of prices:

$$\mathcal{P}_2(\cdot, q_2) > 0 \implies \mu_1 \in [1 - q_2, q_2].$$

Combining these insights, we summarise these conditions in the following proposition:

Proposition 1.6. *Under any equilibrium, the buyer acquires an informative signal $q_2^* \in Q$ in period $t = 2$ if and only if all the following conditions hold. Otherwise, $q_2^* = \emptyset$.*

(i) $q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq \max\{\mu_1, 1 - \mu_1\}$, *or* $\mathcal{P}_2(\cdot, q_2^*) = 0$, *participation constraint;*

(ii) $q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq q_2' - \mathcal{P}_2(\cdot, q_2')$, $\forall q_2' \in Q$, *incentive compatibility constraint;*

(iii) $\mu_1 \in [1 - q_2^*, q_2^*]$, *or* $\mathcal{P}_2(\cdot, q_2^*) = 0$.

Proof. See Appendix 1.A.2. □

The following discusses the economic implications. This characterisation reveals several key insights regarding information demand in dynamic markets. Optimisation by the buyer reflects forward-looking and sophisticated behaviour, indicating that rational agents recognise exactly when information holds value and avoid paying for signals that cannot alter decision outcomes. This rationality gives rise to natural trade-offs between quality and price. The choice amongst signal types depends on the relative surplus each provides, resulting in competition across quality levels when multiple signals are pivotal. However, the pivotality constraint imposes a fundamental limit on information demand that cannot be addressed through pricing alone. Even at very low prices, a buyer will not acquire information that does not affect the final action. These constraints collectively limit the ability of the seller to extract rent from overconfident agents and ensure that the information market operates efficiently, with demand concentrating on signals that contribute genuine value to decision-making.

Understanding the optimisation problem faced by the buyer is essential for analysing the pricing strategy adopted by the seller. The seller must set prices that satisfy the relevant constraints when maximising expected profit. The central question is how the seller can effectively extract value from the willingness to pay of the buyer for informative signals, whilst accounting for the inherent limitations on information demand.

1.4.3 Menu Setting by the Seller in the Second Period

Having characterised the information acquisition strategy of the buyer, the analysis now turns to the pricing problem faced by the seller. This setting reveals a fundamental tension in dynamic information markets. The seller sets prices without access to private signal realisations of the buyer, which results in information asymmetry that significantly affects market outcomes.

At the beginning of period $t = 2$, the seller faces a classic adverse selection problem. Although the seller observes the signal choice q_1 made by the buyer, the corresponding signal realisation s_1 remains unobserved. As a consequence, the seller remains uncertain about the posterior belief μ_1, which determines the magnitude of information rent that can be extracted through second-period signals.

The task for the seller is to formulate a pricing strategy that maximises expected revenue whilst satisfying participation and incentive constraints. Since demand depends on whether a signal is pivotal, and pivotality depends on the unobserved signal realisation, the seller must anticipate the distribution of possible beliefs and choose prices accordingly. This information asymmetry imposes a fundamental constraint on the ability to extract surplus. Unlike a monopolist who can condition prices on observable customer types, the information seller must infer willingness to pay through revealed preferences. This strategic limitation plays a central role in shaping equilibrium outcomes.

The following subsections proceed by analysing the pricing strategy adopted by the seller separately for the two possible first-period signal types: (i) $q_1 = q^H$ and (ii) $q_1 = q^L$.

Case 1: High-Type Signal Previously Acquired $(q_1 = q^H)$

Suppose that the buyer acquires a high-type signal in the first period, $q_1 = q^H$. The seller now faces two possible scenarios, each corresponding to a different signal realisation that they cannot observe: (i) $s_1 = 0$ and (ii) $s_1 = 1$. If $s_1 = 0$ is realised, the buyer will have no incentive to acquire any signal at a strictly positive price in period $t = 2$. To see this formally, first note that the posterior belief is updated to:

$$\mu_1(q^H, 0) = \frac{\pi q^H}{\pi q^H + (1-\pi)(1-q^H)} > q^H.$$

By Proposition 1.4, $\mu_1(q^H, 0) \notin [1 - q_2, q_2]$ holds for any $q_2 \in Q$, which implies that neither type of signal is pivotal. Thus, the buyer will not acquire any signal at a strictly positive price as there exists zero information rent associated with any signal $q_2 \in \{q^H, q^L\}$. The intuition is clear. Given two binary signals with the same precision (e.g., $q_1 = q_2 = q^H$), any Bayesian decision maker forms a posterior belief identical to the prior belief if the two realisations go against each other. For instance, $(s_1, s_2) = (0, 1)$ implies that the posterior belief equals $\pi > \frac{1}{2}$, in which $y = 0$ is preferred. If $s_2 = 0$ is realised instead, the posterior belief will be more extreme towards the state $\theta = 0$, which implies that $y = 0$ is preferred if $q_2 = q^L$ is acquired. As discussed earlier, the variance in the posterior belief resulting from the low-type signal will be smaller, indicating that the posterior belief will be sufficiently solid on the status quo, which implies that $y = 0$ is preferred. Conclusively, if $q_1 = q^H$ is acquired and $s_1 = 0$ is realised, any signal in the second period will not be pivotal and thus with zero information rent. From the perspective of the seller, this represents a dead end, since no further revenue can be extracted regardless of the pricing strategy.

If $s_1 = 1$ is realised, the resulting posterior belief moves closer to $\frac{1}{2}$, creating sufficient uncertainty that may render second-period signals valuable. The buyer considers acquiring additional information, provided that the price does not exceed the associated information rent. More precisely, the buyer has an incentive to acquire a high-type signal in period $t = 2$, as long as the prior belief is sufficiently weak such that the signal is pivotal. According to Proposition 1.4, the posterior belief must not be too extreme, that is, $\mu_1(q^H, 1) \in [1 - q_2, q_2]$.

Formally, the buyer has an incentive to acquire a high-type signal at a strictly positive price in period $t = 2$ only if the following condition holds:

$$\mu_1(q^H, 1) = \frac{\pi(1 - q^H)}{\pi(1 - q^H) + (1 - \pi)q^H} \in \left[1 - q^H, q^H\right],$$

which is equivalent to the more compact expression,

$$\frac{\pi(1 - q^H)}{(1 - \pi)q^H} \in \left[\frac{1 - q^H}{q^H}, \frac{q^H}{1 - q^H}\right] \iff \frac{\pi(1 - q^H)}{(1 - \pi)q^H} \leq \frac{q^H}{1 - q^H} \iff \frac{\pi}{1 - \pi} \leq \left(\frac{q^H}{1 - q^H}\right)^2.$$

The seller recognises a crucial pattern in buyer behaviour: only those who receive contradictory signals, with $s_1 = 1$, return to acquire additional information in the second period. This behavioural response enables a powerful inference mechanism. By merely approaching the seller again, the buyer implicitly reveals private information. More specifically, it can be inferred that the first signal realisation contradicted the initial belief and left the buyer uncertain about the true state. Given this, the seller understands that no revenue can be extracted when $s_1 = 0$, regardless of the price set for the high-type signal. As a result, the seller focuses exclusively on scenarios in which $s_1 = 1$ is realised. In these cases, the seller sets the price equal to the full magnitude of the available information rent, thereby maximising expected revenue from the subset of buyers who find the signal valuable and reveal themselves through their actions.

This revelation solves information problem of the seller. The seller can now extract the full information rent from a returning buyer by setting the price:

$$\mathcal{P}_2(q^H, q^H) = q^H - \max\left\{\mu_1(q^H, 1), 1 - \mu_1(q^H, 1)\right\} = \varphi\left(q^H, \mu_1(q^H, 1)\right),$$

which satisfies the corresponding participation constraint, by Proposition 1.6:

$$q^H - \mathcal{P}_2(q^H, q^H) \geq \max\left\{\mu_1(q^H, 1), 1 - \mu_1(q^H, 1)\right\}.$$

Likewise, the buyer has incentive to acquire a low-type signal in period $t = 2$ only if:

$$\mu_1(q^H, 1) = \frac{\pi(1-q^H)}{\pi(1-q^H) + (1-\pi)q^H} \in \left[1-q^L, q^L\right].$$

However, according to Proposition 1.5, if the low-type signal is pivotal, the high-type signal is also pivotal. When both signals are pivotal, the seller has an incentive to induce the buyer to acquire a high-type signal rather than a low-type one, as the associated information rent is strictly greater. Formally, the difference in information rent between the two signals is given by:

$$\varphi\left(q^H, \mu_1(q^H, 1)\right) - \varphi\left(q^L, \mu_1(q^H, 1)\right) = q^H - q^L > 0.$$

Crucially, this rent differential remains constant at $q^H - q^L$ regardless of the buyer's posterior belief μ_1, indicating the failure of single-crossing in our symmetric framework. In models with single-crossing, the incremental value of higher quality would vary with the buyer's type (here, their posterior belief), enabling the seller to design separating menus where different types self-select into different qualities. However, because matching either state yields identical payoffs in our model, all buyer types value the quality upgrade $q^H - q^L$ identically, making profitable separation impossible.

This implies that the seller earns a strictly higher payoff by offering the high-type signal, without incurring any additional cost. As a result, there is a natural tendency towards quality concentration in information markets. The absence of single-crossing ensures this conclusion continues to hold even when there is strictly positive information rent under the realisation $s_1 = 0$. Moreover, if the seller adopts a separating strategy, offering different signals depending on the realised value of s_1, they can still profitably induce the buyer to acquire a high-type signal in both cases. In particular, the buyer who would otherwise be induced to acquire a low-type signal can instead be offered the high-type signal at a price higher by $q^H - q^L$, without violating participation or incentive compatibility constraints. The following proposition and proof formally establish this claim.

Proposition 1.7. *Under any equilibrium, given any $\mu_1 \in (0,1)$ and $(p, q^H, q^L) \in (\frac{1}{2}, 1)^3$, the seller has no incentive to offer a low-type signal in period $t = 2$.*

Proof. See Appendix 1.A.3. □

To sum up, we have the following proposition given $q_1 = q^H$ is acquired:

Proposition 1.8. *Suppose that* $\frac{\pi}{1-\pi} \in \left(1, \left(\frac{q^H}{1-q^H}\right)^2\right]$ *and* $q_1 = q^H$ *hold. The buyer will have no incentive to acquire either signal type at a strictly positive price if* $s_1 = 0$ *is realised. The buyer will have incentive to acquire a high-type signal at a price lower than* $\varphi\left(q^H, \mu_1(q^H, 1)\right)$ *if* $s_1 = 1$ *is realised.*

Proof. See Appendix 1.A.3. □

We define such strategy of the seller as *high-type only pricing strategy*, since the seller induces the buyer to acquire only high-type signals in both periods. A formal definition will be given in Section 1.3. Next characterise the optimal menu setting strategy. The seller sets the price of a high-type signal such that their participation constraint binds, whilst the price of a low-type signal such that their participation constraint fails.

Proposition 1.9. *Suppose that* $\frac{\pi}{1-\pi} \in \left(1, \left(\frac{q^H}{1-q^H}\right)^2\right]$ *and* $q_1 = q^H$ *hold. The optimal menu setting strategy in period* $t = 2$ *is:*

$$\mathcal{P}_2(q^H, q^H) = \varphi\left(q^H, \mu_1(q^H, 1)\right);$$
$$\mathcal{P}_2(q^H, q^L) \in \varphi\left(q^H, \mu_1(q^H, 1)\right) - \left(q^H - q^L\right), +\infty\Big).$$

More specifically,

$$\left\{\mathcal{P}_2(q^H, q^H), \mathcal{P}_2(q^H, q^L)\right\} \in \left\{q^H - \mu_1(q^H, 1)\right\} \times \left(q^L - \mu_1(q^H, 1), +\infty\right) \;\; \text{if } \pi \geq q^H;$$
$$\left\{\mathcal{P}_2(q^H, q^H), \mathcal{P}_2(q^H, q^L)\right\} \in \left\{q^H - \left(1 - \mu_1(q^H, 1)\right)\right\} \times \left(q^L - \left(1 - \mu_1(q^H, 1)\right), +\infty\right) \;\; \text{if } \pi \leq q^H.$$

Next we investigate the surplus allocation between the two parties. First note that no information rent exists under the realisation $s_1 = 0$, as the posterior belief is sufficiently extreme, $\mu_1(q^H, 0) > q^H$. Consequently, the buyer has no incentive to acquire any signal with a strictly positive price. The expected payoff of the buyer evaluated before the decision is $\mathbb{E}[u_B] = \mu_1(q^H, 0)$. Under the realisation $s_1 = 1$, the seller extracts all the available information rent throughout the trade. As a result, there is no change in the expected payoff of the buyer after the signal acquisition in period $t = 2$. Remarkably, the seller attains the first-best outcome despite the presence of information asymmetry. The strategic behaviour of the buyer, who returns only when the initial signal contradicts the prior belief, effectively reveals private information. This implicit disclosure allows the seller to extract the full information rent without incurring the efficiency losses that are typically associated with adverse selection.

Case 2: Low-Type Signal Previously Acquired $(q_1 = q^L)$

This section examines cases in which the buyer acquires a low-type signal in the first period, $q_1 = q^L$. The analysis is more intricate, as low-type signals generate less extreme posterior beliefs. As a result, the buyer is more likely to return for additional information, regardless of the signal realisation. In contrast to the high-type case, the seller now faces two types of returning buyer: those who observe $s_1 = 0$ and those who observe $s_1 = 1$. These buyers differ in their willingness to pay, yet the seller cannot distinguish between them. This scenario presents a classic monopoly pricing problem. Should the seller set a high price to extract the maximum possible rent from those willing to pay more, or a low price to serve all buyers?

By Proposition 1.4, the buyer has an incentive to acquire a high-type signal in period $t = 2$ at a strictly positive price only if the posterior belief after observing $s_1 = 0$ lies within the pivotal region, that is, $\mu_1(q^L, 0) \in [1 - q^H, q^H]$. This condition is equivalent to:

$$\begin{aligned}\mu_1(q^L, 0) \in \left[1 - q^H, q^H\right] &\iff \frac{\pi q^L}{(1-\pi)(1-q^L)} \in \left[\frac{1-q^H}{q^H}, \frac{q^H}{1-q^H}\right] \\ &\iff \frac{\pi}{1-\pi} \leq \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}.\end{aligned}$$

If this condition fails, the buyer will not have an incentive to acquire any additional signal. If instead $s_1 = 1$ is realised, the buyer may be willing to acquire a high-type signal in period $t = 2$ if it is pivotal. The corresponding posterior belief is given by:

$$\mu_1(q^L, 1) = \frac{\pi(1-q^L)}{\pi(1-q^L) + (1-\pi)q^L}.$$

The buyer will acquire the high-type signal only if:

$$\begin{aligned}\mu_1(q^L, 1) \in \left[1 - q^H, q^H\right] &\iff \frac{\pi(1-q^L)}{(1-\pi)q^L} \in \left[\frac{1-q^H}{q^H}, \frac{q^H}{1-q^H}\right] \\ &\iff \frac{\pi}{1-\pi} \leq \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}.\end{aligned}$$

By Proposition 1.7, the seller has no incentive to induce the buyer to acquire a low-type signal in period $t = 2$. Therefore, it is sufficient to focus on the pricing strategy for the high-type signal.

The relationship between the prior belief of the buyer and the quality of available signals gives rise to three distinct market regimes:

Regime 1: No Demand

$$\frac{\pi}{1-\pi} > \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}.$$

When the buyer holds an extremely strong prior belief, even a contradictory high-type signal fails to generate sufficient uncertainty to make further information acquisition worthwhile. The buyer remains confident regardless of the signal realisation, and no trade occurs in the second period.

Regime 2: Selective Demand

$$\frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H} < \frac{\pi}{1-\pi} \leq \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}.$$

In this intermediate range, only buyers who observe contradictory signals, that is $s_1 = 1$, return for additional information. A confirming signal reinforces the prior belief too strongly, whereas a contradictory signal introduces just enough uncertainty to make further information valuable.

Regime 3: Universal Demand

$$\frac{\pi}{1-\pi} \leq \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}.$$

When the prior belief is relatively weak, further information is valuable, regardless of the first-period signal realisation. The buyer returns in the second period, although their willingness to pay differs depending on the realisation observed. These distinctions motivate the following proposition.

Proposition 1.10. *Given that $q_1 = q^L$ holds. The buyer will have incentive to acquire a signal of high type at a price lower than $\varphi\left(q^H, \mu_1(q^L, s_1)\right)$ if one the following conditions holds:*

(i) $s_1 = 0$ is realised and $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right]$ holds;
(ii) $s_1 = 1$ is realised and $\frac{\pi}{1-\pi} \in \left(1, \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}\right]$ holds.

This analysis reveals why information markets naturally evolve towards quality differentiation. Low-type signals create persistent uncertainty that keeps buyers in the market longer, but this apparent advantage comes with a cost: the seller faces a more complex pricing problem with multiple buyer types. High-type signals, by contrast, create clearer market segmentation where buyer behaviour more reliably reveals their private information. The existence of these three regimes also explains why information providers might strategically choose their quality levels. By offering higher-quality initial information, sellers can better predict and manage second-period demand, potentially achieving higher overall profits despite serving fewer customers in the second period.

The Selective Demand Case: Natural Market Segmentation

The analysis now turns to the characterisation of the optimal pricing strategy adopted by the seller, beginning with the selective demand regime, defined by,

$$\frac{\pi}{1-\pi} \in \left(\frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}, \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}\right].$$

In this regime, the pricing problem is relatively straightforward. Only buyers who observe contradictory signals, namely $s_1 = 1$, return to acquire additional information. Those who observe confirming signals, $s_1 = 0$, remain too confident to place value on further information at any positive price. This outcome results in natural segmentation of the market, which simplifies the pricing decision faced by the seller.

Two strategic insights guide the seller. First, by Proposition 1.7, offering only high-type signals dominates any strategy involving low-type signals, as the surplus gained from higher precision, $q^H - q^L$, outweighs potential demand effects. Second, since only one type of buyer returns for information, the seller is able to extract the entire information rent without encountering adverse selection amongst different types of buyer.

According to the logic established in Proposition 1.9, the seller sets the price of the high-type signal so that the participation constraint of the returning buyer binds exactly. This approach ensures full rent extraction from the only type of buyer willing to acquire second-period information. The strategy is both simple and efficient: charge the entire information rent to the buyer who values it most, whilst excluding those who generate negligible profit. Thus, we have the following proposition.

Proposition 1.11. *Suppose that* $\frac{\pi}{1-\pi} \in \left(\frac{1-q^L}{q^L}\frac{q^H}{1-q^H}, \frac{q^L}{1-q^L}\frac{q^H}{1-q^H}\right]$ *and* $q_1 = q^L$ *hold. The optimal menu setting strategy in period* $t = 2$ *is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L, 1)\right);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi\left(q^H, \mu_1(q^L, 1)\right) - \left(q^H - q^L\right), +\infty\right).$$

More specifically,

$$\left\{\mathcal{P}_2(q^L, q^H), \mathcal{P}_2(q^L, q^L)\right\} \in \left\{q^H - \mu_1(q^L, 1)\right\} \times \left(q^L - \mu_1(q^L, 1), +\infty\right) \;\; \text{if } \pi \geq q^L;$$
$$\left\{\mathcal{P}_2(q^L, q^H), \mathcal{P}_2(q^L, q^L)\right\} \in \left\{q^H - \left(1 - \mu_1(q^L, 1)\right)\right\} \times \left(q^L - \left(1 - \mu_1(q^L, 1)\right), +\infty\right) \;\; \text{if } \pi \leq q^L.$$

This case shows how asymmetry of information can benefit the seller by producing endogenous segmentation in the market. Rather than attempting to price discriminate across buyer types, the seller relies on buyers' own optimisation to reveal willingness to pay through observed purchasing behaviour.

The Universal Demand Case: A Complex Pricing Challenge

When the prior belief held by the buyer is relatively weak, specifically when,

$$\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right],$$

the seller faces a more intricate pricing problem. In this regime, the buyer returns for additional information regardless of the signal observed in the first period, but with different levels of willingness to pay.

The key insight is that although both types of buyer place value on additional information, their willingness to pay differs substantially. Buyers who observe contradictory signals (i.e., $s_1 = 1$) are more uncertain and therefore place higher value on further information than buyers who observe confirming signals (i.e., $s_1 = 0$). This results in an asymmetry in information rents that the seller must navigate strategically. Formally, the information rent is strictly greater when $s_1 = 1$ is observed.

Lemma 1.2. *Given any* $(\pi, q^H, q^L) \in (0, \frac{1}{2})^3$, *the information rent given signal realisation* $s_1 = 1$ *is* *greater than that given signal realisation $s_1 = 0$. That is:*

$$\varphi\left(q^H, \mu_1(q^L,1)\right) = q^H - \max\left\{\mu_1(q^L,1), 1-\mu_1(q^L,1)\right\}$$
$$> q^H - \mu_1(q^L,0) = \varphi\left(q^H, \mu_1(q^L,0)\right),$$

always holds.

Proof. See Appendix 1.A.3. □

Faced with heterogeneity in willingness to pay, the seller must choose between two strategic pricing approaches, each reflecting a different method of extracting value from the market. The first approach is the *aggressive pricing strategy*, which focuses on extracting information rent from the highest-paying buyers, even at the cost of excluding others. The seller sets the price equal to the information rent of the buyer who observed $s_1 = 1$:

$$\mathcal{P}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L,1)\right).$$

This price exceeds the willingness to pay of buyers who observed confirming signals $s_1 = 0$, and thus excludes them from the market. Under such strategy, only buyers with contradictory signals $s_1 = 1$ purchase second-period information, with probability of $\pi(1-q^L) + (1-\pi)q^L$. This approach maximises the ex-post profit by targeting high-value buyers, and is preferred when the difference in information rents between buyer types is substantial.

Proposition 1.12 (Aggressive Pricing Strategy)**.** *Suppose $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right]$ and $q_1 = q^L$. If the aggressive pricing strategy is optimal, then the optimal menu setting in period $t = 2$ is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L,1)\right);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi\left(q^H, \mu_1(q^L,1)\right) - \left(q^H - q^L\right), +\infty\right).$$

More explicitly,

$$\mathcal{P}_2(q^L, q^H) = q^H - \mu_1(q^L,1) \;\; \textit{if} \;\; \pi \geq q^L;$$
$$\mathcal{P}_2(q^L, q^H) = q^H - \left(1 - \mu_1(q^L,1)\right) \;\; \textit{if} \;\; \pi \leq q^L.$$

The price of the low-type signal is set prohibitively high to ensure it is not purchased.

The aggressive strategy reflects a central principle of monopoly pricing: when customers differ in willingness to pay, it may be optimal to serve only those with the highest valuations. This is especially relevant in information markets, where marginal cost is negligible and foregone transactions do not save resources. This strategy also illustrates how asymmetry of information can benefit the seller by generating endogenous segmentation. Rather than requiring complex mechanisms to separate buyer types, the seller relies on own decisions of the buyer to reveal willingness to pay.

The second approach is the *conservative pricing strategy*, which prioritises market coverage over per-unit rent extraction. Under such strategy, the seller sets prices low enough to ensure that all returning buyers purchase additional information, regardless of the signal realisation, which exhibits a fundamental trade-off between profit margins and market penetration. The seller implements this strategy by setting the price equal to the information rent of the least willing buyer:

$$\mathcal{P}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L, 0)\right).$$

This price is deliberately conservative, falling below the maximum willingness to pay of buyers who received contradictory signals. The economic logic is compelling: by ensuring universal participation, the seller eliminates demand uncertainty and captures revenue from the entire market, albeit at reduced margins. Under the conservative pricing strategy, the seller earns a guaranteed payoff $\varphi\left(q^H, \mu_1(q^L, 0)\right)$ in the second period with probability one, regardless of the signal realisation. This contrasts with the aggressive pricing strategy, under which profits are higher per transaction but arise with lower probability.

The optimal strategy choice of the seller between these two strategies reflects a classic monopoly pricing dilemma. The conservative strategy is preferred when:

$$\varphi\left(q^H, \mu_1(q^L, 0)\right) \geq \left[\pi(1 - q^L) + (1 - \pi)q^L\right] \varphi\left(q^H, \mu_1(q^L, 1)\right).$$

Rearranging this condition reveals that the conservative strategy dominates when:

$$q^H \geq \bar{q}^H \equiv \frac{\max\{\pi(1-q^L),(1-\pi)q^L\}}{\pi q^L+(1-\pi)(1-q^L)} + \frac{\pi q^L}{[\pi q^L+(1-\pi)(1-q^L)]^2}.$$

This cut-off point $\bar{q}^H$ represents the critical signal quality where the seller becomes indifferent between the two strategies. The intuition is straightforward: when q^H is close to $\mu_1(q^L, 0)$, the information rent from conservative buyers becomes negligible, making it optimal to focus exclusively on eager buyers through aggressive pricing.

Proposition 1.13 (Conservative Pricing Strategy). *Suppose that $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right]$ and $q_1 = q^L$. If the conservative pricing strategy is optimal, then the optimal menu setting strategy in period $t = 2$ is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L, 0)\right);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi\left(q^H, \mu_1(q^L, 0)\right) - \left(q^H - q^L\right), +\infty\right).$$

The price for the low-type signal is set prohibitively high to prevent its acquisition, consistent with the quality concentration result from Proposition 1.7.

The welfare implications of these pricing strategies reveal a surprising insight about information asymmetry in dynamic markets. Under the aggressive strategy, the seller cannot capture surplus from all buyers due to information asymmetry, since buyers who received confirming signals retain positive information rent that goes unextracted. This creates apparent inefficiency relative to a first-best benchmark where the seller observes all signal realisations. However, this interpretation is changed when the conservative pricing strategy is considered in a dynamic context. Although buyers who received contradictory signals (i.e., $s_1 = 1$) obtain second-period surplus equal to $\varphi\left(q^H, \mu_1(q^L, 1)\right) - \varphi\left(q^H, \mu_1(q^L, 0)\right) > 0$, this inefficiency is only superficial. The seller can extract the expected value of this surplus through higher pricing in the first period, thereby achieving first-best outcomes via intertemporal rent extraction. This reveals a profound insight. Dynamic pricing can overcome information asymmetries that would otherwise generate inefficiencies in static settings. The ability of the seller to adjust prices across time allows for sophisticated rent extraction mechanisms that capture the full value of information provision, even when some buyers retain surplus in individual periods.

The choice between aggressive and conservative pricing reflects deeper strategic considerations regarding market development and buyer engagement. The aggressive strategy maximises short-term profit but may limit long-term participation. The conservative strategy encourages broad participation but requires more sophisticated pricing policies over time. From a market design perspective, this analysis suggests that providers of information face inherent trade-offs between targeting narrow, high-value buyer segments and supporting broad participation through inclusive pricing. The optimal approach depends critically on the quality differential between signals and the distribution of buyer beliefs—factors that help explain the diversity of pricing strategies observed in real-world information markets.

1.4.4 Signal Acquisition by the Buyer in the First Period

Having characterised the second-period equilibrium, the analysis now turns to the information acquisition decision of the buyer in period $t = 1$. This reveals how forward-looking buyers incorporate anticipated future opportunities into current choices, creating dynamic linkages that shape the pricing power of the buyer across periods.

At the beginning of period $t = 1$, the buyer faces a strategic decision that extends beyond the immediate value of information. They must consider not only the direct benefit of first-period information, but also how this choice influences future opportunities to acquire additional signals. As a consequence, this creates a dynamic optimisation problem, where current decisions affect future payoffs. The expected payoff of the buyer consists of two components: the immediate gain from first-period information and the expected surplus from future information acquisition, denoted $V(q_1, \mathcal{P}_2)$. Formally, the buyer solves the following optimisation problem:

$$\max_{q_1 \in Q^0} \mathbb{E}[u_B] = \max_{q_1 \in Q^0} \left\{ \mathbb{1}_{q_1 \in Q} \left\{ \max\{\pi, q_1\} - \mathcal{P}_1(q_1) + V(q_1, \mathcal{P}_2) \right\} + \mathbb{1}_{q_1 = \emptyset}\{\pi\} \right\},$$

The term $V(q_1, \mathcal{P}_2)$ captures the expected net surplus from acquiring second-period information, which depends on the first-period signal choice of the buyer and the anticipated pricing strategy in period $t = 2$. One key insight is that the value function $V(q_1, \mathcal{P}_2)$ must be non-negative for all $q_1 \in Q^0$, since the buyer always has the option to choose $q_2 = \emptyset$ in the second period, which ensures that second-period optimisation is consistent with expectations formed in the first period. The optimal first-period signal choice depends crucially on the nature of second-period pricing strategy of the seller. Two benchmark cases are considered. If the seller implements an aggressive pricing strategy, all available information rent is extracted in period $t = 2$. The buyer anticipates zero surplus from future information acquisition, so that $V(q_1, \mathcal{P}_2) = 0$ for all signal types. In this case, the optimisation problem of the buyer is reduced to:

$$\max_{q_1 \in Q^0} \mathbb{E}[u_B] = \max_{q_1 \in Q^0} \left\{ \mathbb{1}_{q_1 \in Q} \left\{ \max\{\pi, q_1\} - \mathcal{P}_1(q_1) \right\} + \mathbb{1}_{q_1 = \emptyset}\{\pi\} \right\},$$

The solution, denoted as $q_1^* \in Q^0$, must satisfy the following conditions:

$$\begin{aligned} & q_1^* - \mathcal{P}_1\left(q_1^*\right) \geq \pi, \text{ or, } \mathcal{P}_1\left(q_1^*\right) = 0, \\ & q_1^* - \mathcal{P}_1\left(q_1^*\right) \geq q_1' - \mathcal{P}_1\left(q_1'\right), \forall q_1' \in Q, \end{aligned}$$

which suggests that the buyer simply compares the price with the immediate value of information. When the seller extracts all future rents, the buyer becomes entirely myopic. They acquire a signal only if its immediate value exceeds the cost. This imposes a natural upper bound on the first-period pricing power of the seller.

When the seller implements a conservative pricing strategy, the buyer retains surplus in period two. If $q_1 = q^L$ and conservative pricing is optimal, with $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right]$, then the implied expected future surplus is:

$$V(q^L, \mathcal{P}_2) = \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi\left(q^H, \mu_1(q^L, 1)\right) - \varphi\left(q^H, \mu_1(q^L, 0)\right)\right],$$

which represents the expected surplus earned when the buyer receives a contradictory signal and returns for additional information. Thus, the buyer has the following reduced maximisation problem:

$$\max_{q_1 \in Q^0} \mathbb{E}[u_B] = \max_{q_1 \in Q^0} \left\{ \mathbb{1}_{q_1 \in Q} \left\{ \max\left\{\pi, q_1\right\} - \mathcal{P}_1(q_1) + V(q_1, \mathcal{P}_2) \right\} + \mathbb{1}_{q_1 = \emptyset} \left\{\pi\right\} \right\}.$$

The solution, denoted as $q_1^* \in Q^0$, must satisfy the following conditions:

$$\begin{aligned} & q_1^* + V(q_1^*, \mathcal{P}_2) - \mathcal{P}_1(q_1^*) \geq \pi, \text{ or, } \mathcal{P}_1(q_1^*) = 0, \\ & q_1^* + V(q_1^*, \mathcal{P}_2) - \mathcal{P}_1(q_1^*) \geq q_1' + V(q_1', \mathcal{P}_2) - \mathcal{P}_1(q_1'), \forall q_1' \in Q, \end{aligned}$$

which can be regarded as the participation constraint and the incentive compatibility constraint for the problem of the seller. The prospect of second-period surplus creates option value that enhances the attractiveness of first-period signal acquisition. The buyer may pay more upfront to secure access to future value. In summary, these insights yield the following result.

Proposition 1.14 (Buyer's First-Period Optimisation)**.** *Under any equilibrium, the buyer acquires an informative signal $q_1^* \in Q$ in period $t = 1$ if and only if all the following conditions hold. Otherwise, $q_1^* = \emptyset$.*

(i) $q_1^* - \mathcal{P}_1(q_1^*) + V(q_1^*, \mathcal{P}_2) \geq \pi$*, or,* $\mathcal{P}_1(q_1^*) = 0$*, participation constraint;*

(ii) $q_1^* - \mathcal{P}_1(q_1^*) + V(q_1^*, \mathcal{P}_2) \geq q_1' - \mathcal{P}_1(q_1') + V(q_1^*, \mathcal{P}_2)$*,* $\forall q_1' \in Q$*, incentive compatibility constraint;*

(iii) $\mu_1 \in [1 - q_1^*, q_1^*]$*, or,* $\mathcal{P}_1(q_1^*) = 0$*.*

This characterisation reveals several key insights about the strategic nature of dynamic information acquisition. The willingness of the buyer to pay for early information reflects not only immediate value but also the option value of future decision-making, creating intertemporal linkages where strategic complementarities connect behaviour across periods. Signals acquired in period one may provide limited immediate benefit yet create future opportunities that justify the purchase, demonstrating how forward-looking buyers incorporate anticipated future decisions into current optimisation. This strategic sophistication imposes a fundamental limit on the ability of the seller to engage in exploitative pricing, as the buyer will only pay for information that provides genuine value across the entire game horizon. Notably, when future surplus is available to the buyer, the seller can raise prices in early periods to capture this expected value, whereas aggressive future pricing suppresses early demand and restricts overall rent extraction over time. These insights collectively illustrate that dynamic information markets involve complex intertemporal considerations that shape the strategies of both sides of the market, revealing how participants respond strategically to incentives that unfold over multiple periods and giving rise to rich dynamics that differ fundamentally from static information provision.

1.4.5 Menu Setting by the Seller in the First Period

We now arrive at the initial stage of the dynamic game, where the seller selects the first-period pricing strategy with full knowledge of how the game will evolve. This analysis reveals one of the most striking results in the model: it is always optimal for the seller to offer high-type signal in the first period in order to generate profitable opportunities in the second period.

At the beginning of period $t = 1$, the seller must choose prices $\big(\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\big)$ for both signal types to maximise expected profit across both periods. Formally, the seller solves the following problem,

$$\max_{\left(\mathcal{P}_1(q^H),\mathcal{P}_1(q^L)\right)} \mathbb{E}\left[u_S\right] = \mathbb{E}\left[\mathcal{P}_1(q_1^*) + \mathcal{P}_2(q_1^*, q_2^*)\right],$$

where (q_1^*, q_2^*) represents the optimal signal acquisition path of the buyer given the pricing strategy of the seller. The seller anticipates how these decisions will influence immediate profit and future revenue opportunities. This gives rise to a sophisticated intertemporal optimisation problem in which current pricing must account for dynamic consequences. The strategy adopted by the seller must weigh three key considerations: the information rent available in the first period, the likelihood that buyers will return for second-period signals, and the magnitude of surplus that can be extracted from future transactions. The optimal balance amongst these elements depends critically on the strength of the prior belief of the buyer and the quality differential between the available signal types.

Extreme Prior Beliefs: Market Elimination

We begin by analysing the case in which $\frac{\pi}{1-\pi} \in \left[\left(\frac{q^H}{1-q^H}\right)^2, +\infty\right)$, which implies that $\pi > q^H > q^L$. The relationship between the prior belief, π, and the signal types, (q^H, q^L), plays a crucial role in the magnitude of surplus. Given the prior belief for the status quo is sufficiently strong, the buyer has no incentive to acquire any type of signals at any strictly positive price, given any period $t \in \{1, 2\}$ and any signal realisation of $s_1 \in \{0, 1\}$. Intuitively, the prior belief, π, is so extreme that the buyer sticks to the decision of status quo, which will not be reverted by any signal realisation throughout the entire game. Under the most extreme realisation in which the buyer acquires two high-type signals in a row with both signal realisations against the status quo (i.e., $s_1 = s_2 = 1$), the posterior belief is still greater than $\frac{1}{2}$, which implies that signals of both types are not pivotal. The buyer will hold a posterior belief more inclined to $\theta = 0$ with a probability of one, which suggests that the information rent of any signal type is zero. Consequently, given participation constraints hold, the seller should earn zero payoff. Though the buyer will accept offers if the price is zero, the seller has no incentive to provide since they receive zero payoff at the end. Trivially, any price menu is a solution. Thus, we have the following proposition.

Proposition 1.15. *Suppose that* $\frac{\pi}{1-\pi} \in \left[\left(\frac{q^H}{1-q^H}\right)^2, +\infty\right)$ *holds. The optimal menu setting strategy in period* $t = 1$ *is:* $\left(\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right) \in (\mathbb{R}_+)^2$.

Strong Prior Beliefs: High-Type Signal Superiority

Next we consider the case in which $\frac{\pi}{1-\pi} \in \left[\frac{q^L}{1-q^L}\frac{q^H}{1-q^H}, \left(\frac{q^H}{1-q^H}\right)^2\right]$, which again implies that $\pi > q^H > q^L$. The seller recognises that if a low-type signal is offered in period $t = 1$, the buyer will not acquire any further signal in period $t = 2$. The low-type signal is too imprecise to move the posterior belief sufficiently close to the centre of the belief space and therefore fails to generate any information rent in the second period. Formally, this means $\mu_1(q^L, 1) \notin [1-q, q]$, $\forall q \in \{q^H, q^L\}$, which implies that neither type of signal in period $t = 2$ is pivotal, by Proposition 1.4. Furthermore, since $\pi > q^L$, the signal in period $t = 1$ yields zero information rent. Consequently, the seller earns zero surplus in both periods, if they induce the buyer to acquire a low-type signal in period $t = 1$.

By contrast, with some probability, there exists strictly positive information rent in period $t = 2$, if they induce the buyer to acquire a high-type signal in period $t = 1$. More specifically, the buyer has incentive to acquire a high-type signal in period $t = 2$, given that $q_1 = q^H$ is acquired and $s_1 = 1$ is realised, which can be formally verified, $\mu_1(q^H, 1) \in [1-q^H, q^H]$. Intuitively, the high-type signal is sufficiently strong and thus the posterior belief will be pushed further away from the prior belief, which ends up somewhere sufficiently close to the centre. Therefore, the buyer is so uncertain about the true state that the high-type signal in the second period is pivotal, which incentivises them to acquire. The seller can fully extract the information rent in period $t = 2$ by setting the price $\mathcal{P}_2(q^H, q^H) = q^H - \mu_1(q^H, 1)$, as indicated by Proposition 1.14. Compared to the case $q_1 = q^L$ in which the seller earns zero surplus throughout the game, the seller prefers the choice $q_1 = q^H$. Consequently, in order to induce the buyer to acquire $q_1 = q^H$, the seller sets the following menu:

$$\mathcal{P}_1(q^H) = \max\left\{q^H - \pi, 0\right\} = 0;$$
$$\mathcal{P}_1(q^L) \in (0, +\infty).$$

It is interesting that the seller has an incentive to offer a free high-type signal rather than a low-type signal. Intuitively, the seller offers a signal with more precision to introduce a greater variance on the posterior belief. By doing so, there exists a strictly positive probability that the buyer ends up with a posterior belief sufficiently close to the centre, which induces the buyer to acquire a signal associated with strictly positive surplus in period $t = 2$. For notational convenience, a set of strategies are defined as follows.

Definition 1.3. Let $\tau \in \{H, A, C\} \equiv T$ denote a strategy in the pricing game. A strategy is called a *high-type only pricing strategy*, denoted $\tau = H$, if and only if all of the following conditions hold. Under such a strategy, the posted price functions $\widehat{\mathcal{P}}_t$ may not be optimal given the history, but the signal choices q_t^* must be optimal given the beliefs of the buyer and the observed menus.

(i) $\widehat{\mathcal{P}}_1(q^H) = \max\{q^H - \pi, 0\}$;

(ii) $q_1^*(\widehat{\mathcal{P}}_1) = q^H$;

(iii) $\widehat{\mathcal{P}}_2(q^H, q^H) = \varphi\left(q^H, \mu_1(q^H, 1)\right)$;

(iv) $q_2^*(\widehat{\mathcal{P}}_2) = q^H$;

The ex-ante expected payoff of the seller implementing the *high-type only pricing strategy* can be derived as follows. The seller extracts zero information rent in period $t = 1$, as the prior belief is greater than the signal precision (i.e., $\pi > q^H$). The seller fully captures the information rent $\mathcal{P}_2(q^H, q^H) = q^H - \mu_1(q^H, 1)$ with the probability $\mathbb{P}(s_1 = 1 \mid q_1 = q^H) = \pi(1 - q^H) + (1 - \pi)q^H$, and thus the expected payoff of the seller implementing high-type only pricing strategy is:

$$\begin{aligned} u_S^H \equiv \mathbb{E}\left[u_S(\tau = H)\right] &= \left[\pi(1 - q^H) + (1 - \pi)q^H\right] \varphi\left(q^H, \mu_1(q^H, 1)\right) + \max\left\{q^H - \pi, 0\right\} \\ &= \left[\pi(1 - q^H) + (1 - \pi)q^H\right]\left(q^H - \mu_1(q^H, 1)\right). \end{aligned}$$

Moderate Prior Beliefs: Two-Strategy Rivalry

In the case of $\frac{\pi}{1-\pi} \in \left[\frac{1-q^L}{q^L}\frac{q^H}{1-q^H}, \frac{q^L}{1-q^L}\frac{q^H}{1-q^H}\right]$, the seller has two profitable candidates for the pricing strategy. The first one is the high-type only pricing strategy, defined in the previous paragraph, in which the seller induces the buyer to acquire a high-type signal in both periods. The second one is the *aggressive pricing strategy*, in which the seller induces the buyer to acquire a low-type signal in period $t = 1$, followed by a high-type signal in period $t = 2$. Since the buyer acquires a signal in period $t = 2$ only if $s_1 = 1$, there exists no information asymmetry between the two parties. That is, the seller perfectly understands that $s_1 = 1$ must be realised if the buyer approaches them for a second signal. Effectively, the seller is implementing the aggressive pricing strategy. The ex-ante expected payoff of the seller implementing the aggressive pricing strategy is:

$$u_S^A \equiv \mathbb{E}\left[u_S(\tau = A)\right] = \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi\left(q^H, \mu_1(q^L, 1)\right) + \left[\max\left\{q^L - \pi, 0\right\}\right].$$

Definition 1.4. Let $\tau \in T$ denote a strategy in the pricing game. A strategy is called an *aggressive pricing strategy*, denoted $\tau = A$, if and only if all of the following conditions hold. Under such a strategy, the posted price functions $\widehat{\mathcal{P}}_t$ may not be optimal given the history, but the signal choices q_t^* must be optimal given the beliefs of the buyer and the observed menus.

(i) $\widehat{\mathcal{P}}_1(q^L) = \max\{q^L - \pi, 0\}$;

(ii) $q_1^*(\widehat{\mathcal{P}}_1) = q^L$;

(iii) $\widehat{\mathcal{P}}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L, 1)\right)$;

(iv) $q_2^*(\widehat{\mathcal{P}}_2) = q^H$.

With risk neutrality, the seller always prefers the high-type only pricing strategy regardless of the relationship of π, q^H, and q^L. The probability of a trade decreases in q_1 (i.e., $\pi(1-q^H) + (1-\pi)q^H < \pi(1-q^L) + (1-\pi)q^L$), whilst the ex-post surplus increases in q_1 (i.e., $q^H - \mu_1(q^H, 1) > q^H - \mu_1(q^L, 1)$), both in a linear way. The effect of the ex-post surplus dominates over that of the probability of a trade.

Lemma 1.3. *Let* $f(q) \equiv [\pi(1-q) + (1-\pi)q]\left[q^H - \mu_1(q, 1)\right]$*. Then, we have, f is increasing in q,* $\forall q \in \left(\frac{1}{2}, q^H\right)$.

Proof. See Appendix 1.A.4. □

As a result, given $u_S^H = f(q^H)$ and $u_S^A = f(q^L)$, according to Lemma 1.3, the high-type only pricing strategy dominates over the aggressive pricing strategy as $q^H > q^L$ and $f(\cdot)$ is increasing.

Weak Prior Beliefs: Three-Way Strategic Competition

In the case of $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$, there are three profitable candidates for pricing strategies: high-type only, aggressive, and conservative. Note that $\pi > q^H > q^L$ is not possible as this contradicts the inequality $\frac{\pi}{1-\pi} < \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}$. In consequence, there are only two cases: $q^H > \pi > q^L$ and $q^H > q^L > \pi$.

Definition 1.5. Let $\tau \in T$ denote a strategy in the pricing game. A strategy is called a *conservative pricing strategy*, denoted $\tau = C$, if and only if all of the following conditions hold. Under such a strategy, the posted price functions $\widehat{\mathcal{P}}_t$ may not be optimal given the history, but the signal choices q_t^* must be optimal given the beliefs of the buyer and the observed menus.

(i) $\widehat{\mathcal{P}}_1 = \max\{q^L - \pi, 0\} + V(q^L, \widehat{\mathcal{P}}_2)$;

(ii) $q_1^*(\widehat{\mathcal{P}}_1) = q^L$;

(iii) $\widehat{\mathcal{P}}_2(q^L, q^H) = \varphi\left(q^H, \mu_1(q^L, 0)\right)$;

(iv) $q_2^*(\widehat{\mathcal{P}}_2) = q^H$.

The ex-ante expected payoff of the seller implementing the conservative pricing strategy is:

$$
\begin{aligned}
u_S^C &\equiv \mathbb{E}\left[u_S(\tau = C)\right] \\
&= \Big[1\Big]\varphi\left(q^H, \mu_1(q^L,0)\right) \\
&\quad + \Big[\max\{q^L - \pi, 0\} + \Big[\pi(1-q^L) + (1-\pi)q^L\Big]\left(\varphi\left(q^H, \mu_1(q^L,1)\right) - \varphi\left(q^H, \mu_1(q^L,0)\right)\right)\Big] \\
&= \Big[\pi q^L + (1-\pi)(1-q^L)\Big]\left(q^H - \mu_1(q^L,0)\right) + \Big[\pi(1-q^L) + (1-\pi)q^L\Big]\left(q^H - \mu_1(q^L,1)\right).
\end{aligned}
$$

The expected payoff of conservative pricing strategy implementation is strictly greater than aggressive pricing strategy implementation, which implies that the conservative pricing strategy strictly dominates the aggressive pricing strategy. Specifically, the probability of a trade in period $t = 2$ is enhanced to one, since the seller gives up extracting the surplus from the buyer with a signal realisation φ of $s = 1$, in exchange of extra surplus from the buyer with a signal realisation of $s = 0$. However, the seller is able to extract the surplus they give up in period $t = 2$ by charging its expected value in the first period, which does not violate the participation constraint of the buyer. Intuitively, the buyer understands that they might enjoy a positive surplus due to information asymmetry between the two parties in the second period. Consequently, they are willing to be charged up to the expected value of the positive surplus in advance, which is:

$$
V(q^L) = \Big[\pi(1-q^L) + (1-\pi)q^L\Big]\left(\varphi\left(q^H, \mu_1(q^L,1)\right) - \varphi\left(q^H, \mu_1(q^L,0)\right)\right).
$$

As a result, the seller effectively extracts the entire ex-ante expected surplus, whilst the buyer obtains an ex-ante expected surplus of zero. This implies that the conservative pricing strategy strictly dominates the aggressive one, $u_S^C > u_S^A$. See Appendix 1.A.7 for a proof. However, a caveat arises regarding the conservative strategy. The inequality $u_S^C > u_S^A$ does not necessarily imply that the seller prefers $\tau = C$ to $\tau = A$ under the information set in which the seller chooses a menu in the second period. The seller at $t = 2$ overlooks the role of $V(q^L, \cdot)$, as the corresponding surplus has already been obtained in period $t = 1$. Lacking commitment, the seller at $t = 1$ cannot enforce their future self to adopt the conservative strategy in period $t = 2$. Nevertheless, this issue is not problematic, as it will be shown that $\tau = H$ is strictly preferred to $\tau = C$, and therefore $\tau = H$ always constitutes an equilibrium.

Next investigate the preference between the high-type only pricing strategy and the conservative pricing strategy. It can be formally shown that the seller always prefers the high-type only pricing strategy, which strictly dominates the conservative pricing strategy. See Appendix 1.A.6 for a proof. The advantage of selling the high-type signal in period $t = 1$ is the greater variance on signal realisation which generates more surplus to be extracted. On the other hand, the advantage of selling the low-type signal with implementation of the conservative pricing strategy is the greater probability of sale in the second period. However, the expected surplus is lower. Given the assumption that the seller exhibits risk-neutral behaviour, they overlook the risk and only take the magnitude of the expected surplus into account. Consequently, the high-type only strategy is strictly preferred. To sum up concisely, we have, $u_S^H > u_S^C > u_S^A$, given any value of $(\pi, q^H, q^L) \in \left(\frac{1}{2}, 1\right)^3$.

Proposition 1.16. *Suppose that* $\frac{\pi}{1-\pi} \in \left(1, \left(\frac{q^H}{1-q^H}\right)^2\right]$ *holds.*

(i) The optimal menu setting strategy in period $t = 1$ *is:* $\left(\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right) \in \{0\} \times \mathbb{R}_+$.

(ii) The implied expected payoff is: $u_S^H = -\pi(1 - q^H)^2 + (1 - \pi){q^H}^2 > 0$.

The dominance of the high-type only pricing strategy reflects a fundamental trade-off in dynamic information markets. Although high-precision signals reduce the probability of second-period trade, they generate substantially larger information rents when trade occurs. The magnitude effect systematically outweighs the probability effect, making quality concentration optimal. Moreover, the high-type only strategy achieves first-best outcomes through natural buyer revelation, eliminating inefficiencies from information asymmetry whilst maximising expected seller profits.

1.5 Welfare Implication

This section examines the welfare properties of the dynamic information selling mechanism. The analysis addresses four fundamental questions: whether the equilibrium achieves first-best outcomes despite information asymmetry, how the seller should optimally design the information structure, whether expanding signal menus enhances profits, and what role commitment plays in dynamic pricing. These welfare implications reveal surprising insights about the efficiency of information markets and the strategic value of quality concentration over menu diversification.

1.5.1 First-Best Benchmark and the Action-State Paradox

A fundamental principle in mechanism design theory holds that achieving first-best outcomes typically requires the number of available instruments to exceed the number of possible states, providing sufficient degrees of freedom for optimal contracting. In standard information design problems, such a principle manifests as needing multiple signal types to screen different buyer types and extract full surplus. Our model presents a striking departure from such a principle. Despite having only one meaningful action in equilibrium, offering $q^H = 1$, across two possible first-period states $s_1 \in \{0, 1\}$, the seller achieves first-best outcomes. Such an apparent violation of the standard action-state relationship occurs because the dynamic structure fundamentally transforms the nature of the contracting problem.

The key insight is that buyer behaviour functions as an additional instrument that provides the seller with crucial information. When the buyer chooses the high-type signal, $q_1 = q^H$, in the first period, their subsequent return behaviour perfectly reveals the unobserved signal realisation s_1. Specifically, buyers who return for a second signal implicitly reveal that their first signal contradicted their prior belief, whilst non-returning buyers indicate that their signal was confirming. Such a behavioural revelation mechanism operates as follows. Consider a buyer who acquires q^H in the first period. If the signal realisation confirms their prior belief, their posterior belief becomes sufficiently extreme that any second signal provides no positive information value. Conversely, if the signal contradicts the prior belief, sufficient uncertainty remains to justify acquiring additional information. The seller, observing only the buyer return decision, can perfectly infer which scenario occurred.

Such a revelation mechanism creates natural market segmentation which eliminates the need for complex screening menus. The seller can implement a simple strategy: extract full information rent from returning buyers, those who received contradictory signals, whilst correctly anticipating that non-returning buyers have no willingness to pay for additional information. Unlike static mechanisms in which the seller must use multiple instruments to separate buyer types ex-ante, the dynamic structure allows separation through revealed behaviour. The economic intuition is that time and buyer actions serve as substitutes for menu complexity. Rather than offering multiple signal types to screen different buyer types simultaneously, the seller uses the sequential structure to let buyers self-select through their return behaviour. Such an approach transforms a complex multidimensional screening problem into a sequence of simpler participation constraint problems.

In static information selling models, achieving first-best typically requires satisfying multiple incentive compatibility constraints across different buyer types, necessitating complex mechanisms with many instruments. Here, the dynamic structure reduces the problem to simple participation constraints which can be satisfied with a single high-quality signal. The participation constraint for returning buyers becomes straightforward. They will acquire a second signal if and only if the price does not exceed their information rent. Since the seller can identify returning buyers perfectly and knows their information rent precisely, having inferred their signal realisation, setting the price equal to such rent extracts full surplus whilst maintaining participation. The result explains why the standard action-state relationship breaks down in our setting. The seller effectively has access to more instruments than immediately apparent: the first-period signal choice, the second-period signal choice, and the revealed information from the buyer return decision. Such three instruments are sufficient to implement first-best allocations across the two-state space.

The result demonstrates that dynamic structures can overcome the apparent limitations imposed by action-state ratios in mechanism design. When the sequential actions of agents reveal private information, the effective dimensionality of the mechanism expands beyond the number of explicit instruments. The key condition is that the information revelation must be sufficiently rich to allow the principal to tailor mechanisms to each agent type. In our context, the binary nature of the return decision provides exactly the right amount of information to achieve efficiency. The seller learns whether the buyer received confirming or contradictory evidence, which is precisely the information needed to set optimal prices for the second period. Such alignment between the information revealed and the information required for optimal mechanism design drives the efficiency result.

1.5.2 Global Optimum of Information Mechanism Design

This section examines the optimal information mechanism design for the information seller. Given the prior belief, π, how should the seller optimally choose the signal types, (q^H, q^L), to maximise their expected payoff? According to Proposition 1.16, given any $(\pi, q^H, q^L) \in \left(\frac{1}{2}, 1\right)^3$, the seller induces the buyers to acquire q^H in both periods as such strategy maximises their expected payoff.

Consequently, only the precision of the high-type, q^H, matters. It can be proven that the optimality suggests $q^{H^*} = 1$. Thus, we have the following proposition.

Proposition 1.17. *Given any $\pi \in \left(\frac{1}{2}, 1\right)$, the global optimal signal choice is $q^{H^*} = 1$, with the implementation of high-type only pricing strategy.*

Proof. See Appendix 1.A.5. □

Recall that the expected payoff of the seller, given $\tau = H$ is implemented and $q^H > \pi$, is given by:

$$u_S^H = \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi\left(q^H, \mu_1(q^H, 1)\right) + \left(q^H - \pi\right).$$

The probability of a trade is, $\pi(1-q^H) + (1-\pi)q^H$, equivalent to the probability that $s_1 = 1$ is realised. The probability is decreasing in q^H, which implies that a higher precision is associated with a lower chance to trade. However, the ex-post information rent possesses a greater increasing momentum when the precision is higher. More specifically, the posterior belief regarding the status quo decreases whilst the signal precision is higher, which widens the difference between the two. In effect, the magnitude of information rent grows even faster. This can also be verified from the first-order derivative: $\frac{\partial u_S}{\partial q^H} = 2 - (2\pi - 1)(2q^H - 1)$, which is always positive. The increasing momentum becomes higher given q^H further away for the prior belief.

This result shows that although there is a trade-off between the probability of trade and the rent obtained per transaction, the effect of information rent dominates. As a result, the seller prefers to offer the most informative signals available. The optimal information design therefore prioritises quality over quantity, with the seller focusing on highly precise signals rather than maintaining a broad menu of options.

1.5.3 More than Two Signal Types

The baseline model assumes two available signal types. A natural extension considers whether expanding the menu to include additional signal types affects the equilibrium outcome and the expected payoff of the seller. The analysis shows that such expansion brings no benefit, as the optimal outcome remains unchanged regardless of the number of signals offered. To formalise this result, consider an economy with a finite set of signal types denoted by $Q = \{q^1, q^2, \ldots, q^N\}$, where $q^1 > q^2 > \ldots > q^N$. The underlying intuition stems from the nature of information rent extraction in the two-period setting. When multiple signal types are available, the optimal strategy of the seller is still to induce the buyer to acquire only the highest precision signal in both periods, effectively disregarding all other options.

This outcome follows from the relationship between information rent and signal precision. Suppose the seller selects signals q^i and q^j on the equilibrium path, where $q^i > q^j$. Consistent with the Propositions 1.15 and 1.16, the optimal pricing strategy induces $q_1 = q_2 = q^i$ in both periods. The seller has no incentive to offer signals of lower precision, as these generate strictly lower information rent. Since information rent increases with precision, the seller always prefers to extract surplus by encouraging the buyer to acquire the most precise signal available. As long as q^N satisfies $\frac{\pi}{1-\pi} \in \left[1, \left(\frac{q^1}{1-q^1}\right)^2\right]$, follow the proof strategy for Proposition 1.17 and it can be inferred that the seller chooses the highest type, q^1, at the optimum. If the inequality does not hold, by Proposition 1.15, any signal induced and price set is an equilibrium, as there exists no signal associated with strictly positive information rent.

Proposition 1.18. *Given any $\pi \in \left(\frac{1}{2}, 1\right)$ and $(q^1, q^2, ..., q^N) \in \left(\frac{1}{2}, 1\right)^N$, $q^{H*} = \max\left\{q^1, q^2, ..., q^N\right\}$ constitutes an optimal solution, with the implementation of high-type only pricing strategy.*

This result shows that the optimal strategy of the seller reflects a natural form of simplicity. Instead of managing complex menus with multiple signal types, the seller achieves the best outcome by offering only the highest quality information available. This finding has important implications for the study of information markets, as it suggests that competition amongst providers may naturally favour information quality over variety.

1.5.4 Commitment and the Inapplicability of the Coase Conjecture

The dynamic pricing literature extensively examines whether commitment limitations create inefficiencies that reduce seller expected profits. The Coase conjecture suggests that monopolists selling durable goods face time-inconsistency problems because forward-looking buyers anticipate future price cuts, undermining current profits. In information markets, one might expect similar commitment issues to arise when sellers interact with buyers across multiple periods. Our analysis, which follows Perfect Bayesian Equilibrium without commitment, reveals that such concerns are unfounded in the present setting. Under our equilibrium concept, the seller cannot commit to future pricing strategies at the outset. Instead, in each period, the seller re-optimises to choose the menu which maximises their expected payoff given the observed history and their beliefs about buyer behaviour. The absence of commitment means that buyers must form rational expectations about how the seller will behave in future periods, taking into account the incentive of the seller to re-optimise.

The classical Coase problem emerges when a monopolist competes with their own future pricing decisions. Buyers of durable goods anticipate that the seller will reduce prices in subsequent periods to capture additional demand, leading them to delay purchases or demand lower current prices. The anticipation of future price reductions erodes the ability of the seller to extract surplus, creating a time-inconsistency problem that commitment devices could potentially resolve. Three structural features of our model eliminate the conditions which give rise to Coase-style problems. First, information signals are inherently non-durable and period-specific. The information of each period addresses distinct uncertainty, preventing the seller from competing with their own past information provision. A buyer who purchases a signal in period $t = 1$ faces genuinely new uncertainty in period $t = 2$, unlike buyers of durable goods who purchase once for persistent consumption. Second, the seller faces no customer base erosion across periods. In durable goods markets, each sale permanently removes a customer from the potential buyer pool, creating pressure to reduce prices over time. Here, buyers who purchase information in period 1 may retain demand in period $t = 2$ due to residual uncertainty. The seller does not exhaust their market through early sales, eliminating the dynamic which drives price-cutting incentives in traditional Coase problems. Third, the key information asymmetry in our model, namely the inability of the seller to observe signal realisations, affects only the second period. Such timing eliminates the forward-looking strategic behaviour which typically generates time-inconsistency concerns. Buyers make first-period decisions without needing to anticipate how the seller will respond to unobserved information, since the seller learns nothing new between periods about buyer types that would alter their re-optimisation incentives.

The equilibrium mechanism naturally aligns with dynamic incentives even without commitment. When the seller re-optimises in the second period, the optimal strategy remains offering only high-type signals, creating a simple and predictable pricing structure. The seller has no incentive to deviate from such a strategy when re-optimising because doing so would strictly reduce expected profits regardless of first-period outcomes. Moreover, the information revelation mechanism which drives our efficiency result operates effectively without commitment. The ability of the seller to condition second-period prices on first-period choices, whilst re-optimising based on observed buyer behaviour, creates the precise information structure needed for first-best outcomes. Commitment to fixed prices across all contingencies would eliminate the beneficial re-optimisation which allows the seller to tailor second-period menus to the information revealed through buyer actions. The broader implication is that commitment concerns in information markets depend critically on the persistence of information across periods and the structure of information asymmetries. When information needs are renewed each period and the seller's re-optimisation incentives align with efficiency, the standard time-inconsistency problems which motivate commitment devices do not arise. The result clarifies when commitment mechanisms are likely to be valuable in dynamic information provision settings.

1.5.5 The Sustainability Challenge of Full Extraction in Extended Horizons

The remarkable full surplus extraction result in our two-period framework naturally raises the question: does this property extend to longer horizons? This section examines a three-period model in which the seller continues offering only high-type signals q^H. The analysis reveals that whilst full extraction remains theoretically achievable through a specific pricing strategy, its sustainability without commitment becomes problematic, a challenge absent in the two-period case. Consider three periods of potential signal acquisition with only q^H available. A crucial observation is that the posterior belief in any period t is determined by the difference between the number of confirming signals (i.e., $s = 0$) and that of contradictory signals (i.e., $s = 1$). When this difference equals zero, the posterior returns to the prior π. This counting structure implies that in any period, at most one posterior belief value generates pivotal demand, specifically one of the two posteriors closest to the decision threshold $\frac{1}{2}$, one approaching from below and the other one from above.

Final Period Analysis At the beginning of period $t = 3$, the buyer holds posterior belief determined by their signal history. Let δ denote the difference between the number of confirming and contradictory signals:

- $\delta = +2$: Both signals confirmed $(0, 0)$, posterior pushed toward 1;
- $\delta = 0$: One of each $(0, 1)$ or $(1, 0)$, posterior equals π;
- $\delta = -2$: Both contradicted $(1, 1)$, posterior pushed toward 0.

To illustrate the extraction mechanism, consider one possible configuration where $\mu(\delta = -1) < \frac{1}{2} < \mu(\delta = 0)$. In this example, only buyers with $\delta = 0$ find the third-period signal pivotal. With $\delta = 0$, the signal moves them to $\delta \in \{+1, -1\}$, crossing the threshold and altering their decision. These buyers have positive willingness-to-pay equal to their information rent. For buyers with $\delta \in \{+2, -2\}$, an additional signal moves them to $\delta \in \{+3, +1\}$ or $\delta \in \{-1, -3\}$. In this configuration, neither transition crosses the decision threshold. The third-period signal cannot change their optimal decision, yielding zero value. The seller sets $\mathcal{P}_3$ equal to the information rent at the unique pivotal posterior, extracting full surplus from the only type with demand.

The Sustainability Problem in Period 2 Working backwards to period $t = 2$, the buyer knows they will capture zero surplus in period $t = 3$ due to perfect extraction. Given any first-period realisation s_1, the buyer will find the second-period signal pivotal only with specific posterior $\mu_1(s_1)$. Following our running example, this target is $\mu(\delta = -1)$. The naive approach would set $\mathcal{P}_2(s_1) = q^H - \max\{\mu_1(s_1 = 1), 1 - \mu_1(s_1 = 1)\}$, extracting full rent from a pivotal buyer. However, this strategy proves suboptimal. Whilst $s_1 = 1$ buyers purchase as predicted, $s_1 = 0$ buyers, finding the signal non-pivotal, would exit permanently. This premature exit destroys potential future surplus: these buyers might receive $s_2 = 1$, creating valuable third-period demand. To maximise total surplus creation, the seller must ensure all buyer types continue acquiring signals, maintaining the possibility of reaching pivotal states. This requires offering second-period signals for free (i.e., $\mathcal{P}_2 = 0$), sacrificing immediate extraction from pivotal buyers to preserve future opportunities. Given the dynamic structure, the seller is able to recover this foregone revenue through first-period pricing, extracting anticipated future surplus in advance, as in the conservative pricing in the main model. The strategy achieves full extraction if sustainable under Perfect Bayesian Equilibrium. However, sustainability requires the seller not to deviate in period $t = 2$. The question is whether the seller maintains the free signal policy $\mathcal{P}_2 = 0$, or succumbs to the temptation of charging the positive rent $\mathcal{P}_2 = q^H - \max\{\mu_1(s_1 = 1), 1 - \mu_1(s_1 = 1)\}$. The answer depends on parameters. When third-period surplus is small, for instance if $\mu(\delta = 0)$ is sufficiently close to q^H, the immediate gain from deviation may exceed the lost future revenue. Without commitment, the seller faces time inconsistency: the ex-ante optimal strategy with free intermediate signals becomes ex-post suboptimal, which contrasts sharply with the two-period case, where no such intermediate temptation arises.

Generalisation to n Periods The mentioned mechanism can be extended to dynamics with finitely many periods. For full surplus extraction, the following pricing strategy is necessary:

1. **Final period:** Set the price equal to the information rent of the pivotal buyer.

2. **First period:** Set the price equal to expected total surplus from the seller's perspective.
3. **Intermediate periods:** Offer free signals to maintain universal participation.

The strategy maximises surplus creation by ensuring all buyers remain active, maximising the probability of reaching pivotal events. However, sustainability becomes increasingly fragile as horizons extend. Each intermediate period presents a temptation to deviation, and the conditions ensuring time consistency become progressively stringent.

The Role of Signal Diversity By restricting attention to a single signal precision, we ensure at most one pivotal posterior per period. Introducing multiple signal types would create additional pivotal cases, further complicating the surplus extraction problem. With both q^H and q^L available, different posterior beliefs might find different signals optimal, creating heterogeneity the seller cannot screen without single-crossing. As shown in Proposition 1.7, the failure of single-crossing prevents profitable separation even with multiple signals. Moreover, whilst lower-type signals might increase the probability of positive revenue by creating more pivotal events, they also reduce the magnitude of extractable rents. My conjecture is that expected total surplus decreases with signal diversity: lower precision dampens information value creation more than increased pivotal frequency compensates.

The analysis reveals a fundamental tension in multi-period information markets. Whilst full extraction remains theoretically achievable through carefully structured pricing, its implementation requires either commitment mechanisms or parameter configurations that naturally deter deviation. The two-period model's elegance stems partly from avoiding this tension: with no intermediate periods, the commitment problem disappears. Does introducing signals with lower precision ever increase total surplus by sustaining participation across longer horizons? Under the two-period model, the answer is definitively no. With extended horizons, the question becomes more nuanced: might signal diversity help overcome the commitment problem by smoothing incentives across periods? This remains an open question requiring further investigation. The three-period extension illuminates both the potential and limitations of dynamic surplus extraction. Full extraction remains achievable through strategic pricing that maintains universal participation, but sustainability without commitment becomes problematic, a challenge absent in our two-period benchmark. This finding qualifies our main result: whilst quality concentration with high-precision signals remains optimal when feasible, extended horizons introduce implementation challenges that may require commitment devices or alternative mechanisms to resolve. The robustness of full extraction thus depends not only on the information structure but also on the institutional environment supporting dynamic contracting.

1.6 Conclusion

This paper analyses the Perfect Bayesian Equilibrium of a dynamic game between an information seller and a buyer, offering key insights into the structure and efficiency of information markets. The analysis shows that, despite the complexity of dynamic information pricing with multiple signal types and informational asymmetries, the equilibrium outcome displays simplicity and efficiency.

The central result establishes that only the highest precision signal is relevant in equilibrium, with all other signal types having no effect on the seller optimisation problem. This challenges the view that product diversity improves seller outcomes. Instead, the optimal strategy involves offering only the most informative signal, implementing a high-type only approach that excludes lower-quality options. This suggests that information markets tend to prioritise quality over variety, with sellers focusing on the provision of valuable information rather than maintaining complex menus.

Notably, the analysis shows that information asymmetry does not hinder the seller in extracting information rent. Although the seller cannot observe signal realisations, the equilibrium outcome achieves the first-best benchmark. This result follows from the structure of the optimal pricing mechanism, which aligns with efficient information provision. By focusing on high-precision signals, the strategy renders the unobservability of some realisations irrelevant.

The welfare implications extend beyond efficiency. The preference for precise signals introduces a natural incentive for information quality, since more accurate signals enable greater information rent extraction. This alignment between individual incentives and efficiency indicates that market forces may support the emergence of high-quality information provision without requiring regulatory intervention.

The analysis also clarifies the role of commitment. In contrast to many dynamic pricing settings where commitment issues arise due to time inconsistency, this model shows that commitment devices yield no advantage. The absence of self-competition across periods and the non-durable nature of signals ensures that the usual logic behind the Coase Conjecture does not apply. This result informs where commitment concerns are likely to be relevant in information markets.

Several extensions could deepen the understanding of dynamic information markets. First, introducing buyer heterogeneity in prior beliefs or decision problems could reveal how sellers tailor signal design for different users. Second, incorporating competition amongst multiple information sellers may offer insights into market structure and strategic interactions. Third, considering persistent information or signals with intertemporal correlation could uncover dynamics absent from the baseline model. Finally, empirical work testing the model predictions in applied settings would offer valuable evidence and guide further theoretical development.

Appendix 1.A: Omitted Proofs

1.A.1 Proofs of Section 4.1

Proof of Proposition 1.1. The buyer solves the following optimisation problem to maximise their expected payoff (before the realisation of the true state),

$$\begin{aligned} y^* &= \arg\max_{y\in\{0,1\}} \left\{ \mathbb{E}[u_B^y(\theta, y)] - \sum_{t=1}^{2} \mathcal{P}_t(\cdot) \right\} \\ &= \arg\max_{y\in\{0,1\}} \left\{ \mathbb{E}[u_B^y(\theta, y)] \right\} \\ &= \arg\max_{y\in\{0,1\}} \left\{ \mu_2 \cdot \mathbb{1}_{y=0} + (1-\mu_2) \cdot \mathbb{1}_{y=1} \right\}. \end{aligned}$$

Clearly, $y^* = 0$ if $\mu_2 \geq 1 - \mu_2$, and, $y^* = 1$ if $\mu_2 \leq 1 - \mu_2$, which implies that, $y^* = 0$ if $\mu_2 \geq \frac{1}{2}$, and, $y^* = 1$ if $\mu_2 \leq \frac{1}{2}$. The implied expected payoff is, $\max\{\mu_2, 1-\mu_2\}$. Note that the posterior belief, μ_2, depends on both the signal realisation (s_1, s_2), and the signal types, (q_1, q_2). However, the explicit functional form of μ_2 is not relevant to the maximisation problem of the buyer at this stage, as the posterior belief is independent of the choice variable, y. □

1.A.2 Proofs of Section 4.2

Proof of Proposition 1.2. The buyer evaluates the probability of signal realisation $s_2 \in \{0, 1\}$ as $\mathbb{P}(s_2 \mid q_2)$. Given the signal realisation s_2, the implied expected payoff before the true state realisation is:

$$\begin{aligned} \mathbb{E}[u_B \mid q_2 \in \{q^H, q^L\}] &= \mathbb{P}(s_2 = 0|q_2) \max\{\mu_2(q_1, s_1, q_2, 0), 1 - \mu_2(q_1, s_1, q_2, 0)\} \\ &\quad + \mathbb{P}(s_2 = 1|q_2) \max\{\mu_2(q_1, s_1, q_2, 1), 1 - \mu_2(q_1, s_1, q_2, 1)\} \\ &\quad - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) \\ &= [\mu_1 q_2 + (1-\mu_1)(1-q_2)] \max\left\{ \frac{\mu_1 q_2}{\mu_1 q_2 + (1-\mu_1)(1-q_2)}, \frac{(1-\mu_1)(1-q_2)}{\mu_1 q_2 + (1-\mu_1)(1-q_2)} \right\} \\ &\quad + [\mu_1(1-q_2) + (1-\mu_1)q_2] \max\left\{ \frac{\mu_1(1-q_2)}{\mu_1(1-q_2) + (1-\mu_1)q_2}, \frac{(1-\mu_1)q_2}{\mu_1(1-q_2) + (1-\mu_1)q_2} \right\} \\ &\quad - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) \\ &= \max\{\mu_1 q_2, 1 - \mu_1 - q_2 + \mu_1 q_2\} + \max\{\mu_1 - \mu_1 q_2, q_2 - \mu_1 q_2\} - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) \\ &= \max\{\mu_1, 1 - q_2\} + \max\{\mu_1, q_2\} - \mu_1 - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot). \end{aligned}$$

□

Proof of Lemma 1.1. Given the signal realisation $s_2 = 0$, the posterior belief before the decision is:

$$\mu_2(\mu_1, q_2, s_2 = 0) = \frac{\mu_1 q_2}{\mu_1 q_2 + (1-\mu_1)(1-q_2)} = \frac{1}{1 + \frac{(1-\mu_1)(1-q_2)}{\mu_1 q_2}} > \frac{1}{2},$$

since $1 - \mu_1 < q_2$ and $\mu_1 > 1 - q_2$.

Given the signal realisation $s_2 = 1$, the posterior belief before the decision is:

$$\mu_2(\mu_1, q_2, s_2 = 1) = \frac{\mu_1(1-q_2)}{\mu_1(1-q_2) + (1-\mu_1)q_2} = \frac{1}{1 + \frac{(1-\mu_1)q_2}{\mu_1(1-q_2)}} < \frac{1}{2},$$

since $1 - \mu_1 > 1 - q_2$ and $\mu_1 < q_2$. By Definition 1.2, $\mu_2(\mu_1, q_2, s_2 = 0) > \frac{1}{2} > \mu_2(\mu_1, q_2, s_2 = 1)$ implies that q_2 is pivotal. □

Proof of Proposition 1.3. When $\mu_1 \in [1 - q_2, q_2]$, the signal is pivotal, and direct calculation shows that,

$$\mathbb{E}[\max\{\mu_2, 1-\mu_2\} \mid \mu_1, q_2] = \max\{\mu_1, 1-q_2\} + \max\{\mu_1, q_2\} - \mu_1 = q_2.$$

Therefore, $\phi(q_2, \mu_1) = q_2 - \max\{\mu_1, 1-\mu_1\}$.

When $\mu_1 > q_2$, the signal is non-pivotal, and direct calculation shows that,

$$\mathbb{E}[\max\{\mu_2, 1-\mu_2\} \mid \mu_1, q_2] = \max\{\mu_1, 1-q_2\} + \max\{\mu_1, q_2\} - \mu_1 = \mu_1.$$

Therefore, $\phi(q_2, \mu_1) = \mu_1 - \max\{\mu_1, 1-\mu_1\} = 0$.

When $\mu_1 < 1 - q_2$, the signal is non-pivotal, and direct calculation shows that,

$$\mathbb{E}[\max\{\mu_2, 1-\mu_2\} \mid \mu_1, q_2] = \max\{\mu_1, 1-q_2\} + \max\{\mu_1, q_2\} - \mu_1 = 1 - \mu_1.$$

Therefore, $\phi(q_2, \mu_1) = (1-\mu_1) - \max\{\mu_1, 1-\mu_1\} = 0$. □

Proof of Proposition 1.4. Suppose first that $\mu_1 > q_2$. If the buyer acquires signal $q_2 \in Q$, the expected payoff of the buyer is:

$$\mathbb{E}[u_B \mid q_2 \in Q] = \mu_1 - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot).$$

If instead the buyer chooses not to acquire a second signal (i.e., $q_2 = \emptyset$), the expected payoff of the buyer is:

$$\begin{aligned}\mathbb{E}[u_B \mid q_2 = \emptyset] = \max\{\mu_1, 1-\mu_1\} - \mathcal{P}_1(\cdot) &= \mu_1 - \mathcal{P}_1(\cdot) \\ &> \mu_1 - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) = \mathbb{E}[u_B \mid q_2 \in Q],\end{aligned}$$

which implies that $q_2 = \emptyset$ is optimal.

Now suppose that $\mu_1 < 1 - q_2$. We have the following inequality:

$$\begin{aligned}\mathbb{E}[u_B \mid q_2 = \emptyset] = \max\{\mu_1, 1-\mu_1\} - \mathcal{P}_1(\cdot) &= (1-\mu_1) - \mathcal{P}_1(\cdot) \\ &> (1-\mu_1) - \mathcal{P}_2(\cdot, q_2) - \mathcal{P}_1(\cdot) = \mathbb{E}[u_B \mid q_2 \in Q],\end{aligned}$$

which implies that $q_2 = \emptyset$ is optimal. □

Proof of Proposition 1.5. Given that $q_2 = q^L$ is pivotal, by Lemma 1.1, we have $\mu_1 \in [1-q^L, q^L]$. Since $q^H > q^L$, the containment relation $[1-q^L, q^L] \subset [1-q^H, q^H]$ holds, which implies $\mu_1 \in [1-q^H, q^H]$. Therefore, by Lemma 1.1, q^H is also pivotal. □

Proof of Proposition 1.6. We prove this by showing that the optimal choice of the buyer must satisfy all three conditions, and that these conditions are sufficient for optimality.

Necessity: Suppose $q_2^* \in Q$ is the optimal choice of the buyer. We show that each condition must hold.

Condition (i) — Participation Constraint: The buyer chooses q_2^* over the outside option $q_2 = \emptyset$. By the earlier analysis, the expected payoffs are:

$$\begin{aligned}\mathbb{E}[u_B \mid q_2^*] &= q_2^* - \mathcal{P}_2(\cdot, q_2^*) - \mathcal{P}_1(\cdot) \quad \text{(if pivotal)}, \\ \mathbb{E}[u_B \mid q_2 = \emptyset] &= \max\{\mu_1, 1-\mu_1\} - \mathcal{P}_1(\cdot).\end{aligned}$$

Optimality requires, $\mathbb{E}[u_B \mid q_2^*] \geq \mathbb{E}[u_B \mid q_2 = \emptyset]$, which implies,

$$q_2^* - \mathcal{P}_2(\cdot, q_2^*) - \mathcal{P}_1(\cdot) \geq \max\{\mu_1, 1-\mu_1\} - \mathcal{P}_1(\cdot).$$

Simplifying yields condition (i). If $\mathcal{P}_2(\cdot, q_2^*) = 0$, then the buyer is indifferent, so the condition holds trivially.

Condition (ii) — Incentive Compatibility: The buyer chooses q_2^* over any alternative signal $q_2' \in Q$. Optimality requires:

$$\mathbb{E}[u_B \mid q_2^*] \geq \mathbb{E}[u_B \mid q_2'] \quad \text{for all } q_2' \in Q.$$

If both signals are pivotal, this reduces to:

$$q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq q_2' - \mathcal{P}_2(\cdot, q_2').$$

If q_2' is not pivotal, then the expected payoff of the buyer from q_2' is $\max\{\mu_1, 1 - \mu_1\} - \mathcal{P}_2(\cdot, q_2') - \mathcal{P}_1(\cdot)$, which is strictly less than the outside option if $\mathcal{P}_2(\cdot, q_2') > 0$.

Condition (iii) — Pivotal Constraint: By Proposition 1.4, if $\mathcal{P}_2(\cdot, q_2^*) > 0$ and $\mu_1 \notin [1 - q_2^*, q_2^*]$, then $q_2^* = \emptyset$ is optimal. This contradicts our assumption that $q_2^* \in Q$. Therefore, either $\mu_1 \in [1 - q_2^*, q_2^*]$ or $\mathcal{P}_2(\cdot, q_2^*) = 0$.

Sufficiency: Now suppose all three conditions hold for some $q_2^* \in Q$. We show that q_2^* is optimal.

Dominance over the outside option: By condition (i), q_2^* yields at least as high an expected payoff as $q_2 = \emptyset$.

Dominance over alternatives: By condition (ii), q_2^* yields at least as high an expected payoff as any $q_2' \in Q$.

Well-defined payoffs: By condition (iii), either q_2^* is pivotal (so the payoff calculation is valid) or it is offered for free (so the buyer is indifferent and any choice is optimal).

Therefore, q_2^* maximises the expected payoff of the buyer amongst all available options.

Necessity of all conditions: Finally, we show that if any condition fails, then $q_2^* = \emptyset$ is optimal. If (i) fails, the outside option dominates q_2^*. If (ii) fails, some alternative q_2' dominates q_2^*. If (iii) fails, by Proposition 1.4, q_2^* cannot be optimal when priced strictly above zero.

This completes the proof. □

1.A.3 Proofs of Section 4.3

Proof of Proposition 1.7. Suppose that, at the optimum, the seller plays the following pooling strategy. The buyer is induced to acquire $q_2 = q^L$ given any signal realisation s_1. The prices set by the seller, $(\mathcal{P}_2(\cdot, q^H), \mathcal{P}_2(\cdot, q^L))$, must follow the participation constraints by Proposition 1.6,

$$\mathcal{P}_2(\cdot, q^L) \leq q^L - \max\{\mu_1, 1-\mu_1\}.$$

It can be shown that the price, $\mathcal{P}_2(\cdot, q^L) + \varepsilon$, where $\varepsilon \to 0^+$, satisfies the participation constraint for $q_2 = q^H$,

$$\begin{aligned} & \mathcal{P}_2(\cdot, q^L) \leq q^L - \max\{\mu_1, 1-\mu_1\} \\ \iff & \mathcal{P}_2(\cdot, q^L) + \varepsilon \leq q^L + \varepsilon - \max\{\mu_1, 1-\mu_1\} \leq q^H - \max\{\mu_1, 1-\mu_1\}. \end{aligned}$$

The seller strictly improves their payoff, which contradicts the conjectured optimality.

Next investigate if $q_2 = q^L$ can be chosen under a separating strategy. Suppose that, at the optimum, the seller plays the following separating strategy. The buyer is induced to acquire $q^0 \in Q$ given the signal realisation $s_1 = 0$, whilst the buyer is induced to acquire $q^1 \in Q \setminus \{q^0\}$ if $s_1 = 1$ is realised. For notational convenience, the following are defined: $\mu_1^0 \equiv \mu_1(q_1, 0)$ and $\mu_1^1 \equiv \max\{\mu_1(q_1, 1), 1 - \mu_1(q_1, 1)\}$. The prices set by the seller, $(\mathcal{P}_2(\cdot, q^0), \mathcal{P}_2(\cdot, q^1))$, must follow the incentive compatibility constraints by Proposition 1.6:

$$\begin{aligned} q^0 - \mu_1^0 - \mathcal{P}_2(\cdot, q^0) \geq q^1 - \mu_1^0 - \mathcal{P}_2(\cdot, q^1) &\iff q^0 - \mathcal{P}_2(\cdot, q^0) \geq q^1 - \mathcal{P}_2(\cdot, q^1); \\ q^1 - \mu_1^1 - \mathcal{P}_2(\cdot, q^1) \geq q^0 - \mu_1^1 - \mathcal{P}_2(\cdot, q^0) &\iff q^1 - \mathcal{P}_2(\cdot, q^1) \geq q^0 - \mathcal{P}_2(\cdot, q^0). \end{aligned}$$

Combine the two inequalities and we have $q^0 - q^1 = \mathcal{P}_2(\cdot, q^0) - \mathcal{P}_2(\cdot, q^1)$, which implies $q^0 - \mathcal{P}_2(\cdot, q^0) = q^1 - \mathcal{P}_2(\cdot, q^1)$.

Moreover, the prices must satisfy the participation constraints of the buyer: $q^0 - \mu_1^0 - \mathcal{P}_2(\cdot, q^0) \geq 0$ and $q^1 - \mu_1^1 - \mathcal{P}_2(\cdot, q^1) \geq 0$, which implies $q^1 - \mu_1^0 - \mathcal{P}_2(\cdot, q^1) \geq 0$ and $q^0 - \mu_1^1 - \mathcal{P}_2(\cdot, q^0) \geq 0$.

Given that incentive compatibility constraints and participation constraints hold, the buyer with either signal realisation is indifferent between the two signals. The seller can obtain a higher expected payoff if they always induce the buyer to acquire a high-type signal at the price $\mathcal{P}_2(\cdot, q^H)$.

The seller is better off as they increase their expected payoff by $[\pi(1-q^L)+(1-\pi)q^L](q^H-q^L)$.

□

Proof of Proposition 1.8. We prove each claim separately.

No incentive when $s_1 = 0$. When $s_1 = 0$, the posterior belief of the buyer is:

$$\mu_1(q^H, 0) = \frac{\pi q^H}{\pi q^H + (1-\pi)(1-q^H)}.$$

It can be shown that $\mu_1(q^H, 0) > q^H$. Since $\mu_1(q^H, 0) > q^H > q^L > \frac{1}{2}$, we have $\mu_1(q^H, 0) > \max\{q^H, q^L\}$, which implies $\mu_1(q^H, 0) \notin [1-q^H, q^H]$, and $\mu_1(q^H, 0) \notin [1-q^L, q^L]$, so both signals are not pivotal. By Proposition 1.4, the buyer will not acquire either signal type at any strictly positive price.

Incentive when $s_1 = 1$. When $s_1 = 1$, the posterior belief of the buyer is:

$$\mu_1(q^H, 1) = \frac{\pi(1-q^H)}{\pi(1-q^H) + (1-\pi)q^H}.$$

$\mu_1(q^H, 1) \leq q^H$ if and only if:

$$\frac{\pi(1-q^H)}{\pi(1-q^H) + (1-\pi)q^H} \leq q^H,$$

which is equivalent to:

$$\frac{\pi}{1-\pi} \leq \left(\frac{q^H}{1-q^H}\right)^2.$$

This is exactly our assumption. For the lower bound, since $\pi > \frac{1}{2}$ and the signal realisation contradicts the prior, we have $\mu_1(q^H, 1) > 1 - q^H$. Therefore, $\mu_1(q^H, 1) \in [1-q^H, q^H]$, making the high-type signal pivotal. By our earlier analysis, the buyer will acquire the signal if the price is at most $q^H - \max\{\mu_1(q^H, 1), 1-\mu_1(q^H, 1)\}$. □

Proof of Lemma 1.2. The condition $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \cdot \frac{q^H}{1-q^H}\right]$ implies:

$$\mu_1(q^L, 0) = \frac{\pi q^L}{\pi q^L + (1-\pi)(1-q^L)} > \frac{\pi(1-q^L)}{\pi(1-q^L) + (1-\pi)q^L} = \mu_1(q^L, 1).$$

It also implies that:

$$\mu_1(q^L, 0) > 1 - \mu_1(q^L, 1),$$

so that:

$$\mu_1(q^L, 0) > \max\{\mu_1(q^L, 1), 1 - \mu_1(q^L, 1)\},$$

which confirms the claim. □

1.A.4 Proofs of Section 4.5

Proof of Lemma 1.3. By definition:

$$\begin{aligned} f(q) &= [\pi(1-q) + (1-\pi)q]\left[q^H - \mu_1(q,1)\right] \\ &= [\pi(1-q) + (1-\pi)q]\left[q^H - \frac{\pi(1-q)}{\pi(1-q) + (1-\pi)q}\right] \\ &= [\pi(1-q) + (1-\pi)q]\, q^H - \pi(1-q) \\ &= \pi(q^H - 1) + \left[(1-\pi)q^H + (1-q^H)\pi\right] q, \end{aligned}$$

where $(1-\pi)q^H + (1-q^H)\pi > 0$ implies that $f(q)$ is increasing in q. □

1.A.5 Proofs of Section 5.2

Proof of Proposition 1.17. Proof by contradiction. Suppose at the global optimum, $q^H = k \neq 1$. First note that the strategy must yield a strictly positive expected payoff other than zero. If not, given any $\pi \in \left(\frac{1}{2}, 1\right)$, there always exists a $\widehat{q^H}$ which satisfies $\frac{\pi}{1-\pi} \in \left[1, \left(\frac{\widehat{q^H}}{1-\widehat{q^H}}\right)^2\right]$. By Proposition 1.16, $(q^H, q^L) = (\widehat{q^H}, \cdot)$ yields a positive expected payoff, which violates the optimality. Consider $(q^H, q^L) = (k+\varepsilon, k)$, where ε is infinitesimal and positive. By Proposition 1.16, $q_1 = q_2 = k + \varepsilon$ is optimal, which yields a strictly greater expected payoff than inducing $q_1 = q_2 = k$, contradiction. □

1.A.6 Expected Payoff Comparison: $\tau = H$ vs $\tau = C$

This section provides with a formal proof showing the high-type only pricing strategy always dominates over the conservative pricing strategy.

$$
\begin{aligned}
u_S^H &= \Big[[\pi(1-q^H)+(1-\pi)q^H][q^H-\max\{\mu_1(q^H,1),1-\mu_1(q^H,1)\}]\Big]+\Big[\max\{q^H-\pi,0\}\Big]\\
&= \Big[\pi(1-q^H)+(1-\pi)q^H\Big]\Big(q^H-\mu_1(q^H,1)\Big).\\
u_S^C &= \Big[[1][q^H-\mu_1(q^L,0)]\Big]\\
&+\Big[\max\{q^L-\pi,0\}+\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(\mu_1(q^L,0)-\max\{\mu_1(q^L,1),1-\mu_1(q^L,1)\}\Big)\Big]\\
&= \Big[[1][q^H-\mu_1(q^L,0)]\Big]+\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(\mu_1(q^L,0)-\mu_1(q^L,1)\Big).\\
u_S^H-u_S^C &= \Big[\pi(1-q^H)+(1-\pi)q^H-1\Big]q^H-\pi(1-q^H)+\frac{\pi q^L}{\pi q^L+(1-\pi)(1-q^L)}\\
&-\left(\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\frac{\pi q^L}{\pi q^L+(1-\pi)(1-q^L)}-\pi(1-q^L)\right)\\
&= \Big[\pi(1-q^H)+(1-\pi)q^H-1\Big]q^H+\pi(q^H-q^L)+\pi q^L\\
&= \Big[\pi(1-q^H)+(1-\pi)q^H-1+\pi\Big]q^H\\
&= \Big[(2\pi-1)(1-q^H)\Big]q^H>0
\end{aligned}
$$

Thus, it has been proven that, given any $(\pi,q^H,q^L)\in\left(\frac{1}{2},1\right)^2$, $H\succ C$ holds.

1.A.7 Expected Payoff Comparison: $\tau=C$ vs $\tau=A$

This section provides with a formal proof showing the conservative pricing strategy always dominates over the aggressive pricing strategy.

$$
\begin{aligned}
u_S^C &= \Big[[1][q^H-\mu_1(q^L,0)]\Big]\\
&+\Big[\max\{q^L-\pi,0\}+\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(\mu_1(q^L,0)-\max\{\mu_1(q^L,1),1-\mu_1(q^L,1)\}\Big)\Big]\\
&= \Big[[1][q^H-\mu_1(q^L,0)]\Big]+\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(\mu_1(q^L,0)-\mu_1(q^L,1)\Big)\\
&= \Big[[\pi q^L+(1-\pi)(1-q^L)][q^H-\mu_1(q^L,0)]\Big]+\Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(q^H-\mu_1(q^L,1)\Big).\\
u_S^A &= \Big[[\pi(1-q^L)+(1-\pi)q^L][q^H-\max\{\mu_1(q^L,1),1-\mu_1(q^L,1)\}]\Big]\\
&+\Big[\max\{q^L-\pi,0\}\Big]\\
&= \Big[\pi(1-q^L)+(1-\pi)q^L\Big]\Big(q^H-\mu_1(q^L,1)\Big)\\
u_S^C-u_S^A &= \Big[[\pi q^L+(1-\pi)(1-q^L)\Big]\Big(q^H-\mu_1(q^L,0)\Big)>0.
\end{aligned}
$$

Thus, it has been proven that, given any $(\pi,q^H,q^L)\in\left(\frac{1}{2},1\right)^2$, $C\succ A$ holds.

Chapter 2

Dynamic Sale of Information with Risk Aversion

2.1 Introduction

Information markets pervade modern economies, from financial research services and consulting firms to digital platforms selling personalised data and analytics. Yet a fundamental assumption underlying most theoretical analyses that information sellers are risk neutral sits uneasily with empirical evidence that firms exhibit significant risk aversion in their operational decisions. The paper investigates how the seller risk aversion transforms the strategic landscape of dynamic information provision, revealing that the drive for higher quality signals observed under risk neutrality, need not persist when sellers must balance expected returns against revenue volatility.

The importance of this question extends beyond theoretical interest. Real world information providers, including investment research firms and business intelligence companies, face substantial revenue uncertainty when selling information whose value depends on unknown future realisations. A financial research firm, for instance, earns revenue only when its market predictions prove sufficiently uncertain to generate client demand, creating precisely the type of state dependent payoffs that risk-averse agents seek to avoid.

This paper extends the dynamic information selling framework by introducing risk aversion into the preference structure of the seller. Under risk neutrality, the equilibrium pricing strategy relies exclusively on the signal with the highest precision, as it delivers the highest expected information rent. Whilst such a strategy introduces uncertainty in the realised payoff for the seller, the underlying trade off is clear: higher precision leads to greater ex post information rent conditional on a signal realisation unfavourable to the status quo, though the probability of observing such a signal realisation is decreasing in signal precision. Nevertheless, under risk neutrality, the marginal gain in expected information rent from increased precision dominates the marginal loss from its declining likelihood. Consequently, a risk neutral seller strictly prefers signals with higher precision.

Moreover, under risk neutrality, expanding the signal space, either by introducing additional signal types or by allowing for more flexible menu design, does not enhance the expected surplus of the seller. In other words, increased flexibility in pricing does not translate into higher expected payoffs. Furthermore, if the seller is endowed with the ability to design the information structure, they will optimally choose to provide perfectly informative signals in equilibrium. This extreme outcome reflects the alignment between precision and rent extraction when the seller does not bear any cost from uncertainty.

The introduction of risk aversion fundamentally disrupts these simple relationships. Risk-averse sellers do not merely seek to maximise expected revenue; they must also consider the distributional properties of their payoffs. This creates a tension between profit maximisation and revenue smoothing that can reverse established dominance relationships. The central insight emerges from recognising that aggressive pricing strategies, whilst potentially more profitable on average, concentrate revenue in specific scenarios in which buyers exhibit high uncertainty. This represents precisely the type of volatility that risk-averse agents seek to avoid.

The analysis reveals three key departures from the risk neutral benchmark. First, risk-averse sellers do not necessarily prefer signals with higher precision, as they may optimally choose to induce buyers to acquire less precise signals that increase the likelihood of securing information rent through reduced outcome volatility. Second, conservative pricing strategies that ensure universal participation need not be dominated by aggressive approaches that maximise expected returns, since diminishing marginal utility can make sellers prefer strategies that accept lower expected surplus in exchange for higher certainty of positive outcomes. Third, information asymmetry can paradoxically benefit risk-averse sellers in certain parameter regions, as the inability to observe signal realisations enables commitment to revenue smoothing strategies that would be unsustainable under full information.

These findings contribute to several strands of literature. Most directly, the paper extends the growing literature on risk aversion in information markets, which has largely focused on static settings with risk-averse buyers. The dynamic framework developed here reveals how risk aversion affects multi period information provision strategies, showing that many insights from static analysis do not carry over to dynamic settings. The work also contributes to the mechanism design literature by characterising optimal dynamic mechanisms under risk aversion, demonstrating that standard approaches fail when agents exhibit risk aversion.

The paper proceeds as follows. Section 2.2 reviews the relevant literature on risk aversion and information selling. Section 2.3 presents the model and establishes the baseline risk neutral equilibrium. Section 2.4 analyses how risk aversion affects optimal pricing strategies, characterising the conditions under which conservative strategies dominate aggressive approaches. Section 2.5 examines welfare implications and the role of commitment mechanisms. Section 2.6 concludes with a discussion of policy implications and directions for future research.

2.2 Literature Review

The intersection of risk aversion and information selling represents a rapidly developing area that challenges fundamental assumptions in information economics. Whilst the traditional information selling literature assumes risk-neutral agents, introducing risk aversion fundamentally alters market dynamics, requiring new theoretical frameworks and mechanism designs that depart substantially from established benchmarks.

The theoretical literature establishes that risk aversion fundamentally transforms how agents value information. Abbas et al. (2013) demonstrate that risk aversion can either increase or decrease information value depending on decision context, with risk aversion generally decreasing the value of perfect information through complex non-monotonic relationships. Bakır (2015) extend this to show that information selling prices are monotonic in the degree of risk aversion, whilst Gould (1974) provides early foundational work on risk preferences and information demand. Cabrales et al. (2017) develop a unified framework for comparing information values across different risk preferences, and Cabrales et al. (2013) demonstrate that more risk-averse investors value information less because they choose conservative investment strategies, creating feedback loops affecting information production incentives.

The mechanism design literature reveals that standard approaches fail when agents exhibit risk aversion. Maskin and Riley (1984) show that optimal auctions for risk-averse buyers require partial insurance through modified virtual valuation functions, fundamentally departing from risk-neutral optimal auction of Myerson (1981). Esö and Futo (1999) analyse the reverse problem where sellers are risk-averse, demonstrating that risk-neutral buyers can effectively insure risk-averse sellers. Gershkov et al. (2022) find that under constant risk aversion preferences, optimal mechanisms provide full insurance, making agent utility independent of others' reports. Bhalgat et al. (2012) develop algorithms showing that risk-averse sellers prefer mechanisms providing certainty even at the cost of lower expected revenue, whilst Sundararajan and Yan (2020) identify robust mechanisms for risk-averse sellers that are approximately optimal regardless of risk aversion levels.

The limited literature on dynamic information selling with risk aversion reveals significant gaps that this paper addresses. Dai (2021) represents one of the few dynamic analyses, showing that optimal sequential contracts shift from separating to pooling mechanisms as supplier risk aversion increases, fundamentally altering information revelation timing. Benkert (2025) introduce loss aversion into bilateral trade settings, demonstrating that platforms optimally provide partial insurance in ownership but full insurance in monetary dimensions. However, these studies focus on different information asymmetries than the one examined in this paper.

Recent work incorporates behavioural factors beyond standard risk aversion. Andries and Haddad (2020) examine information aversion as preference-based fear of news flows, showing that information-averse agents may prefer less information even when it improves decision-making. Vasserman and Watt (2021) provides a comprehensive survey of risk aversion in auctions, emphasising how risk aversion breaks revenue equivalence and creates heightened significance for auction design decisions. Campo et al. (2011) develop semi-parametric methods for estimating risk aversion in auctions, bridging theoretical mechanism design and empirical implementation.

Despite theoretical advances, the literature exhibits several critical limitations that this paper addresses. First, most existing work focuses on static mechanisms, with limited analysis of dynamic information selling where sellers face uncertainty about signal realisations. When Bergemann et al. (2014) examine optimal menus of Blackwell experiments and Zhong (2022) analyse dynamic information acquisition from the perspective of the buyer, neither addresses the fundamental information asymmetry arising when sellers cannot observe signal realisations in multi-period settings. Second, the risk aversion literature has not adequately examined how unobservable signal realisations interact with risk preferences in dynamic settings. Hörner and Skrzypacz (2016) analyse information sellers with private preferences but assume sellers observe signals, whilst Doval and Skreta (2022) develop tools for dynamic mechanism design with limited commitment but do not address risk aversion or unobservable realisations.

This paper makes three novel contributions to fill these gaps. First, it extends the dynamic information selling literature by analysing how risk aversion affects optimal pricing strategies when sellers cannot observe signal realisations, creating a new form of information asymmetry not previously studied. Second, it demonstrates how risk aversion fundamentally alters the sustainability of commitment mechanisms in dynamic settings, showing that risk-averse sellers may prefer pooling strategies that provide revenue insurance even when separating mechanisms would yield higher expected returns. Third, it provides the first formal analysis of how unobservable signal realisations interact with risk preferences to determine optimal dynamic pricing strategies, revealing that risk aversion can make commitment mechanisms unsustainable when full information eliminates the insurance benefits of pooling. These contributions bridge the gap between the static risk aversion literature and dynamic information selling models, providing new insights into how behavioural factors affect information market design in multi-period settings with incomplete information about signal outcomes.

2.3 Model

This model builds upon the framework developed in the prior chapter. It also considers a dynamic setting involving two agents: a decision-maker (referred to as the buyer) and an information provider (the seller). The buyer is assumed risk neutral as in the previous chapter. However, the seller is now assumed risk-averse. The interaction spans two discrete periods. The objective of the seller is to design a pricing scheme for different types of signals to maximise their expected utility. This task is subject to two key constraints arising from the behaviour of the buyer, which evolves with their posterior beliefs over time. First, a participation constraint ensures that the buyer finds it worthwhile to purchase a signal. Second, an incentive compatibility constraint guarantees that the buyer prefers the seller's intended signal over other available options.

The model assumes a binary state of the world, denoted by $\theta \in \{0,1\} \equiv \Theta$, which determines the ex-post payoff of the buyer. Initially, the buyer believes that the state is $\theta = 0$ (i.e., the *status quo*) with probability $\pi \in \left(\frac{1}{2}, 1\right)$. At the end of period $t = 2$, the buyer chooses an action $y \in \{0,1\}$ to match the true state. The ex-post outcome payoff of the buyer is given by:

$$u_B^y(\theta, y) = \mathbb{1}_{y=\theta}.$$

That is, the buyer receives an outcome payoff of one if the action matches the true state, and zero otherwise. At the beginning of each period $t \in \{1,2\}$, the buyer may acquire an informative binary signal $q_t \in \{q^H, q^L\} \equiv Q$. Each signal provides noisy information about the true state and is independently drawn across periods. A signal realisation is denoted by $s_t \in \{0,1\} \equiv S$. The two signal types differ in precision, with $1 > q^H > q^L > \frac{1}{2}$. Given signal type q_t, the probability of observing a correct signal is:

$$\mathbb{P}(s_t = \theta \mid \theta) = q_t, \quad \mathbb{P}(s_t \neq \theta \mid \theta) = 1 - q_t.$$

The buyer may also choose not to acquire a signal in a given period, denoted $q_t = \emptyset$. Let $Q^0 \equiv Q \cup \{\emptyset\}$ denote the full set of choices, and $S^0 \equiv S \cup \{\emptyset\}$ the set of possible signal realisations. The seller posts a price menu in each period. In the first period, the pricing function is $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$. In the second period, it is $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$. It is assumed that, $\mathcal{P}_1(\emptyset) = \mathcal{P}_2(q_1, \emptyset) = 0 \quad \text{for all } q_1 \in Q^0$.

The game proceeds as follows.

1. In period 1, the seller posts the price menu $\mathcal{P}_1$.
2. The buyer chooses $q_1 \in Q^0$, pays $\mathcal{P}_1(q_1)$, and observes signal $s_1 \in S^0$.
3. In period 2, the seller observes q_1, and posts a menu $\mathcal{P}_2(q_1, \cdot)$.
4. The buyer chooses $q_2 \in Q^0$, pays $\mathcal{P}_2(q_1, q_2)$, and observes signal $s_2 \in S^0$.
5. The buyer selects action $y \in \{0, 1\}$, aiming to match θ.

At the beginning of period $t = 1$, the seller posts a price menu, $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$, with $\mathcal{P}_1(\emptyset) = 0$. Upon observing this menu, the buyer selects a signal $q_1 \in Q^0$. If an informative signal $q_1 \in Q$ is chosen, the buyer pays $\mathcal{P}_1(q_1)$ and observes a signal realisation $s_1 \in \{0, 1\}$. The buyer then updates their belief using Bayes' rule:

$$\mu_1(q_1, s_1) \equiv \mathbb{P}(\theta = 0 \mid q_1, s_1).$$

If no signal is acquired, i.e. $q_1 = \emptyset$, the belief remains at the prior π, and the realisation is defined as $s_1 = \emptyset$. The seller observes the choice of the the buyer, q_1, but crucially not the realisation s_1, and posts a second-period price menu, $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$, with $\mathcal{P}_2(q_1, \emptyset) = 0$. The buyer then chooses a second signal $q_2 \in Q^0$. If an informative signal is chosen, they pay $\mathcal{P}_2(q_1, q_2)$ and observe a second realisation $s_2 \in \{0, 1\}$. The buyer subsequently forms a final posterior belief:

$$\mu_2((q_1, q_2), (s_1, s_2)) \equiv \mathbb{P}(\theta = 0 \mid q_1, q_2, s_1, s_2).$$

Finally, the buyer chooses an action $y \in \{0, 1\}$, and payoffs are realised.

The buyer is assumed risk neutral. As a result, their ex-post payoff to the buyer, denoted u_B, equals the outcome payoff from the final decision minus the total cost of information:

$$u_B(\theta, y, q_1, q_2) = u_B^y(\theta, y) - \mathcal{P}_1(q_1) - \mathcal{P}_2(q_1, q_2).$$

The key difference in terms of assumption is here. Now the seller is inherited with a time-separable and risk-averse preference. The signal generating cost of any type is assumed zero. Thus, the ex-post payoff to the seller, denoted u_S, is assumed:

$$u_S(q_1, q_2) = u\left(\mathcal{P}_1(q_1)\right) + u\left(\mathcal{P}_2(q_1, q_2)\right).$$

Further we assume that the seller exhibits CRRA preferences, $u(x) = x^{1-\sigma}$ with $\sigma \in [0, 1)$. Note that $\sigma = 0$ implies risk neutrality. Zero time discount are assumed for both the seller and the buyer at any point of time throughout the game. It is also assumed that both players rely on Bayes' rule when measuring the probability of any event. Both players are risk neutral and apply no discounting. Beliefs are updated according to Bayes' rule throughout the game.

This model highlights a fundamental tension in dynamic information markets that becomes particularly pronounced when the seller exhibits risk aversion. The introduction of CRRA preferences with $\sigma \in [0, 1)$ creates competing incentives between profit maximisation and revenue smoothing across market scenarios. Under risk neutrality, the seller would focus exclusively on extracting maximum surplus through high-precision signals that enable perfect market segmentation. However, risk aversion introduces preferences for more stable revenue streams, making conservative pricing strategies increasingly attractive as uncertainty aversion intensifies. The seller must balance the higher expected profits from aggressive pricing against the volatility inherent in strategies that concentrate revenue in specific buyer types or market conditions.

The strategic implications extend beyond simple risk-return trade-offs to fundamentally alter market structure and information provision patterns. Risk-averse sellers may deliberately offer lower-precision initial signals to maintain broader buyer engagement and ensure more predictable second-period demand, even though this reduces total extractable surplus. This preference for revenue smoothing across both time periods and buyer types can reverse the dominance relationships established under risk neutrality, potentially favouring conservative pricing strategies that guarantee participation over aggressive approaches that maximise expected profits. The resulting equilibrium reflects a sophisticated balance between information quality, market participation, and risk management that captures essential features of real-world information markets where providers must consider both profitability and operational stability in their strategic decisions.

In the following section, strategies are formally defined and equilibrium outcomes are characterised.

2.4 Equilibrium Characterisation

In this paper, we adopt the equilibrium concept of Perfect Bayesian Equilibrium (PBE), focusing on equilibria in pure strategies. This solution concept proves particularly well-suited to dynamic information markets, where players must form beliefs about unobserved actions and update these beliefs as new information becomes available. A strategy profile satisfies PBE if it meets three fundamental conditions that ensure both individual rationality and collective consistency.

As in the previous chapter, the following mappings define the strategy set of the buyer and seller. The strategy of the buyer consists of three components: a first-period signal choice function, $\mathcal{Q}_1 : \mathbb{R}_+^2 \to Q^0$, which maps the observed menu prices to a signal type; a second-period signal choice function, $\mathcal{Q}_2 : Q^0 \times S^0 \times \mathbb{R}_+^2 \to Q^0$, which maps the observed menu prices in the second period, the first-period signal choice q_1, and its realisation $s_1 \in S^0$, to a signal type; a final action rule, $\mathcal{Y} : (Q^0)^2 \times (S^0)^2 \to \{0, 1\}$, which determines the decision based on all acquired information.

The strategy of the seller consists of two pricing functions that must anticipate buyer behaviour whilst accounting for information asymmetries. The first-period pricing function $\mathcal{P}_1 : Q^0 \to \mathbb{R}_+$ maps each potential signal type to a non-negative price, establishing the initial terms of trade that influence all subsequent interactions. The second-period pricing function $\mathcal{P}_2 : Q^0 \times Q^0 \to \mathbb{R}_+$ maps the observed first-period signal choice of the buyer and each potential second-period signal type to a non-negative price. This function is particularly sophisticated, as it must extract information rent whilst accounting for the inability of the seller to observe signal realisations directly.

An important insight emerges from this strategic structure. The final action rule of the buyer remains identical to that characterised in the previous section, as the decision criterion depends only on posterior beliefs and remains independent of the path through which those beliefs were formed. This observation allows the equilibrium analysis to focus on information acquisition decisions, where the primary strategic tensions arise. The following subsections characterise equilibrium strategies using backward induction, beginning with the final decision and working backward through the dynamic pricing and information acquisition choices that define the market.

2.4.1 Final Decision of the Buyer

The final decision problem of the buyer remains identical to that analysed in the previous chapter, as the decision criterion depends only on posterior beliefs and is independent of the risk attitude of the seller. Since the buyer maintains risk neutrality and faces the same symmetric payoff structure, the optimal decision rule is unchanged regardless of whether the seller exhibits risk aversion or risk neutrality. This invariance allows the equilibrium analysis to focus on the strategic tensions that arise from information acquisition and pricing decisions, where the risk attitude of the seller has their primary impact.

As established in the previous chapter, the buyer chooses action $y = 0$ if their posterior belief, μ_2, is greater than $\frac{1}{2}$, and $y = 1$ otherwise, where $\mu_2 \equiv \mathbb{P}(\theta = 0 \mid (q_1, q_2), (s_1, s_2))$ represents the posterior belief that $\theta = 0$ after observing all signal realisations. The expected outcome payoff is:$\mathbb{E}[u_B^y] = \max\{\mu_2, 1 - \mu_2\}$. The decision rule applies throughout the subsequent analysis without further derivation. Formally, we have the following decision rule of the buyer:

$$y^* \equiv \arg \max_{y \in \{0,1\}} \left\{ \mu_2 \cdot \mathbb{1}_{y=0} + (1 - \mu_2) \cdot \mathbb{1}_{y=1} \right\},$$

which implies the following optimal solution:

$$y^* = 0 \text{ if } \mu_2 \in \left[\frac{1}{2}, 1\right]; \, y^* = 1 \text{ if } \mu_2 \in \left[0, \frac{1}{2}\right].$$

Proposition 2.1. *Given the posterior belief $\mu_2 \in (0, 1)$, the expected outcome payoff of the buyer before the decision is $\mathbb{E}[u_B^y] = \max\{\mu_2, 1 - \mu_2\}$. The expected payoff of the buyer before the decision is:*

$$\mathbb{E}[u_B] = \max\{\mu_2, 1 - \mu_2\} - \mathcal{P}_1(q_1) - \mathcal{P}_2(q_1, q_2).$$

2.4.2 Signal Acquisition by the Buyer in the Second Period

The information acquisition strategy of the buyer in the second period remains fundamentally unchanged from the analysis in the previous chapter. Since the buyer maintains risk neutrality and faces identical decision criteria, the introduction of seller risk aversion does not alter the underlying economics of information demand. The optimisation problem of the buyer, the conditions under which information has value, and the resulting equilibrium behaviour all follow the same logical framework established earlier. Rather than re-deriving these results, we present the key insights and demonstrate their continued relevance in the context of risk-averse sellers. We therefore draw upon the following definitions and results from Section 1.4.2 of Chapter 1.

The Economics of Information Demand. Information acquisition in the second period depends critically on whether signals can influence the final decision of the buyer. A signal is valuable only when it has the potential to change the optimal action, creating what we term a *pivotal* signal. When the posterior belief of the buyer is sufficiently extreme, no additional information can alter their decision, rendering further signals economically worthless regardless of their precision or price.

Definition 2.1. An informative signal $q_2 \in Q$ is pivotal with respect to the posterior belief μ_1 if and only if $\mu_2(\mu_1, q_2, s_2 = 0) > \frac{1}{2} > \mu_2(\mu_1, q_2, s_2 = 1)$. The signal induces different optimal decisions depending on its realisation: $y(s_2 = 0) \neq y(s_2 = 1)$.

The condition for pivotality creates a natural "uncertainty zone" within which information has value. When the posterior belief falls within this zone, the buyer remains sufficiently uncertain about the true state that additional information can meaningfully improve decision-making quality.

Lemma 2.1. *A signal q_2 is pivotal if and only if $\mu_1 \in [1 - q_2, q_2]$.*

Beyond this uncertainty zone, information becomes economically worthless. If the confidence of the buyer in either state becomes too strong, even highly precise signals cannot generate sufficient doubt to justify costly information acquisition. Such boundary conditions impose natural constraints on information demand that persist regardless of seller risk preferences.

Information Rent and Value Creation. The value of information manifests through improved decision-making quality, which we formalise through the concept of information rent. Information rent measures the expected improvement in payoff that results from acquiring a signal, net of the current decision-making capability of the buyer.

Definition 2.2 (Information Rent)**.** Given posterior belief $\mu_1 \in (0,1)$ at the beginning of period $t = 2$, the information rent associated with signal $q_2 \in Q$ is defined as:

$$\varphi(q_2, \mu_1) = \mathbb{E}[\max\{\mu_2, 1-\mu_2\}|\mu_1, q_2] - \max\{\mu_1, 1-\mu_1\},$$

where μ_2 represents the posterior belief after observing signal realisation s_2.

Information rent captures the incremental value created by resolving uncertainty. When signals are pivotal, they enable the buyer to make better decisions on average by providing discriminating information across different states of the world. When signals are not pivotal, they generate zero information rent because they cannot improve decision quality.

Proposition 2.2 (Information Rent Characterisation)**.** *For any informative signal $q_2 \in Q$ and posterior belief $\mu_1 \in (0,1)$:*

(i) If $\mu_1 \in [1-q_2, q_2]$, then $\varphi(q_2, \mu_1) = q_2 - \max\{\mu_1, 1-\mu_1\}$;

(ii) If $\mu_1 \notin [1-q_2, q_2]$, then $\varphi(q_2, \mu_1) = 0$.

Signal Selection and Market Dynamics. The structure of information rent creates clear preferences over signal types when multiple options are available. Higher-precision signals generate greater information rent when both are pivotal, establishing a natural hierarchy in information demand.

Proposition 2.3. *If a low-type signal q^L is pivotal for posterior belief μ_1, then the high-type signal q^H is also pivotal.*

When multiple signals are pivotal, the difference in information rent equals $q^H - q^L > 0$, making high-type signals unambiguously more valuable to buyers and more profitable for sellers. Rational buyers will prefer higher-precision signals when both are available at comparable prices, and rational sellers will focus on providing high-type signals when both generate positive demand.

Proposition 2.4. *Under any equilibrium, the seller has no incentive to offer low-type signals in period $t = 2$.*

The dominance of high-type signals emerges naturally from the structure of information rent rather than from any assumptions about preferences or market power. Even when sellers could potentially segment the market by offering different signal types to different buyers, the economic logic favours concentration on high-precision alternatives.

Equilibrium Constraints and Buyer Behaviour. The optimisation problem of the buyer generates natural constraints that any equilibrium pricing strategy must satisfy. These constraints ensure that information markets operate efficiently by preventing sellers from extracting rent from worthless signals whilst preserving incentives for quality provision.

Proposition 2.5. *Under any equilibrium, positive prices for non-pivotal signals are incompatible with rational buyer behaviour:* $P_2(q_2) > 0$ *and* $\mu_1 \notin [1 - q_2, q_2]$ *implies* $q_2^* = \emptyset$.

When signals are pivotal, the participation and incentive compatibility constraints of the buyer determine the feasible pricing strategies available to sellers.

Proposition 2.6. *Under any equilibrium, the buyer acquires an informative signal* $q_2^* \in Q$ *in period* $t = 2$ *if and only if all the following conditions hold. Otherwise,* $q_2^* = \emptyset$.

(i) $q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq \max\{\mu_1, 1 - \mu_1\}$, *or,* $\mathcal{P}_2(\cdot, q_2^*) = 0$, *participation constraint;*

(ii) $q_2^* - \mathcal{P}_2(\cdot, q_2^*) \geq q_2' - \mathcal{P}_2(\cdot, q_2')$, $\forall q_2' \in Q$, *incentive compatibility constraint;*

(iii) $\mu_1 \in [1 - q_2^*, q_2^*]$, *or,* $\mathcal{P}_2(\cdot, q_2^*) = 0$.

These constraints ensure that buyers participate voluntarily in information markets and select their most preferred signal type from available alternatives. Sellers must respect these constraints when designing pricing strategies, regardless of their own risk preferences. The fundamental economics of information demand thus remain invariant to seller risk aversion. Risk-averse sellers face the same demand conditions and buyer behaviour as their risk-neutral counterparts, but they may respond differently to the revenue uncertainty inherent in information provision. The analysis now turns to examine how the seller risk aversion influences pricing strategies whilst respecting these unchanging demand fundamentals.

2.4.3 Menu Setting by the Seller in the Second Period

The architecture of pricing problem of the seller in the second period parallels the framework established in Chapter 1: buyers have acquired first-period signals, sellers observe these choices but not the underlying realisations, and information asymmetry constrains the ability to extract surplus. Yet the introduction of risk aversion fundamentally transforms how the seller navigate this landscape. Whilst a risk-neutral seller could focus exclusively on maximising expected revenue, a risk-averse seller will now balance profit maximisation against the volatility inherent in different pricing approaches. This creates a richer strategic environment where the same information asymmetries that merely complicated pricing decisions in the previous chapter now become central to risk management. The seller challenge evolves from a pure optimisation problem into a sophisticated portfolio choice between strategies that offer different risk-return profiles.

The seller enters the second period observing which signal type the buyer purchased, q_1, but not what it revealed, s_1. Such information asymmetry creates the central pricing challenge, as the unobserved realisation determines the buyer posterior belief, μ_1, and thus the extractable information rent. The seller designs a pricing menu, $\mathcal{P}_2(q_1, q_2)$, which maximises expected utility whilst anticipating the range of possible buyer beliefs.

The analysis examines two cases that create distinct strategic environments for the risk-averse seller. When the buyer initially acquires high-type signals, natural market segmentation simplifies pricing decisions. When the buyer begins with low-type signals, persistent uncertainty forces the risk-averse seller to choose between aggressive strategies which maximise expected revenue and conservative approaches which smooth income across market scenarios. In both cases, participation constraints cap prices at the available information rent, but risk aversion fundamentally alters how the seller evaluates these trade-offs. The equilibrium characterisation proceeds by examining each case in turn.

Case 1: High-Type Signal Previously Acquired $(q_1 = q^H)$

When the buyer acquires a high-type signal in the first period, the strategic landscape remains remarkably similar to the risk-neutral case analysed in the previous chapter. The high precision of the initial signal creates a natural bifurcation in buyer behaviour which simplifies the seller pricing problem regardless of risk preferences. The economic logic established in the previous chapter continues to hold: buyers who receive confirming signals (i.e., $s_1 = 0$) develop such strong posterior beliefs that no second-period signal can meaningfully influence their decisions. These buyers exit the information market entirely, creating no revenue opportunities for the seller. Conversely, buyers who receive contradictory signals (i.e., $s_1 = 1$) find their confidence sufficiently shaken that high-type signals in the second period become valuable.

The crucial insight from the previous chapter applies with equal force here: the mere act of returning to purchase additional information reveals the buyer private signal realisation. Only buyers with $s_1 = 1$ find second-period signals valuable, which eliminates the information asymmetry that would otherwise complicate pricing decisions. For risk-averse sellers, the revelation mechanism proves particularly attractive because it eliminates uncertainty regarding the buyer type whilst preserving the ability to extract full information rent.

For notational convenience, the following definition regarding information rent under specific pricing strategies is introduced.

Definition 2.3. A mapping $\varphi_\tau : (\frac{1}{2}, 1)^2 \to \mathbb{R}$, $\forall \tau \in \{H\}$, is defined as follows[1]:

$$\varphi_H(q^H) \equiv \varphi\left(q_2 = q^H, \mu_1 = \mu_1(q^H, 1)\right),$$

where $\varphi_\tau(\cdot)$ indicates the ex-post information rent in the second period under the pricing strategy $\tau \in \{H\}$.

The following corollary provides a useful characterisation of this information rent.

Corollary 2.6.1. *Given the high-type signal is pivotal, $q^H \in [1 - \mu_1, \mu_1]$, the information rent in the second period under the pricing strategy $\tau \in \{H\}$ is:*

$$\varphi_H(q^H) = q^H - \max\left\{\mu_1(q^H, 1), 1 - \mu_1(q^H, 1)\right\}.$$

Moreover, note that $q^H \in [1 - \mu_1, \mu_1]$ is equivalent to $\frac{\pi}{1-\pi} \in \left(1, \left(\frac{q^H}{1-q^H}\right)^2\right]$.

With the information rent formally defined, we proceed to characterise the optimal pricing strategy. The optimal pricing strategy mirrors the risk-neutral benchmark: the seller sets the following price,

$$\mathcal{P}_2(q^H, q^H) = \varphi_H(q^H),$$

to bind the participation constraint of returning buyers. Low-type signals are priced prohibitively to ensure they are not purchased, consistent with the quality concentration result established in the previous chapter. Consequently, the key proposition from the previous chapter carry forward unchanged:

[1]We follow the definition of $\tau \in T$ defined in Section 1.4.5 of Chapter 1.

Proposition 2.7. *(High-Type Only Pricing Strategy) Suppose that* $\frac{\pi}{1-\pi} \in \left(1, \left(\frac{q^H}{1-q^H}\right)^2\right]$ *and* $q_1 = q^H$ *hold. The optimal menu setting strategy in period* $t = 2$ *is:*

$$\mathcal{P}_2(q^H, q^H) = \varphi_H(q^H);$$
$$\mathcal{P}_2(q^H, q^L) \in \left(\varphi_H(q^H) - (q^H - q^L), +\infty\right).$$

More explicitly,

$$\{\mathcal{P}_2(q^H, q^H), \mathcal{P}_2(q^H, q^L)\} \in \{q^H - \mu_1(q^H, 1)\} \times (q^L - \mu_1(q^H, 1), +\infty) \ \ if \ \pi \geq q^H;$$
$$\{\mathcal{P}_2(q^H, q^H), \mathcal{P}_2(q^H, q^L)\} \in \{q^H - (1 - \mu_1(q^H, 1))\} \times (q^L - (1 - \mu_1(q^H, 1)), +\infty) \ \ if \ \pi \leq q^H.$$

From a risk management perspective, the result above represents an ideal scenario for risk-averse sellers. The strategy concentrates revenue in specific states (i.e., $s_1 = 1$) but does so predictably, with the seller able to anticipate exactly when revenue will materialise. The absence of buyer-type uncertainty eliminates the need to choose between aggressive and conservative pricing approaches, making risk preferences irrelevant to the strategic calculus. The seller achieves the first-best outcome despite information asymmetry, extracting full information rent from returning buyers whilst correctly anticipating which buyers will return. Risk aversion introduces no additional complexity because the natural market segmentation eliminates the revenue uncertainty that would otherwise concern risk-averse agents.

Case 2: Low-Type Signal Previously Acquired ($q_1 = q^L$)

When the buyer acquires a low-type signal in the first period, the strategic environment becomes considerably more complex than the high-type case. The lower precision of the initial signal creates persistent uncertainty that prevents the clear market segmentation. Unlike high-type signals which generate stark binary outcomes, low-type signals produce more nuanced posterior beliefs that can sustain buyer interest in additional information across multiple scenarios.

The analysis established in the previous chapter reveals that the relationship between the buyer prior belief and signal precision determines whether second-period trade occurs. When the prior belief is extremely strong relative to signal quality, even contradictory evidence fails to generate sufficient uncertainty to justify further information acquisition. However, when the prior belief is moderately strong, different signal realisations can lead to different levels of willingness to pay for additional information, creating the central challenge for seller pricing strategy.

The key insight from the previous chapter applies here: the seller faces heterogeneous buyer types whom they cannot distinguish. Buyers who receive confirming signals (i.e., $s_1 = 0$) may return with relatively low willingness to pay, whilst buyers who receive contradictory signals (i.e., $s_1 = 1$) exhibit higher willingness to pay due to greater uncertainty. The inability to observe signal realisations forces the seller to choose between pricing strategies that serve different segments of this heterogeneous market.

For notational convenience, we introduce definitions for the information rent under different pricing strategies that emerge in this setting.

Definition 2.4. A mapping $\varphi_\tau : (\frac{1}{2}, 1)^2 \to \mathbb{R}$, $\forall \tau \in \{A, C\}$, is defined as follows[2]:

$$\varphi_A(q^H, q^L) \equiv \varphi\left(q_2 = q^H, \mu_1 = \mu_1(q^L, 1)\right);$$
$$\varphi_C(q^H, q^L) \equiv \varphi\left(q_2 = q^H, \mu_1 = \mu_1(q^L, 0)\right),$$

where $\varphi_\tau(\cdot)$ indicates the ex-post information rent in the second period under the pricing strategy $\tau \in \{A, C\}$.

The following corollary provides a useful characterisation of this information rent.

Corollary 2.7.1. *Given the high-type signal is pivotal, $q^H \in [1 - \mu_1, \mu_1]$, the information rent in the second period under the pricing strategy $\tau \in \{A, C\}$ is:*

$$\varphi_A(q^H, q^L) = q^H - \max\{\mu_1(q^L, 1), 1 - \mu_1(q^L, 1)\};$$
$$\varphi_C(q^H, q^L) = q^H - \mu_1(q^L, 0).$$

Moreover, note that $q^H \in [1 - \mu_1, \mu_1]$ is equivalent to: $\frac{\pi}{1-\pi} \in \left(\frac{1-q^L}{q^L}\frac{q^H}{1-q^H}, \frac{q^L}{1-q^L}\frac{q^H}{1-q^H}\right]$ if $\mu_1 = \mu_1(q^L, 1)$; $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$ if $\mu_1 = \mu_1(q^L, 0)$.

With these definitions established, we characterise the optimal pricing strategies. The analysis from the previous chapter demonstrates that the seller optimal choice depends on the strength of the buyer prior belief relative to signal precision. When the prior belief is extremely strong such

[2] We follow the definition of $\tau \in T$ defined in Section 1.4.5 of Chapter 1.

that $\frac{\pi}{1-\pi} \in \left(\frac{q^L}{1-q^L}\frac{q^H}{1-q^H}, +\infty\right)$, the buyer conviction remains intact, preventing any second-period transactions, regardless of signal realisations.

The Selective Demand Case: Risk-Neutral Pricing Persists. When the prior belief is sufficiently strong such that $\frac{\pi}{1-\pi} \in \left(\frac{1-q^L}{q^L}\frac{q^H}{1-q^H}, \frac{q^L}{1-q^L}\frac{q^H}{1-q^H}\right]$, only buyers with contradictory signals (i.e., $s_1 = 1$) return, simplifying the pricing problem to the aggressive strategy characterised in the previous chapter. Consequently, the key proposition from the previous chapter carry forward unchanged:

Proposition 2.8. *Suppose that $\frac{\pi}{1-\pi} \in \left(\frac{1-q^L}{q^L}\frac{q^H}{1-q^H}, \frac{q^L}{1-q^L}\frac{q^H}{1-q^H}\right]$ and $q_1 = q^L$ hold. The optimal menu setting strategy in period $t = 2$ is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi_A(q^H, q^L);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi_A(q^H, q^L) - (q^H - q^L), +\infty\right).$$

More explicitly,

$$\{\mathcal{P}_2(q^L, q^H), \mathcal{P}_2(q^L, q^L)\} \in \{q^H - \mu_1(q^L, 1)\} \times (q^L - \mu_1(q^L, 1), +\infty) \ \text{if} \ \pi \geq q^L;$$
$$\{\mathcal{P}_2(q^L, q^H), \mathcal{P}_2(q^L, q^L)\} \in \{q^L - (1 - \mu_1(q^L, 1))\} \times (q^L - (1 - \mu_1(q^L, 1)), +\infty) \ \text{if} \ \pi \leq q^L.$$

The Universal Demand Case: Risk Aversion Transforms Strategy. The more interesting case emerges when $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$ holds, the prior belief is weaker. In this regime, both types of buyers may return for additional information, but with different reservation prices. The seller must choose between two distinct approaches: the aggressive pricing strategy which targets only high-willingness-to-pay buyers, and the conservative pricing strategy which ensures universal participation at lower margins. Under the aggressive strategy, the seller sets prices to extract maximum rent from buyers with contradictory signals, accepting that buyers with confirming signals will be excluded. The conservative strategy adopts inclusive pricing that captures revenue from all returning buyers, sacrificing per-unit margins for broader market participation. Consequently, the key proposition from the previous chapter carry forward unchanged:

Proposition 2.9 (Aggressive Pricing Strategy)**.** *Suppose $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$ and $q_1 = q^L$. If the aggressive pricing strategy is optimal, $\tau^* = A$, then the optimal menu setting in period $t = 2$ is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi_A(q^H, q^L);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi_A(q^H, q^L) - (q^H - q^L), +\infty\right).$$

Proposition 2.10 (Conservative Pricing Strategy)**.** *Suppose that* $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$ *and* $q_1 = q^L$. *If the conservative pricing strategy is optimal,* $\tau^* = C$*, then the optimal menu setting strategy in period* $t = 2$ *is:*

$$\mathcal{P}_2(q^L, q^H) = \varphi_C(q^H, q^L);$$
$$\mathcal{P}_2(q^L, q^L) \in \left(\varphi_C(q^H, q^L) - (q^H - q^L), +\infty\right).$$

These strategic approaches reflect fundamentally different revenue philosophies. The aggressive strategy prioritises high-value buyers whilst excluding others, maximising per-transaction margins. The conservative strategy ensures universal participation, trading margins for market coverage and revenue predictability.

Under risk neutrality, the choice between these approaches reduces to a straightforward comparison of expected revenues, with the analysis in the previous chapter establishing clear conditions under which each dominates. However, the introduction of risk aversion transforms this calculus in profound ways. Risk-averse sellers do not merely compare expected revenues; they must also weigh the uncertainty inherent in each approach. The aggressive strategy, whilst potentially more profitable on average, concentrates revenue in specific scenarios where buyers exhibit high uncertainty, creating exactly the type of volatility that risk-averse agents seek to avoid. The conservative strategy, by contrast, offers the appeal of guaranteed participation from all returning buyers, smoothing revenue across different market conditions. The question becomes whether the certainty of universal participation can compensate for the sacrifice in expected profits, a trade-off that depends critically on the degree of risk aversion exhibited by the seller.

The risk-averse seller faces a fundamentally different optimisation problem than their risk-neutral counterpart. Rather than simply maximising expected revenue, they must balance profit potential against revenue volatility. Under CRRA preferences with parameter $\sigma \in (0, 1)$, the seller evaluates strategies based on expected utility rather than expected profit alone. The strategic choice crystallises around a key trade-off. The conservative strategy guarantees revenue of $\varphi_C(q^H, q^L)$

with certainty, providing complete income smoothing across buyer types. The aggressive strategy offers higher potential revenue of $\varphi_A(q^H, q^L)$ but only materialises when buyers receive contradictory signals, occurring with probability $\pi(1-q^L)+(1-\pi)q^L < 1$. The decision of the seller hinges on whether the utility from guaranteed income exceeds the expected utility from this riskier but potentially more profitable approach. We can now establish the following proposition.

Proposition 2.11. *Given that $\tau^* \in \{A, C\}$ holds, $\tau = C$ is preferred over $\tau = A$ if and only if the following inequality holds:*

$$\left[\varphi_C(q^H, q^L)\right]^{1-\sigma} \geq \left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_A(q^H, q^L)\right]^{1-\sigma}.$$

This condition reveals the economic forces at play. The left side represents the utility from certain revenue under conservative pricing, whilst the right side captures the expected utility from the probabilistic revenue stream under aggressive pricing. When risk aversion is absent (i.e., $\sigma = 0$), the comparison reduces to expected revenues. As risk aversion increases, the concavity of the utility function increasingly favours the certainty offered by conservative pricing.

The strategic choice depends critically on the quality of available signals, which determines the magnitude of information rents under each strategy. The analysis reveals a natural threshold effect characterised by a cut-off signal quality q^{H^*} at which the seller becomes indifferent between strategies. Above this threshold, risk-averse sellers strictly prefer conservative pricing, whilst below it they may choose aggressive strategies despite their risk preferences.

The economic intuition behind this threshold reflects two competing forces that shape seller preferences. The first force, discussed in the previous chapter under risk neutrality, concerns the absolute magnitude of extractable information rent. When signal quality approaches the lower bound (i.e., q^H approaches $\mu_1(q^L, 0)$ from above), the conservative strategy yields information rent $\varphi_C(q^H, q^L)$ that approaches zero, naturally dampening the seller incentive to pursue guaranteed but minimal returns. The second force emerges from risk aversion itself: the diminishing marginal utility effect makes sellers increasingly reluctant to chase higher expected payoffs when those payoffs come with uncertainty.

Consider the dynamics around the cut-off point. For any given low-type signal precision q^L, the probability of second-period trade under aggressive pricing remains fixed at $\pi(1-q^L)+(1-\pi)q^L$. The strategic choice thus hinges entirely on how the sellers evaluate the information rent differential between strategies. When signal quality increases marginally above the cut-off (i.e., $\widetilde{q^H} = q^{H^*} + \varepsilon$), the marginal utility effect favours conservative pricing decisively. Let φ_C and φ_A denote $\varphi_C(q^H, q^L)$ and $\varphi_A(q^H, q^L)$, respectively. Since $u(\cdot)$ exhibits diminishing marginal utility and $\varphi_C < \varphi_A$, we have $u'(\varphi_C) > u'(\varphi_A)$. Moreover, the probability weighting under aggressive pricing further reduces its attractiveness:

$$u'(\varphi_C) > u'(\varphi_A) > \left[\pi(1-q^L) + (1-\pi)q^L\right] u'(\varphi_A),$$

cementing the preference for conservative pricing. The threshold itself responds systematically to risk aversion. As the extent of risk aversion σ increases, the cut-off point q^{H^*} decreases, expanding the range of signal qualities under which conservative pricing proves optimal. Intuitively, more risk-averse sellers become increasingly willing to sacrifice expected profits for revenue certainty, making the guaranteed participation offered by conservative pricing attractive across a broader spectrum of market conditions. The following theorem formalises this threshold relationship and its dependence on risk aversion.

Proposition 2.12. *For any* $\sigma \in (0, 1]$,

(i) there exists a cut-off point $q^{H^*} \in (\mu_1(q^L, 0), +\infty)$ *such that the seller is indifferent between conservative and aggressive pricing strategy,*

$$\left[\varphi_C(q^{H^*}, q^L)\right]^{1-\sigma} = \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^{H^*}, q^L)\right]^{1-\sigma};$$

(ii) the cut-off point, q^{H^*}*, is decreasing in* σ.

Proof. See Appendix 2.A.1. □

The economic intuition behind this threshold is compelling. As signal quality increases, the conservative strategy becomes more attractive because it guarantees extraction of substantial information rent from all buyer types. The aggressive strategy, whilst offering higher per-transaction profits, becomes relatively less appealing as the guaranteed revenue under conservative pricing grows. Moreover, increased risk aversion lowers this threshold, making conservative pricing optimal over a broader range of signal qualities.

Understanding how risk aversion influences the choice between conservative and aggressive pricing requires examining the relationship between seller preferences and market parameters.

The analysis focuses on how varying degrees of risk aversion affect optimal strategy selection, holding fixed the underlying market structure characterised by parameters (π, q^H, q^L). The clearest insights emerge from considering two polar cases: complete risk neutrality and extreme risk aversion.

The Risk-Neutral Benchmark. When sellers exhibit no risk aversion (i.e., $\sigma = 0$), strategic choice reduces to a straightforward comparison of expected revenues. The seller simply asks: which strategy generates higher expected profits? The answer depends critically on the informativeness of the initial low-type signal, which determines how much buyers learn from their first-period experience.

Consider first the case where low-type signals are highly informative. When q^L approaches its upper bound, buyers who receive confirming evidence (i.e., $s_1 = 0$) develop such strong posterior beliefs that second-period information becomes nearly worthless. The information rent under conservative pricing, $\varphi_C(q^H, q^L)$, approaches zero, making the strategy economically unattractive. In such environments, aggressive pricing dominates by focusing exclusively on the buyers who remain genuinely uncertain after receiving contradictory evidence.

The opposite scenario unfolds when low-type signals are barely informative. As q^L approaches $\frac{1}{2}$, signal realisations provide minimal information, leaving buyer beliefs largely unchanged regardless of what they observe. The information rents under both strategies converge (i.e., $\varphi_C(q^H, q^L) \approx \varphi_A(q^H, q^L)$), but the conservative strategy gains a decisive advantage through its guarantee of universal participation. Whilst aggressive pricing succeeds with probability approximately $\frac{1}{2}$, conservative pricing ensures trade with certainty, making it the clear winner even without risk aversion considerations.

The Extreme Risk Aversion Case. At the opposite extreme, when risk aversion becomes severe (i.e., $\sigma \to 1$), the strategic calculus transforms completely. The utility function becomes so concave that sellers become nearly indifferent to the magnitude of positive payoffs, caring primarily about their probability of occurrence. Any strictly positive surplus yields utility close to one, whilst zero surplus provides no utility whatsoever. Under such extreme preferences, the conservative pricing strategy becomes universally optimal, regardless of signal parameters or information rent structures. The guarantee of positive returns with certainty dominates any uncertain alternative, no matter how potentially profitable. The seller essentially becomes a satisficer rather than an optimiser, prioritising the avoidance of zero outcomes over the pursuit of maximum returns.

Strategic Transitions in the Intermediate Risk Aversion Range. Between these extremes lies the most interesting analytical territory, where moderate risk aversion creates genuine strategic trade-offs. The transition from aggressive to conservative pricing preferences occurs gradually as risk aversion increases, with the critical threshold depending on underlying market parameters. When market conditions favour aggressive pricing under risk neutrality, there exists a critical level of risk aversion $\sigma^*_{CA} \in (0, 1)$ at which the seller becomes indifferent between strategies. Below this threshold, the lure of higher expected profits outweighs concerns about revenue uncertainty. Above it, the appeal of guaranteed returns dominates profit maximisation considerations. Conversely, when market conditions already favour conservative pricing under risk neutrality, risk aversion merely reinforces this preference without creating interesting threshold effects.

The following theorem formalises these insights:

Proposition 2.13. *Let $\sigma \in [0, 1)$ denote the degree of risk aversion of the seller. Then,*

(i) Given that $\varphi_C(q^H, q^L) \leq \left[\pi(1 - q^L) + (1 - \pi)q^L\right] \varphi_A(q^H, q^L)$, there exists a cut-off point $\sigma^*_{CA} \in [0, 1)$ such that the seller prefers the conservative pricing strategy if and only if $\sigma > \sigma^*_{CA}$.

(ii) Given that $\varphi_C(q^H, q^L) \geq \left[\pi(1 - q^L) + (1 - \pi)q^L\right] \varphi_A(q^H, q^L)$, the seller always prefers the conservative pricing strategy, regardless of the degree of risk aversion $\sigma \in [0, 1)$.

Proof. See Appendix 2.A.1. □

The distribution of surplus between seller and buyer depends critically on which pricing strategy is implemented and reveals important insights about the welfare effects of information asymmetry in dynamic settings. When the seller adopts aggressive pricing, surplus allocation follows a stark binary pattern. Buyers who receive contradictory signals (i.e., $s_1 = 1$) face prices that extract the full information rent, leaving them with zero surplus from second-period transactions. The seller captures all available value from these high-uncertainty buyers who desperately need additional information. Conversely, buyers who receive confirming signals (i.e., $s_1 = 0$) find themselves priced out of the second-period market entirely. Whilst these buyers would derive positive value from high-type signals as $q^H > \mu_1(q^L, 0)$ ensures strictly positive information rent, the seller cannot profitably serve them due to information asymmetry. The inability to observe signal realisations prevents optimal price discrimination, leaving potential surplus unexploited. In a hypothetical first-best environment where the seller could observe s_1, this inefficiency would disappear. The seller could set different prices for each buyer type, extracting information rent from both groups and achieving higher expected profits. The information asymmetry thus creates a genuine welfare loss under aggressive pricing.

Conservative pricing generates a more nuanced surplus allocation that initially appears to suffer from similar inefficiencies. Buyers who receive contradictory signals (i.e., $s_1 = 1$) obtain strictly positive surplus equal to $\varphi_A(q^H, q^L) - \varphi_C(q^H, q^L)$, since the conservative price lies below their maximum willingness to pay. At first glance, this surplus retention suggests that information asymmetry prevents the seller from achieving first-best outcomes. However, this appearance proves deceptive. The key insight is that sellers can extract the expected value of this buyer surplus through intertemporal pricing adjustments. By increasing first-period prices by exactly $\mathbb{P}(s_1 = 1)[\varphi_A(q^H, q^L) - \varphi_C(q^H, q^L)]$, the seller captures the anticipated second-period surplus without violating buyer participation constraints. Forward-looking buyers willingly pay this premium because they anticipate receiving compensating surplus in the future. Through such mechanism, conservative pricing achieves the remarkable result of eliminating welfare losses from information asymmetry. The seller extracts the same total surplus as in the first-best benchmark, merely reallocating its timing across periods rather than sacrificing it entirely. The dynamic structure of the game thus provides a natural solution to the adverse selection problem that would otherwise constrain seller profits. A complete analysis of this intertemporal surplus extraction mechanism follows in Section 2.4.5, which characterises optimal first-period pricing strategies.

2.4.4 Signal Acquisition by the Buyer in the First Period

The analysis now turns to the information acquisition decision of the buyer in period $t = 1$. A crucial insight emerges immediately: the strategic calculus of the buyer remains fundamentally unchanged from the analysis in the previous chapter. Since the buyer maintains risk neutrality and faces identical decision criteria, the introduction of seller risk aversion does not alter the underlying economics of information demand from the perspective of the buyer. The invariance of buyer behaviour to seller risk preferences reflects a deeper economic principle. The optimisation problem of the buyer depends solely on anticipated prices and expected payoffs, not on the psychological motivations underlying seller pricing decisions. Whether sellers choose aggressive or conservative strategies due to profit maximisation or risk management considerations proves irrelevant to buyer decision-making, provided the resulting prices and surplus opportunities remain accessible.

At the beginning of period $t = 1$, the buyer faces the same fundamental trade-off analysed in the previous chapter. They weigh immediate information value against price, whilst account for future surplus opportunities. The expected payoff of the buyer decomposes into immediate gains from first-period information and expected surplus from future information acquisition, denoted $V(q_1, \mathcal{P}_2)$, exactly as before. The critical insight from the previous chapter applies with equal force here. The value function, $V(q_1, \mathcal{P}_2)$, depends on the nature of second-period pricing strategies, not on the underlying motivations that drive choices for the seller. Under aggressive pricing, the buyer anticipates zero future surplus regardless of whether the seller chooses this strategy to maximise expected profits or to concentrate revenue in specific scenarios. Under conservative pricing, buyers benefit from positive expected surplus whether they adopt such approach to ensure broad participation or to manage revenue uncertainty.

The key difference does not lie in the behaviour of the buyer but in the frequency with which different pricing strategies emerge. A Risk-averse seller proves to be more likely to adopt conservative approaches, creating more opportunities for buyers to capture future surplus. However, when such opportunities arise, the buyer responses follow the same strategic logic established in the previous chapter. Accordingly, rather than re-deriving the optimisation problem of the buyer, we draw directly upon the results established in the previous chapter. The following characterisation applies without further modification:

Proposition 2.14 (Buyer's First-Period Optimisation)**.** *Under any equilibrium, the buyer acquires an informative signal $q_1^* \in Q$ in period $t = 1$ if and only if all the following conditions hold. Otherwise, $q_1^* = \emptyset$.*

(i) $q_1^* - \mathcal{P}_1(q_1^*) + V(q_1^*, \mathcal{P}_2) \geq \pi$, *participation constraint;*

(ii) $q_1^* - \mathcal{P}_1(q_1^*) + V(q_1^*, \mathcal{P}_2) \geq q_1' - \mathcal{P}_1(q_1') + V(q_1', \mathcal{P}_2)$ *for all* $q_1' \in Q$, *incentive compatibility constraint;*

(iii) $\mu_1 \in [1 - q_1^*, q_1^*]$, *pivotal signal constraint.*

The strategic insights from the previous chapter carry forward unchanged. The willingness to pay of the buyer for early information reflects both immediate value and option value of future decision-making. The forward-looking buyer incorporates anticipated future decisions into current optimisation, creating intertemporal linkages that constrain pricing power of the seller across periods.

Although the risk aversion of the seller does not directly affect the decision-making of the buyer, it generates significant indirect effects by shaping the equilibrium pricing strategies. The risk-averse seller tends to favour conservative pricing, thereby increasing the likelihood that buyers encounter second-period environments with positive surplus opportunities. These environments enhance the value of first-period signals by generating value that extends beyond the immediate informational benefit.

The ability of the seller to extract this enhanced value through higher first-period prices demonstrates the sophisticated nature of dynamic rent extraction. Risk-averse sellers who choose conservative second-period pricing do not sacrifice profits; instead, they shift surplus extraction across time whilst managing revenue uncertainty. Buyers, recognising this pattern, remain willing to pay premium prices for access to future opportunities, ensuring that changes in seller risk preferences ultimately prove neutral for buyer welfare whilst affecting only the timing and distribution of surplus extraction.

The fundamental economics of information demand thus remain unchanged by seller risk aversion. Risk-averse sellers face the same demand conditions and behaviour of the buyer as their risk-neutral counterparts, yet they may respond differently to the revenue uncertainty inherent in information provision. The analysis now turns to examine how the seller risk aversion influences pricing strategies, whilst adhering to these invariant demand fundamentals.

2.4.5 Menu Setting by the Seller in the First Period

This section investigates the optimal pricing scheme of the seller in the first period, where the fundamental tension between risk and return reaches its strategic culmination. The seller chooses a price menu $\mathcal{P}_1(q^H, q^L)$ at the start of period $t = 1$ which maximises their ex-ante expected utility whilst accounting for the interactions that will unfold over both periods. This decision proves particularly intricate because the seller must simultaneously consider three interconnected factors: the immediate revenue from first-period sales, the probability and magnitude of future trading opportunities, and the risk profile of the resulting revenue stream.

The problem of the seller extends far beyond simple price setting, as the first-period menu fundamentally shapes the entire market dynamic. By choosing which signals to offer and at what prices, the seller determines not only who will participate in the initial market but also the composition and behaviour of buyers who return for second-period information. This creates a sophisticated intertemporal optimisation problem where current pricing decisions must account for their impact on future market structure, buyer beliefs, and the ability of the seller to extract rent in subsequent periods. The introduction of risk aversion adds another layer of complexity, as the seller must now balance the higher expected profits from strategies that concentrate revenue in specific market scenarios against the volatility inherent in such approaches.

The analysis reveals that different configurations of prior beliefs and signal precisions give rise to qualitatively different strategic environments. When buyers hold very strong prior beliefs, information has little value regardless of quality, creating a trivial pricing problem. However, as prior beliefs become more moderate, the seller faces increasingly complex trade-off between offering high-type signals that enable superior rent extraction and lower-type signals that generate more stable demand patterns. The optimal strategy depends critically on the degree of risk aversion, with more risk-averse sellers potentially preferring approaches that sacrifice expected profits in favour of revenue stability and broader market participation.

The following analysis proceeds through three distinct parametric regimes that together exhaust all possible configurations of prior beliefs and signal precisions. Each regime reveals fundamentally different strategic considerations for a risk-averse seller.

Extreme Prior Beliefs: Market Breakdown

This regime captures scenarios where buyer prior beliefs are so extreme that information becomes economically worthless regardless of signal quality or pricing strategy. The seller faces a degenerate pricing problem where no positive surplus can be extracted under any circumstances, making risk preferences irrelevant. This case serves as a useful benchmark demonstrating how sufficiently strong priors can completely eliminate information markets.

As in the previous chapter, we first examine the case in which $\frac{\pi}{1-\pi} \in \left[\left(\frac{q^H}{1-q^H}\right)^2, +\infty\right)$ holds, which implies that $\pi > q^H > q^L$. This parameter region represents scenarios where buyers hold extremely strong prior beliefs that render all available information economically irrelevant, creating a degenerate market environment that is independent of risk preferences.

When the prior belief for the status quo is sufficiently strong, the buyer has no incentive to acquire any type of signal at any positive price, regardless of the period $t \in \{1, 2\}$ or the signal realisation $s_1 \in \{0, 1\}$. The economic intuition is straightforward, The initial confidence of the buyer in the status quo is so overwhelming that even the most contradictory evidence cannot generate sufficient uncertainty to justify costly information acquisition. Consider the most extreme possible scenario where the buyer acquires two high-precision signals consecutively, both contradicting the status quo (i.e., $s_1 = s_2 = 1$). Even under these circumstances, the posterior belief remains above $\frac{1}{2}$, confirming that signals of both types fail to be pivotal throughout the entire game. This extreme parameter region illustrates a fundamental property of information markets. When prior beliefs are sufficiently strong, information loses its economic value entirely, which creates a scenario where trade becomes economically unviable for both parties. Although the buyer would accept free signals, the seller has no incentive to provide them since the expected payoff remains zero regardless of the pricing strategy employed.

The implications for optimal pricing are trivial yet instructive. Since no positive surplus can be extracted under any circumstances, the pricing problem of the seller becomes vacuous. Any price menu satisfies the equilibrium conditions, as market forces eliminate all trading opportunities. This result holds regardless of the degree of risk aversion, as the absence of extractable surplus makes risk considerations irrelevant.

Proposition 2.15. *Suppose that* $\frac{\pi}{1-\pi} \in \left[\left(\frac{q^H}{1-q^H}\right)^2, +\infty\right)$ *holds. Given any* $\sigma \in (0, 1]$, *the optimal menu setting strategy in* $t = 1$ *is:*

$$(\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)) \in (\mathbb{R}_+)^2.$$

This degenerate case serves as a useful benchmark for understanding the role of prior beliefs in information markets. It demonstrates that extremely strong priors can completely eliminate the value of information, regardless of signal quality or market structure. The subsequent analysis will show how relaxing this extreme assumption leads to richer strategic interactions where risk preferences play a decisive role in determining optimal pricing strategies.

Strong Prior Beliefs: High-Type Signal Dominance

In this intermediate regime, prior beliefs remain strong but not extreme enough to eliminate all information value. The analysis reveals a stark asymmetry between signal types: low-type signals fail to generate any economic value across both periods, whilst high-type signals create profitable second-period trading opportunities. The seller optimally offers free high-type signals to induce strategic information acquisition, with this dominance holding regardless of the degree of risk aversion.

Next, consider the case in which $\frac{\pi}{1-\pi} \in \left[\frac{q^L}{1-q^L}\frac{q^H}{1-q^H}, \left(\frac{q^H}{1-q^H}\right)^2\right]$, which implies $\pi > q^H > q^L$. This parameter region captures an intermediate scenario where prior beliefs are strong, but not extreme enough to nullify the value of information. The analysis uncovers a striking asymmetry between high-type and low-type signals that fundamentally shapes the optimal pricing strategy.

The seller recognises that offering a low-type signal in the first period leads to a strategic dead end. The low-type signal is too imprecise to shift posterior beliefs towards the centre, thereby failing to generate any information rent in the second period, regardless of the subsequent signal realisation. Formally, $\mu_1(q^L, 1) \notin [1-q, q]$ for all $q \in \{q^H, q^L\}$, which implies that no second-period signal is pivotal, as established in Lemma 2.1. Economically, even when the low-type signal contradicts the prior belief, the resulting posterior remains too extreme to induce genuine uncertainty. Moreover, since $\pi > q^L$, the first-period signal also yields zero information rent. As a result, the seller earns no surplus in either period when the buyer is induced to acquire a low-type signal initially.

By contrast, offering a high-type signal in the first period creates valuable strategic opportunities. With positive probability, the buyer returns in the second period and pays strictly positive information rent—specifically when $q_1 = q^H$ and $s_1 = 1$. This follows from the fact that $\mu_1(q^H, 1) \in [1-q^H, q^H]$ under the assumed parameters. The high-type signal is sufficiently precise to shift the posterior belief significantly when contradictory evidence arises, moving the posterior towards the centre and generating enough uncertainty to make the signal pivotal.

The seller can extract the entire information rent in the second period by setting the price $\mathcal{P}_2(q^H, q^H) = \varphi_H(q^H)$, as shown in Proposition 2.7. This pricing strategy ensures the buyer's participation constraint binds exactly, allowing the seller to capture the full surplus. Compared to the case $q_1 = q^L$, where the seller earns zero, offering a high-type signal strictly dominates for all $\sigma \in (0, 1]$. The superiority of this strategy is robust to risk preferences, as the dominance arises purely from the strategic value of information.

Proposition 2.16. *Suppose that* $\frac{\pi}{1-\pi} \in \left[\frac{q^L}{1-q^L}\frac{q^H}{1-q^H}, \left(\frac{q^H}{1-q^H}\right)^2\right]$. *Then, for any* $\sigma \in (0, 1]$, *the optimal first-period pricing strategy is:*

$$\mathcal{P}_1(q^H) = 0; \quad \mathcal{P}_1(q^L) \in (0, +\infty).$$

The ex-ante expected utility from implementing the *high-type only pricing strategy*, as defined in the previous chapter, is as follows. Since $\pi > q^H$, the seller earns zero in period one. However, they extract the full information rent by setting the price $\mathcal{P}_2(q^H, q^H) = q^H - \mu_1(q^H, 1) = \varphi_H(q^H)$ in period two with probability $\pi(1-q^H) + (1-\pi)q^H$. Under the CRRA utility function with risk aversion parameter $\sigma \in (0, 1]$, the expected utility of the seller is:

$$u_S^H(\sigma) \equiv \mathbb{E}[u_S(\tau = H) \mid \sigma] = \left[\pi(1-q^H) + (1-\pi)q^H\right] \left[\varphi_H(q^H)\right]^{1-\sigma}.$$

Moderate Prior Beliefs: Strategic Trade-offs and Risk Aversion

This regime presents the most intricate strategic environment, $\frac{\pi}{1-\pi} \in \left(1, \frac{q^L}{1-q^L} \frac{q^H}{1-q^H}\right]$, where multiple pricing strategies are viable and risk preferences become decisive. The analysis naturally separates into two sub-cases according to the seller's degree of risk aversion. For less risk-averse sellers, the choice is between high-type only strategies that maximise expected profits and aggressive strategies that increase the likelihood of trade at the expense of revenue concentration. In contrast, more risk-averse sellers may favour conservative strategies that ensure broader market participation, offering a viable alternative to high-type only approaches. The optimal strategy in each sub-case depends critically on the degree of risk aversion and the underlying parameter values.

Low Risk Aversion: High-Type Only versus Aggressive Strategies. When the seller exhibits low risk aversion, $\sigma \leq \sigma_{CA}^*$, the strategic landscape reveals a fascinating choice between two fundamentally different approaches to information provision. The seller faces a portfolio-like decision: pursue the high-expected-return strategy of offering only high-type signals across both periods, or adopt the more diversified approach of combining low-type signals initially with high-type signals subsequently.

The high-type only strategy represents the classic quality-maximisation approach established in the previous chapter. Under this strategy, the seller induces the buyer to acquire high-type signals in both periods, concentrating on extracting maximum surplus from the most informative signals available. The alternative approach follows a more nuanced path: the seller induces the buyer to acquire a low-type signal in the first period, followed by a high-type signal in the second period. The economic beauty of this second approach lies in its information revelation properties. Since the buyer acquires a signal in the second period only if $s_1 = 1$, the seller faces no information asymmetry when designing second-period pricing. The seller perfectly understands that $s_1 = 1$ must be realised if the buyer approaches them. The ex-ante expected payoff of the seller implementing the aggressive pricing strategy captures this strategic insight:

$$u_S^A(\sigma) \equiv \mathbb{E}[u_S(\tau = A \mid \sigma)] = \left[\pi(1-q^L) + (1-\pi)q^L\right]\left[\varphi_A(q^H, q^L)\right]^{1-\sigma} + \left[\max\{q^L - \pi, 0\}\right]^{1-\sigma}.$$

Under risk neutrality, the choice between these strategies was unambiguous: the seller always preferred the high-type-only pricing strategy (i.e., $\tau^* = H$) as it maximised expected surplus. The previous chapter established this result through straightforward expected value comparisons. However, the introduction of risk aversion fundamentally alters this calculus by making sellers sensitive to the probability distribution of outcomes, not merely their expected values.

The transformation becomes most apparent when sellers exhibit extreme risk aversion (i.e., $\sigma \to 1$). In such cases, sellers become concerned primarily with the probability of receiving strictly positive surplus across the two periods, rather than the magnitude of that surplus. This shift in priorities can reverse the preference ordering established under risk neutrality, making the aggressive pricing strategy attractive precisely because it increases the likelihood of positive returns.

In this context, the relationship between the prior belief, π, and the signal precisions, (q^H, q^L), becomes crucial. We distinguish the following three cases for further analysis: (I) $\pi > q^H > q^L$; (II) $q^H > q^L > \pi$; (III) $q^H > \pi > q^L$.

(I) Strong Prior Beliefs: $\pi > q^H > q^L$. Since the prior belief exceeds both signal precisions, the seller is unable to extract any surplus in the first period, regardless of the pricing strategy implemented. The seller therefore focuses solely on the probability of generating surplus in the second period. Given that $\pi(1-q^L) + (1-\pi)q^L > \pi(1-q^H) + (1-\pi)q^H$ always holds, the aggressive pricing strategy is strictly preferred. A concise technical derivation is provided below.

When $\pi > q^H > q^L$, the implementation of the aggressive pricing strategy increases the probability of a trade, $\mathbb{P}(s_1 = 1)$, as follows:

$$\begin{aligned}\mathbb{P}(s_1 = 1 \mid \tau = A) &= \pi(1-q^L) + (1-\pi)q^L \\ &> \pi(1-q^H) + (1-\pi)q^H = \mathbb{P}(s_1 = 1 \mid \tau = H),\end{aligned}$$

whilst the ex-post information rent is lower:

$$\begin{aligned}\varphi_A(q^H, q^L) &= q^H - \mu_1(q^L, 1) \\ &< q^H - \mu_1(q^H, 1) = \varphi_H(q^H).\end{aligned}$$

Effectively, by implementing $\tau = A$, the seller opts for a less risky lottery. In summary, inducing the acquisition of a high-type signal is not necessarily optimal when risk neutrality is not assumed. There exists a cut-off point $\sigma^*_{HA}(\pi, q^H, q^L)$ at which the seller is indifferent between the two pricing strategies. If $\sigma \leq \sigma^*_{HA}$, the seller is relatively less risk-averse and thus prefers to induce the acquisition of high-type signals only. Accordingly, we proceed to establish the following proposition.

Lemma 2.2. *For any* $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ *such that* $\pi > q^H > q^L$ *holds, there exists a unique cut-off point* $\sigma^*_{HA} \in (0, 1)$ *such that:*

(i) the seller is indifferent between $\tau = H$ *and* $\tau = A$ *if* $\sigma = \sigma^*_{HA}$. *That is,* $u^H_S(\sigma^*_{HA}) = u^A_S(\sigma^*_{HA})$ *holds at the cut-off point;*

(ii) the seller prefers $\tau = H$ *over* $\tau = A$ *when they exhibit sufficiently low risk aversion. That is,* $u^H_S(\sigma) \geq u^A_S(\sigma)$ *holds if and only if* $\sigma \in [0, \sigma^*_{HA}]$.

Proof. See Appendix 2.A.2. □

(II) Weak Prior Beliefs: $q^H > q^L > \pi$. Here, both signal precisions exceed the prior belief, and thus the seller receives a strictly positive surplus in the first period under any pricing strategy, with probability one. Thus, the expected utility from the first-period surplus is constant across strategies. Likewise, the preference of the seller is governed by the probability of second-period trade, leading to a strict preference for the aggressive pricing strategy due to its higher trade probability. A concise technical derivation is provided below.

When $q^H > q^L > \pi$, the seller also takes into account the surplus generated in the first period, given by $\max\{q_1 - \pi, 0\}$. As a result, the high-type only pricing strategy, $\tau = H$, becomes more appealing, since the marginal utility derived from the first-period information rent is more pronounced and $q^H - \pi > q^L - \pi$ holds. Nonetheless, the analogous argument applies: there exists a unique cut-off point, denoted $\sigma^*_{HA}(\pi, q^H, q^L)$, at which the seller is indifferent between the two pricing strategies. Accordingly, we proceed to establish the following proposition.

Lemma 2.3. *For any $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ such that $q^H > q^L > \pi$ holds, there exists a unique cut-off point $\sigma^*_{HA} \in (0, 1)$ such that:*

(i) the seller is indifferent between $\tau = H$ and $\tau = A$ if $\sigma = \sigma^*_{HA}$. That is, $u^H_S(\sigma^*_{HA}) = u^A_S(\sigma^*_{HA})$ holds at the cut-off point;

(ii) the seller prefers $\tau = H$ over $\tau = A$ when they exhibit sufficiently low risk aversion. That is, $u^H_S(\sigma) \geq u^A_S(\sigma)$ holds if and only if $\sigma \in [0, \sigma^*_{HA}]$.

Proof. See Appendix 2.A.2. □

(III) Intermediate Prior Beliefs: $q^H > \pi > q^L$. In this case, the seller obtains strictly positive surplus in the first period only if the high-type-only pricing strategy is implemented, since $q^H > \pi$ whilst $q^L < \pi$. For the seller to prefer the aggressive pricing strategy, the difference in second-period trade probability must compensate for the loss of guaranteed surplus in the first period. However, the compensation is infeasible. Specifically, the difference in trade probability between the two strategies is given by $\pi(1 - q^L) + (1 - \pi)q^L - \pi(1 - q^H) + (1 - \pi)q^H = (2\pi - 1)(q^H - q^L) < \frac{1}{2}$, which is strictly less than one and insufficient to offset the marginal utility from guaranteed first-period surplus. Accordingly, in this case, the high-type-only pricing strategy is always strictly preferred. A concise technical derivation is provided below.

When $q^H > \pi > q^L$, the seller always prefers the high-type only pricing strategy, for any $\sigma \in (0, 1]$, since a strictly positive surplus of $q^H - \pi$ is obtained in period $t = 1$ only under this strategy. Intuitively, as the marginal utility evaluated at zero goes to positive infinity, the seller experiences a sharp increase in utility from gaining a strictly positive surplus, making the first-period benefit particularly salient. In contrast, when $q^H > q^L > \pi$ or $\pi > q^H > q^L$, this utility gain is less pronounced, since the seller earns a lower surplus of $q_1 - \pi$ or none at all in the first period. Consequently, the effect of marginal utility is weaker than in the case where $q^H > \pi > q^L$. We therefore proceed to establish the following proposition.

Lemma 2.4. *For any $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ such that $q^H > \pi > q^L$ holds, the seller always prefers $\tau = H$ over $\tau = A$. That is, $u^H_S(\sigma) \geq u^A_S(\sigma)$ holds for any $\sigma \in [0, 1)$.*

Proof. See Appendix 2.A.2. □

The subsequent proposition encapsulates the core result derived from the preceding discussion.

Proposition 2.17. *Suppose the following conditions hold: $\sigma \leq \sigma^*_{CA}$, so that the seller optimally selects $\tau^* = A$ in the second period if $q_1 = q^L$ is chosen, and $\frac{\pi}{1-\pi} \in \left(1, \frac{q^L}{1-q^L} \cdot \frac{q^H}{1-q^H}\right]$, ensuring that each pricing strategy $\tau \in \{H, A\}$ satisfies both the participation and incentive compatibility constraints of the buyer. Then, the optimal pricing strategy of the seller in the second period is characterised as follows:*

(i) $\tau^* = H$*, if* $q^H > \pi > q^L$*, or,* $\sigma \leq \sigma^*_{HA}$*;*

(ii) $\tau^* = A$*, otherwise,*

where,

$$\begin{aligned}
\left\{\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right\} &\in \left\{\max\left\{q^H - \pi, 0\right\}\right\} \times \left(\max\left\{q^L - \pi\right\}, +\infty\right), \ \ \textit{if} \ \ \tau = H; \\
\left\{\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right\} &\in \left(\max\left\{q^H - \pi\right\}, +\infty\right) \times \left\{\max\left\{q^L - \pi, 0\right\}\right\}, \ \ \textit{if} \ \ \tau = A.
\end{aligned}$$

The analysis reveals that risk aversion fundamentally alters the strategic calculus in dynamic information markets. Whilst risk-neutral sellers unambiguously prefer high-type only strategies, risk-averse sellers face genuine trade-offs between expected returns and revenue certainty. The existence of cut-off points σ^*_{HA} demonstrates that moderate risk aversion can reverse optimal strategies, with sellers increasingly favouring aggressive pricing approaches that sacrifice per-unit margins for higher probability of positive outcomes. The configuration of prior beliefs relative to signal precisions determines whether these trade-offs create meaningful strategic choices or maintain the dominance of high-type only approaches, illustrating how risk preferences interact with market fundamentals to shape information provision strategies.

High Risk Aversion: High-Type Only versus Conservative Strategies. In the case of $\sigma \geq \sigma^*_{CA}$, there are two profitable candidates for pricing strategies: high-type only pricing strategy, and conservative pricing strategy.

We start the discussion with a risk-neutral seller (i.e., $\sigma = 0$). The seller always prefers the high-type only pricing strategy, which strictly dominates the conservative pricing strategy. The advantage of selling the high-type signal in $t = 1$ is the greater variance in signal realisation, which generates more surplus to be extracted. In contrast, the advantage of selling the low-type signal under the conservative pricing strategy lies in the higher probability of a second-period sale. However, the expected surplus is lower. Given that the seller exhibits risk-neutral behaviour, they disregard risk and focus solely on the magnitude of expected surplus. Consequently, the high-type only strategy is strictly preferred. Moreover, the conservative pricing strategy dominates the aggressive pricing strategy. The reason is straightforward: although both strategies generate the same ex-post surplus, the seller captures the full surplus under the former, whereas they forgo part of the surplus (i.e., when the signal realisation is $s_1 = 1$) under the latter. To summarise, the following ordering holds if $\sigma = 0$: $H \succ C \succ A$. However, with sufficient risk aversion, the seller changes their preference. First, consider the comparison between H and C. A sufficiently risk-averse seller prefers C over H. For any $(\pi, q^H, q^L) \in \left(\frac{1}{2}, 1\right)^3$ such that $q^H > \max\{q^L, \pi\}$, there exists a cut-off level $\sigma^*_{HC}(\pi, q^H, q^L)$ at which the seller is indifferent between the two strategies. If $\sigma > \sigma^*_{HC}$, the seller is sufficiently risk-averse and therefore chooses to reduce risk exposure at the cost of some expected surplus by implementing the conservative pricing strategy. The ex-ante expected payoff of the seller implementing the conservative pricing strategy is:

$$
\begin{aligned}
u_S^C(\sigma) &\equiv \mathbb{E}[u_S(\tau = C) \mid \sigma] \\
&= \left[1\right] \left[\varphi_C(q^H, q^L)\right]^{1-\sigma} \\
&+ \left[\max\left\{q^L - \pi, 0\right\} + \left[\pi(1-q^L) + (1-\pi)q^L\right] \left(\varphi_A(q^H, q^L) - \varphi_C(q^H, q^L)\right)\right]^{1-\sigma}.
\end{aligned}
$$

Lemma 2.5. *For any $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ such that $q^H > \max\{\pi, q^L\}$ holds, there exists a unique cut-off point $\sigma^*_{HC} \in (0, 1)$ such that:*

(i) the seller is indifferent between $\tau = H$ and $\tau = C$ if $\sigma = \sigma^*_{HC}$. That is, $u_S^H(\sigma^*_{HC}) = u_S^C(\sigma^*_{HC})$ holds at the cut-off point;

(ii) the seller prefers $\tau = H$ over $\tau = C$ when they exhibit sufficiently low risk aversion. That is, $u_S^H(\sigma) \geq u_S^C(\sigma)$ holds if and only if $\sigma \in [0, \sigma^*_{HC}]$.

Proof. See Appendix 2.A.2. □

The subsequent proposition encapsulates the core result derived from the preceding discussion.

Proposition 2.18. *Suppose the following conditions hold: $\sigma \geq \sigma^*_{CA}$, so that the seller optimally selects $\tau^* = C$ in the second period if $q_1 = q^L$ is chosen, and, $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \frac{q^H}{1-q^H}\right]$, ensuring that each pricing strategy $\tau \in \{H, C\}$ satisfies both the participation and incentive compatibility constraints of the buyer. Then, the optimal pricing strategy o the seller in the second period is characterised as follows:*

(i) $\tau^* = H$, if $\sigma \leq \sigma^*_{HC}$;

(ii) $\tau^* = C$, *otherwise.*

where,

$$\left\{\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right\} \in \left\{\max\left\{q^H - \pi, 0\right\}\right\} \times \left(\max\left\{q^L - \pi\right\}, +\infty\right), \ \textit{if } \tau = H;$$

$$\left\{\mathcal{P}_1(q^H), \mathcal{P}_1(q^L)\right\} \in \left(\max\left\{q^H - \pi\right\}, +\infty\right) \times \left\{\max\left\{q^L - \pi, 0\right\} + \left(\varphi_A(q^H, q^L) - \varphi_C(q^H, q^L)\right)\right\}, \ \textit{if } \tau = C.$$

The analysis demonstrates that high risk aversion fundamentally reshapes strategic preferences in favour of conservative pricing approaches. The existence of cut-off point σ^*_{HC} reveals that sufficiently risk-averse sellers will abandon the high-expected-return strategy of offering only high-type signals in favour of conservative approaches that guarantee universal participation and smoother revenue streams. Unlike the low risk aversion case where aggressive strategies compete with high-type only approaches, high risk aversion creates a starker choice between profit maximisation and risk management, with conservative pricing emerging as the dominant strategy when uncertainty aversion becomes sufficiently pronounced.

2.5 Welfare Implication

The introduction of risk aversion into dynamic information markets creates profound welfare implications that extend far beyond the immediate effects on pricing strategies and market participation. Whilst the previous analysis focused on how risk preferences shape equilibrium outcomes, this section examines the broader welfare consequences of these strategic changes, addressing fundamental questions about market efficiency, surplus distribution, and the desirability of different information environments. The analysis reveals that risk aversion generates complex trade-offs between static efficiency and dynamic stability, with implications for both market participants and potential regulatory interventions. Three key welfare dimensions emerge as particularly important: the optimal design of information mechanisms under risk aversion, the comparison between second-best outcomes under information asymmetry and first-best outcomes under full information, and the role of commitment devices in enabling revenue smoothing through dynamic pricing. Each dimension illuminates different aspects of how risk preferences interact with information asymmetries to shape market outcomes, revealing that the welfare effects of risk aversion prove far more nuanced than simple efficiency losses might suggest.

2.5.1 Global Optimum of Information Mechanism Design

The introduction of risk aversion fundamentally transforms the information mechanism design problem, creating rich interactions between signal quality choices and pricing strategies that were absent under risk neutrality. The analysis examines how a risk-averse seller should optimally design the information structure, addressing the central question of whether the drive for higher-quality signals observed under risk neutrality persists when sellers must balance expected returns against revenue volatility.

Under risk neutrality, as demonstrated in the previous chapter, the mechanism design problem yields a stark and elegant result: the seller should always offer the highest possible signal quality $q^H = 1$ by implementing a high-type only pricing strategy. The conclusion emerged from straightforward optimisation where expected payoff increased monotonically with signal precision, making perfect signals the dominant choice. The underlying economic logic was compelling: better signals enable more precise rent extraction without imposing any costs on the risk-neutral seller, who remains indifferent to the variability of outcomes.

Risk aversion disrupts this simple relationship by introducing preferences for revenue smoothing that can conflict with profit maximisation. A risk-averse seller may deliberately choose lower-quality signals if doing so generates more predictable revenue streams, even when high-quality alternatives offer superior expected returns. Such preferences create a fundamental tension between the discriminatory power of information and the financial preferences of the provider regarding risk and uncertainty.

The analysis in Section 2.4.5 revealed that each of the three pricing strategies (i.e., high-type only, aggressive, and conservative) emerge as optimal under different parametric conditions, depending on the degree of risk aversion and the strength of buyer prior beliefs. The multiplicity of equilibrium strategies suggests that the information design problem has become considerably more complex. The seller now considers not only which signal qualities maximise expected profits, but also how these choices interact with the risk profile of different pricing strategies and the resulting implications for market structure and participation.

The key insight is that risk aversion can reverse the quality-maximisation imperative that characterises the risk-neutral case. When sellers prioritise revenue stability over expected returns, they may prefer information structures which support more predictable trading patterns, even if these structures sacrifice some precision and rent extraction capability. Such fundamental shift in priorities transforms information mechanism design from a pure optimisation problem into a sophisticated risk management exercise, where quality choices must be evaluated alongside their implications for market dynamics and revenue uncertainty.

Optimal Signal Quality Under Aggressive and Conservative Strategies. We first investigate the set of equilibria associated with optimal pricing strategies of either A or C. Suppose at the global optimum, $\tau^* \in \{A, C\}$, at which the seller chooses to induce the buyer to acquire a low-type signal in the first period, followed by a high-type signal in the second period. Under these conditions, the seller chooses $q^{H^*} = 1$, since the expected utility functions are increasing in q^H. The economic intuition is straightforward. The probability distribution regarding the signal realisation $s_1 = 1$ is independent of q^H. The uncertainty arises only from q^L. The precision of the high-type signal, q^H, only determines the magnitude of the ex-post surplus extractable in the second period. Consequently, without participation constraints, the seller has an incentive to set q^H as high as possible.

We now examine whether $q^H = 1$ satisfies the buyer participation constraint when pricing strategy A or C is implemented. Fortunately, the answer is positive. When the seller sets the signal precision as $q_2 = q^H = 1$, the buyer always has incentive to acquire it as long as their posterior belief before the acquisition is not one (i.e., $\mu_1(\cdot, \cdot) < 1$), which always holds given the prior belief and the signal precision in $t = 1$ is not one (i.e., $(\pi, q_1) \in (\frac{1}{2}, 1)$). The economic logic is compelling: as the buyer is uncertain about the state at $t = 0$ (i.e., $\pi \neq 1$), and such uncertainty cannot be resolved by the signal in the first period (i.e., $q_1 \neq 1$), a perfect signal in the second period (i.e., $q_2 = q^H = 1$) always further enhances the buyer payoff, from which the seller can extract surplus.

Lemma 2.6. *Given $q^H \to 1$, $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L} \frac{q^H}{1-q^H}\right]$ holds.*

Technically, $\frac{q^H}{1-q^H}$ will diverge towards positive infinity as $q^H \to 1$, which indicates that any prior belief (i.e., π) satisfies the participation constraint mentioned above. To sum up, it is necessary that the seller chooses $q^{H^*} = 1$ to maximise their ex-ante expected utility, if either A or C is implemented under equilibrium.

Lemma 2.7. $\tau^* \in \{A, C\} \implies q^{H^*} = 1$. *If the optimal strategy is A or C, the seller must offer a perfect signal, $q^{H^*} = 1$, in the second period at the optimum.*

Preference Between Aggressive and Conservative Strategies. We now investigate the preference between the aggressive pricing strategy and the conservative pricing strategy with restrictive attention on the cases in which $q^H = 1$ is chosen. Note that the two strategies share the same expected utility level if $q^H = q^L = 1$ is chosen: $u_S(C; 1, 1) = u_S(A; 1, 1) = (1 - \pi)^{1-\sigma}$. However, the following lemma shows that the aggressive pricing strategy is not sustainable in the second period, given any $q^L \in (\frac{1}{2}, 1)$ and $q^H = 1$. The economic intuition is that when $q^H = 1$, the conservative strategy guarantees full surplus extraction with certainty, whilst the aggressive strategy creates the same surplus only probabilistically. For risk-averse sellers, the certainty of conservative pricing dominates the uncertain but potentially equivalent returns of aggressive pricing.

Lemma 2.8. $q^H = 1 \implies C \succ A$, $\forall q^L \in (\frac{1}{2}, 1)$. *Given that* $q^H = 1$ *holds, C is preferred in the second period, regardless of the choice on* q^L. *Thus,* $\tau^* \neq A$.

Proof. The following inequality always holds given any $q^L \in (\frac{1}{2}, 1)$:

$$\left[\varphi_C(q^H = 1, q^L)\right]^{1-\sigma} \geq \left[\pi(1 - q^L) + (1 - \pi)q^L\right] \left[\varphi_A(q^H = 1, q^L)\right]^{1-\sigma}.$$

□

Elimination of High-Type Only Strategies. We now demonstrate that any high-type only pricing strategy is never optimal. Suppose at the global optimum, H is implemented with the choice of signal type $q^H = q^{H^*}$. The expected utility of the seller is:

$$\begin{aligned}\mathbb{E}[u_S(H; q^{H^*})] = &\left[\left[\pi(1 - q^{H^*}) + (1 - \pi)q^{H^*}\right] \left[q^{H^*} - \max\left\{\mu_1(q^{H^*}, 1), 1 - \mu_1(q^{H^*}, 1)\right\}\right]\right]^{1-\sigma} \\ &+ \left[\max\left\{q^{H^*} - \pi, 0\right\}\right]^{1-\sigma}.\end{aligned}$$

The seller is always strictly better off if they opt for the following alternative strategy: $\tau = A$ with the choice of signal types $(q^H, q^L) = (1, q^{H^*})$. The implied expected utility is:

$$\begin{aligned}\mathbb{E}[u_S(A; (1, q^{H^*}))] = &\left[\left[\pi(1 - q^{H^*}) + (1 - \pi)q^{H^*}\right] \left[1 - \max\left\{\mu_1(q^{H^*}, 1), 1 - \mu_1(q^{H^*}, 1)\right\}\right]\right]^{1-\sigma} \\ &+ \left[\max\left\{q^{H^*} - \pi, 0\right\}\right]^{1-\sigma},\end{aligned}$$

which is strictly greater than $\mathbb{E}[u_S(H; q^{H^*})]$. The economic intuition is straightforward: the seller relies on the signal type of q^{H^*} in both periods if H is implemented. However, in the second period, the seller always has incentive to enhance the precision to one (i.e., $q_2 = 1$) for a better payoff, as q_2 does not affect the probability distribution on ex-post payoff realisation. However, by Lemma 2.8, it can be inferred that $\tau = C$ with $(q^H, q^L) = (1, q^{H^*})$ generates a higher payoff, which is sustainable equilibrium strategy proven by Lemma 2.8 itself.

Lemma 2.9. *Given any $\pi \in (\frac{1}{2}, 1)$ and $\sigma \in [0, 1)$, $\tau^* \neq H$.*

Proof. By Lemma 2.8, the following inequality holds given any $q^L = q^{H^*}$ which implies the desired result:

$$\Big[\varphi_C(q^H = 1, q^L = q^{H^*})\Big]^{1-\sigma} \geq \Big[\pi(1 - q^L) + (1 - \pi)q^L\Big]\Big[\varphi_A(q^H = 1, q^L)\Big]^{1-\sigma}.$$

□

The Optimal Mechanism Design. According to Lemma 2.8 and 2.9, it can be inferred that the conservative pricing strategy is optimal, $\tau^* = C$, associated with the choice $q^{H^*} = 1$.

Proposition 2.19. *Given any $\pi \in (\frac{1}{2}, 1)$ and $\sigma \in [0, 1)$, the global optimality suggests the conservative pricing strategy $\tau^* = C$ with the following signal precision:*

$$\begin{aligned} q^{H^*} = 1; q^{L^*} = \arg\max_{q^L} \Big\{ &\Big[\varphi_C(1, q^L)\Big]^{1-\sigma} \\ &+ \Big[\max\Big\{q^L - \pi, 0\Big\} + \Big[\pi(1 - q^L) + (1 - \pi)q^L\Big]\Big(\varphi_A(1, q^L) - \varphi_C(1, q^L)\Big)\Big]^{1-\sigma}\Big\}. \end{aligned}$$

The result reveals a striking conclusion: despite the complexity introduced by risk aversion, the optimal mechanism design converges to a conservative pricing strategy with perfect second-period signals. Risk aversion does not lead to quality degradation in the conventional sense, but rather to a preference for pricing strategies that guarantee universal participation and predictable revenue streams.

Single Signal Type Scenario

This section discusses the case when the seller is restricted to use only one signal type. If so, we have a case equivalent to the one in which the seller implements a high-type only pricing strategy. Unlike the previous chapter in which the seller is assumed risk neutral and thus optimality suggests $q^H = 1$, $q^H = 1$ is no longer necessary if the preference of the seller exhibits sufficient risk aversion. The seller may have an incentive to reduce q^H in order to enhance the probability of a trade in the second period. More specifically, the seller understands that positive information rent in the second period exists only if the signal realisation in $t = 1$ is against the status quo (i.e., $s_1 = 1$). Therefore, there exists a trade-off between the probability of trade (decreasing in q^H) and the ex-post utility (increasing in q^H). The seller maximises their expected utility, u_S^H, by choosing an optimal signal type q^H. The choice regarding q^L is irrelevant, as long as it is lower than q^{H^*}. Consequently, we have the following proposition:

Proposition 2.20. *If the seller is restricted to use only one signal type, $q^{H^*} = 1$ is no longer optimal.*

Proof. Proof by contradiction. Suppose that $q^H = 1$ is implemented. The implied expected utility of the seller is:

$$u_S^H(q^H \to 1) = \left[\pi(1 - q^H) + (1 - \pi)q^H\right] \left[\varphi_H(q^H)\right]^{1-\sigma} + \left[q^H - \pi\right]^{1-\sigma}.$$

Given that $\varphi_H(q^H = 1) = 0$, the expected utility of the seller converges to, $\left[q^H - \pi\right]^{1-\sigma}$. Consider the alternative pricing strategy, $\widetilde{q^H} = q^H - \varepsilon$, where $\varepsilon \to 0^+$. The implied expected utility of the seller is higher due to the positively infinite marginal utility effect from (nearly) zero to strictly positive information rent:

$$\begin{aligned}
u_S^H(\widetilde{q^H}) &= \left[\pi(1 - \widetilde{q^H}) + (1 - \pi)\widetilde{q^H}\right] \left[\varphi_H(\widetilde{q^H})\right]^{1-\sigma} + \left[\widetilde{q^H} - \pi\right]^{1-\sigma} \\
&= \left[\pi(1 - \widetilde{q^H}) + (1 - \pi)\widetilde{q^H}\right] \left[\varphi_H(q^H) + \delta\right]^{1-\sigma} + \left[\widetilde{q^H} - \pi\right]^{1-\sigma} \\
&> \left[\pi(1 - \widetilde{q^H}) + (1 - \pi)\widetilde{q^H}\right] \left[\varphi_H(q^H)\right]^{1-\sigma} + \left[\widetilde{q^H} - \pi\right]^{1-\sigma} \\
&\approx \left[\pi(1 - q^H) + (1 - \pi)q^H\right] \left[\varphi_H(q^H)\right]^{1-\sigma} + \left[q^H - \pi\right]^{1-\sigma} \\
&= u_S^H(q^H).
\end{aligned}$$

□

2.5.2 Information Transparency and Welfare Effects

This section directly addresses the applicability concern by examining a fundamental hypothetical: what would happen if sellers could observe first-period signal realisations? This section examines whether risk-averse sellers benefit from information transparency by comparing outcomes when signal realisations are observable versus unobservable. The analysis contrasts the equilibrium under information asymmetry, where sellers cannot observe buyer signal realisations, with the hypothetical first-best benchmark where sellers possess complete information about buyer types. Whilst conventional wisdom might suggest that additional information always benefits decision-makers, the results reveal a more nuanced relationship. For risk-averse sellers, the inability to observe signal realisations can paradoxically provide welfare benefits by enabling commitment to revenue-smoothing strategies that would prove unsustainable under full information transparency.

When the prior belief is sufficiently extreme (i.e., $\frac{\pi}{1-\pi} > \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}$) such that the buyer has no incentive to acquire any signal type in $t = 2$ when $s_1 = 0$ is realised, the seller cannot extract surplus from buyers who observe $s_1 = 0$. The seller obtains surplus of $\max\{q_1 - \pi, 0\}$ in the first period and ex-post surplus of $\varphi_{\tau^*}(\cdot)$ with probability $[\pi(1 - q_1) + (1 - \pi)q_1]$ in the second period, where $q_1 = q^H$ if $\tau^* = H$ and $q_1 = q^L$ if $\tau^* = A$. The introduction of signal realisation transparency does not affect optimality, which implies the outcome is equivalent to one under information transparency.

When the prior belief is sufficiently moderate (i.e., $\frac{\pi}{1-\pi} \leq \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}$), aggressive and conservative pricing strategies are no longer sustainable under full information transparency. Strictly positive information rent now exists when $s_1 = 0$ is realised from a low-type signal acquisition, which the seller can extract. The question becomes whether such transparency benefits the seller.

If $\tau^* = H$ is optimally chosen under information asymmetry, the seller will not benefit from signal transparency as zero information rent can be found under the realisation $s_1 = 0$. If $\tau^* = A$ is optimally chosen under information asymmetry, the seller unambiguously benefits from information transparency as they can now extract the information rent from buyers with $s_1 = 0$ that was previously inaccessible due to information asymmetry. If $\tau^* = C$ is optimally chosen under information asymmetry, the answer is ambiguous. Under information transparency, the seller will set the price of a high-type signal in the second period as $\varphi_A(q^H, q^L)$ instead of $\varphi_C(q^H, q^L)$. However, the total expected surplus is not increased, as the seller now forgoes the expected future benefit, $V(q^L)$, collected in the first period. Effectively, the seller is forced to reallocate $V(q^L)$ towards the second period. Now we examine whether the seller is better off with the compulsory surplus transfer.

First examine the case in which $q^H > \pi > q^L$. Under the environment of information asymmetry (IA), the seller obtains a surplus of $V(q^L)$ in $t = 1$, and an ex-post surplus of $\varphi_C(q^H, q^L)$ with probability of one in $t = 2$. Under the information environment of information transparency (IT), the seller obtains zero surplus in period $t = 1$, and, an ex-post surplus of $\varphi_A(q^H, q^L)$ with probability of $\pi(1 - q^L) + (1 - \pi)q^L$; $\varphi_C(q^H, q^L)$ with probability of $\pi q^L + (1 - \pi)(1 - q^L)$ in $t = 2$, which makes the seller worse off. Formally, we have:

$$
\begin{aligned}
u_S(C; IA) &= (V(q^L))^{1-\sigma} + [1](q^H - \mu_1(q^L, 0))^{1-\sigma} \\
&= (V(q^L))^{1-\sigma} + [\pi(1 - q^L) + (1 - \pi)q^L](q^H - \mu_1^0)^{1-\sigma} \\
&+ [\pi q^L + (1 - \pi)(1 - q^L)](q^H - \mu_1^0)^{1-\sigma} \\
&> 0^{1-\sigma} + [\pi(1 - q^L) + (1 - \pi)q^L](q^H - \mu_1^1)^{1-\sigma} + [\pi q^L + (1 - \pi)(1 - q^L)](q^H - \mu_1^0)^{1-\sigma},
\end{aligned}
$$

where $\mu_1^{s_1} \equiv \mu_1(q^L, s_1)$. Intuitively, the reallocation of $V(q^L)$ from $t = 1$ to $t = 2$ makes the seller worse off since they no longer benefit from the highest possible marginal utility at zero.

Second, assume that $q^H > q^L > \pi$. Under information asymmetry, the seller obtains surplus of $(q^L - \pi) + V(q^L)$ in period $t = 1$ and $\varphi_C(q^H, q^L)$ with certainty in period $t = 2$. Under information transparency, the seller obtains surplus of $(q^L - \pi)$ in period $t = 1$ and state-contingent surplus in period $t = 2$. Unlike the previous case, the welfare comparison is ambiguous. If q^L is sufficiently close to q^H, the seller benefits from the information environment of information transparency, enjoying higher expected utility. Two factors drive this result. First, when q^L is higher, the marginal utility effect of $V(q^L)$ in period $t = 1$ weakens since the base consumption $q^L - \pi$ is greater, reducing the seller's sensitivity to additional surplus in that period. Second, higher q^L implies lower information rent when $s_1 = 0$ (i.e., $\varphi_C(q^H, q^L)$ decreases), which enhances the marginal utility effect in period $t = 2$. Consequently, sellers facing sufficiently high q^L may prefer reallocating first-period surplus towards the second period, which information transparency facilitates by enabling state-contingent pricing. We therefore state the following proposition:

Proposition 2.21. *Given that* $\frac{\pi}{1-\pi} \in \left(1, \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}\right]$ *holds:*

(i) if $\tau = H$ *is implemented under information asymmetry, the seller is indifferent to information*

transparency;

(ii) if $\tau = A$ is implemented under information asymmetry, the seller is better off under information transparency;

(iii) if $\tau = C$ is implemented under information asymmetry, the seller is worse off under information transparency when the low-type signal is sufficiently informative (i.e., q^L is sufficiently high).

These results reveal the complex relationship between information transparency and welfare in dynamic markets with risk-averse sellers. Whilst transparency typically benefits aggressive sellers by enabling state-contingent pricing, it can harm conservative sellers by forcing them to reallocate surplus across periods in ways that reduce marginal utility. The analysis demonstrates that information asymmetry, typically viewed as a market imperfection, can paradoxically serve as a commitment device that enables risk-averse sellers to implement revenue-smoothing strategies. Such findings challenge conventional policy prescriptions favouring transparency, suggesting that optimal regulatory approaches must account for the risk preferences of market participants and the strategic value of controlled information environments.

For completeness, the technical characterisation of equilibrium strategies under the information environment of information transparency is provided below. In summary, there are two equilibrium classes under the information environment of information transparency (IT), given $\frac{\pi}{1-\pi} \leq \frac{1-q^L}{q^L}\frac{q^H}{1-q^H}$ holds:

(i) $\big(\mathcal{P}_1(q^H), \mathcal{P}_2(q^H, q^H)\big) = \big(q^H - \pi, q^H - \mu_1(q^H, s_1)\big)$, with $(q_1, q_2) = (q^H, q^H)$, equivalent to $\tau = H$;

(ii) $\big(\mathcal{P}_1(q^L), \mathcal{P}_2(q^L, q^H)\big) = \big(\max\{q^L - \pi, 0\}, q^H - \max\{\mu_1(q^L, s_1), 1 - \mu_1(q^L, s_1)\}\big)$, with $(q_1, q_2) = (q^L, q^H)$.

Let $\tau = IT_H$ and $\tau = IT_L$ denote the strategies described in (i) and (ii), respectively. The seller prefers $\tau = FB_L$ under information transparency when $\tau \in \{A, C\}$ is optimal under information asymmetry. However, when $\tau = H$ is optimal under information asymmetry, whether the seller prefers $\tau = IT_H$ over $\tau = IT_L$ under information transparency remains ambiguous.

2.5.3 Commitment and Revenue Smoothing Through Dynamic Pricing

Commitment mechanisms may enhance the welfare of the seller when there exists scope for intertemporal revenue reallocation, particularly when the seller faces unequal marginal expected utility across periods. This opportunity arises most prominently when the seller anticipates zero surplus in the first period, a situation that occurs when the prior belief which the buyer possesses is sufficiently extreme that no signal commands a strictly positive price in the first period. Under such circumstances, the seller faces a stark asymmetry: substantial expected revenue in the second period but no immediate income. Risk aversion amplifies the cost of revenue concentration, as the marginal utility of second-period income diminishes whilst the seller receives no utility from first-period operations.

A commitment device enables the seller to smooth this revenue profile by pre-selling future information services. Specifically, the seller can commit to offering second-period signals at lower prices in exchange for higher first-period payments. Effectively, the arrangement functions as an implicit loan from the buyer to the seller, with the committed repayment taking the form of discounted future information access. This commitment proves sustainable because it aligns with the intertemporal optimisation of the buyer. The buyer recognises that accepting higher upfront costs in exchange for guaranteed future access represents good value, particularly when the alternative involves uncertain future pricing. The arrangement satisfies both participation and incentive compatibility constraints precisely because both parties benefit from the revenue smoothing. However, the reverse arrangement, where the seller would prefer to shift first-period surplus to the second period, proves infeasible under the assumption of non-commitment of the buyer. Since the buyer retains the option to exit in the second period, they will refuse any arrangement where second-period prices exceed the corresponding information rent. The asymmetry in feasible commitment directions reflects the fundamental difference in bargaining power between periods.

The mechanism effectively allows the seller to engage in consumption smoothing through the pricing structure itself, transforming volatile information rents into a more stable revenue stream that better suits risk-averse preferences. Rather than accepting the uncertainty inherent in period-by-period pricing, the seller can use commitment to create a more predictable income flow that reduces the welfare costs associated with revenue volatility. The following proposition formalises these insights about optimal commitment mechanisms and their welfare properties.

Proposition 2.22 (Revenue Smoothing Through Commitment)**.** *Consider a risk-averse seller with CRRA utility parameter $\sigma \in (0,1)$ facing the parameter regime where $\frac{\pi}{1-\pi} \in \left(\left(\frac{q^H}{1-q^H}\right)^2, +\infty\right)$, such that $\mathcal{P}_1(q^H) = \mathcal{P}_1(q^L) = 0$. Define a commitment mechanism, $(\mathcal{P}_1^C, \mathcal{P}_2^C)$, where the seller commits to* *second-period pricing before the first-period decision of the buyer. Then:*

(i) *Forward commitment is feasible: there exist prices* $(\mathcal{P}_1^C(q^H), \mathcal{P}_2^C(q^H, q^H))$ *with* $\mathcal{P}_1^C(q^H) > 0$ *and* $\mathcal{P}_2^C(q^H, q^H) < \varphi_H(q^H)$ *such that both participation constraint and incentive compatibility hold.*

(ii) *Optimality condition: the optimal commitment satisfies,*

$$\lambda_1(\mathcal{P}_1^C(\cdot))^{-\sigma} = \lambda_2 \mathbb{E}\left[(\mathcal{P}_2^C(\cdot))^{-\sigma}\right],$$

where λ_1, λ_2 *are the marginal utilities of consumption in periods* $t = 1$ *and* $t = 2$*, respectively.*

2.6 Conclusion

This analysis demonstrates that the introduction of risk aversion fundamentally transforms the strategic landscape of dynamic information markets, revealing that the drive for higher-quality signals observed under risk neutrality need not persist when sellers must balance expected returns against revenue volatility. The central insight emerges from the tension between profit maximisation and revenue smoothing: whilst risk-neutral sellers focus exclusively on extracting maximum surplus through high-precision signals that enable perfect market segmentation, risk-averse sellers may rationally sacrifice expected profits in favour of more stable revenue streams.

The findings reveal three key departures from the risk-neutral benchmark. First, risk-averse sellers do not necessarily prefer signals with higher precision, as they may optimally choose to induce buyers to acquire less precise signals that increase the likelihood of securing information rent through reduced outcome volatility. Second, conservative pricing strategies that reallocate surplus across periods through commitment mechanisms need not dominate aggressive pricing approaches, since diminishing marginal utility can make sellers prefer strategies that accept lower expected surplus in exchange for higher marginal utility from concentrated payoffs. Third, information asymmetry can paradoxically benefit risk-averse sellers in certain parameter regions, as the inability to observe signal realisations enables commitment to revenue smoothing strategies that would be unsustainable under full information.

This preference for consumption smoothing can reverse established dominance relationships, potentially favouring conservative pricing strategies that guarantee participation over aggressive approaches that maximise expected returns. The resulting equilibrium reflects a sophisticated balance between information quality, market participation, and risk management that captures essential features of real-world information markets where providers must consider both profitability and operational stability in their strategic decisions. The paper thus contributes to our understanding of information provision by showing that seller risk preferences constitute a crucial determinant of market structure, signal quality, and pricing strategies, with important implications for regulatory policy and market design.

Several avenues for future research emerge from this analysis. First, introducing buyer heterogeneity in risk preferences or prior beliefs would reveal how sellers optimally segment markets and whether risk-averse providers naturally select particular client types. Second, competition amongst multiple information sellers could illuminate whether market forces amplify or mitigate the effects of risk aversion on information quality and pricing strategies. Third, extending the framework to multi-period horizons would enable examination of reputation building and learning dynamics, particularly how risk-averse sellers balance short-term revenue smoothing against long-term market positioning. Fourth, incorporating alternative risk preference specifications such as ambiguity aversion or loss aversion could provide deeper insights into how different forms of uncertainty affect information provision decisions. Finally, empirical work calibrating the model using data from real information markets such as financial research services, consulting firms, or digital information platforms would provide valuable evidence on the quantitative importance of these mechanisms and guide further theoretical development. These extensions would deepen our understanding of how risk preferences shape information market dynamics whilst connecting the theoretical insights to observable market phenomena.

Appendix 2.A: Omitted Proofs

2.A.1 Proofs of Section 4.3

Proof of Proposition 2.12. Given $q^H = \mu_1(q^L, 0)$, we have:

$$\begin{aligned} 0^{1-\sigma} &< \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\mu_1(q^L,0) - \max\{\mu_1(q^L,1), 1-\mu_1(q^L,1)\}\right]^{1-\sigma} \\ \iff &\left[\varphi_C(q^H,q^L)\right]^{1-\sigma} - \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^H,q^L)\right]^{1-\sigma} < 0. \end{aligned}$$

Given $q^H \to +\infty$, by Mean Value Theorem, we have: $\left[\varphi_A(q^H,q^L)\right]^{1-\sigma} - \left[\varphi_C(q^H,q^L)\right]^{1-\sigma} \to 0$, which implies:

$$\begin{aligned} &\left[\varphi_A(q^H,q^L)\right]^{1-\sigma} - \left[\varphi_C(q^H,q^L)\right]^{1-\sigma} - \left[\pi q^L + (1-\pi)(1-q^L)\right] \left[\varphi_A(q^H,q^L)\right]^{1-\sigma} < 0 \\ \iff &\left[\varphi_C(q^H,q^L)\right]^{1-\sigma} - \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^H,q^L)\right]^{1-\sigma} > 0. \end{aligned}$$

By Intermediate Value Theorem, there exists a cut-off point at which the seller is indifferent between the two strategies, denoted $q^{H^*} \in (\mu_1(q^L,0), +\infty)$ such that:

$$\left[\varphi_C(q^{H^*},q^L)\right]^{1-\sigma} - \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^{H^*},q^L)\right]^{1-\sigma} = 0.$$

Take natural log on both sides, and we have: $(1-\sigma)\log\varphi_C(\cdot) = \log[\cdot] + (1-\sigma)\log\varphi_A(\cdot)$. Take the total derivative on both side, and we have:

$$\begin{aligned} [-\log\varphi_C]\, d\sigma + \frac{1}{\varphi_C}\frac{\partial\varphi_C}{\partial q^{H^*}} dq^{H^*} &= [-\log\varphi_A]\, d\sigma + \frac{1}{\varphi_A}\frac{\partial\varphi_A}{\partial q^{H^*}} dq^{H^*} \\ \implies [-\log\varphi_C]\, d\sigma + \frac{1}{\varphi_C} dq^{H^*} &= [-\log\varphi_A]\, d\sigma + \frac{1}{\varphi_A} dq^{H^*} \\ \implies \frac{dq^{H^*}}{d\sigma} = \frac{\log\varphi_A - \log\varphi_C}{\frac{1}{\varphi_A} - \frac{1}{\varphi_C}} &= \frac{\log\frac{\varphi_A}{\varphi_C}}{\frac{\varphi_C - \varphi_A}{\varphi_A\varphi_C}} < 0. \end{aligned}$$

□

Proof of Proposition 2.13. First define that:

$$\Delta_{CA}(\sigma) \equiv \left[1\right] \left[\varphi_C(q^H, q^L)\right]^{1-\sigma} - \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^H, q^L)\right]^{1-\sigma}.$$

Let φ_C and φ_A denote $\varphi_C(q^H, q^L)$ and $\varphi_A(q^H, q^L)$, respectively. The first-order derivative with respect to σ is:

$$\frac{\partial \Delta_{CA}}{\partial \sigma} = \varphi_C^{1-\sigma} \log \varphi_C \,(-1) - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log \varphi_A \,(-1)\,.$$

The first-order condition holds if and only if:

$$\varphi_C^{1-\sigma} \log \varphi_C = \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log \varphi_A.$$

The second-order derivative with respect to σ is:

$$\frac{\partial^2 \Delta_{CA}}{\partial \sigma^2} = \varphi_C^{1-\sigma} \log^2 \varphi_C - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log^2 \varphi_A.$$

The following result demonstrates that the corresponding second-order derivative at any critical point is positive, thereby implying that $\sigma = \sigma^{FOC}$ constitutes a local minimum:

$$\begin{aligned}
\frac{\partial^2 \Delta_{CA}}{\partial \sigma^2}(\sigma^{FOC}) &= \varphi_C^{1-\sigma} \log^2 \varphi_C \\
&= \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log \varphi_A \log \varphi_C \\
&> \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log^2 \varphi_A,
\end{aligned}$$

which suggests that $\sigma = \sigma^{FOC}$ is a local minimum. Given that any critical point must correspond to a local minimum, the function admits at most one such point. Suppose $\Delta_{CA}(\sigma = 0) < 0$. Combined with the fact that $\Delta_{CA}(\sigma \to 1) > 0$, it follows that the cut-off point is unique. Suppose $\Delta_{CA}(\sigma = 0) > 0$. Then, the first-order derivative at $\sigma = 0$ is positive, since we have:

$$\begin{aligned}
\frac{\partial \Delta_{CA}}{\partial \sigma}(0) &= \varphi_C \log \varphi_C \,(-1) - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A \log \varphi_A \,(-1) \\
&> \varphi_C \log \varphi_A \,(-1) - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A \log \varphi_A \,(-1) \\
&= \log \varphi_A \,(-1) \left\{\varphi_C - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A\right\} \\
&= \log \varphi_A \,(-1)\, \Delta_{CA}(\sigma = 0) > 0.
\end{aligned}$$

Combined with the facts that $\Delta_{CA}(\sigma \to 1) > 0$, it follows that the cut-off point is unique. $\square$

2.A.2 Proofs of Section 4.5

Proof of Lemma 2.2. First we show that:

$$\Delta_{HA}(\sigma) \equiv \left[\pi(1-q^H) + (1-\pi)q^H\right] \left[\varphi_H(q^H)\right]^{1-\sigma} - \left[\pi(1-q^L) + (1-\pi)q^L\right] \left[\varphi_A(q^H, q^L)\right]^{1-\sigma},$$

contains only one critical point in $\mathbb{R}$. Let φ_H and φ_A denote $\varphi_H(q^H)$ and $\varphi_A(q^H, q^L)$, respectively. The first-order derivative with respect to σ is:

$$\frac{\partial \Delta_{HA}}{\partial \sigma} = \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log \varphi_H \, (-1) - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma} \log \varphi_A \, (-1) \, .$$

The first-order condition holds if and only if:

$$\left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma^{FOC}} \log \varphi_H = \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A^{1-\sigma^{FOC}} \log \varphi_A,$$

which suggests the following unique critical point:

$$\sigma^{FOC} = 1 - \frac{\log \left[\frac{\left[\pi(1-q^L)+(1-\pi)q^L\right] \log \varphi_A}{\left[\pi(1-q^H)+(1-\pi)q^H\right] \log \varphi_H} \right]}{\log \left[\frac{\varphi_H}{\varphi_A} \right]},$$

For any $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ and $\sigma = 0$ such that $\pi > q^H > q^L$ holds, we have:

$$\begin{aligned}
& \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H > \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A \\
\iff & \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H - \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_A > 0 \\
\iff & \Delta_{HC}(\sigma = 0) > 0,
\end{aligned}$$

which has been proven in the previous chapter.

Given $\sigma \to 1$, we have:

$$\begin{aligned}
&\lim_{\sigma\to 1^-} \left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma} - \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma} \\
&\quad= \left[\pi(1-q^H)+(1-\pi)q^H\right] - \left[\pi(1-q^L)+(1-\pi)q^L\right] = (1-2\pi)\left(q^H-q^L\right) < 0 \\
&\iff \Delta_{HC}(\sigma \to 1) < 0.
\end{aligned}$$

By Intermediate Value Theorem, there exists a cut-off point, $\sigma^*_{HA} \in (0,1)$ such that:

$$\left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma^*_{HA}} = \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma^*_{HA}}.$$

The uniqueness of σ^*_{HA} follows from the fact that the function possesses at most one critical point, permitting only a single crossing of the horizontal axis. □

Proof of Lemma 2.3. First show that:

$$\begin{aligned}
\Delta_{HA}(\sigma) \equiv &\left[\pi(1-q^H)+(1-\pi)q^H\right]\left[\varphi_H(q^H)\right]^{1-\sigma} + \left[q^H-\pi\right]^{1-\sigma} \\
&- \left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_A(q^H,q^L)\right]^{1-\sigma} - \left[q^L-\pi\right]^{1-\sigma},
\end{aligned}$$

contains only one critical point in $\mathbb{R}$. Let φ_H and φ_A denote $\varphi_H(q^H)$ and $\varphi_A(q^H,q^L)$, respectively. The first-order derivative with respect to σ is:

$$\begin{aligned}
\frac{\partial \Delta_{HA}}{\partial \sigma} = &\left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma}\log\varphi_H\,(-1) + \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right](-1) \\
&- \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log\varphi_A\,(-1) - \left[q^L-\pi\right]^{1-\sigma}\log\left[q^L-\pi\right](-1).
\end{aligned}$$

The first-order condition holds if and only if:

$$\begin{aligned}
&\left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma}\log\varphi_H + \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right] \\
&= \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log\varphi_A + \left[q^L-\pi\right]^{1-\sigma}\log\left[q^L-\pi\right].
\end{aligned}$$

The second-order derivative with respect to σ is:

$$\begin{aligned}\frac{\partial^2 \Delta_{HA}}{\partial \sigma^2} &= \left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma}\log^2\varphi_H + \left[q^H-\pi\right]^{1-\sigma}\log^2\left[q^H-\pi\right] \\ &- \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log^2\varphi_A - \left[q^L-\pi\right]^{1-\sigma}\log^2\left[q^L-\pi\right].\end{aligned}$$

The following result demonstrates that the corresponding second-order derivative at any critical point is negative, thereby implying that $\sigma = \sigma^{FOC}$ constitutes a local maximum:

$$\begin{aligned}\frac{\partial^2 \Delta_{HA}}{\partial \sigma^2}(\sigma^{FOC}) &= \left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma}\log^2\varphi_H + \left[q^H-\pi\right]^{1-\sigma}\log^2\left[q^H-\pi\right] \\ &< \left[\pi(1-q^H)+(1-\pi)q^H\right]\varphi_H^{1-\sigma}\log\varphi_H\log\varphi_A \\ &+ \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\log\left[q^L-\pi\right] \\ &= \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log^2\varphi_A + \left[q^L-\pi\right]^{1-\sigma}\log\left[q^L-\pi\right]\log\varphi_A \\ &- \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\log\varphi_A + \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\log\left[q^L-\pi\right] \\ &= \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log^2\varphi_A + \left[q^L-\pi\right]^{1-\sigma}\log\left[q^L-\pi\right]\log\varphi_A \\ &- \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\log\varphi_A + \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\log\left[q^L-\pi\right] \\ &= \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log^2\varphi_A + \left[q^L-\pi\right]^{1-\sigma}\log\left[q^L-\pi\right]\log\varphi_A \\ &+ \left[q^H-\pi\right]^{1-\sigma}\log\left[q^H-\pi\right]\left\{\log\left[q^L-\pi\right]-\log\varphi_A\right\} \\ &< \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_A^{1-\sigma}\log^2\varphi_A - \left[q^L-\pi\right]^{1-\sigma}\log^2\left[q^L-\pi\right],\end{aligned}$$

which suggests that $\sigma = \sigma^{FOC}$ is a local maximum. Given that any critical point must correspond to a local maximum, the function admits at most one such point. Combined with the facts that $\Delta_{HA}(\sigma = 0) > 0$ and $\Delta_{HA}(\sigma = 1) < 0$, it follows that the cut-off point is unique.

□

Proof of Lemma 2.4. The following provides a direct proof:

$$\begin{aligned}u_S^H(\sigma) &= \left[\pi(1-q^H)+(1-\pi)q^H\right]\left[\varphi_H(q^H)\right]^{1-\sigma} + \left[q^H-\pi\right]^{1-\sigma} \\ &= \left[\pi(1-q^H)+(1-\pi)q^H\right]\left[q^H - \frac{(1-\pi)q^H}{\pi(1-q^H)+(1-\pi)q^H}\right]^{1-\sigma} + \left[q^H-\pi\right]^{1-\sigma} \\ &= \left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\left[\left[\pi(1-q^H)+(1-\pi)q^H\right]q^H-(1-\pi)q^H\right]^{1-\sigma} + \left[q^H-\pi\right]^{1-\sigma} \\ &= \left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\left[(1-2\pi)q^Hq^H+(2\pi-1)q^H\right]^{1-\sigma} + \left[q^H-\pi\right]^{1-\sigma}\end{aligned}$$

$$
\begin{aligned}
&\geq \left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\left[(1-2\pi)q^Hq^H+(2\pi-1)q^H+(q^H-\pi)\right]^{1-\sigma}\\
&\quad+\left[1-\left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\right]\left[q^H-\pi\right]^{1-\sigma}\\
&=\left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\left[(1-2\pi)q^Hq^H+(2q^H-1)\pi\right]^{1-\sigma}\\
&\quad+\left[1-\left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\right]\left[q^H-\pi\right]^{1-\sigma}\\
&\geq \left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\left[(1-2\pi)q^Hq^L+(q^H+q^L-1)\pi\right]^{1-\sigma}\\
&\quad+\left[1-\left[\pi(1-q^H)+(1-\pi)q^H\right]^{\sigma}\right]\left[q^H-\pi\right]^{1-\sigma}\\
&\quad\geq \left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_A(q^H,q^L)\right]^{1-\sigma}=u_S^A(\sigma).
\end{aligned}
$$

We also verify the boundary cases. For any $(\pi, q^H, q^L) \in (\frac{1}{2}, 1)^3$ and $\sigma = 0$ such that $q^H > \pi > q^L$ holds, we have:

$$
\begin{aligned}
u_S^H(0)-u_S^A(0)=\left[\pi(1-q^H)+(1-\pi)q^H\right]\left[\varphi_H(q^H)\right]+\left[q^H-\pi\right]&\\
-\left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_A(q^H,q^L)\right]&>0,
\end{aligned}
$$

which has been proven in the previous chapter.

Given $\sigma \to 1$, we have:

$$
\begin{aligned}
&\lim_{\sigma\to 1^-} u_S^H(\sigma)-u_S^A(\sigma)\\
&=\lim_{\sigma\to 1^-}\left[\pi(1-q^H)+(1-\pi)q^H\right]\left[\varphi_H(q^H)\right]^{1-\sigma}+\left[q^H-\pi\right]^{1-\sigma}\\
&-\left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_A(q^H,q^L)\right]^{1-\sigma}\\
&=\left[\pi(1-q^H)+(1-\pi)q^H\right]-\left[\pi(1-q^L)+(1-\pi)q^L\right]+1=(1-2\pi)\left(q^H-q^L\right)+1>0.
\end{aligned}
$$

□

Proof of Lemma 2.5. First show that:

$$
\begin{aligned}
\Delta_{HC}(\sigma)\equiv\left[\pi(1-q^H)+(1-\pi)q^H\right]\left[\varphi_H(q^H)\right]^{1-\sigma}+\left[q^H-\pi\right]^{1-\sigma}&\\
-\left[\pi(1-q^L)+(1-\pi)q^L\right]\left[\varphi_C(q^H,q^L)\right]^{1-\sigma}-\left[\max\left\{q^L-\pi,0\right\}+V(q^L)\right]^{1-\sigma}&,
\end{aligned}
$$

contains only one critical point in $\mathbb{R}$. Let φ_H and φ_C denote $\varphi_H(q^H)$ and $\varphi_C(q^H, q^L)$, respectively. The first-order derivative with respect to σ is:

$$\begin{aligned}\frac{\partial \Delta_{HC}}{\partial \sigma} &= \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log \varphi_H \, (-1) + \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] (-1) \\ &- \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log \varphi_C \, (-1) - \left[V(q^L)\right]^{1-\sigma} \log\left[V(q^L)\right] (-1) .\end{aligned}$$

The first-order condition holds if and only if:

$$\begin{aligned}&\left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log \varphi_H + \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \\ &= \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log \varphi_C + \left[V(q^L)\right]^{1-\sigma} \log\left[V(q^L)\right] .\end{aligned}$$

The second-order derivative with respect to σ is:

$$\begin{aligned}\frac{\partial^2 \Delta_{HC}}{\partial \sigma^2} &= \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log^2 \varphi_H + \left[q^H - \pi\right]^{1-\sigma} \log^2\left[q^H - \pi\right] \\ &- \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log^2 \varphi_C - \left[V(q^L)\right]^{1-\sigma} \log^2\left[V(q^L)\right] .\end{aligned}$$

The following result demonstrates that the corresponding second-order derivative at any critical point is negative, thereby implying that σ^{FOC} constitutes a local maximum:

$$\begin{aligned}\frac{\partial^2 \Delta_{HC}}{\partial \sigma^2}(\sigma^{FOC}) &= \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log^2 \varphi_H + \left[q^H - \pi\right]^{1-\sigma} \log^2\left[q^H - \pi\right] \\ &< \left[\pi(1-q^H) + (1-\pi)q^H\right] \varphi_H^{1-\sigma} \log \varphi_H \log \varphi_C \\ &+ \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \log\left[V(q^L)\right] \\ &= \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log^2 \varphi_C + \left[V(q^L)\right]^{1-\sigma} \log\left[V(q^L)\right] \log \varphi_C \\ &- \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \log \varphi_C + \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \log\left[V(q^L)\right] \\ &= \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log^2 \varphi_C + \left[V(q^L)\right]^{1-\sigma} \log\left[V(q^L)\right] \log \varphi_C \\ &- \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \log \varphi_C + \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \log\left[V(q^L)\right] \\ &= \left[\pi(1-q^L) + (1-\pi)q^L\right] \varphi_C^{1-\sigma} \log^2 \varphi_C + \left[V(q^L)\right]^{1-\sigma} \log\left[V(q^L)\right] \log \varphi_C \\ &+ \left[q^H - \pi\right]^{1-\sigma} \log\left[q^H - \pi\right] \left\{\log\left[V(q^L)\right] - \log \varphi_C\right\}\end{aligned}$$

$$< \left[\pi(1-q^L)+(1-\pi)q^L\right]\varphi_C^{1-\sigma}\log^2\varphi_C - \left[V(q^L)\right]^{1-\sigma}\log^2\left[V(q^L)\right],$$

which suggests that $\sigma = \sigma^{FOC}$ is a local maximum. Given that any critical point must correspond to a local maximum, the function admits at most one such point. Combined with the facts that $\Delta_{HC}(\sigma = 0) > 0$ and $\Delta_{HC}(\sigma = 1) < 0$, it follows that the cut-off point is unique.

□

Chapter 3

Deliberation and Voting with Opposed Preferences

3.1 Introduction

Information aggregation through committee deliberation represents a fundamental mechanism for enhancing collective decision-making under uncertainty. The theoretical foundation for such mechanisms rests on the premise that agents, despite potentially heterogeneous preferences regarding different types of errors, share a common objective of reaching decisions that align with available evidence. However, this assumption may prove inadequate in organisational contexts where agents possess fundamentally opposed preferences regarding decision outcomes. Consider corporate governance, where directors serving on boards may hold financial positions that directly conflict with shareholder interests. A director maintaining short positions in the firm's equity benefits from decisions that diminish share value, creating incentives to mislead the committee towards suboptimal investment choices through strategic information transmission during deliberation. Such scenarios give rise to a critical question in mechanism design: under what conditions does establishing a deliberative committee improve decision quality when some agents possess preferences that are systematically opposed to the objectives of the principal?

The analytical challenge intensifies when accounting for the strategic responses of well-intentioned agents. Recognising that some committee members may transmit false information, agents aligned with the principal's interests may themselves choose to withhold or distort their private signals to counteract potential manipulation. The resulting equilibrium behaviour can fundamentally undermine the information aggregation benefits that deliberative mechanisms are designed to provide. The paper develops a formal model of committee decision-making in which agents engage in cheap talk communication before casting votes under the majority rule. The analysis characterises three distinct equilibrium configurations: truthfully revealing equilibrium in which all agents honestly transmit their private signals, non-truthful equilibrium in which good agents reveal truthfully whilst bad agents systematically misrepresent their information, and babbling equilibrium in which no meaningful information transmission occurs.

The central contribution lies in characterising the conditions under which deliberative mechanisms enhance or diminish collective decision-making efficiency when preferences are fundamentally opposed rather than merely heterogeneous. The efficiency analysis demonstrates that deliberative mechanisms improve upon decision-making based solely on prior beliefs only when signal precision exceeds the prior probability and the likelihood of recruiting well-intentioned agents is sufficiently high. When either condition fails, principals achieve superior outcomes by foregoing information gathering entirely rather than investing in deliberative committee processes. These results carry significant implications for organisational design and the structure of collective decision-making institutions, suggesting that deliberative mechanisms, whilst potentially beneficial under ideal conditions may prove counterproductive when preference alignment cannot be ensured.

The paper proceeds as follows. Section 3.2 reviews the relevant literature on information aggregation and deliberation with opposed preferences. Section 3.3 presents the model and establishes the baseline framework with strategic messaging and voting. Section 3.4 characterises the equilibrium conditions, analysing the conditions under which truthful revelation, systematic deception, or communication breakdown emerge. Section 3.5 examines the efficiency implications of each equilibrium, establishing when deliberation improves upon pure voting and when it enhances decision-making relative to prior beliefs alone. Section 3.6 concludes with a discussion of institutional design implications and directions for future research.

3.2 Literature Review

The literature on information aggregation through deliberation and voting can be broadly categorised into three streams: studies examining voting-only mechanisms with heterogeneous preferences, analyses of deliberation with standard preference assumptions, and research departing from traditional monotonicity constraints.

Early research on committee decision-making focused on voting mechanisms without deliberation, examining when agents truthfully reveal private information through their votes. Whilst maximal efficiency requires all agents to vote according to their private signals, Austen-Smith and Banks (1996) demonstrate that such nonstrategic voting may fail to constitute Nash equilibrium behaviour under general assumptions. Feddersen and Pesendorfer (1998) extend this framework to jury settings, showing that no Nash equilibrium exists in which all jurors employ nonstrategic voting under unanimity rules, highlighting fundamental tensions between individual incentives and collective information aggregation.

Building on cheap talk communication models (e.g., Austen-Smith, 1990; Crawford and Sobel, 1982; Farrell and Rabin, 1996), subsequent research introduced deliberation stages preceding voting. Coughlan (2000) provides the foundational framework, analysing jury deliberation under publicly known preferences and demonstrating that truthful communication emerges only when juror preferences are sufficiently similar. Austen-Smith and Feddersen (2005) and Austen-Smith and Feddersen (2006) advance this analysis by incorporating private preference information, showing that truthful deliberation can persist under non-unanimous voting rules despite potential ex-post disagreement amongst committee members. Further developments include Gerardi and Yariv (2007), who establish that all voting rules except unanimity generate identical equilibrium outcomes from a mechanism design perspective, and Van Weelden (2008), who demonstrate that sequential rather than simultaneous deliberation eliminates full information revelation under any voting rule. Deimen et al. (2015) generalise these results to richer state and signal spaces, whilst Le Quement and Yokeeswaran (2015) show that subgroup deliberation can Pareto improve upon plenary deliberation outcomes.

The most directly relevant research is Meirowitz (2007), who introduces the concept of opposed preferences that this paper adopts and extends. Their framework represents a fundamental departure from traditional jury models by allowing agents whose preferences systematically conflict with truth-seeking objectives. In their model, 'good' agents prefer decisions that match the true state whilst 'bad' agents prefer decisions that contradict available evidence. However, Meirowitz (2007) includes the assumption that all agents might share identical preferences with some positive probability, creating scenarios in which preference conflict disappears entirely. Under their framework, truthful deliberation exists only when each agent believes that the majority of other agents share their preference type, and larger committees make truthful deliberation less sustainable.

This paper advances beyond the analysis of Meirowitz (2007) in several key respects. First, the model eliminates the assumption that agents might share identical preferences, ensuring that preference conflict is always present and providing a more stringent test of deliberative mechanisms. Second, whilst Meirowitz (2007) provides only sufficient conditions for truthful equilibrium existence, conditions which are often far from tight, this paper delivers complete and precise characterisation of equilibrium existence conditions. Third, and most significantly, Meirowitz (2007) completely omits the analysis of babbling equilibrium, despite its fundamental importance as a benchmark configuration. Babbling equilibrium always exists regardless of parameter values, effectively reducing deliberative mechanisms to pure voting without information transmission. This paper recognises babbling equilibrium as the natural comparison point for evaluating when deliberation adds value, using the expected payoff from babbling as a crucial benchmark against which to measure other equilibria.

The efficiency analysis represents the primary theoretical advance of this paper. Whilst Meirowitz (2007) establishes incomplete conditions for truthful deliberation to exist, they provide no analysis of whether such deliberation actually improves decision quality relative to alternatives. This paper conducts systematic efficiency comparisons that are entirely absent from Meirowitz (2007), establishing precise parameter conditions under which deliberative mechanisms improve upon decision-making based purely on prior beliefs and upon the babbling equilibrium benchmark. The analysis reveals a fundamental tension: conditions that favour truthful equilibrium existence typically conflict with conditions that enable efficiency gains. This implies that for most realistic parameter combinations, organisations face a choice between sustainable equilibria that provide no efficiency benefits and potentially beneficial but unsustainable configurations.

Additional literature examines related aspects of strategic communication in group settings. Gerardi et al. (2009) and Wolinsky (2002) employ mechanism design approaches to information extraction under conflicts of interest, whilst Galeotti et al. (2013) analyse multi-agent communication networks, demonstrating that truthful revelation depends jointly on preference composition and the number of honest agents. Hagenbach et al. (2014) examine information revelation with certifiable communication, and Jackson and Tan (2013) show that full revelation can be approximated in large societies regardless of voting rules. Visser and Swank (2007) incorporate reputational concerns, finding that agents manipulate information strategically when preferences diverge substantially.

3.3 Model

This analysis follows the model from Meirowitz (2007). Assume that the *true state* is binary, $\omega \in \Omega = \{0, 1\}$ associated with a *prior*, in which $\mathbb{P}(\omega = 0) = p \in (\frac{1}{2}, 1)$ and $\mathbb{P}(\omega = 1) = 1 - p$. There is a principal who is making a *decision* between two alternatives, $d \in \mathcal{D} = \{0, 1\}$. Instead of purely relying on the prior, p, they set up a committee to aggregate information from agents in order to enhance the probability of making a correct decision. Without loss of generality, the analysis assumes that the correct decision is the one which matches the true state. The committee is composed of n agents, $i \in \mathcal{I} = \{1, 2, \dots, n\}$. For simplicity, the number of agents, n, is odd. Each agent possesses a *preference*, $\theta_i \in \Theta = \{1, -1\}$, in which $\theta_i = 1$ indicates the agent i is a 'good' agent who shares the same preference with the principal and thus prefers to match the decision to the state, whilst $\theta_i = -1$ indicates the agent i is a 'bad' agent who prefers to have the decision unmatched. Preferences are independently and identically generated from a publicly known distribution, $\mathbb{P}(\theta_i = 1) = \alpha \in (\frac{1}{2}, 1)$ and $\mathbb{P}(\theta_i = -1) = 1 - \alpha$. In addition, preferences are private information known by themselves only. Before deliberation, each agent receives a private *signal* regarding the state, $s_i \in \mathcal{S} = \{0, 1\}$, which is independently and identically generated from a conditional distribution, $\mathbb{P}(s_i = \omega \mid \omega) = q \in (\frac{1}{2}, 1)$ and $\mathbb{P}(s_i \neq \omega \mid \omega) = 1 - q$. The signal generating distribution is known by the public. After receiving their private signal, s_i, each agent simultaneously sends a *message*, $m_i \in \mathcal{M} = \{0, 1\}$, during the stage of deliberation. The resulting *message profile*, $\boldsymbol{m} \equiv (m_1, m_2, \dots, m_n)$, is publicly observed by all agents. Each agent then simultaneously casts a *vote*, $v_i \in \mathcal{V} = \{0, 1\}$. The decision, $d \in \{0, 1\}$, is made according to the majority rule. That is, $d(\boldsymbol{v}) = 0$ if $\#\{v_i = 0\} > \frac{n}{2}$, in which $\boldsymbol{v} \equiv (v_1, v_2, \dots, v_n)$ is the *voting profile*. The ex-post *payoff* of an agent i with $\theta_i = 1$ is $\pi_i = 1$ if $d = \omega$, and $\pi_i = 0$ otherwise, whilst that of an agent i with $\theta_i = -1$ is $\pi_i = 1$ if $d \neq \omega$, and $\pi_i = 0$ otherwise.

This paper focuses on symmetric pure message and voting strategies. Each agent observes their preference and private signal, and then chooses a message for deliberation. Thus, a symmetric pure *message strategy* is defined as a mapping: $m : \mathcal{S} \times \Theta \to \mathcal{M}$. Afterwards, each agent further observes the *messages from other agents*, i.e., *message profile other than i*, $\boldsymbol{m_{-i}} \equiv \boldsymbol{m} \setminus \{m_i\}$, and updates the *posterior belief* regarding the true state, $\mu_i \equiv \mathbb{P}(\omega = 0 \mid s_i, \boldsymbol{m_{-i}})$. Based on the posterior belief, they cast a vote, v_i. Thus, a symmetric pure *voting strategy* is defined as a mapping: $v : \mathcal{S} \times \mathcal{M}^{n-1} \times \Theta \to \mathcal{V}$. The equilibrium concept follows the definition of pure Perfect Bayesian Equilibrium. An *equilibrium* is a pair of message strategy and voting strategy (m, v) which satisfies the following two requirements: given the message strategy, no agent can improve their expected payoff by deviating from their voting strategy whilst given the voting strategy, no agent can improve their expected payoff by deviating from their message strategy. Formally, we define the *voting profile other than i*, i.e., *votes from other agents*, as $\boldsymbol{v_{-i}} \equiv \boldsymbol{v} \setminus \{v_i\}$.

Definition 3.1 (Equilibrium concept). Let $m_i^* \equiv m^*(s_i, \theta_i)$ and $v_i^* \equiv v^*(s_i, \boldsymbol{m_{-i}}, \theta_i)$. A pair (m^*, v^*)

is an equilibrium if and only if,

(i) no agent can improve their payoff by unilaterally deviating from their voting strategy, taking voting strategies as given:

$$\forall v_i' \in \mathcal{V},\ \mathbb{E}[\pi_i(\omega, v_i^*, \boldsymbol{v_{-i}}, \theta_i)] \geq \mathbb{E}[\pi_i(\omega, v_i', \boldsymbol{v_{-i}}, \theta_i)];$$

(ii) no agent can improve their payoff by unilaterally deviating from their message strategy, accounting for how this affects subsequent voting of any other agent:

$$\begin{aligned}\forall m_i' \in \mathcal{M},\ \mathbb{E}[\pi_i(\omega, v_i^*, \{v_j^*(\cdot, (m_1, ..., m_i^*, ..., m_n), \cdot)\}_{j \neq i}, \theta_i)] \\ \geq \mathbb{E}[\pi_i(\omega, v_i^*, \{v_j^*(\cdot, (m_1, ..., m_i', ..., m_n), \cdot)\}_{j \neq i}, \theta_i)],\end{aligned}$$

where $v_j^* = v(s_j, \boldsymbol{m_{-j}}, \theta_j),\ \forall j \in \mathcal{I}$.

These equilibrium conditions capture the strategic interaction inherent in this two-stage game. The first condition ensures that no agent wishes to deviate from their prescribed voting behaviour, taking as given both the messaging strategy and the voting behaviour of all other agents. The second condition is more subtle: it requires that no agent benefits from sending a different message, accounting for how such a deviation would influence not only their own subsequent voting decision but also the voting decisions of all other agents who observe the altered message profile. Such a condition recognises that messages serve as signals to other committee members, and any deviation in messaging will trigger corresponding adjustments in the voting behaviour of others, creating a chain of strategic responses that the deviating agent must anticipate when evaluating alternative messaging strategies.

This analysis examines three equilibrium configurations, each characterised by a distinct message strategy combined with sincere voting: truthfully revealing equilibrium, non-truthful equilibrium, and babbling equilibrium. The following presents the formal definitions of these message strategies.

Definition 3.2 (Selected Message Strategies)**.** This analysis examines three distinct message strategies:

(i) *Truthfully revealing message strategy*: each agent of any preference sends their private signal during deliberation, $m(s_i, \theta_i) = s_i, \ \forall i \in \mathcal{I}$.

(ii) *Non-truthful message strategy*: a good agent reveals their private signal, $m(s_i, \theta_i = 1) = s_i$, whilst a bad agent sends a false message, $m(s_i, \theta_i = -1) = 1 - s_i, \ \forall i \in \mathcal{I}$.

(iii) *Babbling message strategy*: each agent sends a null message, $m(s_i, \theta_i) = 0, \ \forall i \in \mathcal{I}$.

Note that semi-separating strategies, i.e., one type plays babbling whilst the other plays non-babbling, and mixed strategies are excluded from this paper, since the implied efficiency of any of them should be a convex combination of the three considered message strategies. The rationale for this exclusion rests on the observation that any semi-separating equilibrium in which one agent type employs truthful revelation whilst the other employs babbling would yield efficiency outcomes that lie between those of the pure strategy equilibria. Similarly, any mixed strategy equilibrium in which agents randomise between different messaging approaches would produce expected efficiency measures that represent weighted averages of the pure strategy outcomes. Since efficiency comparison constitutes a primary objective of this analysis, examining these intermediate cases would not provide additional insights beyond understanding the performance boundaries established by the pure strategy equilibria. Therefore, it is sufficient to restrict attention to the three pure strategy configurations for efficiency comparison purposes. The analysis derives the conditions on parameters for the existence of these three equilibria and subsequently examines their relative efficiency properties.

The posterior belief regarding the state based on their signal s_i and messages from other agents, $\boldsymbol{m_{-i}}$ is:

$$\mu_i \equiv \mu(s_i, \boldsymbol{m_{-i}}) \equiv \mathbb{P}(\omega = 0 \mid s_i, \boldsymbol{m_{-i}}) = \frac{\mathbb{P}(\omega = 0, s_i, \boldsymbol{m_{-i}})}{\mathbb{P}(\omega = 0, s_i, \boldsymbol{m_{-i}}) + \mathbb{P}(\omega = 1, s_i, \boldsymbol{m_{-i}})}.$$

For algebraic simplicity, the analysis introduces a new mapping, $\phi : \mathcal{S} \times \mathcal{M}^{n-1} \to (0, 1)$, and a new variable, ϕ_i, to evaluate the posterior belief,

Definition 3.3 (Posterior Belief)**.** Let ϕ_i denote the posterior odds ratio possessed by the agent i given their signal s_i and message profile of the others $\boldsymbol{m_{-i}}$,

$$\phi_i \equiv \phi(s_i, \boldsymbol{m_{-i}}) \equiv \frac{\mathbb{P}(\omega = 0, s_i, \boldsymbol{m_{-i}})}{\mathbb{P}(\omega = 1, s_i, \boldsymbol{m_{-i}})},$$

where $\phi : \mathcal{S} \times \mathcal{M}^{n-1} \to (0, 1)$ denotes the posterior odds ratio mapping.

Note that $\mu_i \geq \frac{1}{2}$ and $\phi_i \geq 1$ are equivalent. Henceforth, the analysis refers to ϕ_i as the posterior belief regarding the true state.

Regarding the voting stage, this paper examines the *sincere voting strategy* under each equilibrium. Sincere voting differs from naive signal-based voting in that agents take into account all information revealed during the deliberation stage. Formally, sincere voting is defined as a voting strategy in which each good agent votes for the alternative corresponding to a posterior belief exceeding $\frac{1}{2}$, or equivalently, a posterior odds ratio exceeding one. We now provide the formal definition.

Definition 3.4 (Sincere Voting Strategy)**.** Under *sincere voting*, an agent i votes for the alternative associated with the more probable state according to their posterior odds ratio,

$$\phi(s_i, \boldsymbol{m_{-i}}) \geq 1 \implies v_i = 0; \;\; \phi(s_i, \boldsymbol{m_{-i}}) < 1 \implies v_i = 1.$$

Strategic voting in committee settings involves a fundamental tension between information aggregation and preference heterogeneity. Rational agents who play strategically focus primarily on events in which they serve as pivotal voters, recognising that their individual vote determines the collective outcome only in these decisive moments. The optimal voting decision in such pivotal events may differ substantially from what the agent posterior belief would suggest, as strategic considerations must account for the informational content revealed by others voting behaviour.

The literature typically examines weakly undominated voting strategies, in which an agent weakly prefers to cast a particular vote v_i regardless of the voting profile of other agents, $\boldsymbol{v_{-i}}$. Under a truthfully revealing equilibrium, sincere voting emerges as the natural strategy choice. Since all private information has been completely shared during the deliberation stage, agent votes cannot convey additional information beyond what has already been revealed through messages. The voting stage thus becomes purely implementational: agents vote according to their posterior beliefs without concern for signalling or information extraction. In such environments, an agent optimal voting decision when pivotal coincides exactly with the decision indicated by their posterior belief, making sincere voting both individually rational and collectively efficient.

However, the strategic landscape changes dramatically under non-truthful and babbling equilibria. In these configurations, the voting actions of other agents, $\boldsymbol{v_{-i}}$, contain valuable information that was not fully transmitted during the deliberation stage. Bad agent votes may reveal their true private signals despite their earlier strategic messaging, whilst the pattern of votes across all agents provides additional evidence about the underlying state. Consequently, agents must solve a complex pivotal voting problem, weighing their posterior beliefs against the informational inferences they can draw from observing other agent voting behaviour. The analysis of these strategic voting equilibria requires careful attention to the information revelation properties of votes themselves. The subsequent analysis provides a comprehensive characterisation of equilibrium existence conditions.

3.4 Equilibrium Characterisation

In this section, we characterise the conditions on the parameter quadruple (α, p, q, n) governing the existence of the three equilibrium types by examining two distinct strategic incentives that agents face. The analysis focuses on potential deviations from equilibrium behaviour: first, whether agents wish to deviate from sincere voting given the conjectured message strategy, and second, whether agents wish to deviate from their prescribed messaging behaviour given that all agents vote sincerely.

The underlying logic rests on the pivotal player principle. Rational agents recognise that their individual actions matter only in situations in which they are decisive for the collective outcome. Consequently, strategic agents concentrate their decision-making calculus exclusively on these pivotal events, carefully analysing the probability that such decisive moments will arise and optimising their strategies to maximise their expected payoff conditional on being pivotal. The pivotal reasoning manifests differently across the two stages of the game. In the voting stage, an agent considers only events in which they cast the decisive vote, understanding that their voting decision determines whether the committee chooses alternative zero or alternative one. The agent weighs their posterior belief about the true state against the strategic implications of their vote, particularly when voting behaviour itself conveys information to other committee members. In the messaging stage, the strategic calculus becomes more complex. An agent considers only events in which they serve as a pivotal sender, in which their message choice fundamentally alters the information available to other committee members and thereby influences the final collective decision. Such pivotal messaging events are less immediately apparent than pivotal voting events, as they require the agent to trace through how their message affects posterior beliefs of others, which in turn affects their voting behaviour, which ultimately determines the collective outcome. The agent must therefore engage in sophisticated forward reasoning, anticipating the chain of responses that their messaging choice will trigger amongst other committee members.

3.4.1 Full Information Transmission: Truthfully Revealing Equilibrium

Under a truthfully revealing equilibrium, deliberative democracy achieves its theoretical ideal. Consider a committee in which every member abandons strategic calculation and simply reports what they observe. Such transparency transforms the group into a collective sensing mechanism, pooling individual observations to construct a shared understanding of reality. The mathematical structure mirrors the conceptual appeal. Each message, $m(s_i, \theta_i) = s_i$, serves as an unfiltered window into private knowledge, creating perfect information aggregation through voluntary revelation.

The strategic beauty of this equilibrium lies in its simplicity. Since all agents commit to honesty, each participant can decipher the message profile perfectly to infer the exact signals observed by every other committee member. A message of zero unambiguously indicates that the sender received signal $s_i = 0$, whilst a message of one reveals that the sender received signal $s_i = 1$. Such perfect mapping between messages and underlying private information transforms the committee into a collective information processing unit, in which individual signals are pooled transparently to form a comprehensive picture of the evidence regarding the true state. The power of this transparent system becomes evident in how agents process information. When a committee member i observes the complete message profile, they face a straightforward belief updating problem. Given that they understand exactly which signals their colleagues observed, they apply Bayes' rule to form their posterior belief, $\phi : \mathcal{S} \times \mathcal{M}^{n-1} \to \mathbb{R}_+$,

$$\phi_i \equiv \phi(s_i, \boldsymbol{m_{-i}}) = \frac{pq^k(1-q)^{n-k}}{(1-p)(1-q)^k q^{n-k}} = \frac{p}{1-p}\left(\frac{q}{1-q}\right)^{2k-n},$$

where k is the number of null messages in the message profile, $k \equiv \#\{m_i = 0 \mid \boldsymbol{m}\}$. Also, let $\boldsymbol{m}^k \in \{\boldsymbol{m} \in \mathcal{M}^n \mid \#\{m_i = 0\} = k\}$ denote a message profile containing k null messages. An agent believes that $\omega = 0$ is more likely to be the true state if $\phi_i \geq 1$. The intuition behind this threshold is straightforward. Since ϕ_i represents the probability ratio, a value of $\phi_i \geq 1$ indicates that the agent assigns at least equal probability to state zero relative to state one. When $\phi_i > 1$, the agent considers state zero strictly more likely, whilst $\phi_i < 1$ indicates that state one appears more probable. The threshold of unity thus serves as the natural decision boundary. Since the ex-post payoff of a correct decision (i.e., $d = \theta$) is assumed symmetric given any true state, rational agents with preference $\theta = 1$ prefer the alternative that corresponds to the more likely state, leading them to favour decision $d = 0$ when $\phi_i \geq 1$ and decision $d = 1$ when $\phi_i < 1$. Note that ϕ_i is increasing in k, the number of null messages. The monotonic relationship reflects the informational content of messages. Since null messages correspond to null signals under truthful revelation, observing more null messages provides stronger evidence in favour of $\omega = 0$. Let k^* denote the cut-off point such that $\phi_i \geq 1$ if and only if $k \geq k^*$, and $\phi_i < 1$ otherwise. We therefore introduce the following formal definition.

Definition 3.5 (Posterior Belief Cut-off)**.** Let $k^* \in \mathbb{N}$ denote the minimum number of null messages required for an agent to believe that $\omega = 0$ is more likely than $\omega = 1$,

$$k^* \equiv \operatorname*{argmin}_{k \in \mathbb{N}} \left\{ \frac{p}{1-p} \left(\frac{q}{1-q} \right)^{2k-n} \mid \frac{p}{1-p} \left(\frac{q}{1-q} \right)^{2k-n} \geq 1 \right\}.$$

The threshold k^* emerges naturally from this updating process, representing the tipping point at which accumulated evidence overcomes prior scepticism. Imagine a committee initially biased towards believing state one is more likely. As null messages accumulate during deliberation, each additional report of null signal chips away at this bias until, at exactly k^* messages, the scales tip decisively towards state zero. The threshold k^* thus serves as a critical tipping point in the collective assessment: when the number of null messages reaches or exceeds this threshold, the accumulated evidence becomes sufficiently compelling to shift the balance of belief in favour of state zero.

The economic interpretation of k^* reveals important insights about the interaction between prior beliefs and signal precision. It is possible that $k^* = 0$ when the prior is sufficiently unequal in favour of state zero, which implies that signals and messages make no substantial impact on the posterior belief. In such cases, the prior dominates: agents enter the deliberation already convinced that $\omega = 0$ is more likely, and even contradictory evidence from signals fails to overturn this strong initial conviction. Such situations highlight a fundamental limitation of deliberative mechanisms when prior beliefs are heavily skewed. The information aggregation process becomes largely ceremonial rather than genuinely informative. Conversely, when priors are more balanced and signals are highly precise, k^* takes on positive values, meaning that a substantial consensus of null messages is required to tip the scales towards state zero. Such scenarios represent the deliberative ideal. Agents genuinely update their beliefs based on the collective evidence revealed through deliberation, with the committee serving as an effective information processing mechanism. After the stage of deliberation, all agents converge to thinking that $\omega = 0$ is more likely to be true, but this convergence reflects genuine learning rather than mere confirmation of pre-existing biases.

In the voting stage, we conjecture that all agents follow the sincere voting strategy: each good agent (i.e., $\theta_i = 1$) votes for the alternative associated with a higher posterior belief, whilst each bad agent (i.e., $\theta_i = -1$) votes for the alternative associated with a lower posterior belief. The intuitive justification is that sincere voting constitutes an equilibrium because truthful signal revelation during deliberation eliminates any informational advantage from strategic voting. Since all private information has already been shared and incorporated into each agent posterior belief, observing others votes yields no further insight. The voting stage therefore becomes implementational rather than informational: agents translate posterior beliefs directly into votes without concern for signalling or inference.

Formally, any deviation from sincere voting can be beneficial only in pivotal cases, where an individual vote determines the collective outcome. However, under truthful communication, an agent posterior belief conditional on being pivotal remains identical to the unconditional belief, since all relevant information has already been revealed. Therefore, the optimal decision in pivotal cases coincides with the sincere voting strategy. Strategic considerations such as signalling or inference become irrelevant once deliberation achieves full information aggregation. This transparency ensures that sincere voting is both individually rational and collectively efficient, and that the voting stage faithfully implements the information gathered during deliberation.

Now we investigate how the final decision is determined. According to majority rule, the final decision is determined as,

$$d(\boldsymbol{v}) = 0, \text{if } \#\{v_i = 0 \mid v_i \in \boldsymbol{v}\} > \frac{n}{2}; \;\; d(\boldsymbol{v}) = 1, \text{otherwise.}$$

The final decision d is effectively determined by the composition of preferences rather than the informational content of signals. The alternative associated with a higher posterior belief will be chosen if there are more good agents, whilst the alternative associated with a lower posterior belief becomes the outcome if there are more bad agents. Such outcome reflects a fundamental tension in the truthfully revealing equilibrium: whilst information aggregation functions perfectly through deliberation, the voting stage reintroduces preference-based distortions that can override the collective assessment of evidence.

Since the posterior belief is derived from the collection of signals and the final decision, $d \in \{0,1\}$, depends on the composition of preferences, we can establish a direct mapping from the underlying fundamentals (i.e., signals and preferences) to the collective decision. The decision outcome depends on two distinct factors operating through different channels. The informational content of signals determines which alternative agents believe is correct, whilst the preference composition determines which belief actually prevails in the voting. Let $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, ..., \theta_n)$ denote the *preference profile*. We therefore have the following reduced-form mapping from signals and preferences to the decision:

Lemma 3.1 (Decision Rule)**.** *Given any signal profile $\boldsymbol{s} \in \{0,1\}^n$ and preference profile $\boldsymbol{\theta} \in \{0,1\}^n$, the truthfully revealing equilibrium yields the final decision $d = 0$ if and only if one of the following two conditions holds:*

(i) null signals are sufficiently numerous and good agents constitute the majority:

$$\#\{s_i = 0 \mid s_i \in \boldsymbol{s}\} \geq k^* \text{ and } \#\{\theta_i = 1 \mid \theta_i \in \boldsymbol{\theta}\} > \frac{n}{2};$$

(ii) null signals are sufficiently few and good agents constitute the minority:

$$\#\{s_i = 0 \mid s_i \in \boldsymbol{s}\} < k^* \text{ and } \#\{\theta_i = 1 \mid \theta_i \in \boldsymbol{\theta}\} < \frac{n}{2}.$$

The theorem reveals the dual nature of decision-making under opposed preferences: the committee reaches decision $d = 0$ either when the evidence genuinely supports this choice and good agents implement it, or when the evidence points against it but bad agents strategically vote in the opposite direction. Such characterisation highlights how preference heterogeneity can produce correct decisions through two fundamentally different mechanisms, either through aligned incentives when information favours the decision, or through misaligned incentives that paradoxically lead to the right outcome when information opposes it. The presence of bad agents thus creates scenarios in which inferior evidence can yield superior decisions, demonstrating the complex interplay between information aggregation and preference conflict in committee settings.

To further simplify the notation and facilitate the discussion, we define the *type* of an agent, x_i, as the pair of their signal and preference, $x_i \equiv (s_i, \theta_i)$. Hence, there are four types of agents, $x_i \in \mathcal{X} = \{(0,1), (1,1), (0,-1), (1,-1)\}$. For better comprehensibility, we relabel the elements such that $\mathcal{X} = \{G_0, G_1, B_0, B_1\}$, in which $\{G, B\}$ indicates the preference of the agent, and the subscript $\{0, 1\}$ indicates the signal received by the agent. For example, an agent with $\theta_i = 1$ who receives $s_i = 0$ is associated with type G_0. We now define a *type profile* as $\boldsymbol{n} \equiv (n'_{G_0}, n'_{G_1}, n'_{B_0}, n'_{B_1}) \in \mathcal{N}' \equiv \left\{\boldsymbol{x} \in \mathbb{N}^4 \mid \sum_{i=1}^4 x_i = n\right\}$, in which n'_x indicates the number of agents of type x in the committee. The type profile thus characterises the complete type composition of the committee. Clearly, $n'_{G_0} + n'_{G_1} + n'_{B_0} + n'_{B_1} = n$. Consequently, the decision rule can be reduced to a mapping from the type profile:

Proposition 3.1 (Decision Rule: Truthfully Revealing Equilibrium). *Given the type profile, $\boldsymbol{n} \in \mathcal{N}'$, and the cut-off point, $k^* \in \mathbb{N}$, the truthfully revealing equilibrium yields the final decision $d = 0$ if and only if one of the following two conditions holds:*

(i) null signals are sufficiently numerous and good agents represent the majority: $n'_{G_0} + n'_{B_0} \geq k^*$ and $n'_{G_0} + n'_{G_1} \geq \frac{n}{2}$;

(ii) null signals are sufficiently few and good agents represent the minority: $n'_{G_0} + n'_{B_0} < k^*$ and $n'_{G_0} + n'_{G_1} < \frac{n}{2}$.

Sustainability of Sincere Voting. Now we examine whether an agent has incentive to deviate from sincere voting. First note that sincere voting under a truthfully revealing equilibrium is an undominated voting strategy since no further information is transmitted through the voting profile from other agents, $\boldsymbol{v_{-i}}$. Therefore, regardless of how other agents cast their votes, voting according to the posterior belief is optimal. In terms of pivotal voting, when an agent i considers the events in which they are a pivotal voter, the conditional posterior belief regarding the true state is exactly the same as $\phi(s_i, \boldsymbol{m_{-i}})$. To formalise the argument, define an *event* with respect to an agent i as a *type profile other than i*, denoted as $\boldsymbol{n_{-i}} \equiv (n_{G_0}, n_{G_1}, n_{B_0}, n_{B_1}) \in \mathcal{N} \equiv \left\{\boldsymbol{x} \in \mathbb{N}^4 \mid \sum_{i=1}^4 x_i = n - 1\right\}$, in which n_x denotes the number of agent type x, excluding i. The agent understands that they are a pivotal voter if and only if other agents are evenly divided in their votes, which occurs if and only if others are evenly divided in their preference. That is, $n_{G_0} + n_{G_1} = n_{B_0} + n_{B_1} = \frac{n-1}{2}$. Let $\mathcal{N}_V^{TRE} \subseteq \mathcal{N}$ denote the set of events in which the agent i is a pivotal voter. Therefore, the *posterior belief conditional on pivotal voting events*, denoted as $\widetilde{\phi} : \mathcal{S} \times \mathcal{M}^{n-1} \to \mathbb{R}_+$, is given by,

$$\widetilde{\phi}_i \equiv \widetilde{\phi}(s_i, \boldsymbol{m_{-i}}) \equiv \frac{\sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_V^{TRE}} \mathbb{P}(\omega = 0, s_i, \boldsymbol{n_{-i}})}{\sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_V^{TRE}} \mathbb{P}(\omega = 1, s_i, \boldsymbol{n_{-i}}).}$$

It is worth noting that the conditional posterior belief equals the unconditional one. We construct the following lemma.

Lemma 3.2 (Pivotal Posterior Equivalence). *Under truthfully revealing equilibrium, the posterior belief conditional on pivotal voting events equals the unconditional posterior belief. That is,*

$$\widetilde{\phi}_i = \phi_i.$$

Consequently, sincere voting is always sustainable, regardless of (α, p, q, n).

Proof. See Appendix 3.A.1. □

The economic intuition behind this result is straightforward. Since all private information has already been revealed through truthful messaging, the voting stage contains no additional informational content. When an agent conditions on being pivotal, they learn nothing new about the true state beyond what they already knew from observing the message profile. The act of being pivotal simply confirms that preferences are equally split, but this knowledge about preference composition provides no additional insight into the underlying state of the world. Consequently, the agent's optimal voting decision when pivotal remains identical to their optimal decision based purely on the posterior belief formed after deliberation. Therefore, they have no incentive to deviate from sincere voting. Sincere voting is sustainable under a truthfully revealing equilibrium.

Sustainability of Truthfully Revealing Messaging. We now turn to the more complex question of messaging incentives. The key insight lies in understanding when an agent becomes a pivotal sender, a position in which their message choice fundamentally alters the collective decision of the committee. When contemplating whether to deviate from truthful messaging, a strategic agent focuses exclusively on scenarios in which their message matters. In all other circumstances, deviating from the conjectured strategy produces no change in the final outcome, rendering such deviations pointless. The agent must therefore identify precisely those events in which their message tips the balance.

Consider an agent of type $x_i = G_0$, who recognises that truthful messaging will add another null message to the profile, potentially pushing the total count to the critical threshold k^*. Should this occur, all committee members will form posterior beliefs $\phi_i \geq 1$. As someone aligned with truth-seeking, our agent naturally prefers this outcome when the evidence genuinely supports it.

Yet the following strategic consideration emerges: whilst the evidence may point towards $d = 0$, the ultimate decision depends not merely on information but on preference composition. Under sincere voting, good agents will vote for $d = 0$ whilst bad agents vote for $d = 1$. The final outcome therefore hinges on which group commands the majority. The agent becomes pivotal precisely when their message determines whether the committee reaches the threshold k^*. Formally, this occurs when the event, $\boldsymbol{n_{-i}}$, satisfies $n_{G_0} + n_{B_0} = k^* - 1$ and $n_{G_1} + n_{B_1} = (n-1) - (k^* - 1)$. In such circumstances, the choice of the pivotal agent between truthful messaging and deception reverses informational assessment and, consequently, the voting behaviour of all members. Formally, let $\mathcal{N}_M^{TRE}(x_i)$ denote the set of events in which the agent i of type x_i is pivotal at the messaging stage. The decision of the agent hinges on a straightforward question. Conditional on being pivotal, does their side command the majority? If good agents predominate, truthful messaging yields the preferred outcome $d = 0$. Conversely, if bad agents hold sway, truthful messaging paradoxically leads to the undesired outcome $d = 1$, since bad agents vote against the evidence. Formally, let $\mathcal{N}_{x_i}^+$ ($\mathcal{N}_{x_i}^-$) denote the set of events in which the agent i of type x_i belongs to the majority (minority). Specifically, for example, $\mathcal{N}_{G_0}^+ = \{\boldsymbol{x} \in \mathcal{N} \mid n_{G_0} + n_{G_1} > \frac{n-1}{2}\}$ and $\mathcal{N}_{G_0}^- = \{\boldsymbol{x} \in \mathcal{N} \mid n_{B_0} + n_{B_1} > \frac{n-1}{2}\}$. The condition for sustainable truthful messaging requires that the expected payoff from honesty weakly dominates that from deception,

$$\begin{aligned}
&\sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_M^{TRE}(x_i) \cap \mathcal{N}_{x_i}^+} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n_{-i}}) + \sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_M^{TRE}(x_i) \cap \mathcal{N}_{x_i}^-} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n_{-i}}) \\
\geq &\sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_M^{TRE}(x_i) \cap \mathcal{N}_{x_i}^-} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n_{-i}}) + \sum_{\boldsymbol{n_{-i}} \in \mathcal{N}_M^{TRE}(x_i) \cap \mathcal{N}_{x_i}^+} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n_{-i}}).
\end{aligned}$$

When our agent finds themselves in a pivotal position, the strategic calculation becomes clear. Should they send the truthful message $m_i = 0$, they obtain the expected payoff corresponding to the left-hand side of our inequality. Conversely, deviating to $m_i = 1$ yields the payoff shown on the right-hand side. The first term in each expression captures the payoff when the committee ultimately chooses $d = 0$, whilst the second term represents the payoff under decision $d = 1$. Truthful messaging remains optimal when the left-hand side weakly dominates. Simplifying this condition reveals its economic essence. After rearranging terms to isolate the preference composition effects, the sustainability condition reduces to,

$$\mathbb{P}(i \text{ pivotal sender and in the majority}) \geq \mathbb{P}(i \text{ pivotal sender and in the minority}).$$

The interpretation proves illuminating. Applying identical reasoning to agents of types G_1, B_0, and B_1 yields a complete characterisation. The analysis reveals an elegant symmetry: the sustainability conditions for types G_0 and G_1 prove identical, as do those for types B_0 and B_1. Such symmetry reflects the fundamental alignment within each preference group, despite differences in private signals. Following algebraic manipulation of the preceding expressions, we establish the following proposition.

Proposition 3.2 (Sustainability: Truthfully Revealing Equilibrium)**.** *Given $k^* \neq 0$, a truthfully revealing equilibrium exists if and only if $(\alpha, p, q, n) \in [\frac{1}{2}, 1]^3 \times \mathbb{N}$ (n odd) satisfies,*

$$\sum_{i=0}^{k^*-1} \binom{k^*-1}{i} \left[- \sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*} \binom{n-k^*}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} \right] \geq 0,$$

$$\sum_{i=0}^{k^*-1} \binom{k^*-1}{i} \left[- \sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*} \binom{n-k^*}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} \right] \geq 0.$$

Given $k^* = 0$, truthfully revealing equilibrium always exists.

Proof. Given $k^* \neq 0$, see Appendix 3.A.1 for full derivation. Given $k^* = 0$, no pivotal events exist for any agent type, since the constraint $n_{G_0} + n_{G_1} = -1$ proves impossible to satisfy. Consequently, the sustainability conditions hold trivially for any parameter (α, p, q, n) such that $k^* = 0$ holds. □

A remarkable feature emerges from these conditions for non-trivial cases (i.e., $k^* \neq 0$). The derived conditions depend solely on the preference composition parameter α and committee size n, proving independent of both prior beliefs p and signal precision q, which suggests that the sustainability of truthful revelation hinges entirely on the balance between good and bad agents, not on the informational environment they face. This feature persists even when the assumption of symmetric payoffs is relaxed. Accordingly, we obtain the following corollary:

Corollary 3.2.1 (Robustness to Payoff Asymmetry)**.** *Consider an asymmetric payoff structure where $u(d = \omega = 0) = \lambda$ and $u(d = \omega = 1) = 1$ for any $\lambda \geq 1$. The sustainability conditions for truthful revelation established in Proposition 3.2 remain unchanged. They depend solely on preference composition α and committee size n, remaining independent of both prior beliefs p and signal precision q regardless of the value of λ.*

Proof. The proof follows directly from the observation that payoff asymmetry scales expected utilities proportionally across all strategic configurations. See Appendix 3.A.1 for formal derivation. □

Such robustness to payoff asymmetry reveals the fundamental mechanism underlying truthful revelation sustainability. Consider an agent evaluating whether to report truthfully when pivotal. Under truthful messaging, the agent benefits identically from two scenarios: commanding a majority whilst reporting truthfully (leading to their preferred outcome), and forming a minority whilst misreporting (overturning an adverse decision). Crucially, the relative probability of these scenarios depends exclusively on the distribution of preference types, not on the magnitude of state-contingent payoffs. The parameter λ affects absolute payoff levels but preserves the strategic equivalence between truth-telling in the majority and deception in the minority. When λ increases, the value of correctly matching state zero rises proportionally across all decision paths, leaving unchanged the fundamental trade-off between leveraging majority influence through honesty versus exploiting minority position through strategic misrepresentation. The independence from informational parameters (p, q) thus persists regardless of payoff asymmetry, as the strategic calculation reduces to comparing their likelihood of commanding influence, determined entirely by preference composition α and committee size n. Such structural invariance demonstrates that the sustainability of truthful communication hinges not on the stakes involved in different states, but on the beliefs about their relative influence within the committee structure.

The analysis of truthful revelation reveals a profound paradox at the heart of deliberative institutions with opposed preferences. The sustainability of honest communication depends not on the quality of information available to committee members, but solely on their belief about whether they represent the majority in pivotal moments. When agents expect their preference group to predominate, they willingly share private information, confident that the resulting collective decision will align with their interests. Conversely, when they anticipate being outnumbered, the temptation to manipulate information through strategic messaging becomes irresistible.

This creates a striking independence result: the informational sophistication of committee members, captured by signal precision q, and their collective prior knowledge, reflected in p, prove entirely irrelevant to the survival of truthful deliberation. Instead, the mechanism's success hinges on a delicate balance of preference composition and committee size. Larger committees make truthful revelation increasingly fragile, as the influence of each agent diminishes and the probability of commanding a majority in pivotal scenarios declines.

Perhaps most remarkably, the presence of bad agents can occasionally improve decision quality through a perverse channel: when good agents expect to be outnumbered, their strategic deception can paradoxically align collective choices with the true state. Such findings challenge conventional wisdom about deliberative design, suggesting that the institutional architecture must account not merely for information aggregation capabilities, but for the strategic incentives created by preference heterogeneity itself.

3.4.2 Zero Information Transmission: Babbling Equilibrium

In the starkest equilibrium, deliberation completely fails to serve its intended purpose. Under a babbling equilibrium, each agent sends an uninformative message regardless of their private signal or preference type. The term 'babbling' captures the essence of this communication breakdown: agents speak, yet convey no meaningful information about the underlying state. Without loss of generality, we model this scenario by having each agent send a null message, $(m(s_i, \theta_i) = 0)$ for all agents. The specific message content matters little. The crucial feature lies in the complete absence of informational content. Messages become mere noise, indistinguishable from random utterances. When deliberation provides no information, each agent forms beliefs using only their private signal and the common prior. The posterior belief calculation becomes strikingly simple,

$$\phi(s_i = 0, \boldsymbol{m_{-i}}) = \frac{pq}{(1-p)(1-q)} > 1; \;\; \phi(s_i = 1, \boldsymbol{m_{-i}}) = \frac{p(1-q)}{(1-p)q}.$$

The relationship between prior strength and signal precision determines how agents behave in the voting stage. When the prior dominates signal precision (i.e., $p \geq q$), agents effectively ignore their private information. Both $\phi(s_i = 0, \boldsymbol{m_{-i}})$ and $\phi(s_i = 1, \boldsymbol{m_{-i}})$ exceed unity, leading all agents to believe that state zero is more likely regardless of their signal. Such prior dominance creates a stark voting pattern: good agents consistently cast $v_i = 0$ whilst bad agents consistently cast $v_i = 1$. The irony becomes apparent. Agents receive costly private information yet completely disregard it when making decisions. Conversely, when signals prove more informative than the prior (i.e., $p < q$), agents rely primarily on their private information. Good agents vote according to their signals, whilst bad agents vote against them. The resulting coalition structure depends on signal realisations rather than preference types alone, introducing genuine uncertainty about outcomes. We formalise this decision pattern in the following preposition,

Proposition 3.3 (Decision Rule: Babbling Equilibrium)**.** *Given the type profile, $\boldsymbol{n} \in \mathcal{N}'$, the babbling equilibrium yields the final decision $d = 0$ if and only if one of the following two conditions holds:*

(i) $p \geq q$ *and* $n'_{G_0} + n'_{G_1} > \frac{n}{2}$;

(ii) $p < q$ *and* $n'_{G_0} + n'_{B_1} > \frac{n}{2}$.

What sustains this communication breakdown as an equilibrium outcome? The mechanism operates through a self-reinforcing cycle of expectations. When an agent anticipates that others will send uninformative messages, any unilateral attempt to convey meaningful information becomes futile. Their deviation cannot influence others' beliefs or voting decisions since other agents rationally ignore all messages. The collective decision remains unchanged, rendering any informational effort pointless. The coordination around silence possesses a compelling strategic logic. Each agent faces the choice between costly truth-telling and costless babbling. When truth-telling provides no strategic advantage, since others ignore messages anyways, babbling becomes the dominant strategy. This creates a coordination equilibrium in which rational individual behaviour produces collective irrationality.

Perhaps the most striking feature of babbling equilibrium lies in its universal existence. The sincere voting strategy constitutes an equilibrium since agents of any type form conditional posterior beliefs on pivotal events that are identical to their unconditional beliefs. In other words, regardless of parameter values, whether agents are predominantly good or bad, whether signals are precise or noisy, whether priors are strong or weak, babbling always represents a viable equilibrium outcome. We formalise this insight in the following proposition.

Proposition 3.4 (Sustainability: Babbling Equilibrium)**.** *A babbling equilibrium exists for all* $(\alpha, p, q, n) \in [\frac{1}{2}, 1]^3 \times \mathbb{N}$ *(n odd).*

Proof. See Appendix 3.A.2. □

The proof proceeds by examining potential deviations from equilibrium behaviour along two dimensions. First, we demonstrate that no agent benefits from deviating from sincere voting by showing that posterior beliefs conditional on pivotal events equal unconditional beliefs (i.e., $\widetilde{\phi}_i = \phi_i$). The key insight relies on a symmetry argument: the conditional probability structure ensures that being pivotal provides no additional information beyond what agents already possess from their private signals. Second, we establish that deviating from the babbling message strategy cannot improve the payoff of any agent. Since messages convey no information, altering one's own message cannot influence others' posterior beliefs or voting decisions. The decision outcome remains invariant to message content, eliminating any strategic incentive for truthful revelation.

The economic intuition behind this universal existence reveals a fundamental tension in deliberative settings. Babbling emerges not from agent irrationality, but from perfectly rational responses to the expected behaviour of others. The equilibrium represents a coordination failure: whilst all agents would benefit from mutual information sharing, no individual agent can unilaterally break the cycle of silence.

From a theoretical perspective, babbling equilibrium serves as the natural benchmark for evaluating deliberative mechanisms. It represents the worst-case scenario: a committee that functions no better than a pure voting mechanism without any information aggregation benefits. The universal existence of this equilibrium carries a sobering message, even sophisticated deliberative institutions can collapse into meaningless communication patterns. The existence of babbling equilibrium reminds us that institutional design must go beyond merely creating opportunities for communication. Effective deliberation requires mechanisms that align individual incentives with collective information sharing goals. Sometimes, the most rational individual response to strategic uncertainty is collective silence, transforming deliberation from an information aggregation tool into an elaborate charade.

3.4.3 Distorted Information Transmission: Non-Truthful Equilibrium

The non-truthful equilibrium is perhaps the most strategically sophisticated and counterintuitive configuration in our model, challenging conventional wisdom about the role of deception in collective decision-making. In this equilibrium, the committee effectively bifurcates along preference lines: good agents continue to truthfully reveal their private signals whilst bad agents systematically lie, sending messages that directly contradict their actual information. This creates a fascinating strategic ecosystem where truth and deception coexist in a stable pattern. At first glance, such systematic lying might seem to undermine any possibility of meaningful information aggregation completely. Why would good agents continue to share honestly when they know that bad agents are deliberately misleading the committee? The answer lies in the predictable nature of the deception. When bad agents lie systematically, always sending the opposite of their true signal, their messages paradoxically become informative again. A sophisticated good agent can decode the lies: if a bad agent sends message zero, the good agent infers that the bad agent actually observed signal one. Yet this adversarial communication system proves remarkably fragile. The sustainability of non-truthful equilibrium depends on a delicate balance of strategic forces. Good agents must find it worthwhile to continue truth-telling despite the presence of liars, whilst bad agents must benefit from systematic deception rather than either truth-telling or complete silence. Meanwhile, both types must find sincere voting optimal given the distorted information environment created by partial lying.

The existence of an equilibrium hinges on the complex interplay between signal precision, prior beliefs, and preference intensity. When signals are highly precise, agents rely primarily on their own private information, making others' messages, whether truthful or deceptive, relatively unimportant. Conversely, when prior beliefs dominate, message content becomes largely irrelevant regardless of its veracity. The most interesting and problematic case emerges in the intermediate range where messages retain significant influence but deception becomes strategically valuable, potentially destroying the equilibrium altogether. The non-truthful equilibrium thus represents a knife-edge phenomenon: a strategic configuration that can theoretically improve upon pure babbling by maintaining some information transmission, but whose existence requires very specific parameter conditions that may rarely occur in practice.

Under a non-truthful equilibrium, the information structure becomes more complex than in other configurations. Good agents truthfully reveals their signal, $m(s_i, \theta_i = 1) = s_i$, whilst a bad agents lie, $m(s_i, \theta_i = -1) = 1 - s_i$. A null message $m_j = 0$ must originate from either a good agent who observed signal zero (i.e., $x_i = G_0$) or a bad agent who observed signal one but lies (i.e., $x_i = B_1$). The probability that an agent sends a null message conditional on the state $\omega = 0$ is, $\mathbb{P}(x_i \in \{G_0, B_1\} \mid \omega = 0) = \alpha q + (1-\alpha)(1-q)$, denoted $\widetilde{q}$, whilst the conditional probability on the state $\omega = 1$ is, $\mathbb{P}(x_i \in \{G_0, B_1\} \mid \omega = 1) = 1 - \widetilde{q}$, which creates a new effective signal precision $\widetilde{q}$ that differs from the assumed signal precision q. Given the signal s_i and messages from other agent $\boldsymbol{m_{-i}} \in \mathcal{N}^{n-1}$, the induced posterior belief is as follows:

$$\phi(s_i = 0, \boldsymbol{m_{-i}}) = \frac{p}{1-p}\frac{q}{1-q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k_{-i}-n+1};$$
$$\phi(s_i = 1, \boldsymbol{m_{-i}}) = \frac{p}{1-p}\frac{1-q}{q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k_{-i}-n+1},$$

where k_{-i} denote the number of null messages observed in $\boldsymbol{m_{-i}}$.

Crucially, different agent types form different posterior beliefs from the same message profile. Recall that we have let $\boldsymbol{m}^k \in \{\boldsymbol{m} \in \mathcal{M}^n \mid \#\{m_i = 0\} = k\}$ denote a message profile containing k null messages. Given any $\boldsymbol{m}^k \in \mathcal{M}^n$, the posterior beliefs satisfy the ordering,

$\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$, where $\phi^k_{x_i}$ denote the posterior belief of agent type x_i when $\boldsymbol{m}^k$ is observed. See Appendix 3.A.3, Lemma 3.4 for proof. The hierarchy reflects how different agent types interpret the same evidence: bad agents of type $x_i = B_0$ form the highest posterior belief favouring $\omega = 0$ because they observe signal $s_i = 0$and know all null messages come from agents who received zero signals, whilst bad agents of type $x_i = B_1$ form the lowest belief because they received contradictory evidence.

The existence of non-truthful equilibrium requires both sustainable messaging strategies and sincere voting behaviour. The complete characterisation involves examining five distinct posterior belief orderings that can arise depending on the number of null messages observed, each creating different strategic incentives for deviation. The analysis reveals that sustainability depends critically on the relationship between signal precision, prior beliefs, and the effective informativeness created by systematic deception. Non-truthful equilibrium exists when signals are either highly informative or when priors dominate, but fails for intermediate parameter values where strategic manipulation becomes too profitable. Formally, the existence conditions depend on the posterior beliefs induced by extreme message profiles (i.e., all agents sending zero versus all sending one),

Proposition 3.5 (Sustainability: Non-Truthful Equilibrium)**.** *A non-truthful equilibrium exists if and only if* $(\alpha, p, q, n) \in [\frac{1}{2}, 1]^3 \times \mathbb{N}$ *(n odd) satisfies,*

(i) If $\frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{n-1} > 1$, $\frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{1-n} > 1$ *and* $p \geq q$.

(ii) If $\frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{n-1} < 1$, $\frac{p}{1-p}\frac{q}{1-q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{1-n} > 1$.

Proof. See Appendix 3.A.3. □

The non-truthful equilibrium requires very specific parameter conditions. When signals are highly precise, agents rely primarily on private information, making others' messages relatively unimportant and reducing the strategic value of deception. Conversely, when prior beliefs dominate, message content becomes largely irrelevant regardless of veracity, again making deception sustainable. The problematic region lies in the intermediate range where messages significantly influence beliefs but deception provides strategic advantages. Here, good agents' incentives to continue truth-telling conflict with bad agents' desires to mislead, potentially destroying the equilibrium altogether. When it exists, the non-truthful equilibrium enables partial information aggregation despite systematic deception. Unlike babbling equilibrium where communication breaks down completely, the predictable nature of lies allows some information extraction. However, restrictive existence conditions mean this configuration rarely improves upon simpler alternatives. The efficiency analysis demonstrates that non-truthful equilibrium typically yields better outcomes than babbling but worse than truthful revelation. The parameter regions where it both exists and provides efficiency gains prove extremely narrow, limiting practical relevance for institutional design. The equilibrium represents a fragile middle ground between complete truth-telling and communication breakdown. Whilst theoretically more beneficial than babbling, sustainability under only narrow conditions that rarely occur in practice limits its practical value.

3.4.4 Comparative Equilibrium Analysis: Existence and Sustainability

This section synthesises the existence conditions for all three equilibria by combining the analytical results derived previously with computational illustrations. The equilibrium structure is examined under varying parameter constellations (α, p, q, n), with Figures 3.1 to 3.4 providing visual partitions of the parameter space. Each shaded region in these diagrams corresponds to a distinct equilibrium regime, thereby facilitating intuitive interpretation of how the proportion of good agents, α, and signal precision, q, jointly determine equilibrium sustainability. Specifically, all equilibria exist for all (α, q) in the lightest region; only a truthfully revealing equilibrium and a babbling equilibrium exist for all (α, q) in the second lightest region; only a non-truthful equilibrium and a babbling equilibrium exist for all (α, q) in the second darkest region; only a babbling equilibrium exists for all (α, q) in the darkest region.

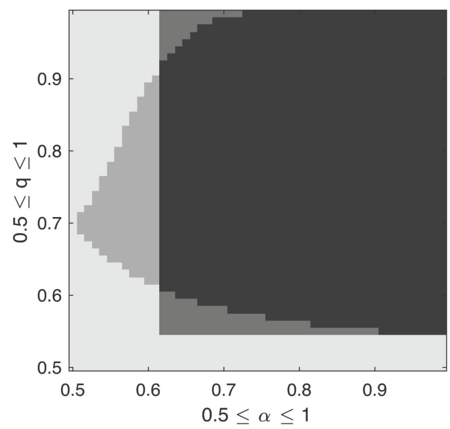

Figure 3.1: Existence: $p = 0.7; n = 5$

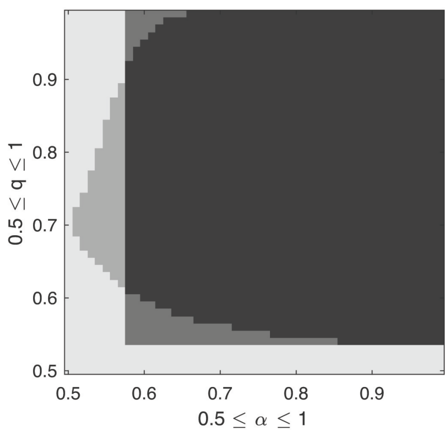

Figure 3.2: Existence: $p = 0.7; n = 7$

Figure 3.1 to 3.4 indicate that a truthfully revealing equilibrium exists if and only if (α, q) falls in

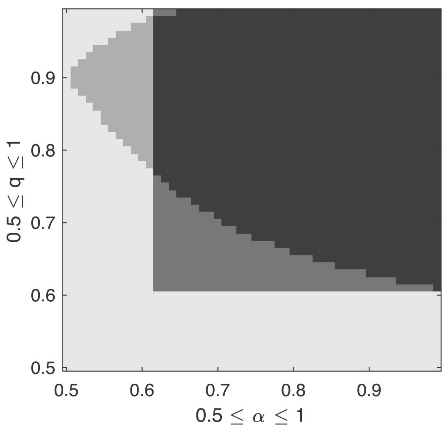

Figure 3.3: Existence: $p = 0.9$; $n = 5$

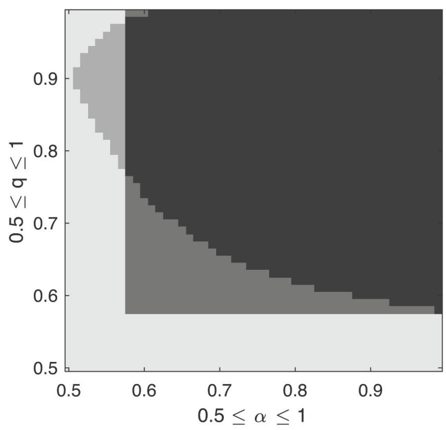

Figure 3.4: Existence: $p = 0.9$; $n = 7$

the 'L-shaped' region. In other words, it is sustainable if either α or q is sufficiently moderate (i.e., close to $\frac{1}{2}$). When α is sufficiently moderate, both good and bad agents think that they represent the majority when they are pivotal voters. Therefore, they are willing to truthfully reveal their signals since they are not much concerned about the possibility that agents of the other type exploit their signal to distort the final decision.

Recall that in Proposition 3.2, the results show that the conditions of the existence are independent of p and q, which can be proven again by Figure 3.1 to 3.4. It is clear that, given the same number of agents, the two corresponding figures associated with different priors share exactly the same vertical boundary. On the other hand, a sufficiently moderate q indicates that the signal provides limited information. Therefore, all types of agents ignore their signals and messages from others. Instead, they rely on the prior to form their posterior belief. Eventually, good agents cast a vote $v_i = 0$ and bad agents cast $v_i = 1$ independent of messages from other agents and their private signals, which implies that there exists no incentive to deviate from any conjectured message strategy and voting strategy. The horizontal boundary in each diagram indicates the highest q for the negligence of signals.

Moreover, the prior belief makes an impact on the set of equilibrium existence. By comparing Figure 3.1 and 3.3 (also Figure 3.2 and 3.4), it is clear that, with a more moderate p, we need a more moderate q to sustain the equilibrium. This is because a moderate p indicates that prior belief is not strong enough to make agents neglect their signals. In other words, with a more moderate prior, agents are willing to neglect their signals if and only if the precision q is lower.

It is also found that, with more agents (i.e., a higher n), we need a more moderate α to sustain the equilibrium. From the perspective of a bad agent, the probability that the bad represent the majority under pivotal events is decreasing in n. Roughly speaking, without conditioning on pivotal events, the probability that a bad agent thinks the bad represent the majority is:

$$\begin{aligned}\mathbb{P}\left(\#\{\theta_i = -1 \mid \boldsymbol{\theta}\} \geq \frac{n}{2}\right) &= \sum_{i=\frac{n-1}{2}}^{n-1} \binom{n-1}{i}(1-\alpha)^i \alpha^{(n-1)-i} \\ &= \begin{cases} \binom{2}{1}(1-\alpha)^1\alpha^1 + \binom{2}{2}(1-\alpha)^2\alpha^0, & \text{if } n = 3 \\ \binom{4}{2}(1-\alpha)^2\alpha^2 + \binom{4}{3}(1-\alpha)^3\alpha^1 + \binom{4}{4}(1-\alpha)^4\alpha^0, & \text{if } n = 5 \end{cases}.\end{aligned}$$

This gives the maximal value of $\alpha_{\max} = 0.7071$ such that $\mathbb{P}(\#\{\theta_i = -1 \mid \boldsymbol{\theta}\} \geq \frac{1}{2})$ when $n = 3$ and $\alpha_{\max} = 0.6143$ when $n = 5$. When $n = 11$, the value drops to $\alpha_{\max} = 0.5483$. Intuitively, with fewer agents, say $n = 3$, conditional on knowing herself as a bad agent, it is more likely that the bad is associated with the majority, since the weight of herself in the committee is $\frac{1}{3}$. With more agents, the weight decreases and thus, given the same α, the probability of being the majority decreases.

Given any $\alpha \in \left[\frac{1}{2}, 1\right]$, a non-truthful equilibrium exists if q is sufficiently moderate or more extreme. When we have a more moderate q, the signal and messages are not informative to some extent. Consequently, agents of all types tend to rely on prior belief which implies that there exists no change in posterior relation: $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ always holds for any $k \in \{0, 1, 2, ..., n\}$. When we have a more extreme q, signals are more informative. Thus, agents now stick to their signals: an agent who receives $s_i = 0$ (i.e., $x_i \in \{G_0, B_0\}$) forms a posterior belief $\phi_i > 1$, whilst an agent who receives $s_i = 1$ (i.e., $x_i \in \{G_1, B_1\}$) forms a posterior belief $\phi_i < 1$, regardless of the observed messages from others. Thus, the posterior belief is $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{0, 1, 2, ..., n\}$, which again implies that there exists no change in posterior relation. Since no one can bias the posterior belief such that some types vote differently, there is no incentive to deviate from the non-truthful message strategy.

As for the incentive to deviate from sincere voting, when we have a moderate q, which results in the posterior relation $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ for all k, by Lemma 3.7a, once $p \geq q$ is satisfied, there is no incentive for any type to deviate. Since we consider a moderate q here, $p \geq q$ holds, and so does the sustainability of sincere voting. When we have an extreme q, which results in the posterior relation $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all k, by Lemma 3.7c, there is no incentive for any type to deviate.

In addition, a non-truthful equilibrium exists when α is sufficiently moderate. When we have a moderate α, $\widetilde{q}$ converges to $\frac{1}{2}$, which implies that the informativeness of a message disappears. To see this, recall that the posterior belief induced by a message is $\mathbb{P}(\omega = 0 \mid m_i = 0) = \widetilde{q}$ and $\mathbb{P}(\omega = 0 \mid m_i = 1) = 1 - \widetilde{q}$. With $\widetilde{q}$ close to $\frac{1}{2}$, we have $\mathbb{P}(\omega = 0 \mid m_i = 0) = \mathbb{P}(\omega = 0 \mid m_i = 1) = \frac{1}{2}$. Hence, agents tend to stick to their signals and the prior belief only. For the same reason, if $p > q$, $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ is likely to be induced for all k; if $q > p$, $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ is likely to be induced for all k. Under either posterior relation, no type has incentive to deviate from the non-truthful message and sincere voting.

By Proposition 3.4, given any (α, p, q, n), a babbling equilibrium always exists. In fact, the mechanism under a babbling equilibrium is equivalent to a voting-only mechanism: each agent knows their preference and receives a signal and directly cast a vote without any forms of deliberation. She plays strategically and thus considers the pivotal voting events for a vote. Hence, the babbling equilibrium effectively serves as a benchmark to evaluate the benefit of the introduction of deliberation into the mechanism.

The conclusions indicate that a truthfully revealing equilibrium arises only when either α or q is sufficiently moderate. When α is close to $\frac{1}{2}$, both types of agents perceive themselves as likely pivotal voters, and thus truthfully transmit their private signals. A low q, in contrast, renders signals uninformative, prompting agents to ignore messages and rely on the prior belief. In both cases, strategic misrepresentation becomes unprofitable. Additionally, for a fixed n, the vertical boundaries remain constant across different priors, reflecting the independence of existence conditions from q. However, higher values of n tighten the existence region, suggesting that larger committees require more moderate α to sustain deliberation-based equilibria. These comparative statics highlight the fragility of information aggregation when either belief polarisation or signal precision intensifies.

3.5 Efficiency Analysis

This section evaluates the efficiency of the deliberation-voting mechanism by examining how likely the committee is to reach a decision that matches the true state of the world. Efficiency, in this setting, refers to the principal ex-ante expected payoff. That is, it captures the probability that the final majority decision coincides with the actual state, before any agent observes their signal or preference. This measure reflects how effectively the mechanism transforms dispersed private information into an accurate collective choice.

To assess this efficiency across strategic environments, the analysis compares the principal expected payoff under three benchmark equilibrium configurations. These are the truthfully revealing equilibrium, the non-truthful equilibrium, and the babbling equilibrium. Each of these represents a distinct pattern of communication and, consequently, a different channel through which information is aggregated. The evaluation proceeds by first characterising the decision rule under each equilibrium. That is, the mapping from signal and preference profiles, through message transmission and voting, to the final collective outcome. For each case, the expected probability that the decision matches the true state is computed by integrating over all possible combinations of agent types and signal realisations. This method provides a clean comparison of how different strategic behaviours affect the mechanism capacity to synthesise private information into accurate group decisions.

Efficiency under each equilibrium is formally defined as the ex-ante expected payoff of the principal. Let $\psi(m, v)$ denote the probability that the committee selects the alternative which matches the true state, given the message strategy m and the voting strategy v,

$$\psi_{(m,v)} = \mathbb{P}(d = \omega \mid m, v).$$

For each equilibrium configuration, the efficiency function is derived by integrating over all realisations of agent types and signal draws, accounting for how messages and votes jointly determine the final decision. Figures 3.5 to 3.7 illustrate the efficiency function ψ under the three equilibrium types: truthfully revealing equilibrium (TRE), non-truthful equilibrium (NTE), and babbling equilibrium (BE). These diagrams fix $(p, n) = (0.7, 7)$ and vary the remaining parameters to examine how equilibrium behaviour interacts with belief updating and strategic communication.

In the babbling case, agents transmit no messages. Each agent votes based only on their private signal and prior belief, conditioning their decision on the expected distribution of preferences. This configuration serves as a baseline in which information aggregation is limited to decentralised signal-based voting. Comparisons with the other two equilibria reveal the informational value of deliberation, particularly when truthful or structured deception affects belief formation.

For any given parameter quadruple (α, p, q, n), efficiency typically satisfies the ordering $\psi_{TRE} \geq \psi_{NTE} \geq \psi_{BE}$. In rare cases, however, when the message informativeness q lies sufficiently close to the prior p and the signal precision α approaches $\frac{1}{2}$, the non-truthful equilibrium may yield higher efficiency than the truthfully revealing one. That is, $\psi_{NTE} \geq \psi_{TRE}$. Yet, in these cases, the specified parameters fail to support equilibrium existence. Figure 3.8 confirms this tension: the region enclosed by the grey contour corresponds to parameter values for which $\psi_{NTE} \geq \psi_{TRE}$, but this region lies strictly within the parameter space where the non-truthful equilibrium does not exist. As a result, although a non-truthful equilibrium might appear more efficient in certain parameter ranges, such configurations are not sustainable in equilibrium. Therefore, the efficiency comparison can be restricted without loss of generality to the truthfully revealing and babbling equilibria.[1]

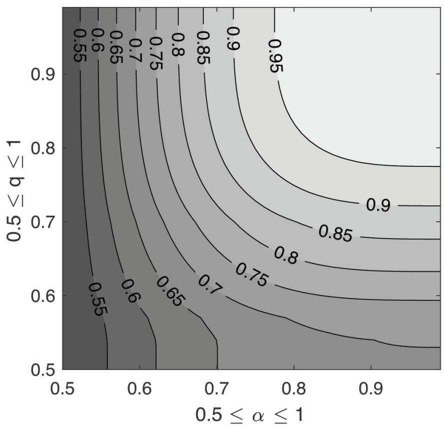

Figure 3.5: ψ_{TRE}: $p = 0.7; n = 7$

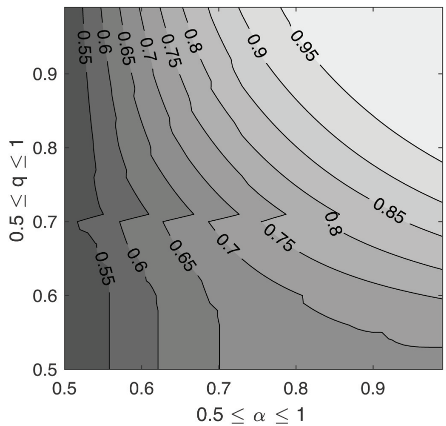

Figure 3.6: ψ_{NTE}: $p = 0.7; n = 7$

3.5.1 Optimal Information Aggregation: Truthfully Revealing Equilibrium

Under the truthfully revealing equilibrium, deliberative mechanisms achieve their theoretical ideal of perfect information aggregation. Each agent truthfully reports their private signal $m(s_i, \theta_i) = s_i$, enabling complete information sharing within the committee. Such transparency transforms the voting stage into pure implementation: agents vote according to posterior beliefs formed from the complete signal profile, with good agents supporting the more likely state and bad agents voting in opposition. The efficiency analysis of the equilibrium serves multiple theoretical purposes. First, it establishes an upper bound for deliberative performance under opposed preferences,

[1]Note that due to discretisation of (α, q) and interpolation by the computational software, the bounded set appears to include some (α, q) combinations for which all equilibria exist (i.e., the lightest region). However, the underlying matrix results remain consistent with the interpretation presented above.

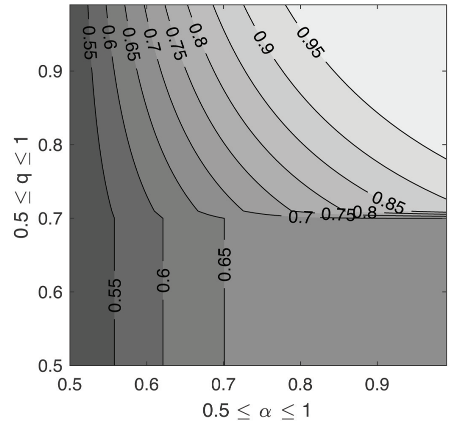

Figure 3.7: ψ_{BE}: $p = 0.7; n = 7$

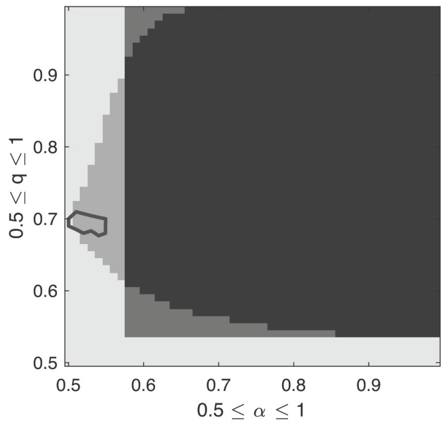

Figure 3.8: No existence for $\psi_{NTE} > \psi_{TRE}$: $p = 0.7; n = 7$

providing a benchmark against which other configurations can be evaluated. Second, it reveals how preference heterogeneity fundamentally alters the information aggregation process compared to standard jury models. Third, the mathematical structure illuminates the precise conditions under which deliberation provides value over prior-based decision-making.

The efficiency of truthfully revealing equilibrium depends on the complex interaction between signal realisations, preference composition, and the decision threshold k^*. Each possible type profile generates a probability-weighted contribution to overall efficiency, accounting for both the likelihood of correct information aggregation and the preference-based voting which determines final outcomes. Note that the symmetric case $\alpha = \frac{1}{2}$ merits particular attention given the existence conditions established previously. As demonstrated in the earlier analysis and illustrated in the equilibrium existence figures (i.e., Figures 3.1 to 3.4), truthfully revealing equilibrium exists if and only if either the preference parameter α or the signal precision q is sufficiently moderate. Understanding mechanism performance when α approaches $\frac{1}{2}$ is therefore essential, since the boundary characterises the region in which truthful revelation remains sustainable. The efficiency implications are captured in the following theorem:

Proposition 3.6 (Efficiency: Truthfully Revealing Equilibrium)**.** *Given (α, p, q, n), the efficiency under a truthfully revealing equilibrium is:*

$$\psi_{TRE}(\alpha, p, q, n) = \textit{[Detailed expression in Appendix 3.7]}$$

Moreover, when preference composition is symmetric (i.e., $\alpha = \frac{1}{2}$), the efficiency simplifies to,

$$\psi_{TRE}\left(\frac{1}{2}, p, q, n\right) = \frac{1}{2}.$$

The result holds for any parameter values (p, q, n) and represents a fundamental symmetry property of deliberative mechanisms under opposed preferences.

Proof. See Appendix 3.A.4. □

The result reveals a crucial insight about information aggregation under preference opposition. When good and bad agents are equally represented, the informational benefits from perfect signal sharing exactly offset the distortions introduced by systematic preference opposition. Such balance creates a natural efficiency benchmark which any deliberative mechanism must exceed to justify its adoption.

The general efficiency function demonstrates that improvements over the benchmark require either informational advantages (i.e., signals more precise than priors) or compositional advantages through sufficient prevalence of well-intentioned agents. However, the existence conditions derived in the previous section create a fundamental tension: the parameter regions supporting truthful revelation often coincide with those providing minimal efficiency gains, highlighting the practical limitations of theoretically optimal mechanisms.

3.5.2 Baseline Performance: Babbling Equilibrium

Under babbling equilibrium, deliberation completely fails to serve its informational purpose. Agents send uninformative messages regardless of their private signals or preference types, effectively reducing the mechanism to pure voting without any information aggregation benefits. The term 'babbling' captures the essence of communication breakdown: agents speak but convey no meaningful information about the underlying state.

The babbling equilibrium serves as the natural baseline for evaluating deliberative mechanisms. Since it represents the worst-case scenario in which communication provides no value, any improvement over babbling efficiency demonstrates the informational benefits of structured deliberation. Moreover, the universal existence of babbling equilibrium across all parameter configurations makes it a robust benchmark against which other equilibria can be assessed. When messages convey no information, each agent forms posterior beliefs using only their private signal and the common prior. The voting behaviour depends critically on the relationship between prior strength and signal precision. When the prior dominates signal informativeness (i.e., $p \geq q$), all agent types effectively ignore their private information and vote according to preference type alone. Conversely, when signals prove more informative than the prior (i.e., $p < q$), agents vote according to their signal realisations, creating coalitions based on evidence rather than preferences.

The resulting decision patterns reflect the complete absence of information aggregation through deliberation. Good agents vote for the alternative they believe most likely to be correct, whilst bad agents vote for the opposite, but without any coordination or information sharing to improve collective assessment. The analysis of these voting patterns yields the following efficiency characterisation:

Proposition 3.7 (Efficiency: Babbling Equilibrium)**.** *Given (α, p, q, n), the efficiency under babbling equilibrium is:*

$$\psi_{BE}(\alpha, p, q, n) = \begin{cases} p - (2p-1)\sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i}\alpha^i(1-\alpha)^{n-i}, & \text{if } p \geq q \\ \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i}\widetilde{q}^i(1-\widetilde{q})^{n-i}, & \text{if } p < q \end{cases}$$

where $\widetilde{q} = \alpha q + (1-\alpha)(1-q)$.

Proof. See Appendix 3.A.4. □

The babbling equilibrium establishes the efficiency floor for deliberative mechanisms under opposed preferences. Since the equilibrium exists universally across all parameter combinations, it represents the guaranteed outcome when more sophisticated communication strategies prove unsustainable. The efficiency analysis reveals that pure voting without deliberation can still achieve reasonable performance when preference composition is favourable or when signal-preference alignment creates effective information aggregation through voting patterns alone. However, the analysis also demonstrates the fundamental limitations of mechanisms without information sharing. When signals are only moderately informative or when preference composition is unfavourable, babbling efficiency can fall substantially below the theoretical optimum, highlighting the potential value of successful deliberative institutions whilst simultaneously illustrating the risks when such institutions fail to function as intended.

3.5.3 The Value of Deliberation

The central question for any information aggregation mechanism concerns whether deliberation improves upon simpler alternatives. Whilst truthfully revealing equilibrium achieves optimal information sharing and babbling equilibrium represents pure voting without communication, understanding when the former dominates the latter illuminates the informational benefits of successful deliberation under opposed preferences. The efficiency comparison between truthful revelation and babbling reveals the upper bound of deliberation benefits. Since truthful revelation achieves perfect information aggregation whilst babbling eliminates such aggregation entirely, the difference between their efficiency levels captures the maximum value that communication can provide when strategic considerations are taken into account.

The analysis demonstrates that truthful revelation consistently dominates babbling when both equilibria exist. The intuition proves straightforward: good agents benefit from aggregating private signals through deliberation rather than relying solely on individual information, whilst the majority status of good agents ($\alpha \geq \frac{1}{2}$) ensures that coordinated information sharing yields more precise collective beliefs than isolated decision-making. From the perspective of mechanism design, truthful communication allows good agents to coordinate effectively, mitigating the noise introduced by bad agents who vote against posterior beliefs. The informational advantages from structured deliberation systematically outweigh the distortions created by preference opposition, provided that the deliberative process itself remains sustainable.

A good agent naturally prefers to aggregate private signals through deliberation rather than rely solely on their own signal. Although bad agents vote against the posterior belief and could potentially undermine the benefit of truth-telling, the assumption that $\alpha \geq \frac{1}{2}$ ensures that good agents constitute the majority in expectation. Therefore, exchanging signals through deliberation yields more precise collective beliefs than simply pooling individual signals via voting alone. Accordingly, we have the following efficiency comparison:

Proposition 3.8 (Deliberation Efficiency Gains)**.** *Given any (α, p, q, n) supporting truthfully revealing equilibrium,*

$$\psi_{TRE}(\alpha, p, q, n) > \psi_{BE}(\alpha, p, q, n).$$

The efficiency improvement reflects the superior information aggregation achieved through credible *communication relative to pure voting mechanisms.*

Proof. See Appendix 3.A.4. □

The efficiency differential between truthful revelation and babbling depends on the specific decision patterns that emerge under each equilibrium. Under truthful revelation, the committee reaches decision $d = 0$ when either the evidence genuinely supports such choice and good agents implement it, or when the evidence points against it but bad agents strategically vote in the opposite direction. Such dual pathway to correct decisions reflects the complex interplay between information aggregation and preference conflict. In contrast, babbling equilibrium produces decisions based purely on preference composition when priors dominate (i.e., $p \geq q$), or on signal-preference alignment when signals prove more informative (i.e., $p < q$). The absence of information sharing eliminates the sophisticated belief updating that characterises successful deliberation, reducing collective decision-making to mechanical aggregation of individual assessments.

The dominance of truthful revelation over babbling establishes a fundamental result about deliberative mechanisms under opposed preferences. Despite the presence of agents who systematically vote against available evidence, the informational benefits of communication prove robust when truthful revelation can be sustained. Such robustness suggests that the traditional emphasis on preference alignment in mechanism design, whilst important, may not be as critical as ensuring credible information transmission. However, the sustainability conditions for truthful revelation create important caveats to this optimistic assessment. The efficiency gains from deliberation materialise only when the underlying strategic environment supports honest communication, requiring either moderate preference heterogeneity or limited signal precision. When these conditions fail, the mechanism defaults to babbling equilibrium, eliminating any informational advantages whilst retaining the costs of deliberative processes.

The analysis thus reveals deliberation as a double-edged institutional tool: highly beneficial when functioning properly, but potentially wasteful when strategic considerations undermine honest communication. Such characterisation underscores the importance of institutional design choices that promote truthful revelation rather than merely facilitating communication opportunities.

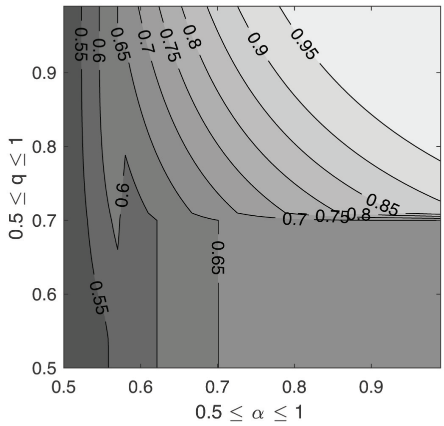

Figure 3.9: Highest possible efficiency given existence: $p = 0.7$; $n = 7$

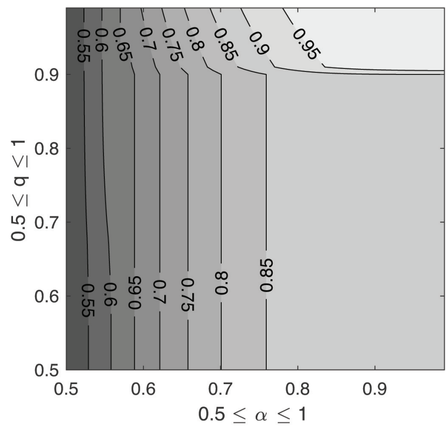

Figure 3.10: Highest possible efficiency given existence: $p = 0.9$; $n = 7$

3.5.4 The Limits of Deliberation

The ultimate test of any information aggregation mechanism lies not merely in whether it outperforms alternative institutional arrangements, but in whether it improves upon the status quo of relying solely on existing knowledge. In our context, this benchmark is represented by the prior belief fo the principal p, which reflects their ex-ante assessment of the decision environment without consulting external agents. The economic intuition suggests a fundamental trade-off: whilst deliberative mechanisms can harness dispersed private information to enhance decision quality, they also introduce strategic distortions when agents possess opposed preferences. The presence of bad agents creates a coordination problem that may offset the informational benefits of deliberation. Moreover, the costs of establishing and operating a committee, both explicit monetary costs and implicit coordination costs, are only justified if the mechanism delivers superior outcomes relative to unilateral decision-making based on prior beliefs alone.

Now we investigate whether the mechanism yields a higher efficiency compared to the prior, p. Note that the prior distribution represents the ex-ante expected payoff of the principal (i.e., the efficiency) if they decide to stick to the prior and thus choose $d = 0$ without reliance on the mechanism to aggregate any further information from agents. If the mechanism does not yield a better efficiency greater than p, then there is no reason for the principal to seek help from outside agents.

Figures 3.9 and 3.10 show the highest possible efficiency under an equilibrium such that the equilibrium is sustainable given (α, q). The figures demonstrate that the deliberation-voting mechanism improves efficiency in decision-making if and only if $q \geq p$. Additionally, it requires that the likelihood of recruiting a good agent, α, is sufficiently high. For example, in a 7-agent committee with a prior $p = 0.7$, the principal will definitely not benefit from this mechanism if α is less than approximately 0.61.

By Proposition 3.7, if the informativeness of a signal does not exceed the prior (i.e., $p \geq q$), the efficiency under babbling equilibrium is lower than the prior (i.e., $\psi_{BE} < p$). Therefore, the principal should not establish a committee for decision-making if a truthfully revealing equilibrium does not exist, since babbling equilibrium yields worse expected payoff than adhering to the prior. Nevertheless, as discussed in the previous section, truthfully revealing equilibrium exists if and only if either α or q is moderate. Given moderate α or q, it remains unlikely that the improvement in efficiency mentioned in Proposition 3.8 ensures that efficiency exceeds the prior. Hence, we can conclude that the mechanism is unlikely to enhance efficiency in decision-making when $p \geq q$.

In contrast, if $p < q$, truthfully revealing equilibrium exists if and only if α is sufficiently close to $\frac{1}{2}$, which implies that even if truthfully revealing equilibrium exists, by Proposition 3.6, the corresponding efficiency remains close to $\frac{1}{2} < p$. Alternative equilibria exist if either α is sufficiently moderate, yielding efficiency lower than p, or q is exceptionally close to 1. For example, suppose that $n = 7$ and $p = 0.7$. An alternative equilibrium with efficiency greater than p exists only if $q > 0.98$ and $\alpha \in (0.6, 0.65)$. Beyond this narrow region, there exists no (α, q) such that truthfully revealing equilibrium with greater-than-prior efficiency exists. Unless we aim to focus on such extreme cases, analysing efficiency under babbling equilibrium proves sufficient.

By Proposition 3.7, the efficiency under a babbling equilibrium when $p < q$ increases in α, since $\widetilde{q}$ is an increasing function of α. Note that $\widetilde{q} = \frac{1}{2}$ if $\alpha = \frac{1}{2}$, which implies that the efficiency under a babbling equilibrium with $\alpha = \frac{1}{2}$ is:

$$\psi_{BE}(\alpha = \frac{1}{2}, p, q, n) = \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^{i} \left(\frac{1}{2}\right)^{n-i} = \frac{1}{2}.$$

Besides, note that $\widetilde{q} = q$ if $\alpha = 1$, which implies that the efficiency under a babbling equilibrium when $\alpha = 1$ is:

$$\psi_{BE}(\alpha = 1, p, q, n) = \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q^{i} (1-q)^{n-i} > q > p.$$

See Appendix for a formal proof (Lemma 3.15). Hence, by continuity, given any $q \in (p, 1]$, there exists a cut-off point, denoted as $\alpha^*(q)$, such that the efficiency under a babbling equilibrium is greater than p if and only if $\alpha \geq \alpha^*(q)$. Consequently, the mechanism improves the efficiency under a babbling equilibrium only if α is sufficiently large (i.e., greater than $\alpha^*(q)$). In addition, as shown in Figure 3.7, $\alpha^*(q)$ is decreasing in q, which indicates that we need a higher α with less informativeness of a signal.

Proposition 3.9 (Conditions for Mechanism Efficiency)**.** *Consider the deliberation-voting mechanism with opposed preferences. The mechanism yields efficiency strictly greater than the prior belief, p, if and only if the following conditions are satisfied:*

(i) signal informativeness exceeds the prior: $q > p$;

(ii) the proportion of good agents exceeds a signal-dependent threshold: $\alpha \geq \alpha^*(q)$, where $\alpha^*(q)$ is decreasing in q and satisfies $\alpha^*(q) \in \left(\frac{1}{2}, 1\right)$ for all $q \in (p, 1]$.

Moreover, when $q \leq p$, the mechanism fails to improve upon the prior for any $\alpha \in \left[\frac{1}{2}, 1\right]$, except in the degenerate case where q is exceptionally close to 1 and α lies within a narrow interval near $\frac{1}{2}$.

Proof. The proof follows directly from the efficiency analysis in the previous section. When $q \leq p$, Preposition 3.7 establishes that $\psi_{BE} < p$ under babbling equilibrium, whilst conditions for truthfully revealing and non-truthful equilibria become restrictive, yielding efficiency gains only in exceptional parameter regions. When $q > p$, the continuity argument establishes the existence of the threshold $\alpha^*(q)$, with the decreasing property following from the monotonicity of $\widetilde{q}$ in α. The boundary conditions $\alpha^*(\cdot) \in \left(\frac{1}{2}, 1\right)$ follow from the efficiency expressions evaluated at $\alpha = \frac{1}{2}$ and $\alpha = 1$. □

The analysis culminates in a sobering conclusion about the practical value of deliberative mechanisms in environments with opposed preferences. The fundamental theorem establishing conditions for mechanism efficiency captures the essence of the trade-off between information aggregation benefits and strategic distortions. The requirement that signal informativeness exceed the prior ($q > p$) reflects the basic principle that private information must be sufficiently valuable to justify the risks of strategic manipulation—if agents' signals are no better than existing knowledge, there is no reason to expose the decision process to potential deception. The second condition, requiring a sufficiently high proportion of good agents ($\alpha \geq \alpha^*(q)$), embodies the core tension in collective decision-making: the informational benefits from truthful participants must outweigh the harm from those who actively mislead. The decreasing relationship between $\alpha^*(q)$ and signal quality reveals an intuitive economic principle—more precise signals reduce the required proportion of well-intentioned agents, as high-quality information becomes sufficiently valuable to overcome moderate levels of strategic distortion. Yet even this relationship offers limited practical comfort, as the efficiency gains prove modest except under restrictive parameter combinations, suggesting that in many realistic scenarios characterised by moderate signal precision and heterogeneous preferences, principals achieve superior outcomes by foregoing deliberation entirely and relying on their prior beliefs.

3.6 Conclusion

This paper addresses whether deliberative mechanisms enhance decision quality when participants possess opposed preferences. The analysis reveals that preference opposition fundamentally alters the conditions under which deliberation succeeds or fails. The equilibrium analysis identifies three strategic configurations. Truthfully revealing equilibria emerge when either the proportion of good agents is moderate or signal precision approaches neutrality. Non-truthful equilibria exist under restrictive conditions, enabling partial information aggregation through predictable deception. Babbling equilibria emerge universally when more informative equilibria prove unsustainable. The efficiency analysis provides sobering insights. Whilst truthfully revealing equilibria dominate alternatives when they exist, deliberation improves upon prior-based decision-making only under restrictive conditions: signal informativeness must exceed prior probability and the likelihood of recruiting well-intentioned agents must be sufficiently high. When either condition fails, principals achieve superior outcomes by foregoing deliberation entirely.

These findings carry profound implications for organisational design. Deliberative mechanisms may prove counterproductive in settings where preference alignment cannot be guaranteed. Consider corporate boards: extensive deliberation benefits aligned directors with high-quality information, but generates inferior outcomes when directors hold conflicting positions. The analysis reveals a fundamental tension between inclusivity and effectiveness. Deliberative mechanisms perform best when participants share common objectives—precisely when deliberation may be least necessary. For institutional designers, the results suggest focusing on participant selection and information quality rather than assuming more voices improve decisions.

This paper advances understanding of information aggregation under strategic communication. The opposed preference framework departs significantly from standard jury models, allowing systematic analysis of environments where agents prefer decisions contradicting available evidence. The complete characterisation of equilibrium existence provides precise boundaries for when different forms of deliberation succeed. The efficiency analysis constitutes the primary advance. Previous research established incomplete conditions for truthful deliberation but provided no analysis of whether such deliberation improves decision quality. This study establishes precise conditions under which deliberative mechanisms improve upon alternatives, revealing the tension between equilibrium sustainability and efficiency gains.

The implications of this analysis extend far beyond the theoretical framework to challenge fundamental assumptions about collective decision-making. The conventional wisdom that 'more deliberation is better' is fundamentally flawed. This research demonstrates that in realistic organisational settings with conflicting interests, deliberation often makes decisions worse, not better, because the conditions required for deliberation to help require perfect information sharing and overwhelming majorities of well-intentioned participants, circumstances which rarely occur in practice. The practical implication is stark: organisations should be highly selective about when to use deliberative processes, investing instead in better information gathering and participant screening rather than defaulting to committees and consensus-building. When leaders cannot ensure both high-quality information and aligned incentives, making decisions based on existing knowledge often produces superior outcomes than exposing the process to strategic manipulation. Such findings fundamentally challenge how we think about democratic decision-making and organisational governance, suggesting that inclusive deliberation, whilst normatively appealing, may systematically underperform more selective or hierarchical approaches in environments with significant preference conflicts.

The framework opens several promising research directions. Semi-separating equilibria may offer intermediate possibilities through asymmetric strategies. Multidimensional preferences may fundamentally alter strategic calculus. Empirical validation may provide crucial insights through laboratory experiments and field studies.

Appendix 3.A: Omitted Proofs

3.A.1 Truthfully Revealing Equilibrium: Existence

Proof of Lemma 3.2. The following shows the results that the posterior belief conditional on pivotal voting, denoted as $\widetilde{\phi} : \mathcal{S} \times \mathcal{M}^{n-1} \to \mathbb{R}_+$, is equal to the unconditional posterior belief under truthfully revealing equilibrium.

$$
\begin{aligned}
\widetilde{\phi}_i \equiv \widetilde{\phi}(s_i, \boldsymbol{m}_{-i}) &\equiv \frac{\sum_{\boldsymbol{n}_{-i} \in \mathcal{N}_V^{TRE}} \mathbb{P}(\omega = 0, s_i, \boldsymbol{n}_{-i})}{\sum_{\boldsymbol{n}_{-i} \in \mathcal{N}_V^{TRE}} \mathbb{P}(\omega = 1, s_i, \boldsymbol{n}_{-i})} \\
&= \frac{p\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}} q^k (1-q)^{n-k}}{(1-p)\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}(1-q)^k q^{n-k}} \\
&= \frac{pq^k(1-q)^{n-k}}{(1-p)(1-q)^k q^{n-k}} = \phi(s_i, \boldsymbol{m}_{-i}),
\end{aligned}
$$

where k denotes the number of null messages.

□

Proof of Proposition 3.2. Recall that the condition that an agent of type G_0 has no incentive to deviate from truthfully deliberating is:

$$
\begin{aligned}
&\sum_{\substack{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_0} = k^* - 1; \\ n_{G_1} + n_{B_1} = (n-1)(k^*-1); \\ n_{G_0} + n_{G_1} \geq \frac{n-1}{2}\}}} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n}_{-i}) + \sum_{\substack{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_0} = k^* - 1; \\ n_{G_1} + n_{B_1} = (n-1)(k^*-1); \\ n_{B_0} + n_{B_1} > \frac{n-1}{2}\}}} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n}_{-i}) \\
\geq &\sum_{\substack{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_0} = k^* - 1; \\ n_{G_1} + n_{B_1} = (n-1)(k^*-1); \\ n_{B_0} + n_{B_1} > \frac{n-1}{2}\}}} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n}_{-i}) + \sum_{\substack{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_0} = k^* - 1; \\ n_{G_1} + n_{B_1} = (n-1)(k^*-1); \\ n_{G_0} + n_{G_1} \geq \frac{n-1}{2}\}}} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n}_{-i}),
\end{aligned}
$$

which implies:

$$
\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}} \binom{n-1}{k^*-1}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}} pq\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}n_{B_1}} q^{k^*-1}(1-q)^{(n-1)-(k^*-1)} \\
+ &\sum_{\{\boldsymbol{n}_{-i}|\cdot\}} \binom{n-1}{k^*-1}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}} (1-p)(1-q)\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}} q^{(n-1)-(k^*-1)}(1-q)^{k^*-1} \\
\geq &\sum_{\{\boldsymbol{n}_{-i}|\cdot\}} \binom{n-1}{k^*-1}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}} pq\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}} q^{k^*-1}(1-q)^{(n-1)-(k^*-1)}
\end{aligned}
$$

$$+\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{n-1}{k^*-1}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}(1-p)(1-q)\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}q^{(n-1)-(k^*-1)}(1-q)^{k^*-1}.$$

which further implies:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}pq\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}q^{k^*-1}(1-q)^{(n-1)-(k^*-1)}\\
&+\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}(1-p)(1-q)\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}q^{(n-1)-(k^*-1)}(1-q)^{k^*-1}\\
&\geq\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}pq\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}q^{k^*-1}(1-q)^{(n-1)-(k^*-1)}\\
&+\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}(1-p)(1-q)\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}q^{(n-1)-(k^*-1)}(1-q)^{k^*-1}.
\end{aligned}$$

which further implies:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}[pq\cdot q^{k^*-1}(1-q)^{(n-1)-(k^*-1)}\\
&-(1-p)(1-q)q^{(n-1)-(k^*-1)}(1-q)^{k^*-1}]\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}\\
\geq&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}[pq\cdot q^{k^*-1}(1-q)^{(n-1)-(k^*-1)}\\
&-(1-p)(1-q)q^{(n-1)-(k^*-1)}(1-q)^{k^*-1}]\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}.
\end{aligned}$$

which finally implies:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}\\
\geq&\sum_{\{\boldsymbol{n}_{-i}|\cdot\}}\binom{k^*-1}{n_{G_0}}\binom{(n-1)-(k^*-1)}{n_{G_1}}\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}.
\end{aligned}$$

Note that the LHS of the simplified expression is the probability that good agents represent the majority whilst the RHS is the probability that the bad agents represent the majority, when the agent i is a pivotal sender. Following the same argument, the conditions that an agent of type G_1, B_0, and B_1 has no incentive to deviate are:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i}|n_{G_0}+n_{B_0}=k^*;n_{G_1}+n_{B_1}=(n-1)-k^*;n_{G_0}+n_{G_1}\geq\frac{n-1}{2}\}}\binom{k^*}{n_{G_0}}\binom{(n-1)-k^*}{n_{G_1}}\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}}\\
\geq&\sum_{\{\boldsymbol{n}_{-i}|n_{G_0}+n_{B_0}=k^*;n_{G_1}+n_{B_1}=(n-1)-k^*;n_{B_0}+n_{B_1}>\frac{n-1}{2}\}}\binom{k^*}{n_{G_0}}\binom{(n-1)-k^*}{n_{G_1}}\alpha^{n_{G_0}+n_{G_1}}(1-\alpha)^{n_{B_0}+n_{B_1}},
\end{aligned}$$

$$
\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i} | n_{G_0}+n_{B_0}=k^*-1; n_{G_1}+n_{B_1}=(n-1)-(k^*-1); n_{B_0}+n_{B_1} \geq \frac{n-1}{2}\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} \\
\geq &\sum_{\{\boldsymbol{n}_{-i} | n_{G_0}+n_{B_0}=k^*-1; n_{G_1}+n_{B_1}=(n-1)-(k^*-1); n_{G_0}+n_{G_1} > \frac{n-1}{2}\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}},
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{\{\boldsymbol{n}_{-i} | n_{G_0}+n_{B_0}=k^*; n_{G_1}+n_{B_1}=(n-1)-k^*; n_{B_0}+n_{B_1} \geq \frac{n-1}{2}\}} \binom{k^*}{n_{G_0}} \binom{(n-1)-k^*}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} \\
\geq &\sum_{\{\boldsymbol{n}_{-i} | n_{G_0}+n_{B_0}=k^*; n_{G_1}+n_{B_1}=(n-1)-k^*; n_{G_0}+n_{G_1} > \frac{n-1}{2}\}} \binom{k^*}{n_{G_0}} \binom{(n-1)-k^*}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}},
\end{aligned}
$$

respectively. Hence, we have the following explicit expressions for the four conditions.

$$
\begin{aligned}
&\sum_{i=0}^{k^*-1} \binom{k^*-1}{i} \left[-\sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*} \binom{n-k^*}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} \right] \geq 0, \\
&\sum_{i=0}^{k^*} \binom{k^*}{i} \left[-\sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*-1}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*-1} \binom{n-k^*-1}{j} \alpha^{i+j} (1-\alpha)^{n-1-i-j} \right] \geq 0, \\
&\sum_{i=0}^{k^*-1} \binom{k^*-1}{i} \left[-\sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*} \binom{n-k^*}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} \right] \geq 0, \\
&\sum_{i=0}^{k^*} \binom{k^*}{i} \left[-\sum_{j=0}^{\frac{n-1}{2}-i-1} \binom{n-k^*-1}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} + \sum_{j=\frac{n-1}{2}-i}^{n-k^*-1} \binom{n-k^*-1}{j} (1-\alpha)^{i+j} \alpha^{n-1-i-j} \right] \geq 0.
\end{aligned}
$$

Now we show that the first two inequalities and the last two inequalities are equivalent, respectively. We first rearrange the LHS of an inequality by sorting $\alpha^i(1-\alpha)^{n-1-i}$ for each i. We investigate the absolute value of the coefficients of $\alpha^i(1-\alpha)^{n-1-i}$ for each i for the first two inequalities and find that they are exactly the same. Since, for each i, $\alpha^i(1-\alpha)^{n-1-i}$ shares the same sign in the two inequalities, it is sufficient to investigate the absolute value of the coefficients. First suppose that $i \leq k^*-1$. The coefficient for the first inequality can be expressed as $\sum_{j=0}^{i} \binom{k^*-1}{j}\binom{n-k^*}{i-j}$, whilst for the second inequality it is $\sum_{j=0}^{i} \binom{k^*}{j}\binom{n-k^*-1}{i-j}$. By the Binomial Theorem, we can show that the two inequalities are equivalent:

$$
\begin{aligned}
\sum_{j=0}^{i} \binom{k^*-1}{j}\binom{n-k^*}{i-j} &= \sum_{j=0}^{i-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j} + \binom{k^*-1}{i}\binom{n-k^*}{0} \\
&= \sum_{j=0}^{i-1} \binom{k^*-1}{j} \left[\binom{n-k^*-1}{i-j} + \binom{n-k^*-1}{i-j-1} \right] + \binom{k^*-1}{i}\binom{n-k^*-1}{0}
\end{aligned}
$$

$$= \sum_{j=0}^{i} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=0}^{i-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j-1}.$$

$$\begin{aligned}
\sum_{j=0}^{i} \binom{k^*}{j}\binom{n-k^*-1}{i-j} &= \sum_{j=1}^{i} \binom{k^*}{j}\binom{n-k^*-1}{i-j} + \binom{k^*}{0}\binom{n-k^*-1}{i} \\
&= \sum_{j=1}^{i} \left[\binom{k^*-1}{j}\binom{k^*-1}{j-1}\right]\binom{n-k^*-1}{i-j} + \binom{k^*-1}{0}\binom{n-k^*-1}{i} \\
&= \sum_{j=0}^{i} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=1}^{i} \binom{k^*-1}{j-1}\binom{n-k^*-1}{i-j} \\
&= \sum_{j=0}^{i} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=0}^{i-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j-1}.
\end{aligned}$$

Suppose that $k^* \leq i \leq n-1-k^*$. The coefficient for the first inequality can be expressed as $\sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j}$, whilst for the second inequality it is $\sum_{j=0}^{k^*} \binom{k^*}{j}\binom{n-k^*-1}{i-j}$. Thus, we have:

$$\begin{aligned}
\sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j} &= \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\left[\binom{n-k^*-1}{i-j} + \binom{n-k^*-1}{i-j-1}\right] \\
&= \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j-1}.
\end{aligned}$$

$$\begin{aligned}
\sum_{j=0}^{k^*} \binom{k^*}{j}\binom{n-k^*-1}{i-j} &= \binom{k^*}{k^*}\binom{n-k^*-1}{i} + \sum_{j=1}^{k^*-1} \binom{k^*}{j}\binom{n-k^*-1}{i-j} + \binom{k^*}{0}\binom{k^*}{i-k^*} \\
&= \binom{k^*-1}{0}\binom{n-k^*-1}{i} + \sum_{j=1}^{k^*-1} \left[\binom{k^*-1}{j}\binom{k^*-1}{j-1}\right]\binom{n-k^*-1}{i-j} + \binom{k^*-1}{k^*-1}\binom{k^*}{i-k^*} \\
&= \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=1}^{k^*} \binom{k^*-1}{j-1}\binom{n-k^*-1}{i-j} \\
&= \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=0}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j-1}.
\end{aligned}$$

Suppose that $i \geq n-k^*$. The coefficient for the first inequality can be expressed as $\sum_{j=i-(n-k^*)}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j}$, whilst for the second inequality it is $\sum_{j=i-(n-k^*-1)}^{k^*} \binom{k^*}{j}\binom{n-k^*-1}{i-j}$. Thus, we have:

$$\begin{aligned}
\sum_{j=i-(n-k^*)}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j} &= \binom{k^*-1}{i-(n-k^*)}\binom{n-k^*}{n-k^*} + \sum_{j=i-(n-k^*)+1}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*}{i-j} \\
&= \binom{k^*-1}{i-(n-k^*)}\binom{n-k^*}{n-k^*} + \sum_{j=i-(n-k^*)+1}^{k^*-1} \binom{k^*-1}{j}\left[\binom{n-k^*-1}{i-j} + \binom{n-k^*-1}{i-j-1}\right] \\
&= \sum_{j=i-(n-k^*)+1}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j} + \sum_{j=i-(n-k^*)}^{k^*-1} \binom{k^*-1}{j}\binom{n-k^*-1}{i-j-1}.
\end{aligned}$$

$$\sum_{j=i-(n-k^*-1)}^{k^*} \binom{k^*}{j}\binom{n-k^*-1}{i-j} = \sum_{j=i-(n-k^*-1)}^{k^*-1} \binom{k^*}{j}\binom{n-k^*-1}{i-j} + \binom{k^*}{k^*}\binom{n-k^*-1}{i-k^*}$$

$$= \sum_{j=i-(n-k^*-1)}^{k^*-1} \left[\binom{k^*-1}{j} + \binom{k^*-1}{j-1} \right] \binom{n-k^*-1}{i-j} + \binom{k^*-1}{k^*-1} \binom{n-k^*-1}{i-k^*}$$
$$= \sum_{j=i-(n-k^*-1)}^{k^*-1} \binom{k^*-1}{j} \binom{n-k^*-1}{i-j} + \sum_{j=i-(n-k^*-1)}^{k^*} \binom{k^*-1}{j-1} \binom{n-k^*-1}{i-j}$$
$$= \sum_{j=i-(n-k^*-1)}^{k^*-1} \binom{k^*-1}{j} \binom{n-k^*-1}{i-j} + \sum_{j=i-(n-k^*)}^{k^*-1} \binom{k^*-1}{j} \binom{n-k^*-1}{i-j-1}.$$

Since the coefficients of $\alpha^i(1-\alpha)^{n-1-i}$ are identical, the two inequalities are equivalent. Likewise, it can be proven that the last two inequalities are equivalent as well. Now we show that the inequalities only depend on α and n. That is, they are independent of p and q. The only proof required for this claim is to prove the inequalities are independent of k^*. Suppose that $k^* = \bar{k}$. As mentioned, after rearrangement, we can rewrite the first inequality as:

$$-\sum_{i=0}^{\bar{k}-1} \left[\sum_{j=0}^{i} \binom{\bar{k}-1}{j} \binom{n-\bar{k}}{i-j} \alpha^i (1-\alpha)^{n-1-i} \right] - \sum_{i=\bar{k}}^{\frac{n-1}{2}-1} \left[\sum_{j=0}^{\bar{k}-1} \binom{\bar{k}-1}{j} \binom{n-\bar{k}}{i-j} \alpha^i (1-\alpha)^{n-1-i} \right]$$
$$+ \sum_{i=\frac{n-1}{2}}^{n-\bar{k}-1} \left[\sum_{j=0}^{\bar{k}-1} \binom{\bar{k}-1}{j} \binom{n-\bar{k}}{i-j} \alpha^i (1-\alpha)^{n-1-i} \right] + \sum_{i=n-\bar{k}}^{n-1} \left[\sum_{j=i-(n-\bar{k})}^{\bar{k}-1} \binom{\bar{k}-1}{j} \binom{n-\bar{k}}{i-j} \alpha^i (1-\alpha)^{n-1-i} \right] \geq 0.$$

For notational simplicity, denote the coefficient of $\alpha^i(1-\alpha)^{n-1-i}$ as $\kappa_i(k^*)$. The inequality becomes:

$$\sum_{i=0}^{n-1} \kappa_i(\bar{k}) \alpha^i (1-\alpha)^{n-1-i} \geq 0.$$

Suppose now instead we have $k^* = \bar{k}+1$. Since we have proven the following identities:

$$\sum_{j=0}^{i} \binom{k^*-1}{j} \binom{n-k^*}{i-j} = \sum_{j=0}^{i} \binom{k^*}{j} \binom{n-k^*-1}{i-j},$$

$$\sum_{j=0}^{k^*-1} \binom{k^*-1}{j} \binom{n-k^*}{i-j} = \sum_{j=0}^{k^*} \binom{k^*}{j} \binom{n-k^*-1}{i-j},$$

$$\sum_{j=0}^{k^*-1} \binom{k^*-1}{j} \binom{n-k^*}{i-j} = \sum_{j=0}^{k^*} \binom{k^*}{j} \binom{n-k^*-1}{i-j},$$

we have $\kappa_i(\bar{k}) = \kappa_i(\bar{k}+1)$, which implies the inequality is independent of k^*, and hence independent of p and q.

Likewise, we have the same result for the three remaining inequalities.

□

Proof of Corollary 3.2.1. According to the proof of Proposition 3.2, the condition under which an agent of type G_0 has no incentive to deviate from truthfully deliberating is:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \lambda pq \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} q^{k^*-1} (1-q)^{(n-1)-(k^*-1)} \\
&+ \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} (1-p)(1-q) \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} q^{(n-1)-(k^*-1)} (1-q)^{k^*-1} \\
&\geq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \lambda pq \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} q^{k^*-1} (1-q)^{(n-1)-(k^*-1)} \\
&+ \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} (1-p)(1-q) \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} q^{(n-1)-(k^*-1)} (1-q)^{k^*-1}.
\end{aligned}$$

which further implies:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} [\lambda pq \cdot q^{k^*-1} (1-q)^{(n-1)-(k^*-1)} \\
&- (1-p)(1-q) q^{(n-1)-(k^*-1)} (1-q)^{k^*-1}] \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} \\
&\geq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} [\lambda pq \cdot q^{k^*-1} (1-q)^{(n-1)-(k^*-1)} \\
&- (1-p)(1-q) q^{(n-1)-(k^*-1)} (1-q)^{k^*-1}] \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}}.
\end{aligned}$$

which finally implies:

$$\begin{aligned}
&\sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}} \\
&\geq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{k^*-1}{n_{G_0}} \binom{(n-1)-(k^*-1)}{n_{G_1}} \alpha^{n_{G_0}+n_{G_1}} (1-\alpha)^{n_{B_0}+n_{B_1}}.
\end{aligned}$$

Following the same argument, the conditions that an agent of type G_1, B_0, and B_1 has no incentive to deviate from truthfully deliberating coincide with those stated in Proposition 3.2.

□

3.A.2 Babbling Equilibrium: Existence

Proof of Proposition 3.4. We first examine the incentive to deviate from sincere voting. An agent understands that they are a pivotal voter if and only if, apart from herself, half of the agents cast a vote $v_j = 0$ and the other half cast $v_j = 1$. Mathematically, an event $\boldsymbol{n_{-i}}$ is pivotal if and only if $(n_{G_0}, n_{G_1}, n_{B_0}, n_{B_1})$ satisfies:

$$n_{G_0} + n_{G_1} = n_{B_0} + n_{B_1} = \frac{n-1}{2} \text{ if } p \geq q;$$
$$n_{G_0} + n_{B_0} = n_{G_1} + n_{B_1} = \frac{n-1}{2} \text{ if } p < q.$$

The agent then forms posterior belief conditional on pivotal events, $\widetilde{\phi}_i$. Below we show that this equals ϕ_i. Suppose that $p \geq q$ and $x_i = G_0$. The posterior belief conditional on pivotal events is:

$$\begin{aligned}
\widetilde{\phi}_i = \widetilde{\phi}(s_i, \boldsymbol{m_{-i}}) &= \frac{\sum_{\{\boldsymbol{n_{-i}} | n_{G_0}+n_{G_1}=n_{B_0}+n_{B_1}=\frac{n-1}{2})\}} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n_{-i}})}{\sum_{\{\boldsymbol{n_{-i}} | n_{G_0}+n_{G_1}=n_{B_0}+n_{B_1}=\frac{n-1}{2})\}} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n_{-i}})} \\
&= \frac{pq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{n-1}{\frac{n-1}{2}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} q^{n_{G_0}+n_{B_0}} (1-q)^{n_{G_1}+n_{B_1}}}{(1-p)(1-q) \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{n-1}{\frac{n-1}{2}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} (1-q)^{n_{G_0}+n_{B_0}} q^{n_{G_1}+n_{B_1}}} \\
&= \frac{pq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} q^{n_{G_0}+n_{B_0}} (1-q)^{n_{G_1}+n_{B_1}}}{(1-p)(1-q) \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} (1-q)^{n_{G_0}+n_{B_0}} q^{n_{G_1}+n_{B_1}}}.
\end{aligned}$$

Observe that, if an event (a, b, c, d) is pivotal, then (d, c, b, a) is also pivotal. Furthermore, the conditional probability of the event $(\omega = 0, s_i = 0, \boldsymbol{n_{-i}} = (a, b, c, d))$ is identical to the conditional probability of $(\omega = 1, s_i = 0, \boldsymbol{n_{-i}} = (d, c, b, a))$. To prove this formally, we have:

$$\mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n_{-i}} = (a, b, c, d)) = pq \binom{\frac{n-1}{2}}{a} \binom{\frac{n-1}{2}}{c} q^{a+c} (1-q)^{b+d};$$
$$\mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n_{-i}} = (d, c, b, a)) = (1-p)(1-q) \binom{\frac{n-1}{2}}{d} \binom{\frac{n-1}{2}}{b} (1-q)^{d+b} q^{c+a},$$

which implies:

$$\mathbb{P}(\boldsymbol{n_{-i}} = (a, b, c, d) \mid \omega = 0, s_i = 0) = \mathbb{P}(\boldsymbol{n_{-i}} = (d, c, b, a) \mid \omega = 1, s_i = 0).$$

Therefore, we have:

$$\begin{aligned}
\widetilde{\phi}_i &= \frac{pq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} q^{n_{G_0}+n_{B_0}} (1-q)^{n_{G_1}+n_{B_1}}}{(1-p)(1-q) \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \binom{\frac{n-1}{2}}{n_{G_0}} \binom{\frac{n-1}{2}}{n_{B_0}} (1-q)^{n_{G_0}+n_{B_0}} q^{n_{G_1}+n_{B_1}}} \\
&= \frac{pq \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \mathbb{P}(\boldsymbol{n_{-i}} \mid \omega = 0, s_i = 0)}{(1-p)(1-q) \sum_{\{\boldsymbol{n_{-i}}|\cdot\}} \mathbb{P}(\boldsymbol{n_{-i}} \mid \omega = 1, s_i = 0)}
\end{aligned}$$

$$= \frac{pq}{(1-p)(1-q)} = \phi(s_i, \boldsymbol{m_{-i}}) = \phi_i.$$

Following the same argument, it can be proven that $\widetilde{\phi}_i = \phi_i$ holds for any other type (i.e., G_1, B_0, and B_1). Suppose that $p < q$ and $x_i = G_0$. The posterior belief conditional on pivotal events is:

$$\widetilde{\phi}_i = \widetilde{\phi}(s_i, \boldsymbol{m_{-i}}) = \frac{\sum_{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_1} = n_{B_0} + n_{G_1} = \frac{n-1}{2})\}} \mathbb{P}(\omega = 0, s_i = 0, \boldsymbol{n_{-i}})}{\sum_{\{\boldsymbol{n}_{-i} | n_{G_0} + n_{B_1} = n_{B_0} + n_{G_1} = \frac{n-1}{2})\}} \mathbb{P}(\omega = 1, s_i = 0, \boldsymbol{n_{-i}})}.$$

Likewise, if an event (a, b, c, d) is pivotal, then (d, c, b, a) is also pivotal. Therefore, based on the previous argument, $\widetilde{\phi}_i = \phi_i$. In conclusion, given babbling strategy, the posterior belief conditional on pivotal events, $\widetilde{\phi}_i$, is the same as ϕ_i, which implies that sincere voting is sustainable. Note that if we have any other k-voting rule instead of simple majority, sincere voting will no longer be sustainable in general. For example, suppose that unanimity rule of $d = 1$ (i.e., $d = 1$ if and only if all agents cast $v_i = 1$) is implemented and $p < q$. A bad agent who receives $s_i = 0$ forms $\phi_i > 1$. Sincere voting suggests that they should vote $v_i = 0$. However, in pivotal voting events in which all other agents vote $v_i = 1$, their posterior belief should rather be:

$$\widetilde{\phi}_i = \frac{p}{1-p} \frac{q}{1-q} \left(\frac{\alpha(1-q) + (1-\alpha)q}{\alpha q + (1-\alpha)(1-q)} \right)^{n-1} = \frac{p}{1-p} \frac{q}{1-q} \left(\frac{1-\widetilde{q}}{\widetilde{q}} \right)^{n-1},$$

which can be $\widetilde{\phi}_i < 1 < \phi_i$ if, say, α is sufficiently high. As a result, the agent deviates from sincere voting and thus sincere voting is not sustainable.

Now we investigate if any type has incentive to deviate from babbling strategy, given sincere voting. Since messages offer no clue on private signals from other agents, deviation from sending a null message does not influence the posterior belief and thus voting actions of any type. Consequently, the decision is not affected by the sent message, which concludes that there exists no incentive for any type of agents to deviate from the babbling strategy.

□

3.A.3 Non-Truthful Equilibrium: General

This appendix provides the complete technical analysis supporting the equilibrium characterisation of non-truthful equilibrium in Section 3.4.2. The analysis proceeds in three parts: establishing the information structure under systematic deception, deriving sustainability conditions for sincere voting, and characterising the conditions for sustainable messaging strategies. Under a non-truthful equilibrium, good agents truthfully reveal their signals, $m(s_i, \theta_i = 1) = s_i$, whilst bad agents systematically lie, $m(s_i, \theta_i = -1) = 1 - s_i$. This creates a complex information environment where message interpretation depends on understanding the systematic deception pattern.

Message Generation Process: A null message $m_j = 0$ originates from agents of type $x_i \in \{G_0, B_1\}$, whilst a message $m_j = 1$ comes from agents of type $x_i \in \{G_1, B_0\}$. The conditional probabilities become:

$$\begin{aligned}
&\mathbb{P}(x_i \in \{G_0, B_1\} \mid \omega = 0) = \alpha q + (1-\alpha)(1-q) \equiv \widetilde{q};\\
&\mathbb{P}(x_i \in \{G_0, B_1\} \mid \omega = 1) = 1 - \widetilde{q};\\
&\mathbb{P}(x_i \in \{G_1, B_0\} \mid \omega = 0) = 1 - \widetilde{q};\\
&\mathbb{P}(x_i \in \{G_1, B_0\} \mid \omega = 1) = \widetilde{q}.
\end{aligned}$$

This creates an effective signal precision $\widetilde{q}$ that differs from the original precision q.

Posterior Belief Formation: When the agent i of type x_i receives signal s_i and observes k null messages in the message profile $\boldsymbol{m} \in \mathcal{M}^n$, their posterior belief, denoted $\phi^k_{x_i}$ is:

$$\begin{aligned}
\phi^k_{G_0} &= \frac{p}{1-p}\frac{q}{1-q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k-(n+1)}; \ \phi^k_{G_1} = \frac{p}{1-p}\frac{1-q}{q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k-(n-1)};\\
\phi^k_{B_0} &= \frac{p}{1-p}\frac{q}{1-q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k-(n-1)}; \ \phi^k_{B_1} = \frac{p}{1-p}\frac{1-q}{q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{2k-(n+1)}.
\end{aligned}$$

Lemma 3.3 (Monotonicity of induced posterior belief)**.** *Given any message profile $\boldsymbol{m}^k$ any posterior belief increases in $k \equiv \#\{m_i = 0 \mid \boldsymbol{m}\}$.*

Proof. Denote a message profile with k null messages as, $\boldsymbol{m}^k \in \{\boldsymbol{m} \mid \#\{m_i = 0 \mid \boldsymbol{m}\} = k\} \subset \mathcal{M}^n$. Moreover, denote a message profile other than i with k null messages as, $\boldsymbol{m}^k_{-i} \in \{\boldsymbol{m}_{-i} \mid \#\{m_j = 0 \mid \boldsymbol{m}_{-i}\} = k\} \subset \mathcal{M}^{n-1}$. Then, it can be shown that:

$$\phi(s_i, \boldsymbol{m}^a_{-i}) > \phi(s_i, \boldsymbol{m}^b_{-i}) \iff a > b.$$

Note that $\widetilde{q} > \frac{1}{2}$ since, $\widetilde{q} = \alpha q + (1-\alpha)(1-q) = \alpha \cdot \frac{1}{2} + (1-\alpha)\cdot\frac{1}{2} + \alpha(q - \frac{1}{2}) + (1-\alpha)(1-q-\frac{1}{2})$ $= \frac{1}{2} + (2\alpha - 1)(q - \frac{1}{2}) > \frac{1}{2}$. Hence, $\frac{\widetilde{q}}{1-\widetilde{q}} > 1$.

($\Leftarrow$): $a > b \implies \phi(0, \boldsymbol{m}^a_{-\boldsymbol{i}}) = \frac{p}{1-p}\frac{q}{1-q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{a-(n-1-a)} > \frac{p}{1-p}\frac{q}{1-q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{b-(n-1-b)} = \phi(0, \boldsymbol{m}^b_{-\boldsymbol{i}})$; $\phi(1, \boldsymbol{m}^a_{-\boldsymbol{i}}) = \frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{a-(n-1-a)} > \frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{b-(n-1-b)} = \phi(1, \boldsymbol{m}^b_{-\boldsymbol{i}})$.

($\Rightarrow$): If $\phi(s_i, \boldsymbol{m}^a_{-\boldsymbol{i}}) > \phi(s_i, \boldsymbol{m}^b_{-\boldsymbol{i}})$, $\frac{\phi(s_i, \boldsymbol{m}^a_{-\boldsymbol{i}})}{\phi(s_i, \boldsymbol{m}^b_{-\boldsymbol{i}})} = \frac{\frac{p}{1-p}(\frac{q}{1-q})^{2s_i-1}(\frac{\widetilde{q}}{1-\widetilde{q}})^{a-(n-1-a)}}{\frac{p}{1-p}(\frac{q}{1-q})^{2s_i-1}(\frac{\widetilde{q}}{1-\widetilde{q}})^{b-(n-1-b)}} = (\frac{\widetilde{q}}{1-\widetilde{q}})^{2(a-b)} > 1$ $\implies a > b$. □

Lemma 3.4 (Same message, different posterior)**.** *Let $\phi^k_{x_i}$ denote the posterior belief of an agent of type x_i observing k null messages in the message profile $\boldsymbol{m} \in \mathcal{M}^n$. Given any message profile $\boldsymbol{m}$ with any k, the induced posterior belief of all types satisfies the following relationship: $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$.*

Proof. Recall that we denote the posterior belief of type x_i as $\phi_{x_i}(\boldsymbol{m})$, where the message profile is $\boldsymbol{m}$. Mathematically, we have:

$$\phi_{x_i}(\boldsymbol{m}) \equiv \phi(s_i, \boldsymbol{m}_{-\boldsymbol{i}}) \text{ s.t. } m(x_i) = m(s_i, \theta_i) = m_i.$$

Suppose a message profile $\boldsymbol{m}$ contains k null messages. We have,

$$\phi^k_{G_0} = \phi(0, \boldsymbol{m}^{k-1}_{-\boldsymbol{i}})\,;\, \phi^k_{G_1} = \phi(1, \boldsymbol{m}^{k}_{-\boldsymbol{i}})\,;\, \phi^k_{B_0} = \phi(0, \boldsymbol{m}^{k}_{-\boldsymbol{i}})\,;\, \phi^k_{B_1} = \phi(1, \boldsymbol{m}^{k-1}_{-\boldsymbol{i}}).$$

By Lemma 3.3, we have:

$$\phi^k_{B_0} > \phi^k_{G_0}\,;\, \phi^k_{G_1} > \phi^k_{B_1}.$$

Moreover, we have the following inequality:

$$\frac{\phi^k_{G_0}}{\phi^k_{G_1}} = \frac{\frac{p}{1-p}\frac{q}{1-q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{k-1-(n-1-(k-1))}}{\frac{p}{1-p}\frac{1-q}{q}(\frac{\widetilde{q}}{1-\widetilde{q}})^{k-(n-1-k)}} = (\frac{q}{1-q})^2(\frac{\widetilde{q}}{1-\widetilde{q}})^{-2} > 1,$$

which implies, $\phi^k_{G_0} > \phi^k_{G_1}$. Therefore, we have, $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. □

Posterior Belief Relations and Voting Sustainability

The posterior beliefs relative to unity create five possible orderings that determine voting behaviour:

(a) $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$;
(b) $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$;

(c) $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$;
(d) $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$;
(e) $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$

Lemma 3.5 (Monotonicity in posterior relation)**.** *Let $\phi^k_{x_i}$ denote the posterior belief of an agent of type x_i observing k null messages in the message profile $\boldsymbol{m} \in \mathcal{M}^n$. Let $\mathcal{K}^a \equiv \{k \mid \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1\}$; $\mathcal{K}^b \equiv \{k \mid \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}\}$; $\mathcal{K}^c \equiv \{k \mid \phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}\}$; $\mathcal{K}^d \equiv \{k \mid \phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}\}$; $\mathcal{K}^e \equiv \{k \mid 1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}\}$. Then, we have, $a > b > c > d > e, \forall a \in \mathcal{K}^a, \forall b \in \mathcal{K}^b, \forall c \in \mathcal{K}^c, \forall d \in \mathcal{K}^d, \forall e \in \mathcal{K}^e$.*

Proof. If $\phi^a_{B_0} > \phi^a_{G_0} > \phi^a_{G_1} > \phi^a_{B_1} > 1$ and $\phi^b_{B_0} > \phi^b_{G_0} > \phi^b_{G_1} > 1 > \phi^b_{B_1} \implies \phi_{B_1}(\boldsymbol{m}^a) > \phi^b_{B_1} \implies \phi(s_i = 1, \boldsymbol{m}^{a-1}_{-i}) > \phi(s_i = 1, \boldsymbol{m}^{b-1}_{-i})$. By Lemma 3.3, $a - 1 > b - 1$ and thus $a > b$.

Likewise, $\phi^b_{B_0} > \phi^b_{G_0} > \phi^b_{G_1} > 1 > \phi^b_{B_1}$ and $\phi^c_{B_0} > \phi^c_{G_0} > 1 > \phi^c_{G_1} > \phi^c_{B_1} \implies b > c$; $\phi^c_{B_0} > \phi^c_{G_0} > 1 > \phi^c_{G_1} > \phi^c_{B_1}$ and $\phi^d_{B_0} > 1 > \phi^d_{G_0} > \phi^d_{G_1} > \phi^d_{B_1} \implies c > d$; $\phi^d_{B_0} > 1 > \phi^d_{G_0} > \phi^d_{G_1} > \phi^d_{B_1}$ and $1 > \phi^e_{B_0} > \phi^e_{G_0} > \phi^e_{G_1} > \phi^e_{B_1} \implies d > e$. □

Lemma 3.6. *The following lemmas examine some relevant conditions regarding change in posterior relation induced by $\boldsymbol{m}^{k-1}$, $\boldsymbol{m}^k$, and $\boldsymbol{m}^{k+1}$.*

Lemma 3.6a. *Given $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$, it cannot be that, $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1} > 1$, or, $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > 1 > \phi^{k-1}_{B_1}$.*

Proof. $1 < \phi^k_{G_1} = \phi^{k+1}_{B_1}$. By Lemma 3.4, $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1} > 1$.

$1 > \phi^k_{B_1} = \phi^{k-1}_{G_1}$. □

Lemma 3.6b. *Given $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$, it cannot be that, $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$, or, $\phi^{k+1}_{B_0} > 1 > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1}$.*

Proof. $1 > \phi^k_{G_0} = \phi^{k-1}_{B_0}$. By Lemma 3.4, $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$.

$1 < \phi^k_{B_0} = \phi^{k+1}_{G_0}$. □

Lemma 3.6c. *Given $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$, it cannot be that, $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$, or, $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1} > 1$.*

Proof. $1 < \phi^k_{G_0} = \phi^{k-1}_{B_0}; 1 > \phi^k_{G_1} = \phi^{k+1}_{B_1}$. □

Lemma 3.6d. *Given* $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$*, it cannot be that,* $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$*,* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$*, or,* $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > 1 > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$*.*

Proof. $1 < \phi^k_{B_1} = \phi^{k-1}_{G_1}$. □

Lemma 3.6e. *Given* $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$*, it cannot be that,* $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1} > 1$*,* $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > \phi^{k+1}_{G_1} > 1 > \phi^{k+1}_{B_1}$*, or,* $\phi^{k+1}_{B_0} > \phi^{k+1}_{G_0} > 1 > \phi^{k+1}_{G_1} > \phi^{k+1}_{B_1}$*.*

Proof. $1 > \phi^k_{B_0} = \phi^{k+1}_{G_0}$. □

Sustainability of Sincere Voting

The sustainability of sincere voting depends critically on which posterior relation holds and the resulting pivotal voting incentives.

Lemma 3.7. *The following lemmas examine some relevant conditions regarding sustainability of sincere voting.*

Lemma 3.7a. *Given any* $k \in \{0, 1, 2, ..., n\}$ *such that* $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$*,* $p \geq q$ *must hold for sustainability of sincere voting.*

Proof. An agent of type x_i understands that they are a pivotal voter if and only if $\#\{v_j = 0 \mid \boldsymbol{v_{-i}}\} = \#\{v_j = 1 \mid \boldsymbol{v_{-i}}\}$, or equivalently, $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} = n_{B_0} + n_{B_1} = \frac{n-1}{2}$. Moreover, they also know that $n_{G_0} + n_{B_1} = k_{x_i}(\boldsymbol{m}^k)$ and $n_{G_1} + n_{B_0} = (n-1) - k_{x_i}(\boldsymbol{m}^k)$. Therefore, the agent is a pivotal voter if and only if $\boldsymbol{n_{-i}}$ satisfies:

$$n_{G_0} + n_{G_1} = n_{B_0} + n_{B_1} = \frac{n-1}{2};\ n_{G_0} + n_{B_1} = k_{x_i}(\boldsymbol{m}^k);\ n_{G_1} + n_{B_0} = (n-1) - k_{x_i}(\boldsymbol{m}^k).$$

Observe that, if $\boldsymbol{n_{-i}} = (a, b, c, d)$ is pivotal, then (d, c, b, a) is also pivotal. Define the conditional probability of $\boldsymbol{n_{-i}}$ given (ω, s_i) as $\mathbb{P}(\boldsymbol{n_{-i}} \mid \omega, s_i)$. We have:

$$\mathbb{P}((a, b, c, d) \mid 0, 0) = \mathbb{P}((a, b, c, d) \mid 0, 1) = \mathbb{P}((d, c, b, a) \mid 1, 0) = \mathbb{P}((d, c, b, a) \mid 1, 1),$$

since:

$$\begin{aligned}\mathbb{P}((a,b,c,d)\mid 0,0) = \mathbb{P}((a,b,c,d)\mid 0,1) &= \binom{a+d}{a}\binom{b+c}{b}\alpha^{a+b}(1-\alpha)^{c+d}q^{a+c}(1-q)^{b+d} \\ &= \binom{d+a}{d}\binom{c+b}{c}\alpha^{d+c}(1-\alpha)^{b+a}(1-q)^{d+b}q^{c+a} \\ &= \mathbb{P}((d,c,b,a)\mid 1,0) = \mathbb{P}((d,c,b,a)\mid 1,1).\end{aligned}$$

The posterior belief of type x_i restricted on pivotal voting events is:

$$\begin{aligned}\widetilde{\phi}^k_{x_i} &= \frac{\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\omega=0,s_i,\boldsymbol{n_{-i}})}{\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\omega=1,s_i,\boldsymbol{n_{-i}})} = \frac{pq^{1-s_i}(1-q)^{s_i}\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=0,s_i)}{(1-p)q^{s_i}(1-q)^{1-s_i}\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=1,s_i)} \\ &= \frac{pq^{1-s_i}(1-q)^{s_i}}{(1-p)q^{s_i}(1-q)^{1-s_i}}\frac{\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=0,s_i)}{\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=1,s_i)},\end{aligned}$$

where $\boldsymbol{n_{-i}}$ satisfies $n_{G_0}+n_{G_1} = n_{B_0}+n_{B_1} = \frac{n-1}{2}$, $n_{G_0}+n_{B_1} = k_{x_i}(\boldsymbol{m}^k)$, and $n_{G_1}+n_{B_0} = (n-1) - k_{x_i}(\boldsymbol{m}^k)$. Since $\mathbb{P}((a,b,c,d)\mid 0,0) = \mathbb{P}((d,c,b,a)\mid 1,0)$ and $\mathbb{P}((a,b,c,d)\mid 0,1) = \mathbb{P}((d,c,b,a)\mid 1,1)$, $\sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=0,s_i) = \sum_{\boldsymbol{n_{-i}}}\mathbb{P}(\boldsymbol{n_{-i}}\mid\omega=1,s_i)$.

Thus, we have:

$$\widetilde{\phi}^k_{x_i} = \frac{pq^{1-s_i}(1-q)^{s_i}}{(1-p)q^{s_i}(1-q)^{1-s_i}} = \begin{cases}\frac{pq}{(1-p)(1-q)}, & \text{if } x_i\in\{G_0,B_1\} \\ \frac{p(1-q)}{(1-p)q}, & \text{if } x_i\in\{G_1,B_0\}\end{cases}.$$

For types $x_i \in \{G_0, B_1\}$, they have no incentive to deviate since $\widetilde{\phi}^k_{x_i} \geq 1$ and $\phi^k_{x_i} \geq 1$, unconditionally. For types $x_i \in \{G_1, B_0\}$, they have no incentive to deviate if and only if $p \geq q$ so that $\widetilde{\phi}^k_{x_i} \geq 1$ and $\phi^k_{x_i} \geq 1$, which implies that $p \geq q$ is a necessary condition for the sustainability.

□

Lemma 3.7b. *Given any $k \in \{0, n\}$ such that $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^k$.*

Proof. Suppose that $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$. Note that there exist only type G_0 and B_1 if $\boldsymbol{m}^n$ is observed. Conditional on observing $\boldsymbol{m}^n$, an agent of type $x_i \in \{G_0, B_1\}$ understands that they are a pivotal voter if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0}+n_{G_1}+n_{B_1} = n_{B_0} = \frac{n-1}{2}$. Given $\boldsymbol{m}^n$, there exists no type B_0 since no $m_j = 1$ is observed, which implies there exists no $\boldsymbol{n_{-i}}$ such that $n_{B_0} = \frac{n-1}{2}$. Hence, no pivotal event exists which implies that sustainability holds.

Suppose that $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > 1 > \phi^0_{B_1}$. Note that there exist only type G_1 and B_0 if $\boldsymbol{m}^0$ is observed. Conditional on observing $\boldsymbol{m}^0$, a agent of type $x_i \in \{G_1, B_0\}$ understands that they are a pivotal voter if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_1} = n_{B_0} = \frac{n-1}{2}$. Therefore, we have,

$$\begin{aligned}
\widetilde{\phi}^0_{G_1} &= \frac{p(1-q)\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}}{(1-p)q\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}} = \frac{p(1-q)}{(1-p)q} \\
&> \frac{p}{1-p}\frac{1-q}{q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{1-n} = \phi^0_{G_1} > 1; \\
\widetilde{\phi}^0_{B_0} &= \frac{pq\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}}{(1-p)(1-q)\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}} = \frac{pq}{(1-p)(1-q)} \\
&> \frac{p}{1-p}\frac{1-q}{q}\left(\frac{\widetilde{q}}{1-\widetilde{q}}\right)^{1-n} = \phi^0_{B_0} > 1.
\end{aligned}$$

Thus, there is no incentive to deviate for type G_1 and B_0. □

Lemma 3.7c. *If $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^k$.*

Proof. Conditional on observing $\boldsymbol{m}^k$, an agent of type x_i is a pivotal voter only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{B_1} = n_{G_1} + n_{B_0} = \frac{n-1}{2}$, $n_{G_0} + n_{B_1} = k_{x_i}(\boldsymbol{m}^k)$, and $n_{G_1} + n_{B_0} = (n-1) - k_{x_i}(\boldsymbol{m}^k)$.

If $k_{x_i}(\boldsymbol{m}^k) = \frac{n-1}{2}$, they are a pivotal voter for all possible $\boldsymbol{n_{-i}}$, which implies $\phi^k_{x_i} = \widetilde{\phi}^k_{x_i}$. Therefore, they have no incentive to deviate from sincere voting. If $k_{x_i}(\boldsymbol{m}^k) \neq \frac{n-1}{2}$, they are not a pivotal voter for each $\boldsymbol{n_{-i}}$. Since there exists no pivotal event, they have no incentive to deviate from sincere voting. □

Lemma 3.7d. *If $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^0$.*

Proof. Suppose that $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$. Note that there exist only type G_1 and B_0 if $\boldsymbol{m}^0$ is observed.

Conditional on observing $\boldsymbol{m}^0$, an agent of type $x_i \in \{G_1, B_0\}$ understands that they are a pivotal voter only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_0} = n_{B_1} = \frac{n-1}{2}$. Given $\boldsymbol{m}^0$, there exists no type B_1 since no $m_j = 0$ is observed, which implies there exists no $\boldsymbol{n_{-i}}$ such that $n_{B_1} = \frac{n-1}{2}$. Hence, no pivotal event exists.

□

Lemma 3.7e. *If* $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ *holds for some* $k \in \{0, 1, 2, ..., n\}$, *the sustainability of sincere voting fails.*

Proof. Observe that, if $\boldsymbol{n_{-i}} = (a, b, c, d)$ is pivotal, (d, c, b, a) is also pivotal. Following the same argument in the proof for Lemma 3.7a, the posterior belief conditional on pivotal events is,

$$\widetilde{\phi}^k_{x_i} = \begin{cases} \frac{pq}{(1-p)(1-q)}, & \text{if } x_i \in \{G_0, B_0\} \\ \frac{p(1-q)}{(1-p)q}, & \text{if } x_i \in \{G_1, B_1\} \end{cases}.$$

Type G_0 and B_0 always have incentive to deviate since $\widetilde{\phi}^k_{x_i} \geq 1$ whilst $\phi^k_{x_i} < 1$. Equilibrium fails. □

Sustainability of Message Strategy

An agent becomes a pivotal message sender when their deviation changes the posterior belief relation and thereby alters voting outcomes sufficient to flip the collective decision.

Lemma 3.8 (Change in posterior relation by one message). *Suppose that there is a change in posterior belief from* $\boldsymbol{m}^k$ *to* $\boldsymbol{m}^{k-1}$, *or, from* $\boldsymbol{m}^{k-1}$ *to* $\boldsymbol{m}^k$, *where* $k \in \{1, 2, ..., n\}$. *One of the following statement is true.*

(i) $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$, *and* $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > 1 > \phi^{k-1}_{B_1}$;

(ii) $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$, *and* $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > 1 > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$;

(iii) $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$, *and* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$;

(iv) $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$, *and* $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$;

(v) $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$, *and* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$.

Proof. Suppose that $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$. By Lemma 3.5, we cannot have $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$, $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$, or $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > 1 > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. Thus, $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > 1 > \phi^{k-1}_{B_1}$.

Suppose that $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$. By Lemma 3.5, we cannot have $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1} > 1$. By Lemma 3.6a, we cannot have $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. Thus, we have $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > 1 > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$ or $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$.

Suppose that $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$. By Lemma 3.5, we cannot have $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1} > 1$ or $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > 1 > \phi^{k-1}_{B_1}$. By Lemma 3.6c, we cannot have $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. Thus, we have $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$.

Suppose that $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. By Lemma 3.6b, we have $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$.

Suppose that $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. By Lemma 3.5, there exists no possible change in posterior relation from $\boldsymbol{m}^k$ to $\boldsymbol{m}^{k-1}$.

Suppose that $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1} > 1$. By Lemma 3.5, there exists no possible change in posterior relation from $\boldsymbol{m}^{k-1}$ to $\boldsymbol{m}^k$.

Suppose that $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > 1 > \phi^{k-1}_{B_1}$. By Lemma 3.6a, we have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$.

Suppose that $\phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > 1 > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. By Lemma 3.5, we cannot have $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$, or $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. By Lemma 3.6c, we cannot have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$. Thus, we have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$.

Suppose that $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. By Lemma 3.5, we cannot have $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. By Lemma 3.6b, we cannot have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$. Thus, we have $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ or $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$.

Suppose that $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$. By Lemma 3.6e, we cannot have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$, $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$, or $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$. Thus, we have $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$. □

Critical Insight: The analysis reveals that certain transition patterns between posterior relations create systematic incentives for good agents to deviate from the non-truthful messaging strategy.

Non-Empty Set of Pivotal Events

Lemma 3.9. *The following lemmas examine key conditions related to the non-empty set of pivotal events.*

Lemma 3.9a. *If $\phi^{k^{(i)}}_{B_0} > \phi^{k^{(i)}}_{G_0} > \phi^{k^{(i)}}_{G_1} > \phi^{k^{(i)}}_{B_1} > 1$ and $\phi^{k^{(i)}-1}_{B_0} > \phi^{k^{(i)}-1}_{G_0} > \phi^{k^{(i)}-1}_{G_1} > 1 > \phi^{k^{(i)}-1}_{B_1}$ hold for some $k^{(i)} \in \{2, 3, ..., n\}$, there always exists a non-empty set of pivotal events where type G_0 is a pivotal sender conditional on observing $\boldsymbol{m}^{k^{(i)}}$, and a non-empty set pivotal events where type G_1 is a pivotal sender conditional on observing $\boldsymbol{m}^{k^{(i)}-1}$. Moreover, the two sets are equivalent.*

Proof. Conditional on observing $\boldsymbol{m}^{k^{(i)}}$, an agent of type G_0 understands that they are a pivotal sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{G_0} + n'_{G_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $(n_{G_0} + 1) + n_{G_1} + n_{B_1} > \frac{n}{2} > (n_{G_0} + 1) + n_{G_1}$. Along with the constraint that they observe $k_{G_0}(\boldsymbol{m}^{k^{(i)}}) = k^{(i)} - 1$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{G_0} + n_{G_1}$, $n_{G_0} + n_{B_1} = k^{(i)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(i)} - 1)$.

Likewise, conditional on observing $\boldsymbol{m}^{k^{(i)}-1}$, an agent of type G_1 is a pivotal sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{G_0} + n'_{G_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + (n_{G_1} + 1) + n_{B_1} > \frac{n}{2} > n_{G_0} + (n_{G_1} + 1)$. Along with the constraint that they observe $k_{G_1}(\boldsymbol{m}^{k^{(i)}-1}) = k^{(i)} - 1$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{G_0} + n_{G_1}$, $n_{G_0} + n_{B_1} = k^{(i)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(i)} - 1)$.

Since the conditions of a pivotal events are identical, $\boldsymbol{n_{-i}}$ is a pivotal event of type G_0 if and only if $\boldsymbol{n_{-i}}$ is a pivotal event of type G_1, which concludes that the two sets are equivalent.

Denote the pivotal events as $\boldsymbol{n}^{(i)}_{\boldsymbol{-i}} \in \{\boldsymbol{n_{-i}} \mid n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{G_0} + n_{G_1}, n_{G_0} + n_{B_1} = k^{(i)} - 1, n_{G_1} + n_{B_0} = (n-1) - (k^{(i)} - 1)\}$. □

Lemma 3.9b. *If* $\phi^{k^{(ii)}}_{B_0} > \phi^{k^{(ii)}}_{G_0} > \phi^{k^{(ii)}}_{G_1} > 1 > \phi^{k^{(ii)}}_{B_1}$ *and* $\phi^{k^{(ii)}-1}_{B_0} > \phi^{k^{(ii)}-1}_{G_0} > 1 > \phi^{k^{(ii)}-1}_{G_1} > \phi^{k^{(ii)}-1}_{B_1}$ *hold for some* $k^{(ii)} \in \{1, 2, ..., \frac{n-1}{2}\}$*, there always exists a non-empty set of pivotal events where type* G_0 *is a pivotal sender conditional on observing* $\boldsymbol{m}^{k^{(ii)}}$*, and a non-empty set pivotal events where type* G_1 *is a pivotal sender conditional on observing* $\boldsymbol{m}^{k^{(ii)}-1}$*. Moreover, the two sets are equivalent.*

Proof. Conditional on observing $\boldsymbol{m}^{k^{(ii)}}$, an agent of type G_0 is a pivotal message sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{G_0} + n'_{B_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $(n_{G_0} + 1) + n_{G_1} + n_{B_1} > \frac{n}{2} > (n_{G_0} + 1) + n_{B_1}$. Along with the constraint that they observe $k_{G_0}(\boldsymbol{m}^{k^{(ii)}}) = k^{(ii)} - 1$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(ii)} - 1)$.

Likewise, conditional on observing $\boldsymbol{m}^{k^{(ii)}-1}$, an agent of type G_1 is a pivotal message sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{G_0} + n'_{B_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + (n_{G_1} + 1) + n_{B_1} > \frac{n}{2} > n_{G_0} + n_{B_1}$. Along with the constraint that they observe $k_{G_1}(\boldsymbol{m}^{k^{(ii)}-1})$

$= k^{(ii)} - 1$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(ii)} - 1)$.

Since the conditions of a pivotal events are identical, $\boldsymbol{n_{-i}}$ is a pivotal event of type G_0 if and only if $\boldsymbol{n_{-i}}$ is a pivotal event of type G_1, which concludes that the two sets are equivalent.

Denote the pivotal events as $\boldsymbol{n}_{-i}^{(ii)} \in \{\boldsymbol{n_{-i}} \mid n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1, n_{G_0} + n_{B_1} = k^{(ii)} - 1, n_{G_1} + n_{B_0} = (n-1) - (k^{(ii)} - 1)\}$. □

Lemma 3.9c. *If $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > 1 > \phi^1_{B_1}$ and $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ hold, there always exists a non-empty set of pivotal events where type G_0 is a pivotal sender conditional on observing $\boldsymbol{m}^1$, and a non-empty set pivotal events where type G_1 is a pivotal sender conditional on observing $\boldsymbol{m}^0$. Moreover, the two sets are equivalent.*

Proof. Conditional on observing $\boldsymbol{m}^1$, an agent of type G_0 is a pivotal message sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{B_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $(n_{G_0} + 1) + n_{G_1} + n_{B_1} > \frac{n}{2} > n_{B_1}$. Along with the constraint that they observe $k_{G_0}(\boldsymbol{m}^1) = 1 - 1 = 0$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{B_1} - 1$, $n_{G_0} + n_{B_1} = 0$, and $n_{G_1} + n_{B_0} = n - 1$, which is equivalent to $n_{G_1} > \frac{n}{2} - 1$, $n_{G_0} = n_{B_1} = 0$, and $n_{G_1} + n_{B_0} = n - 1$.

Likewise, conditional on observing $\boldsymbol{m}^0$, an agent of type G_1 is a pivotal message sender if the type profile, $\boldsymbol{n}$, satisfies $n'_{G_0} + n'_{G_1} + n'_{B_1} > \frac{n}{2} > n'_{B_1}$. Equivalently, $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + (n_{G_1} + 1) + n_{B_1} > \frac{n}{2} > n_{B_1}$. Along with the constraint that they observe $k_{G_1}(\boldsymbol{m}^0) = 0$ null messages from other agents of type G_0 and B_1, $\boldsymbol{n_{-i}}$ is a pivotal event only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{B_1} - 1$, and $n_{G_0} + n_{B_1} = 0$, and $n_{G_1} + n_{B_0} = n - 1$, which is equivalent to $n_{G_1} > \frac{n}{2} - 1$, $n_{G_0} = n_{B_1} = 0$, and $n_{G_1} + n_{B_0} = n - 1$.

Since the conditions of a pivotal events are identical, $\boldsymbol{n_{-i}}$ is a pivotal event of type G_0 if and only if $\boldsymbol{n_{-i}}$ is a pivotal event of type G_1, which concludes that the two sets are equivalent.

Denote the pivotal events as $\boldsymbol{n}_{-i}^{(v)} \in \{\boldsymbol{n_{-i}} \mid n_{G_1} > \frac{n}{2} - 1, n_{G_0} = n_{B_1} = 0, n_{G_1} + n_{B_0} = n - 1\}$. □

Deviation if There Exist Pivotal (i), (ii), (v)

Lemma 3.10 (Either type of good agents deviates)**.** *Suppose that* $\nexists k \in \{2, 3, ..., n\}$ *such that* $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ *and* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$*, or,* $\phi^1_{B_0} > \phi^1_{G_0} > 1 > \phi^1_{G_1} > \phi^1_{B_1}$ *and* $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$*.*

If at least one of the following statements is true, either type G_0 *or* G_1 *will deviate.*

(i) $\exists k^{(i)} \in \{2, 3, ..., n\}$ *such that* $\phi^{k^{(i)}}_{B_0} > \phi^{k^{(i)}}_{G_0} > \phi^{k^{(i)}}_{G_1} > \phi^{k^{(i)}}_{B_1} > 1$ *and* $\phi^{k^{(i)}-1}_{B_0} > \phi_{G_0}(\boldsymbol{m}^{k^{(i)}-1}) > \phi^{k^{(i)}-1}_{G_1} > 1 > \phi^{k^{(i)}-1}_{B_1}$*.*

(ii) $\exists k^{(ii)} \in \{1, 2, 3, ..., \frac{n-1}{2}\}$ *such that* $\phi^{k^{(ii)}}_{B_0} > \phi^{k^{(ii)}}_{G_0} > \phi^{k^{(ii)}}_{G_1} > 1 > \phi^{k^{(ii)}}_{B_1}$ *and* $\phi^{k^{(ii)}-1}_{B_0} > \phi^{k^{(ii)}-1}_{G_0} > 1 > \phi^{k^{(ii)}-1}_{G_1} > \phi^{k^{(ii)}-1}_{B_1}$*.*

(iii) $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > 1 > \phi^1_{B_1}$ *and* $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$*.*

Proof. By Lemma 3.9a, 3.9b, and 3.9c, there exists a pivotal event. An agent of type G_0 has incentive to deviate from $m_i = 0$ to $m_i = 1$ if and only if,

$$\begin{aligned}
\Delta_{G_0} \equiv\ & pq \sum_{\{\boldsymbol{n}^{(i)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(i)}_{-i} \mid 0,0) - (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(i)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(i)}_{-i} \mid 1,0) \\
& - pq \sum_{\{\boldsymbol{n}^{(ii)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(ii)}_{-i} \mid 0,0) + (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(ii)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(ii)}_{-i} \mid 1,0) \\
& - pq \sum_{\{\boldsymbol{n}^{(v)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(v)}_{-i} \mid 0,0) + (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(v)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(v)}_{-i} \mid 1,0) > 0.
\end{aligned}$$

However, an agent of type G_1 has incentive to deviate from $m_i = 0$ to $m_i = 1$ if and only if,

$$\begin{aligned}
\Delta_{G_1} \equiv & -pq \sum_{\{\boldsymbol{n}^{(i)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(i)}_{-i} \mid 0,0) + (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(i)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(i)}_{-i} \mid 1,0) \\
& + pq \sum_{\{\boldsymbol{n}^{(ii)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(ii)}_{-i} \mid 0,0) - (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(ii)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(ii)}_{-i} \mid 1,0) \\
& + pq \sum_{\{\boldsymbol{n}^{(v)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(v)}_{-i} \mid 0,0) - (1-p)(1-q) \sum_{\{\boldsymbol{n}^{(v)}_{-i}\}} \mathbb{P}(\boldsymbol{n}^{(v)}_{-i} \mid 1,0) > 0
\end{aligned}$$

Given any $(\{\boldsymbol{n}^{(i)}_{-i}\}, \{\boldsymbol{n}^{(ii)}_{-i}\}), \{\boldsymbol{n}^{(v)}_{-i}\})$, it is clear that $\Delta_{G_0} = -\Delta_{G_1}$ from the expression above. Therefore, if either $\{\boldsymbol{n}^{(i)}_{-i}\}$, $\{\boldsymbol{n}^{(ii)}_{-i}\}$, or $\{\boldsymbol{n}^{(v)}_{-i}\}$ is non-empty, either type G_0 or G_1 will deviate and thus non-truthful message strategy is not sustainable. □

Conditions for Equilibrium Failure

Several configurations guarantee that non-truthful equilibrium cannot be sustained.

Lemma 3.11. *The following lemmas examine key conditions related to failure of non-truthful equilibrium.*

Lemma 3.11a. *If* $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ *and* $1 > \phi^{k-1}_{B_0} > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$ *hold for some* $k \in \{1, 2, ..., n\}$*, equilibrium fails.*

Proof. By Lemma 3.7e, the conjectured voting strategy is not sustainable if $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1}$ $> \phi^k_{B_1}$ for some $k \in \{1, 2, ..., n\}$. Hence, equilibrium fails. □

Lemma 3.11b. *If* $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ *and* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$ *hold for some* $k \in \{2, 3, ..., n\}$*, equilibrium fails.*

Proof. By Lemma 3.6b, $1 > \phi^{k-2}_{B_0} > \phi^{k-2}_{G_0} > \phi^{k-2}_{G_1} > \phi^{k-2}_{B_1}$. By Lemma 3.7e, the conjectured voting strategy is not sustainable. Hence, equilibrium fails. □

Lemma 3.11c. *If* $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$ *and* $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$ *hold for some* $k \in \{1, 2, 3, ..., n\}$*, equilibrium fails.*

Proof. Suppose that $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > 1 > \phi^k_{B_1}$ and $\phi^{k-1}_{B_0} > 1 > \phi^{k-1}_{G_0} > \phi^{k-1}_{G_1} > \phi^{k-1}_{B_1}$, where $k \in \{2, 3, ..., n\}$. By Lemma 3.6b, We have $1 > \phi^{k-2}_{B_0} > \phi^{k-2}_{G_0} > \phi^{k-2}_{G_1} > \phi^{k-2}_{B_1}$ By Lemma 3.7e, equilibrium fails since sincere voting is not sustainable.

Suppose that $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > 1 > \phi^1_{B_1}$ and $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ By Lemma 3.6a, we have $\phi^2_{B_0} > \phi^2_{G_0} > \phi^2_{G_1} > \phi^2_{B_1} > 1$. By Lemma 3.10, non-truthful message strategy is not sustainable. □

Conditions for Equilibrium: Empty Set of Pivotal Events or No Incentive

When equilibrium failure conditions are avoided, sustainability depends on the boundary posterior relations.

Lemma 3.12. *The following lemmas examine key conditions when a pivotal event does not exist, or, no type has incentive to deviate.*

Lemma 3.12a. *If* $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > \phi^1_{B_1} > 1$ *and* $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > 1 > \phi^0_{B_1}$ *hold, there exists no pivotal event regarding the change in posterior relation from* $\boldsymbol{m}^1$ *to* $\boldsymbol{m}^0$*, or, from* $\boldsymbol{m}^0$ *to* $\boldsymbol{m}^1$*, for type* $x_i \in \{G_0, G_1, B_0\}$*; type* B_1 *have no incentive to deviate.*

Proof. Conditional on observing $\boldsymbol{m}^1$, an agent of type G_0 is a pivotal sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{G_0} + n_{G_1}$, $n_{G_0} + n_{B_1} = 0$, and $n_{G_0} + n_{B_1} = n - 1$, which implies that $n_{G_1} > \frac{n}{2} - 1 > n_{G_1}$. Therefore, there exists no pivotal event for G_0.

Conditional on observing $\boldsymbol{m}^0$, an agent of type G_1 is a pivotal sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} - 1 > n_{G_0} + n_{G_1}$, $n_{G_0} + n_{B_1} = 0$, and $n_{G_0} + n_{B_1} = n - 1$, which implies that $n_{G_1} > \frac{n}{2} - 1 > n_{G_1}$. Therefore, there exists no pivotal event for G_1.

Conditional on observing $\boldsymbol{m}^0$, an agent of type B_0 is a pivotal message sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} > n_{G_0} + n_{G_1}$, $n_{G_0} + n_{B_1} = 0$, and $n_{G_0} + n_{B_1} = n - 1$, which implies that $n_{G_1} > \frac{n}{2} > n_{G_1}$. Therefore, there exists no pivotal event for B_0.

Conditional on observing $\boldsymbol{m}^1$, an agent of type B_1 is a pivotal message sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + (n_{B_1} + 1) > \frac{n}{2} > n_{G_0} + n_{G_1}$ and $n_{G_0} + n_{B_1} = 0$, which implies that $n_{G_1} > \frac{n}{2} - 1 > n_{G_1} - 1$. The only pivotal event is where $n_{G_0} = n_{B_1} = 0$ and $n_{G_1} = n_{B_0} = \frac{n-1}{2}$. Hence, the posterior belief conditional on pivotal events is,

$$\hat{\phi}_{B_1} \equiv \frac{p(1-q)\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}}{(1-p)q\binom{n-1}{\frac{n-1}{2}}\alpha^{\frac{n-1}{2}}(1-\alpha)^{\frac{n-1}{2}}(1-q)^{\frac{n-1}{2}}q^{\frac{n-1}{2}}} = \frac{p(1-q)}{(1-p)q}.$$

Recall that, by Lemma 3.7a, $p \geq q$ must hold for sustainability of sincere voting if $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > \phi^1_{B_1} > 1$. Therefore, we have $\hat{\phi}_{B_1} > 1$. Since type B_1 prefer to mismatch the state, they prefer the decision $d = 1$. If they deviate from $m_i = 0$ to $m_i = 1$, all type B_1 will believe that $\omega = 1$ is more likely and cast a vote $v_j = 0$. Thus, the final decision is changed from $d = 1$ to $d = 0$ in pivotal events, which results in reduction in their expected payoff. Therefore, they have no incentive to deviate. □

Lemma 3.12b. *If $\phi^{k^{(ii)}}_{B_0} > \phi^{k^{(ii)}}_{G_0} > \phi^{k^{(ii)}}_{G_1} > 1 > \phi^{k^{(ii)}}_{B_1}$ and $\phi^{k^{(ii)}-1}_{B_0} > \phi^{k^{(ii)}-1}_{G_0} > 1 > \phi^{k^{(ii)}-1}_{G_1} > \phi^{k^{(ii)}-1}_{B_1}$ hold for some $k^{(ii)} \in \{\frac{n+1}{2}+1, \frac{n+1}{2}+2, ..., n\}$, there exists no pivotal event regarding the change in posterior relation from $\boldsymbol{m}^{(ii)}$ to $\boldsymbol{m}^{k^{(ii)}-1}$, or, from $\boldsymbol{m}^{k^{(ii)}-1}$ to $\boldsymbol{m}^{k^{(ii)}}$, for all type $x_i \in \{G_0, G_1, B_0, B_1\}$.*

Proof. Conditional on observing $\boldsymbol{m}^{k^{(ii)}}$, an agent of type G_0 is a pivotal sender if and only if $\boldsymbol{n_{-i}}$ satisfies $(n_{G_0} + 1) + n_{G_1} + n_{B_1} > \frac{n}{2} > (n_{G_0} + 1) + n_{B_1}$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(ii)} - 1)$, which implies that $\frac{n}{2} > (n_{G_0} + 1) + n_{B_1} = k^{(ii)} \geq \frac{n+1}{2} + 1$. Therefore, there exists no pivotal event for G_0.

Conditional on observing $\boldsymbol{m}^{k^{(ii)}-1}$, an agent of type G_1 is a pivotal sender if and only if $\boldsymbol{n_{-i}}$

satisfies $n_{G_0} + (n_{G_1} + 1) + n_{B_1} > \frac{n}{2} > n_{G_0} + n_{B_1}$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_1} + n_{B_0} = (n-1) - (k^{(ii)} - 1)$, which implies that $\frac{n}{2} > n_{G_0} + n_{B_1} = k^{(ii)} - 1 \geq \frac{n+1}{2}$. Therefore, there exists no pivotal event for G_1.

Conditional on observing $\boldsymbol{m}^{k^{(ii)}-1}$, an agent of type B_0 is a pivotal message sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + n_{B_1} > \frac{n}{2} > n_{G_0} + n_{B_1}$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_0} + n_{B_1} = (n-1) - (k^{(ii)} - 1)$, which implies that $\frac{n}{2} > n_{G_0} + n_{B_1} = k^{(ii)} - 1 \geq \frac{n+1}{2}$. Therefore, there exists no pivotal event for B_0.

Conditional on observing $\boldsymbol{m}^{k^{(ii)}}$, an agent of type B_1 is a pivotal message sender if and only if $\boldsymbol{n_{-i}}$ satisfies $n_{G_0} + n_{G_1} + (n_{B_1} + 1) > \frac{n}{2} > n_{G_0} + (n_{B_1+1})$, $n_{G_0} + n_{B_1} = k^{(ii)} - 1$, and $n_{G_0} + n_{B_1} = (n-1) - (k^{(ii)} - 1)$, which implies that $\frac{n}{2} > n_{G_0} + (n_{B_1+1}) = k^{(ii)} \geq \frac{n+1}{2} + 1$. Therefore, there exists no pivotal event for B_1. □

Lemma 3.12c. *If $\phi^1_{B_0} > \phi^1_{G_0} > 1 > \phi^1_{G_1} > \phi^1_{B_1}$ and $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ hold, there exists no pivotal event regarding the change in posterior relation from $\boldsymbol{m}^1$ to $\boldsymbol{m}^0$, or, from $\boldsymbol{m}^0$ to $\boldsymbol{m}^1$.*

Proof. Recall that $\boldsymbol{n_{-i}}$ is a pivotal event regarding the change in posterior relation from $\boldsymbol{m}^1$ to $\boldsymbol{m}^0$, or, from $\boldsymbol{m}^0$ to $\boldsymbol{m}^1$ only if $n_{G_0} + n_{B_1} > \frac{n}{2}$. Nevertheless, according to non-truthful message strategy, here we have $n_{G_0} + n_{B_1} \in \{1, 0\}$, which implies $n_{G_0} + n_{B_1} < \frac{n}{2}$. Hence, there exists no pivotal event. □

Complete Characterisation via Boundary Analysis

The sustainability analysis focuses on extreme message profiles m^n and m^0 as these capture the boundary conditions determining equilibrium existence.

Lemma 3.13. *The following lemmas examine implication of posterior relation given a all-one message profile.*

Lemma 3.13a. *If $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1} > 1$ holds, non-truthful equilibrium exists if and only if $p \geq q$. In addition, $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$.*

Proof. By Lemma 3.5, $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ for all $k \in \{0, 1, 2, ..., n\}$, which implies that there exists no change in posterior relation. Therefore, no type has incentive to deviate from non-truthful message strategy. By Lemma 3.7a, sustainability of sincere voting holds if and only if $p \geq q$. □

Lemma 3.13b. *Suppose that* $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > 1 > \phi^0_{B_1}$ *holds. Non-truthful equilibrium exists if and only if* $p \geq q$*. In addition,* $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$.

Proof. By Lemma 3.6a, we have $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > \phi^1_{B_1} > 1$. By Lemma 3.5, we have $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ for all $k \in \{1, 2, ..., n\}$. By Lemma 3.12a, there exists no incentive to deviate from non-truthful message strategy regarding the change in posterior relation from $\boldsymbol{m}^1$ to $\boldsymbol{m}^0$, or, from $\boldsymbol{m}^0$ to $\boldsymbol{m}^1$. By Lemma 3.7a, sustainability of sincere voting holds if and only if $p \geq q$. By Lemma 3.7b, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^0$. □

Lemma 3.13c. *If* $\phi^0_{B_0} > \phi^0_{G_0} > 1 > \phi^0_{G_1} > \phi^0_{B_1}$ *holds, non-truthful equilibrium exists if and only if* $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$ *or* $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$.

Proof. By Lemma 3.5, we cannot have $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ or $\phi^k_{B_0} > 1 > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{0, 1, 2, ..., n\}$.

If $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$, by Lemma 3.5, we have $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{0, 1, 2, ..., n\}$, which implies that there exists no change in posterior relation. Therefore, no type has incentive to deviate from non-truthful message strategy. By Lemma 3.7c, sustainability of sincere voting always holds.

If $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$, by Lemma 3.5, we cannot have $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1} > 1$; by Lemma 3.6a, we cannot have $1 > \phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$ or $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > 1 > \phi^{n-1}_{B_1}$. Hence, we have $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > 1 > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$ or $\phi^{n-1}_{B_0} > 1 > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$.

Nevertheless, if $\phi^{n-1}_{B_0} > 1 > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$, by Lemma 3.6b, we have $1 > \phi^{n-2}_{B_0} > \phi^{n-2}_{G_0} > \phi^{n-2}_{G_1} > \phi^{n-2}_{B_1}$, which implies that $1 > \phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ by Lemma 3.5. Contradiction.

Hence, we have $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > 1 > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$, and thus, by Lemma 3.5, $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{0, 1, 2, ..., n-1\}$. By Lemma 3.7b, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^n$. By Lemma 3.7c, sustainability of sincere voting always holds. □

Lemma 3.13d. *If* $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ *holds, non-truthful equilibrium exists if and only if* $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$ *or* $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$.

Proof. By Lemma 3.5, we cannot have $1 > \phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{0, 1, 2, ..., n\}$. By Lemma 3.6b, we cannot have $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > \phi^1_{B_1} > 1$ and $\phi^1_{B_0} > 1 > \phi^1_{G_0} > \phi^1_{G_1} > \phi^1_{B_1}$. Hence, we have $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > 1 > \phi^1_{B_1}$ or $\phi^1_{B_0} > \phi^1_{G_0} > 1 > \phi^1_{G_1} > \phi^1_{B_1}$.

If $\phi^1_{B_0} > \phi^1_{G_0} > \phi^1_{G_1} > 1 > \phi^1_{B_1}$, by Lemma 3.6a, we have $\phi_{B_0}(\boldsymbol{m}^2) > \phi_{G_0}(\boldsymbol{m}^2) > \phi_{G_1}(\boldsymbol{m}^2) > \phi_{B_1}(\boldsymbol{m}^2) > 1$. By Lemma 3.10, either type G_0 or G_1 has incentive to deviate from non-truthful message strategy, which implies that non-truthful equilibrium does not exist.

If $\phi^1_{B_0} > \phi^1_{G_0} > 1 > \phi^1_{G_1} > \phi^1_{B_1}$, by Lemma 3.12c, there exists no incentive to deviate from non-truthful message strategy regarding the change in posterior relation from $\boldsymbol{m}^1$ to $\boldsymbol{m}^0$, or, from $\boldsymbol{m}^0$ to $\boldsymbol{m}^1$. Hence, non-truthful message strategy is sustainable only if $\phi^1_{B_0} > \phi^1_{G_0} > 1 > \phi^1_{G_1} > \phi^1_{B_1}$.

By Lemma 3.5, we have $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$, $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$, or $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$.

If $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$, by Lemma 3.5, we have $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{1, 2, ..., n\}$. Therefore, no type has incentive to deviate from non-truthful message strategy. Equilibrium exists.

If $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$, by Lemma 3.6a, we cannot have $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > \phi^{n-1}_{G_1} > 1 > \phi^{n-1}_{B_1}$. Hence, we have $\phi^{n-1}_{B_0} > \phi^{n-1}_{G_0} > 1 > \phi^{n-1}_{G_1} > \phi^{n-1}_{B_1}$; otherwise, Lemma 3.5 is violated. By Lemma 3.12b, there exists no incentive to deviate from non-truthful message strategy regarding the change in posterior relation from $\boldsymbol{m}^n$ to $\boldsymbol{m}^{n-1}$, or, from $\boldsymbol{m}^{n-1}$ to $\boldsymbol{m}^n$. By Lemma 3.5, we have $\phi^k_{B_0} > \phi^k_{G_0} > 1 > \phi^k_{G_1} > \phi^k_{B_1}$ for all $k \in \{1, 2, ..., n-1\}$. Therefore, no type has incentive to deviate from non-truthful message strategy. By Lemma 3.7b, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^n$. By Lemma 3.7c, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^k$ for all $k \in \{1, 2, ..., n-1\}$. By Lemma 3.7d, no type has incentive to deviate from sincere voting conditional on observing $\boldsymbol{m}^0$. Equilibrium exists.

If $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$, by Lemma 3.5 and 3.6c, $\exists k' \in \{2, 3, ..., n-1\}$ such that $\phi_{B_0}(\boldsymbol{m}^{k'}) > \phi^{k'}_{G_0} > \phi^{k'}_{G_1} > 1 > \phi^{k'}_{B_1}$. By Lemma 3.10, either type G_0 or G_1 has incentive to deviate from non-truthful message strategy, which implies that non-truthful equilibrium does not exist. □

Lemma 3.13e. *If* $1 > \phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$ *holds, non-truthful equilibrium does not exist.*

Proof. By Lemma 3.7e, sincere voting is not sustainable. Thus, non-truthful equilibrium does not exist. □

Key Result: The analysis establishes that equilibrium existence depends entirely on the posterior relations induced by the all-zero message profile m^n and all-one message profile m^0, with the additional requirement that $p \geq q$ when agents form sufficiently high posterior beliefs.

Lemma 3.14. *A non-truthful equilibrium exists if and only if the posterior relation induced by* $\boldsymbol{m}^n$ *and* $\boldsymbol{m}^0$ *satisfies one of the following conditions, and,* $p \geq q$ *if* $\phi^k_{B_0} > \phi^k_{G_0} > \phi^k_{G_1} > \phi^k_{B_1} > 1$ *holds for some* $k \in \{0, 1, .., n\}$.

(i) $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$, *and* $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1} > 1$.

(ii) $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > \phi^n_{B_1} > 1$, *and* $\phi^0_{B_0} > \phi^0_{G_0} > \phi^0_{G_1} > 1 > \phi^0_{B_1}$.

(iii) $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$, *and* $\phi^0_{B_0} > \phi^0_{G_0} > 1 > \phi^0_{G_1} > \phi^0_{B_1}$.

(iv) $\phi^n_{B_0} > \phi^n_{G_0} > \phi^n_{G_1} > 1 > \phi^n_{B_1}$, *and* $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$.

(v) $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$, *and* $\phi^0_{B_0} > \phi^0_{G_0} > 1 > \phi^0_{G_1} > \phi^0_{B_1}$.

(vi) $\phi^n_{B_0} > \phi^n_{G_0} > 1 > \phi^n_{G_1} > \phi^n_{B_1}$, *and* $\phi^0_{B_0} > 1 > \phi^0_{G_0} > \phi^0_{G_1} > \phi^0_{B_1}$.

Proof. By Lemma 3.13a to 3.13e, we have Lemma 3.14. □

This complete technical analysis establishes the foundation for Proposition 3.5 in the main text, demonstrating that non-truthful equilibrium exists when signals are either highly informative (making agents rely primarily on private information) or when priors dominate (making message content largely irrelevant), but fails for intermediate parameter values where strategic manipulation becomes profitable whilst messages retain significant influence on decision-making.

Proof of Proposition 3.5. By Lemma 3.7a and 3.14, the sufficient and necessary condition for a non-truthful equilibrium is:

(i) if $\phi^n_{B_1} > 1, \phi^0_{G_1} > 1$ and $p \geq q$;

(ii) if $\phi^n_{B_1} < 1, \phi^0_{B_0} > 1$.

which implies:

(i) If $\frac{p}{1-p}\frac{1-q}{q}(\frac{\tilde{q}}{1-\tilde{q}})^{n-1} > 1$, $\frac{p}{1-p}\frac{1-q}{q}(\frac{\tilde{q}}{1-\tilde{q}})^{1-n} > 1$ and $p \geq q$.

(ii) If $\frac{p}{1-p}\frac{1-q}{q}(\frac{\tilde{q}}{1-\tilde{q}})^{n-1} < 1$, $\frac{p}{1-p}\frac{q}{1-q}(\frac{\tilde{q}}{1-\tilde{q}})^{1-n} > 1$.

□

3.A.4 Efficiency

Proposition 3.7 (Efficiency: Truthfully Revealing Equilibrium)**.** *Given (α, p, q, n), the efficiency under a truthfully revealing equilibrium is:*

$$\begin{aligned}
\psi_{TRE}(\alpha, p, q, n) = p & \left[\sum_{k=0}^{k^*-1} \left\{ \sum_{i=0}^{k} \xi(i,k)\chi(i,k,0,i) + \sum_{i=k+1}^{\frac{n-1}{2}} \xi(i,k)\chi(i,k,0,k) \right\} \right. \\
& + \sum_{k=k^*}^{\frac{n-1}{2}} \left\{ \sum_{i=\frac{n+1}{2}}^{n-k} \xi(i,k)\chi(i,k,0,k) + \sum_{i=n-k+1}^{n} \xi(i,k)\chi(i,k,k-(n-i),k) \right\} \\
& \left. + \sum_{k=\frac{n+1}{2}}^{n} \left\{ \sum_{i=\frac{n+1}{2}}^{k} \xi(i,k)\chi(i,k,k-(n-i),i) + \sum_{i=k+1}^{n} \xi(i,k)\chi(i,k,k-(n-i),k) \right\} \right] \\
+ (1-p) & \left[\sum_{k=0}^{k^*-1} \left\{ \sum_{i=\frac{n+1}{2}}^{n-k} \xi(i,n-k)\chi(i,k,0,k) + \sum_{i=n-k+1}^{n} \xi(i,n-k)\chi(i,k,k-(n-i),k) \right\} \right. \\
& + \sum_{k=k^*}^{\frac{n-1}{2}} \left\{ \sum_{i=0}^{k} \xi(i,n-k)\chi(i,k,0,i) + \sum_{i=k+1}^{\frac{n-1}{2}} \xi(i,k)\chi(i,k,0,k) \right\} \\
& \left. + \sum_{k=\frac{n+1}{2}}^{n} \left\{ \sum_{i=0}^{n-k-1} \xi(i,n-k)\chi(i,k,0,i) + \sum_{i=n-k+1}^{\frac{n-1}{2}} \xi(i,n-k)\chi(i,k,k-(n-i),i) \right\} \right],
\end{aligned}$$

where $\xi(i,k) \equiv \alpha^i(1-\alpha)^{n-i}q^k(1-q)^{n-k}$ and $\chi(i,k,j_o,j_f) \equiv \binom{n}{i}\sum_{j=j_o}^{j_f}\binom{i}{j}\binom{n-i}{k-j}$. Moreover, given any (p,q,n), the efficiency is $\frac{1}{2}$ if $\alpha = \frac{1}{2}$,

$$\psi_{TRE}(\frac{1}{2}, p, q, n) = \frac{1}{2}.$$

Proof of Proposition 3.6. Under a truthfully revealing equilibrium, good agents cast $v_i = 0$ and bad agents cast $v_i = 1$ if and only if $k \geq k^*$. If $k < k^*$, good agents cast $v_i = 1$ and bad agents cast $v_i = 0$. Moreover, given truthfully deliberating, we have $n'_{G_0} + n'_{B_0} = k$. Hence, the efficiency of a truthfully revealing equilibrium can be expressed as,

$$\begin{aligned}
\psi_{TRE}(\boldsymbol{n}) = & \mathbb{P}(\omega = 0, n'_{G_0} + n'_{G_1} > \frac{n}{2}, n'_{G_0} + n'_{B_0} \geq k^*) + \mathbb{P}(\omega = 0, n'_{G_0} + n'_{G_1} < \frac{n}{2}, n'_{G_0} + n'_{B_0} < k^*) \\
& + \mathbb{P}(\omega = 1, n'_{G_0} + n'_{G_1} > \frac{n}{2}, n'_{G_0} + n'_{B_0} < k^*) + \mathbb{P}(\omega = 1, n'_{G_0} + n'_{G_1} < \frac{n}{2}, n'_{G_0} + n'_{B_0} \geq k^*).
\end{aligned}$$

which implies the following proposition regarding efficiency under a truthfully revealing equilibrium.

Note that, conditional on $\omega = 0$, a type profile $\boldsymbol{n} = (a, b, c, d)$ induces $d = 0$ if and only if $\boldsymbol{n} = (c, d, a, b)$ induces $d = 1$. In addition, note that the probability of the type profile conditional on the state is, $\mathbb{P}((a, b, c, d) \mid \omega = 0) = \alpha^{a+b}(1-\alpha)^{c+d}q^{a+c}(1-q)^{b+d}\binom{n}{a+b}\binom{a+b}{a}\binom{c+d}{c}$. Given that $\alpha = \frac{1}{2}$ and $q = \frac{1}{2}$, $\mathbb{P}((a, b, c, d) \mid \omega = 0) = (\frac{1}{2})^n(\frac{1}{2})^n\binom{n}{a+b}\binom{a+b}{a}\binom{c+d}{c}$. Likewise, $\mathbb{P}((c, d, a, b) \mid \omega = 0) = (\frac{1}{2})^n(\frac{1}{2})^n\binom{n}{c+d}\binom{c+d}{c}\binom{a+b}{c}$. Since $\binom{n}{a+b} = \binom{n}{c+d}$, $\mathbb{P}((a, b, c, d) \mid \omega = 0)$ $= \mathbb{P}((c, d, a, b) \mid \omega = 0)$. Hence, if we sum over the conditional probability of $d = 0$ and $d = 1$, we have, $\sum \mathbb{P}((a, b, c, d) \mid \omega = 0) = \sum \mathbb{P}((c, d, a, b) \mid \omega = 0) \implies \mathbb{P}(d = 0 \mid \omega = 0)$ $= \mathbb{P}(d = 1 \mid \omega = 0) \implies \mathbb{P}(d = 0 \mid \omega = 0) = \frac{1}{2}$. Likewise, $\mathbb{P}(d = 0 \mid \omega = 1) = \mathbb{P}(d = 1 \mid \omega = 1)$ $= \frac{1}{2}$. Therefore, $\mathbb{P}(d = \omega) = \mathbb{P}(\omega = 0)\mathbb{P}(d = 0 \mid \omega = 0) + \mathbb{P}(\omega = 1)\mathbb{P}(d = 1 \mid \omega = 1)$ $= p \cdot \frac{1}{2} + (1-p) \cdot \frac{1}{2} = \frac{1}{2}$. □

Proof of Proposition 3.7. In Section 3.4.2, it is shown that, under a babbling equilibrium, good agents (i.e., types G_0 and G_1) cast $v_i = 0$, and bad agents (i.e., types B_0 and B_1) cast $v_i = 1$ if $p \geq q$, regardless of their private signals. When $p < q$, each agent believes that the state indicated by their own signal is more likely to be true. In this case, types G_0 and B_1 cast $v_i = 0$, whilst types G_1 and B_0 cast $v_i = 1$. Hence, the efficiency of a babbling equilibrium can be expressed as follows.

$$\psi_{BE}(\boldsymbol{n}) = \begin{cases} \mathbb{P}(\omega = 0, n'_{G_0} + n'_{G_1} > \frac{n}{2}) + \mathbb{P}(\omega = 1, n'_{G_0} + n'_{G_1} < \frac{n}{2}), & \text{if } p \geq q; \\ \mathbb{P}(\omega = 0, n'_{G_0} + n'_{B_1} > \frac{n}{2}) + \mathbb{P}(\omega = 1, n'_{G_0} + n'_{B_1} < \frac{n}{2}), & \text{if } p < q, \end{cases}$$

which suggests the analytical solution of efficiency under a babbling equilibrium as in the proposition.

Suppose that $p \geq q$. Type G_0 and G_1 cast $v_i = 0$ whilst type G_0 and G_1 cast $v_i = 1$. Hence, the efficiency under a babbling equilibrium is:

$$\begin{aligned}
\psi(\alpha, p, q, n)_{BE} &= \mathbb{P}(\omega = 0)\mathbb{P}(n'_{G_0} + n'_{G_1} > \frac{n}{2} \mid \omega = 0) + \mathbb{P}(\omega = 1)\mathbb{P}(n'_{G_0} + n'_{G_1} < \frac{n}{2} \mid \omega = 1) \\
&= p \sum_{n'_{G_0} + n'_{G_1} = \frac{n+1}{2}}^{n} \binom{n}{n'_{G_0} + n'_{G_1}} \alpha^{n'_{G_0} + n'_{G_1}} (1-\alpha)^{n - (n'_{G_0} + n'_{G_1})} \\
&+ (1-p) \sum_{n'_{G_0} + n'_{G_1} = 0}^{\frac{n-1}{2}} \binom{n}{n'_{G_0} + n'_{G_1}} \alpha^{n'_{G_0} + n'_{G_1}} (1-\alpha)^{n - (n'_{G_0} + n'_{G_1})} \\
&= p \sum_{i=0}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} + (1-2p) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}
\end{aligned}$$

$$= p(\alpha + (1-\alpha))^n - (2p-1) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}$$
$$= p - (2p-1) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}.$$

Suppose that $p < q$. Type G_0 and B_1 cast $v_i = 0$ whilst type G_1 and B_0 cast $v_i = 1$. Hence, the efficiency under a babbling equilibrium is:

$$\begin{aligned}
\psi(\alpha,p,q,n)_{BE} =& \mathbb{P}(\omega=0)\mathbb{P}(n'_{G_0}+n'_{B_1} > \frac{n}{2} \mid \omega=0) + \mathbb{P}(\omega=1)\mathbb{P}(n'_{G_0}+n'_{B_1} < \frac{n}{2} \mid \omega=1) \\
=& p \sum_{n'_{G_0}+n'_{B_1}=\frac{n+1}{2}}^{n} \binom{n}{n'_{G_0}+n'_{B_1}} \Bigg[\sum_{n'_{G_0}=0}^{n'_{G_0}+n'_{B_1}} \binom{n'_{G_0}+n'_{B_1}}{n'_{G_0}} \Big(\alpha q\Big)^{n'_{G_0}} \Big((1-\alpha)(1-q)\Big)^{(n'_{G_0}+n'_{B_1})-n'_{G_0}} \\
& \sum_{n'_{G_1}=0}^{n-(n'_{G_0}+n'_{B_1})} \binom{n-(n'_{G_0}+n'_{B_1})}{n'_{G_1}} \Big(\alpha(1-q)\Big)^{n'_{G_1}} \Big((1-\alpha)q\Big)^{\left(n-(n'_{G_0}+n'_{B_1})\right)-n'_{G_1}} \Bigg] \\
&+(1-p) \sum_{n'_{G_0}+n'_{B_1}=0}^{\frac{n-1}{2}} \binom{n}{n'_{G_0}+n'_{B_1}} \Bigg[\sum_{n'_{G_0}=0}^{n'_{G_0}+n'_{B_1}} \binom{n'_{G_0}+n'_{B_1}}{n'_{G_0}} \Big(\alpha(1-q)\Big)^{n'_{G_0}} \Big((1-\alpha)q\Big)^{(n'_{G_0}+n'_{B_1})-n'_{G_0}} \\
& \sum_{n'_{G_1}=0}^{n-(n'_{G_0}+n'_{B_1})} \binom{n-(n'_{G_0}+n'_{B_1})}{n'_{G_1}} \Big(\alpha q\Big)^{n'_{G_1}} \Big((1-\alpha)(1-q)\Big)^{\left(n-(n'_{G_0}+n'_{B_1})\right)-n'_{G_1}} \Bigg] \\
=& p \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \Bigg[\sum_{j=0}^{i} \binom{i}{j} \Big(\alpha q\Big)^j \Big((1-\alpha)(1-q)\Big)^{i-j} \sum_{k=0}^{n-i} \binom{n-i}{k} \Big(\alpha(1-q)\Big)^k \Big((1-\alpha)q\Big)^{n-i-k} \Bigg] \\
&+(1-p) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \Bigg[\sum_{j=0}^{i} \binom{i}{j} \Big(\alpha(1-q)\Big)^j \Big((1-\alpha)q\Big)^{i-j} \sum_{k=0}^{n-i} \binom{n-i}{k} \Big(\alpha q\Big)^k \Big((1-\alpha)(1-q)\Big)^{n-i-k} \Bigg] \\
=& p \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \Bigg[\Big(\alpha q + (1-\alpha)(1-q)\Big)^i \Big(\alpha(1-q) + (1-\alpha)q\Big)^{n-i} \Bigg] \\
&+(1-p) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \Bigg[\Big(\alpha(1-q) + (1-\alpha)q\Big)^i \Big(\alpha q + (1-\alpha)(1-q)\Big)^{n-i} \Bigg] \\
=& p \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \Bigg[\Big(\tilde{q}\Big)^i \Big(1-\tilde{q}\Big)^{n-i} \Bigg] + (1-p) \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} \Bigg[\Big(1-\tilde{q}\Big)^i \Big(\tilde{q}\Big)^{n-i} \Bigg] \\
=& \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} \Bigg[\Big(\tilde{q}\Big)^i \Big(1-\tilde{q}\Big)^{n-i} \Bigg].
\end{aligned}$$

□

Proof of Proposition 3.8. Denote the decision under a truthfully revealing equilibrium and a babbling equilibrium as d_{TRE} and d_{BE}, respectively. Note that, $d_{TRE} = 0$ and $d_{BE} = 1$ is decided with a type profile $\boldsymbol{n} = (a, b, c, d)$ if and only if $d_{TRE} = 1$ and $d_{BE} = 0$ is decided with (c, d, a, b).

The sum of the impact of the two type profile on the difference in efficiency is:

$$\begin{aligned}\Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)} = & \binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{c}\Big\{p\Big[\alpha^{a+b}(1-\alpha)^{c+d}q^{a+c}(1-q)^{b+d}\Big] \\ & -(1-p)\Big[\alpha^{a+b}(1-\alpha)^{c+d}q^{b+d}(1-q)^{a+c}\Big]\Big\} \\ & +\binom{n}{c}\binom{n-c}{b}\binom{n-c-b}{a}\Big\{-p\Big[\alpha^{c+d}(1-\alpha)^{a+b}q^{c+a}(1-q)^{d+b}\Big] \\ & +(1-p)\Big[\alpha^{c+d}(1-\alpha)^{a+b}q^{d+b}(1-q)^{c+a}\Big]\Big\} \\ = & \binom{n}{a}\binom{n-a}{b}\binom{n-a-b}{c}\Big[\alpha^{a+b}(1-\alpha)^{c+d} - \alpha^{c+d}(1-\alpha)^{a+b}\Big] \\ & \Big[pq^{a+c}(1-q)^{b+d} - (1-p)q^{b+d}(1-q)^{a+c}\Big].\end{aligned}$$

Suppose that $p \geq q$. Then, we have, $n'_{G_0} + n'_{G_1} < \frac{n}{2} \implies a+b < \frac{n}{2}$; $n'_{G_0} + n'_{B_0} < k^* \implies pq^k(1-q)^{n-k} < (1-p)(1-q)^k q^{n-k} \implies pq^{a+c}(1-q)^{b+d} < (1-p)(1-q)^{b+d}q^{a+c}$. Hence, $\Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)} > 0$. Suppose that $p < q$. Then, we have either:

(i) $n'_{G_0} + n'_{G_1} > \frac{n}{2} \implies a+b > \frac{n}{2}$; $n'_{G_0} + n'_{B_0} \geq k^* \implies pq^k(1-q)^{n-k} > (1-p)(1-q)^k q^{n-k} \implies pq^{a+c}(1-q)^{b+d} > (1-p)(1-q)^{b+d}q^{a+c}$. Hence, $\Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)} > 0$, or,

(ii) $n'_{G_0} + n'_{G_1} < \frac{n}{2} \implies a+b < \frac{n}{2}$; $n'_{G_0} + n'_{B_0} < k^* \implies pq^k(1-q)^{n-k} < (1-p)(1-q)^k q^{n-k} \implies pq^{a+c}(1-q)^{b+d} < (1-p)(1-q)^{b+d}q^{a+c}$. Hence, $\Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)} > 0$.

Thus, the difference in efficiency is the summation of $\Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)}$ over the subset of type profiles, denoted as $\widehat{\mathcal{N}} \subset \mathbb{N}^4$, such that any type profile in the subsets induces d_{TRE} and d_{BE}.

$$\sum_{\{\boldsymbol{n}=(a,b,c,d)\in\widehat{\mathcal{N}}\}} \Delta_{(a,b,c,d)} + \Delta_{(c,d,a,b)} > 0.$$

□

Lemma 3.15.

$$\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q^i\,(1-q)^{n-i} > q, \;\; \forall q \in \left(\frac{1}{2}, 1\right], \;\; n \;\; odd.$$

Proof. Note that we have:

$$\frac{\binom{n}{i}q^i(1-q)^{n-i}}{\binom{n}{n-i}q^{n-i}(1-q)^i} = \left(\frac{q}{1-q}\right)^{2i-n} > \frac{q}{1-q}, \ \forall i \geq \frac{n+1}{2} \implies \frac{\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i}q^i\,(1-q)^{n-i}}{\sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i}q^i\,(1-q)^{n-i}} > \frac{q}{1-q}.$$

According to Binomial Theorem, we have:

$$\sum_{i=0}^{n} \binom{n}{i} q^i\,(1-q)^{n-i} = \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q^i\,(1-q)^{n-i} + \sum_{i=0}^{\frac{n-1}{2}} \binom{n}{i} q^i\,(1-q)^{n-i} = 1$$
$$\implies \sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q^i\,(1-q)^{n-i} > q.$$

□

Appendix 3.B: Supplementary Figures

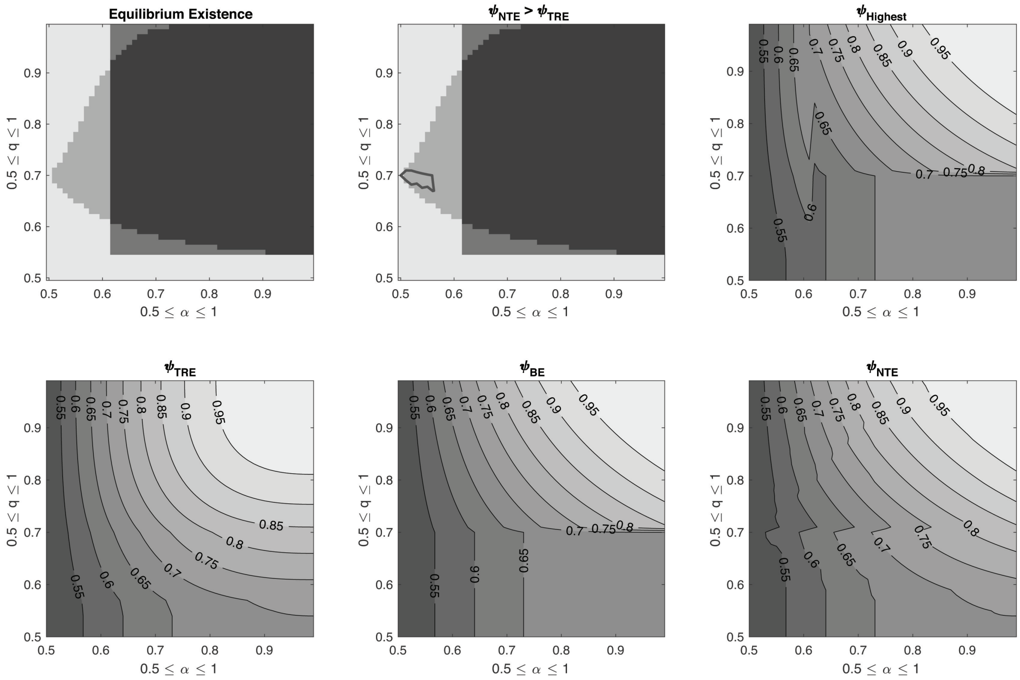

Figure 3.11: Summary of existence and efficiency: $p = 0.7; n = 5$

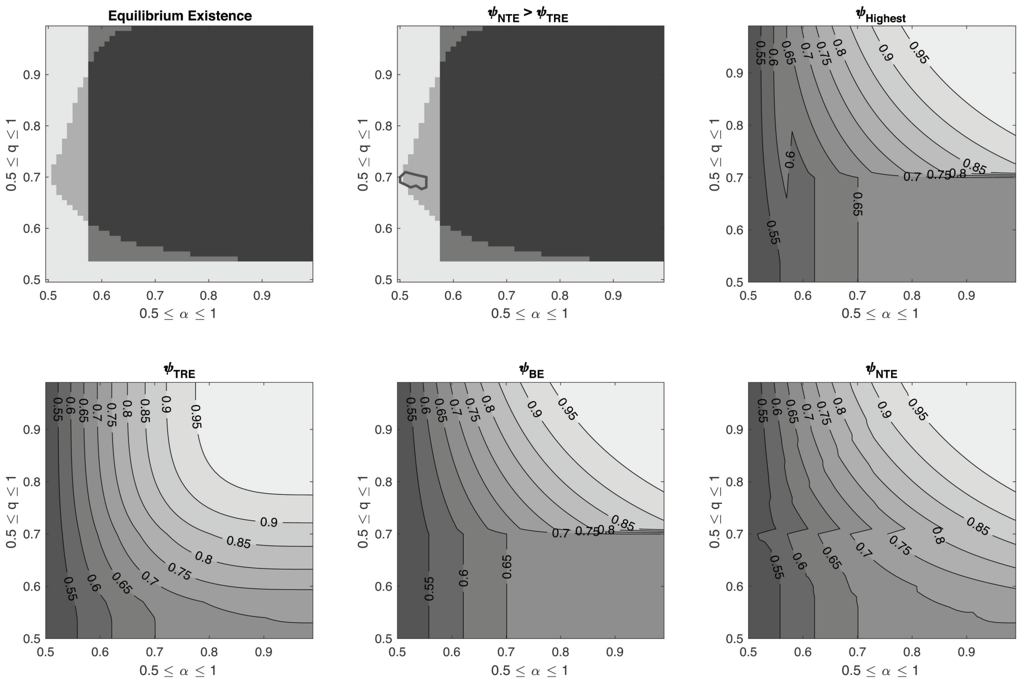

Figure 3.12: Summary of existence and efficiency: $p = 0.7; n = 7$

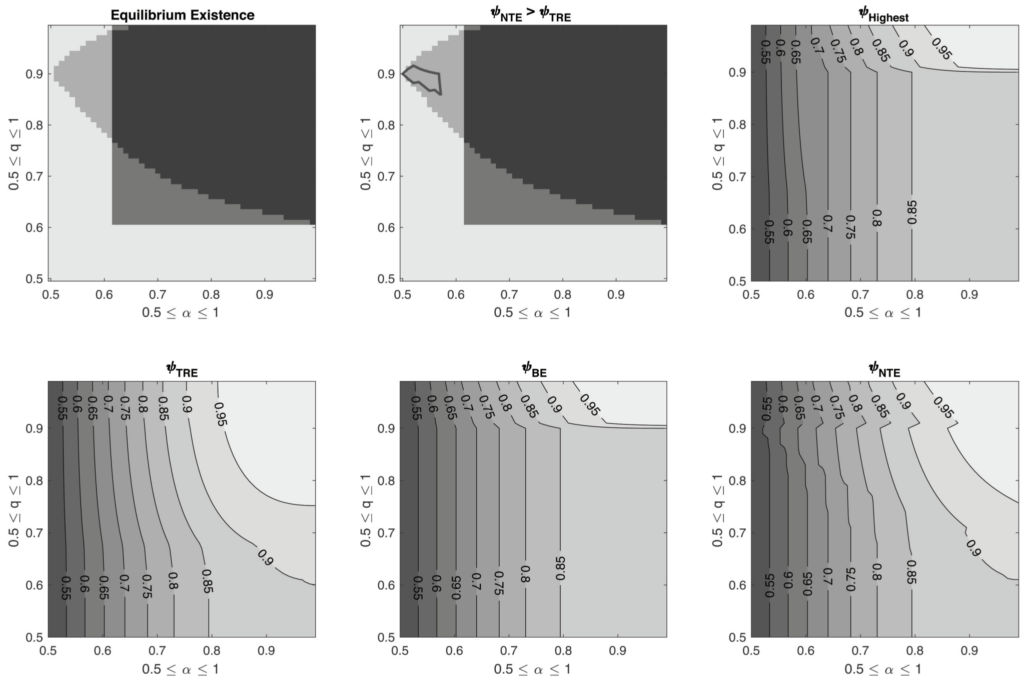

Figure 3.13: Summary of existence and efficiency: $p = 0.9$; $n = 5$

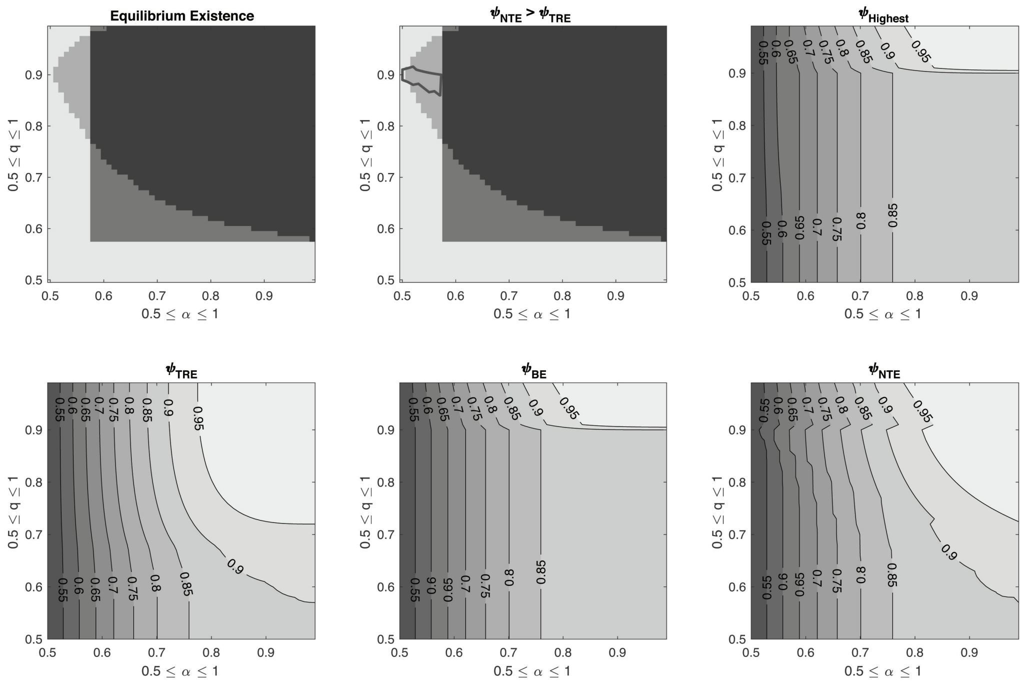

Figure 3.14: Summary of existence and efficiency: $p = 0.9$; $n = 7$

Bibliography

Abbas, A. E., Bakır, N. O., Klutke, G. A., & Sun, Z. (2013). Effects of risk aversion on the value of information in two-action decision problems. *Decision Analysis*, *10*(3), 257–275.

Admati, A. R., & Pfleiderer, P. (1990). Direct and indirect sale of information. *Econometrica*, *58*(4), 901–928.

Andries, M., & Haddad, V. (2020). Information aversion. *Journal of Political Economy*, *128*(5), 1901–1939.

Austen-Smith, D. (1990). Information transmission in debate. *American Journal of Political Science*, *34*(1), 124–152.

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, *90*(1), 34–45.

Austen-Smith, D., & Feddersen, T. J. (2005). Deliberation and voting rules. In D. Austen-Smith & J. Duggan (Eds.), *Social Choice and Strategic Decisions* (pp. 269–316). Springer.

Austen-Smith, D., & Feddersen, T. J. (2006). Deliberation, preference uncertainty, and voting rules. *American Political Science Review*, *100*(2), 209–217.

Bakır, N. O. (2015). Monotonicity of the selling price of information with risk aversion in two action decision problems. *Central European Journal of Economic Modelling and Econometrics*, *7*(2), 71–90.

Benkert, J. M. (2025). Bilateral trade with loss-averse agents. *Economic Theory*, *79*(2), 519–560.

Bergemann, D., & Bonatti, A. (2015). Selling cookies. *American Economic Journal: Microeconomics*, *7*(3), 259–294.

Bergemann, D., Bonatti, A., & Smolin, A. (2014). *Selling experiments: Menu pricing of information* (Cowles Foundation Discussion Paper No. 1952). Cowles Foundation for Research in Economics, Yale University.

Bergemann, D., Bonatti, A., & Smolin, A. (2018). The design and price of information. *American Economic Review*, *108*(1), 1–48.

Bhalgat, A., Chakraborty, T., & Khanna, S. (2012). Mechanism design for a risk averse seller. In P. W. Goldberg (Ed.), *Internet and Network Economics: WINE 2012* (pp. 198–211, Vol. 7695). Springer.

Cabrales, A., Gossner, O., & Serrano, R. (2013). Entropy and the value of information for investors. *American Economic Review*, *103*(1), 360–377.

Cabrales, A., Gossner, O., & Serrano, R. (2017). A normalized value for information purchases. *Journal of Economic Theory*, *170*, 266–288.

Campo, S., Guerre, E., Perrigne, I., & Vuong, Q. (2011). Semiparametric estimation of first-price auctions with risk-averse bidders. *The Review of Economic Studies*, *78*(1), 112–147.

Che, Y. K., & Mierendorff, K. (2019). Optimal dynamic allocation of attention. *American Economic Review*, *109*(8), 2993–3029.

Coase, R. H. (1972). Durability and monopoly. *The Journal of Law & Economics*, *15*(1), 143–149.

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

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–1451.

Dai, C. (2021). Optimal sequential contract with a risk-averse supplier. *Canadian Journal of Economics*, *54*(1), 92–125.

Daskalakis, C., Papadimitriou, C., & Tzamos, C. (2016). Does information revelation improve revenue? In *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 233–250). Association for Computing Machinery.

Deimen, I., Ketelaar, F., & Le Quement, M. T. (2015). Consistency and communication in committees. *Journal of Economic Theory*, *160*, 24–35.

Doval, L., & Skreta, V. (2022). Mechanism design with limited commitment. *Econometrica*, *90*(4), 1463–1500.

Esö, P., & Futo, G. (1999). Auction design with a risk averse seller. *Economics Letters*, *65*(1), 71–74.

Esö, P., & Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, *74*(3), 705–731.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives*, *10*(3), 103–118.

Feddersen, T., & Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, *92*(1), 23–35.

Galeotti, A., Ghiglino, C., & Squintani, F. (2013). Strategic information transmission networks. *Journal of Economic Theory*, *148*(5), 1751–1769.

Gerardi, D., McLean, R., & Postlewaite, A. (2009). Aggregation of expert opinions. *Games and Economic Behavior*, *65*(2), 339–371.

Gerardi, D., & Yariv, L. (2007). Deliberative voting. *Journal of Economic Theory*, *134*(1), 317–338.

Gershkov, A., Moldovanu, B., Strack, P., & Zhang, M. (2022). Optimal auctions: Non-expected utility and constant risk aversion. *The Review of Economic Studies*, *89*(5), 2630–2662.

Gould, J. P. (1974). Risk, stochastic preference, and the value of information. *Journal of Economic Theory*, *8*(1), 64–84.

Hagenbach, J., Koessler, F., & Perez-Richet, E. (2014). Certifiable pre-play communication: Full disclosure. *Econometrica*, *82*(3), 1093–1131.

Hörner, J., & Skrzypacz, A. (2016). Selling information. *Journal of Political Economy*, *124*(6), 1515–1562.

Jackson, M. O., & Tan, X. (2013). Deliberation, disclosure of information, and voting. *Journal of Economic Theory*, *148*(1), 2–30.

Le Quement, M. T., & Yokeeswaran, V. (2015). Subgroup deliberation and voting. *Social Choice and Welfare*, *45*, 155–186.

Liang, A., Mu, X., & Syrgkanis, V. (2022). Dynamically aggregating diverse information. *Econometrica*, *90*(1), 47–80.

Malenko, A., & Malenko, N. (2019). Proxy advisory firms: The economics of selling information to voters. *The Journal of Finance*, *74*(5), 2441–2490.

Maskin, E. S., & Riley, J. G. (1984). Optimal auctions with risk averse buyers. *Econometrica*, *52*(6), 1473–1518.

Meirowitz, A. (2007). In defense of exclusionary deliberation: Communication and voting with private beliefs and values. *Journal of Theoretical Politics*, *19*(3), 301–327.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73.

Nöldeke, G., & Samuelson, L. (2018). The implementation duality. *Econometrica*, *86*(4), 1283–1324.

Sundararajan, M., & Yan, Q. (2020). Robust mechanisms for risk-averse sellers. *Games and Economic Behavior*, *124*, 644–658.

Van Weelden, R. (2008). Deliberation rules and voting. *Quarterly Journal of Political Science*, *3*(1), 83–88.

Vasserman, S., & Watt, M. (2021). Risk aversion and auction design: Theoretical and empirical evidence. *International Journal of Industrial Organization*, *79*, 102758.

Visser, B., & Swank, O. H. (2007). On committees of experts. *The Quarterly Journal of Economics*, *122*(1), 337–372.

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

Zhong, W. (2018). *Selling information* [Preprint]. arXiv: 1809.06770.

Zhong, W. (2022). Optimal dynamic information acquisition. *Econometrica*, *90*(4), 1537–1582.