ESSAYS ON INFORMATION ECONOMICS

by

Nicolas Rodriguez Gonzalez

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2025

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Nicolas Rodriguez Gonzalez, titled *Essays on Information Economics,* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: 06/11/2025
Andreas Blume

Inga Deimen (Jun 12, 2025 09:02 GMT+2)
____ Date: 06/11/2025
Inga Deimen

LMathevet
____ Date: 06/11/2025
Laurent Mathevet

Amanda Friedenberg (Jun 12, 2025 13:38 EDT)
____ Date: 06/11/2025
Amanda Friedenberg

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 06/11/2025
Andreas Blume
Dissertation Committee Co-Chair
Department of Economics

Inga Deimen (Jun 12, 2025 09:02 GMT+2)
____ Date: 06/11/2025
Inga Deimen
Dissertation Committee Co-Chair
Department of Economics

ACKNOWLEDGMENTS

I am deeply grateful to my advisors, Inga Deimen and Andreas Blume, for their guidance, encouragement, patience, and intellectual support throughout my doctoral studies. Their mentorship has played a central role in my development as a researcher.

I would like to extend my gratitude to the other members of my dissertation committee, Amanda Friedenberg and Laurent Mathevet, for their thoughtful feedback, constructive criticism, and valuable discussions that helped strengthen my work.

I am especially thankful to Ernesto Rivera Mora, my coauthor and collaborator, for many stimulating conversations, shared ideas, and continuous support throughout the research process.

I would also like to acknowledge the staff and academic community at the University of Arizona. I am grateful for the supportive environment they foster, the enriching seminars and conversations that contributed to my growth, and the many efforts that help sustain a great academic community.

This dissertation would not have been possible without the contributions, guidance, and generosity of all these individuals. I am deeply appreciative of the time and attention they have devoted to my growth as a scholar.

DEDICATION

To my wife, Luisa Gonzalez.
To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation studies problems in information economics. The first chapter investigates monopolistic information selling when the Seller and the Buyer hold different prior beliefs about the state of the world. The main result shows that the Seller's revenue is maximized by gradually selling information over multiple periods. The second chapter examines optimal communication when multiple Senders share the same information but have access only to a limited language. The main result shows that Senders use their language heterogeneously, and assign non-convex meanings to individual messages while ensuring that the meaning of message profiles remains convex. The final chapter studies strategic communication when a decision-making Receiver must acquire a limited skill set in advance. The main result shows a negative relationship between the informativeness of the Sender's advice and the Receiver's skill capacity.

Chapter 1

Information Selling under Prior Disagreement

with Ernesto Rivera Mora

This chapter studies monopolistic information selling in environments in which (1) the seller has limited commitment power, and (2) the buyer and the seller hold different beliefs about the state of the world. In environments with a common prior, there is no advantage to selling information sequentially; the seller cannot achieve higher revenue than by offering an experiment that fully reveals the state in one period. We find that if, on the other hand, the agents *agree to disagree* about their prior beliefs, the seller achieves a strictly higher revenue by gradually selling information over multiple periods. Moreover, increasing the number of periods of the protocol strictly increases the seller's expected revenue. In multiple environments, it is optimal for the seller to first offer a *free sample test*, i.e., an experiment that partially reveals information, at no charge. While the seller benefits from additional periods to trade, the expected revenue remains bounded even as the number of periods tends to infinity.

1.1 Introduction

Information buyers and sellers often disagree on the value of information. Individuals frequently undervalue recommendations of lawyers, physicians, dietitians, technicians, and financial advisors. As a result, experts often employ strategies that encourage individuals to reassess the value of their advice. A common tactic is the provision of "complimentary consultations"—free initial sessions in which the expert reveals limited information at no charge. These free sessions are designed to persuade individuals about the true value of the expert's advice and ultimately increase their willingness to accept higher prices for further information.

This paper sheds light on these selling strategies. To this end, we study monopolistic information markets in which the seller and the buyer hold different prior beliefs about the state of the world. Our approach departs from traditional models attributing belief differences to asymmetric information (Bergemann, Bonatti, and Smolin, 2018; Hörner and Skrzypacz, 2016). Instead, we study environments in which the common prior assumption is dropped, and agents *agree to disagree* about their beliefs regarding the state of world. For instance, prior disagreement may stem from *overconfidence* (Grubb, 2009), differences in *opinions* (Che and Kartik, 2009), or simply from different *views of the world* (Alonso and Câmara, 2016).

We introduce a general monopolistic framework in which (1) the seller has limited commitment, and (2) the seller and the buyer agree to disagree about their prior beliefs. The buyer faces a decision problem and hence is willing to pay for experiments (in the sense of Blackwell (1951, 1953)) that reveal information about the state. The seller can implement any experiment at no cost and interacts with the buyer sequentially. Before the buyer selects an action in the decision problem, the seller sequentially offers experiments to the buyer. At each period, the seller commits to honor each experiment that is purchased, but cannot further commit to transfers or offers in future periods. This paper shows that the interplay between prior disagreement and limited commitment induces the seller to sell her information gradually over time.

To illustrate our main insights, we introduce a two-period example. Consider a manager (the information buyer) who faces a decision problem involving two potential states of the world: either the firm's data is safe, or the data is vulnerable to cyberattacks. The manager is overconfident in the security of the firm, assigning a high probability to the safe state.[1] A technician (the information seller) has more cautious beliefs, assigning equal probability to each state. The heterogeneity of beliefs is transparent to the agents, leading them to openly disagree about the value of information. The buyer, confident in the data's security, sees little value in purchasing information. The seller, however, recognizes how this information could potentially mitigate costly errors for the buyer.

---

[1] By *overconfident* we mean prior disagreement in the sense of Grubb (2009).

In a one-period interaction, the seller struggles to convey the value of her information, resulting in low revenue from fully disclosing the state. However, by extending the interaction to two periods, the seller can significantly increase her expected revenue. A particularly effective strategy involves offering a free sample test that partially reveals the state in the first period, followed by a fully revealing test at a high price in the second period. Section 2.2 shows that the seller strictly benefits from using the free sample as it, on average, reduces the buyer's confidence in the data's security, thereby enhancing the buyer's subjective value for additional information. Moreover, offering an initial free sample test plus a subsequent fully-revealing test is the unique optimal two-period selling strategy. Furthermore, our results show that the seller can further increase their subjective revenue by extending the interaction to three or more periods.

This paper characterizes the agents' equilibrium payoffs for (1) any decision problem that the buyer faces, (2) any prior beliefs of the agents, and (3) any number of periods for which the agents are allowed to trade. We show that the equilibrium payoffs are unique and that belief disagreement impacts the qualitative behavior of the seller. If the agents share a common prior, there is no advantage in selling information sequentially. The seller maximizes her expected revenue by fully revealing the state in one period. Under prior disagreement, however, selling all information in one period is strictly suboptimal. Intuitively, selling information gradually allows the seller to tailor experiments that *drive* the buyer's posterior belief towards paths in which the seller expects higher future payments. This creates a non-trivial trade-off: revealing too much information diminishes the available information for future sales, and withholding too much information decreases the posterior drift. Consequently, at any point before the last period, the optimal revenue is not achieved at extremes—neither through full disclosure nor complete withholding—but rather through an experiment that partially reveals the state of the world. Our first main result shows that increasing the length of the interaction strictly increases the seller expected revenue.

Our second main result characterizes the seller's marginal value of time. While the seller strictly benefits from having extra periods of trade, the marginal value of an extra period sharply decreases. Moreover, the seller's expected revenue is bounded regardless of the number of periods of the interaction, thereby showing that the seller is not able to infinitely "deceive" the buyer. Intuitively, while a sequential interaction allows the seller to steer the buyer's beliefs toward high-revenue paths, the seller ends up selling most of the 'stock' of information at early periods, making long-run periods almost irrelevant.

Related Literature The common prior assumption has long played a prominent role in economic theory. (See Harsanyi (1968); Aumann (1976); Halpern (2002).) Nevertheless, models without a common prior are fully consistent with rationality, especially if beliefs are interpreted through a "personalistic" or "subjectivist" Bayesian lens. (See Savage (1972); Morris (1995).) Our paper shows that dropping the common prior assumption significantly impacts the behavior of information monopolists.

This paper contributes to the literature on principals with limited commitment. Most of the literature primarily focuses on common-prior environments in which some agents have private information. (See Acharya and Ortner (2017); Bester and Strausz (2001); Krishna and Morgan (2008); Doval and Skreta (2022).) By contrast, we explore the implications of limited commitment with belief heterogeneity, absent any private information.

Our analysis combines tools from the literatures on dynamic programming (Stokey, 1989; Miao, 2020) and information design (Kamenica and Gentzkow, 2011b; Rayo and Segal, 2010), drawing particularly from the literature on dynamic information design (Ely, Frankel, and Kamenica, 2015; Ely, 2017; Renault, Solan, and Vieille, 2017; Ely and Szydlowski, 2020; Bizzotto, Rüdiger, and Vigier, 2021; Escudé and Sinander, 2023). We contribute to this literature by developing a dynamic programming approach for settings with prior disagreement. Our approach reduces the seller's sequential problem into a series of static Bayesian persuasion problems with prior disagreement (Alonso and Câmara, 2016). Our paper closely relates to Che et al. (2023), which explores a dynamic information design setting under limited commitment.

This paper fits into a broad literature that analyzes information markets. Initiated by the seminal work of Arrow (1973); Admati and Pfleiderer (1986) the literature has been recently extended by Hörner and Skrzypacz (2016); Bergemann, Bonatti, and Smolin (2018); Bergemann and Bonatti (2019); Ichihashi (2021); Ali, Haghpanah, Lin, and Siegel (2022); Zhong (2022); Bergemann, Bonatti, and Gan (2022). Among these, the paper closest to this one is Hörner and Skrzypacz (2016). As in their paper, the seller's optimal selling scheme is a sequential procedure that gradually sells imperfect signals. However, there are important differences. Their results apply to a model where (1) the buyer's decision problem has two actions and two states; (2) the seller has private information; (3) the agents share a common prior; and (4) the seller has preferences about the action that the buyer takes. In contrast, our paper studies settings in which: (1) the buyer faces an arbitrary decision problem; (2) there is no private information; (3) there may be no common prior; and (4) the action taken by the buyer has no impact on the seller's payoffs.

Lastly, our paper is related to the literature that studies monopolists offering free-samples of information and data. Drakopoulos and Makhdoumi (2023) analyze a continuous-time environment in which (1) the state is normally distributed, (2) the seller offers signals that follow an exogenous normal distribution, and (3) the buyer communicates his private information to the seller. They restrict to a class of selling strategies that offer (potentially free) signals at a constant rate and sells the entire data set at the end of the interaction. Our paper differs in that (1) beliefs are arbitrary, (2) does not restrict the structure of the signals, and (3) there is no private information. The closest paper is Zheng and Chen (2021) which studies optimal "advertising" of information. They analyze two-period schemes in which agents have prior disagreement and the first-experiment is required to be free. Our paper differs in that we allow for general T-period settings and the seller is allowed to charge positive prices at any period. We show that in multiple environments—though not all—free-sampling is the unique equilibrium strategy of the seller. So, rather than exogenously imposing the first experiment to be free, free-sampling of information endogenously emerges.

Organization of the paper The remainder of the paper is organized as follows. Section 2.2 introduces our leading example. Section 1.3 lays out the model. In Section 1.4, we solve the game as a dynamic programming problem and characterize the equilibrium payoffs. In Section 1.5, we show that the seller strictly benefits from interacting over more periods, but that the revenue is asymptotically bounded. Finally, Section 1.6 discusses some of our modeling assumptions. All proofs are collected in the appendix.

1.2 Example

An information buyer (a manager of a firm) faces a decision problem under uncertainty. There are two possible states: $\underline{\theta}$ (the firm's data is vulnerable) and $\overline{\theta}$ (the firm's data is safe). Denote the state space as $\Theta = \{\underline{\theta}, \overline{\theta}\}$. The set of actions is $A = \{\mathcal{U}, \mathcal{N}\}$, where $\mathcal{U}$ denotes updating the firm's firewall and $\mathcal{N}$ denotes not updating the firewall. The buyer's payoff function $u : A \times \Theta \to \mathbb{R}$ is summarized in the table below.

$u(\theta, a)$	$\mathcal{U}$	$\mathcal{N}$
$\overline{\theta}$	0	1
$\underline{\theta}$	0	-1

Table 1.1. Buyer's utility function

So, updating the system is the right action when the system is vulnerable and non updating it is the right action when the state is not vulnerable.

The buyer is uncertain about the true state of the world. Write $\nu_b \in (0,1)$ for the buyer's prior belief of state $\overline{\theta}$. Observe, absent any information, the buyer prefers $\mathcal{N}$ if $\nu_b \geqslant \frac{1}{2}$ and $\mathcal{U}$ if $\nu_b \leqslant \frac{1}{2}$. The blue line (resp. red line) in Figure 1.1 describes the buyer's expected utility from choosing action $\mathcal{U}$ (resp. choosing $\mathcal{N}$).

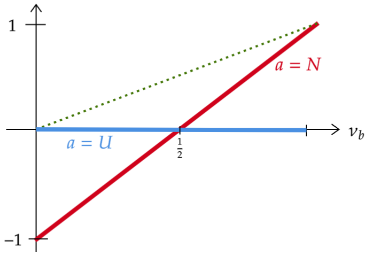


Figure 1.1. Expected utility of the buyer

For each prior belief $\nu_b \in (0,1)$, we can compute the buyer's value of observing the state using Figure 1.1. The value of observing the state is the difference between the dotted line (the value of fully observing the state) and the maximum of the blue and red lines (the expected utility absent any information). So, the buyer's value for fully observing the state is:

$$V(\nu_b) = \begin{cases} \nu_b & \text{if } \nu_b \leqslant \frac{1}{2} \\ 1-\nu_b & \text{otherwise.} \end{cases}$$

Figure 1.2 illustrates the function $V(\cdot)$ in terms of the prior belief ν_b. Notice, the buyer places a higher value for priors that are closer to $\frac{1}{2}$, where the buyer has the highest uncertainty about the state of the world.

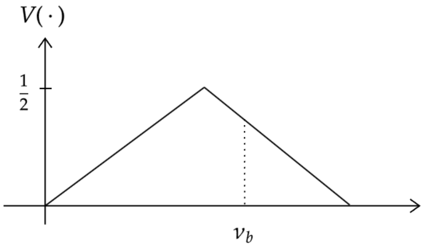


Figure 1.2. Buyer's value of observing the state for prior belief ν_b

A monopolistic information seller has access to "experiments" or "tests" that can fully or partially reveal the state. The seller has a prior belief $\nu_s \in (0,1)$ that the state is $\overline{\theta}$. Importantly, we allow ν_s to be different from ν_b. The agents' prior beliefs are transparent to them and agents have no private information.

Fix a set of signals M with $|M| \geqslant 2$. An experiment is a stochastic mapping $\pi : \Theta \to \Delta M$ that describes the probability of each signal conditional on each state. We impose three assumptions: First, the agents agree about the probabilities described by each experiment π. Second, the realized signal of each experiment is publicly reveal to both agents. Third, the agents are Bayesian, i.e., their posterior beliefs are derived by Bayes' rule according to their subjective prior beliefs.

The agents interact in two periods. In each period $t \in \{1,2\}$, the seller offers an experiment π^t at a fixed price $p^t \geqslant 0$. If the buyer purchases the experiment, she pays the price p^t to the seller, and both agents observe the realized signal $m^t \in M$. The seller has limited commitment. At each period t, the seller commits to honor the priced experiment (π^t, p^t) offered to the buyer. That is, the seller commits to charge p^t and truthfully reveal the realized signal m^t of the experiment π^t. The seller cannot further commit to transfers that are contingent upon the state or the signal realization, or to provide an experiment at a future period.

1.2.1 Common Prior

We start our analysis with a canonical benchmark in which the buyer and the seller agree about their prior beliefs regarding the state.

Proposition 1 *If $\nu_b = \nu_s$, then the seller's maximum expected revenue is $V(\nu_b)$. In particular, it is optimal to offer an experiment that completely reveals the state in the first period.*

A fully revealing experiment in the first period gives the seller a maximum payoff of $V(\nu_b)$. Moreover, under a common prior, the seller cannot exceed this revenue by selling information sequentially. To see this, suppose that there is a selling scheme in which the seller's expected revenue is higher than $V(\nu_b)$. Since the agents share the same prior, the agents agree about the probability of any outcome of the experiment. Consequently, if the seller expects to receive a higher revenue than $V(\nu_b)$, then the buyer expects to pay more than $V(\nu_b)$, i.e., her willingness to pay for full disclosure. Thus, the buyer avoids participating in any scheme that attempts to charge more than $V(\nu_b)$.

1.2.2 Prior Disagreement

We now consider a setting in which the agents have prior disagreement. For instance, disagreement may stem from overconfidence as described by Grubb (2009). Consider the case in which the seller's and the buyer's prior belief about $\overline{\theta}$ are $\nu_s = 0.5$ and $\nu_b = 0.9$, respectively. Observe, in this case, the agents not only disagree about their beliefs, but also disagree about the value of information. From the seller's point of view, the value of information is $V(0.5) = 0.5$. However, the buyer is overconfident and is willing to pay only $V(0.9) = 0.1$ for full revelation of the state. As a result, the maximum price the seller can charge in one period is $p = 0.1$, even though the seller believes that the information has a higher value.

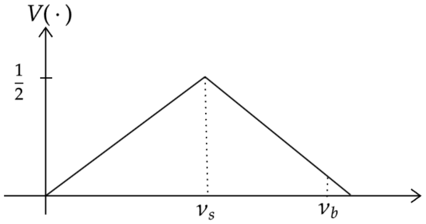


Figure 1.3. Value of information under prior disagreement

We show that the seller can strictly increase her revenue by sequentially selling information in two periods. Moreover, we show that boosting revenue may be achieved by a selling scheme that (1) initially offers some information for free, and (2) subsequently offers to fully reveal the state at a high price.

The first *free-sample* experiment is tailored to maximize the probability that the buyer has the posterior probability $\mu_b = 0.5$ (the posterior belief leading to the highest valuation of information). Fix a set of signals $M = \{\underline{m}, \overline{m}\}$ and consider the signal mapping $\pi : \Theta \to \Delta(M)$ given by the table below:

$\pi(m \mid \theta)$	$\underline{m}$	$\overline{m}$
$\overline{\theta}$	$\frac{1}{9}$	$\frac{8}{9}$
$\underline{\theta}$	1	0

Using Bayes' rule, the posterior probability that the buyer assigns to state $\overline{\theta}$ after each signal is

$$\mu_b(\overline{m}) = \frac{0.9 \cdot \frac{8}{9}}{0.9 \cdot \frac{8}{9} + 0.1 \cdot 0} = 1 \qquad \text{and} \qquad \mu_b(\underline{m}) = \frac{0.9 \cdot \frac{1}{9}}{0.9 \cdot \frac{1}{9} + 0.1 \cdot 1} = \frac{1}{2}.$$

An illustration of the "posterior spread" of the free sample is given in Figure 1.4. Observe, after receiving signal $\overline{m}$, the buyer becomes certain that the state is $\overline{\theta}$, eliminating the need to purchase additional information. However, after receiving signal $\underline{m}$, the buyer's uncertainty increases. Moreover, after signal $\underline{m}$, the buyer is willing to buy a second experiment that fully reveals the state for a price $p = \frac{1}{2}$.

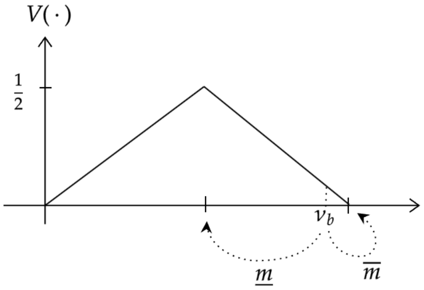


Figure 1.4. Buyer's new beliefs after the free sample experiment

Notice, agents not only disagree about their beliefs about the state, but also about the probability of the signals of the experiment. While the buyer believes that the probability of signal $\underline{m}$ is $\frac{1}{10} \cdot 1 + \frac{9}{10} \cdot \frac{1}{9} = \frac{1}{5}$, the seller believes that such probability is $0.5 \cdot 1 + 0.5 \cdot \frac{1}{9} = \frac{5}{9}$. After such a signal, the buyer purchases a fully revealing experiment at price $\frac{1}{2}$. Hence, the seller's expected revenue under the dynamic selling scheme is $\frac{5}{9} \cdot \frac{1}{2} = \frac{5}{18} \approx 0.27$, which is higher than the revenue from selling the information in a single period, $V(0.9) = 0.1$.

The subjective increase in revenue can be interpreted as an exploitation of the buyer's "incorrect" prior belief.[2] When the information buyer is overconfident, he underestimates the probability of state $\underline{\theta}$ and receiving a low signal $\underline{m}$. Consequently, the buyer underestimates the probability of purchasing the fully-revealing experiment. This, in turn, allows to raise the price of full disclosure in the second period, raising the expected revenue to $\frac{5}{18}$. The results of Section 1.5 will imply that (1) $\frac{5}{18}$ is the maximum revenue that can be achieved in two periods, (2) any two-period scheme that charges for information in the first period is strictly suboptimal, and (3) revenue strictly increases by selling information in three or more periods.

1.3 Model

Throughout the paper, take the following conventions. Endow a compact metric space C with its Borel sigma-algebra. Denote by ΔC the set of probability measures on C and endow ΔC with the topology of weak convergence. Denote the interior of ΔC by $\operatorname{int} \Delta C$.

---

[2] The buyer's beliefs are "incorrect" from the seller's subjective point of view. In the same way as Grubb (2009) the seller's payoffs are computed using the seller's prior belief.

For each $c \in C$ write $\delta_c \in \Delta C$ for the probability measure that assigns probability one to the singleton $\{c\}$.

1.3.1 Environment

There are two agents, the buyer (denoted by b) and the seller (denoted by s). The buyer faces an individual decision problem described by a finite state space Θ, a compact set of actions A, and a continuous utility function $u : A \times \Theta \to \mathbb{R}$. We assume that there are some states $\theta, \theta' \in \Theta$ such that

$$\arg\max_{a \in A} u(a, \theta) \cap \arg\max_{a \in A} u(a, \theta') = \varnothing.$$

In this sense, the problem is not trivial, and the buyer strictly benefits from observing the state.

Each agent $i \in \{s, b\}$ has a prior belief $\nu_i \in \operatorname{int} \Delta\Theta$ about the state. The agents' prior beliefs are *transparent* to them.[3] In particular, the agents have no private information. If $\nu_b = \nu_s$ we say that the agents share a *common prior*. If $\nu_b \neq \nu_s$, we say that the agents *agree to disagree* about their priors.

1.3.2 Interaction

We model the interaction between the agents as a dynamic game in which the seller sequentially offers information to the buyer over multiple periods. Information is sequentially revealed by implementing the Blackwell experiments that the buyer chooses to purchase.

The timing is as follows. Nature first chooses a state of the world $\theta \in \Theta$. Agents do not observe the realization of the state. After the state is realized, there is a finite sequence of periods that are indexed (backward) by $t = T, T-1, \ldots, 1, 0$. The number of periods T is exogenous. In each period $t > 0$ the agents trade information and in the last period $t = 0$, the buyer takes an action. So, t reflects the number of periods left before the buyer faces her decision problem.

---

[3] Formally, the agents' beliefs are transparent if they are described by a type structure $(\Theta, (\mathcal{T}_i, \beta_i)_{i \in \{s,b\}})$ in which (1) each type set $\mathcal{T}_i$ is a singleton, and (2) each belief mapping $\beta_i : \mathcal{T}_i \to \Delta(\Theta \times \mathcal{T}_{-i})$ satisfies $\operatorname{marg}_{\Theta} \beta_i(t_i) = \nu_i$.

There is a finite set of signals M that satisfies $|M| \geqslant |\Theta|$.[4] At each period $t > 0$, the seller makes a take-it-or-leave-it offer to the buyer, consisting of a priced experiment $E^t = (\pi^t, p^t)$, where $\pi^t : \Theta \to \Delta M$ is an experiment, and p^t is the experiment's price. The experiment π^t describes how signals are distributed conditional on the state. Importantly, the agents agree about the conditional distribution of signals π^t. The buyer observes E^t and decides whether to accept or reject it. If E^t is accepted, the buyer pays the price p^t to the seller, the experiment π^t is implemented, both agents observe the realized signal $m^t \in M$, and period $t-1$ starts. If E^t is rejected, no transfer is made, the experiment is not implemented, and period $t-1$ starts. In the last period, $t = 0$, the buyer decides what action $a \in A$ to take. Once the game is over, the state θ is revealed and the buyer receives utility $u(a, \theta)$.

To isolate the strategic effects of selling information, we assume that (1) there is no time-discounting, (2) the seller can implement any sequence of experiments without incurring any cost, and (3) the seller's payoff does not depend either on the state or on the action the receiver takes.[5] The buyer has quasilinear preferences regarding the decision problem and the payments made to the seller. So, at the end of the interaction, the buyer's total utility is $u(a,\theta) - \sum_{t=1}^{T} p^t \mathbb{1}\{E^t \text{ is accepted}\}$. The seller's objective is to maximize the revenue $\sum_{t=1}^{T} p^t \mathbb{1}\{E^t \text{ is accepted}\}$.

The seller has limited commitment power. Within each period t, the seller commits to honor the priced experiment $E^t = (\pi^t, p^t)$, provided that the buyer accepts it. That is, the seller commits to charge p^t and truthfully reveal the realized signal of the experiment π^t. The seller cannot further commit to transfers or future experiments that are contingent upon the state θ or the signal realization m^t.

1.3.3 Equilibrium

Write $\mathcal{E}$ for the set of all priced experiments. A history for the seller at period $T > t \geqslant 1$ is a sequence $h_s^t = \{(E^\tau, c^\tau, m^\tau(c^\tau))\}_{\tau=T}^{t+1}$, where $E^\tau \in \mathcal{E}$ is the experiment offered at time τ, and $c^\tau \in \{\text{accept, reject}\}$ is the buyer's choice. The entry $m^\tau(c^\tau) \in \operatorname{Supp}(\pi^\tau)$ is the signal

---

[4] Assuming that M is finite simplifies the analysis by avoiding updating in zero probability events. Assuming that M is finite does not affect the results. (See Part (i) of Lemma 3.)

[5] Section 1.6.2 discusses the environment in which agents discount the future. Section 1.6.3 discusses the environment in which the seller faces non-zero costs for implementing the experiments.

generated by π^τ provided that c^τ =accept, and $m^\tau(c^\tau) = \varnothing$ otherwise. A history for the buyer at period $T \leqslant t \leqslant 1$ is a sequence $h_b^t = \{h_s^t, E^t\}$, describing the seller's history h_s^t up to that period and the priced experiment E^t that the seller offers at period t. Write H_i^t for i's set of histories at period $t \geqslant 1$ and set $H_s^T = \{\varnothing\}$. Write $\mathcal{H}_i := \bigcup_{t=T}^{1} H_i^t$ for i's set of histories prior to the last period $t = 0$. A **trading history** is a sequence $h^0 = \{(E^t, c^t, m^t(c^t))\}_{t=T}^{1}$. Denote by $\mathcal{H}^0$ the set of all trading histories.

A **behavior strategy for the seller** is a mapping $\sigma : \mathcal{H}_s \to \Delta\mathcal{E}$ that associates a priced experiment with each history in $\mathcal{H}_s$. A **behavior strategy for the buyer** is a pair (c, α) where $c : \mathcal{H}_b \to \Delta\{\text{accept, reject}\}$ is an acceptance rule and $\alpha : \mathcal{H}^0 \to \Delta A$ is a final action rule. So, for each buyer's history $h_b^t = (h_s^t, E^t) \in \mathcal{H}^b$, $c(h_b^t)$ prescribes whether to accept or reject the priced experiment E^t given the seller's history h_s^t; and, for each $h^0 \in \mathcal{H}^0$, $\alpha(h^0)$ prescribes the action to take after acquiring the information induced by the history h^0.

In this game the agents have no private information and all signals are public. Hence, we employ subgame perfect equilibrium (SPE) as solution concept. Discussion 1.6.1 discusses how this solution concept is equivalent to strong perfect Bayesian equilibrium (SPBE).

1.4 A Dynamic Programming Approach

To analyze equilibrium behavior, we employ a dynamic programming approach that characterizes (1) how the agents' beliefs evolve after any sequence of experiments and (2) how equilibrium payoffs and behavior depend on such beliefs at any history.

1.4.1 Posterior Dynamics

Before describing the strategic behavior of the agents, we first describe the dynamics of the agents' posteriors after observing any sequence of signal realizations. For each period $t \in \{T, \ldots, 1\}$, write $\mu_i^t \in \Delta\Theta$ for i's beliefs at the beginning of period t. Hence, each experiment π^t and each signal $m^t \in \operatorname{Supp} \pi^t$ induce an agent i's posterior belief μ_i^{t-1}, following Bayesian updating. For each belief $\mu \in \Delta\Theta$, write $\mathrm{PS}[\mu] := \{\tau \in \Delta(\Delta\Theta) : \mathbb{E}_\tau[\mu'] = \mu\}$ for the set of *posterior spreads* of μ that are Bayes plausible. So, assuming that agent i has belief μ_i^t at time t, each experiment $E^t = (\pi^t, p^t)$ induces a posterior spread $\tau_i \in \mathrm{PS}[\mu_i^t]$ for agent i. Moreover, each posterior spread $\tau_i \in \mathrm{PS}[\mu_i^t]$ such that $|\mathrm{Supp}\,(\tau_i)| \leqslant M$ is induced by some experiment $\pi^t : \Theta \to \Delta(M)$. (See Kamenica and Gentzkow (2011b).)

For each $\theta \in \Theta$, write $r(\theta) := \frac{\nu_s(\theta)}{\nu_b(\theta)}$ for the agents' likelihood ratio of state θ. Notice, given the agents' prior beliefs are in the interior of the simplex, the vector of likelihood ratios $r := (r(\theta))_{\theta \in \Theta}$ is well defined. Let $g : \Delta\Theta \to \Delta\Theta$ be given by

$$g(\mu)(\theta) := \frac{r(\theta)\mu(\theta)}{r \cdot \mu}.$$

Notice, the mapping g is the identity if and only if the agents share a common prior. This mapping describes the relation of the agents' beliefs along the entire interaction. To see this, consider an experiment $E^T = (\pi^T, p^T)$ in the first period T. The mapping g links the seller's posterior with the buyer's posterior. That is, after any signal realization $m^T \in \mathrm{Supp}\,\pi^T$, the agents' posterior beliefs satisfy $\mu_s^{T-1} = g(\mu_b^{T-1})$. (See Proposition 1 in Alonso and Câmara (2016).) Furthermore, since any sequence of experiments is itself an experiment, the relation $\mu_s^t = g(\mu_b^t)$ holds at each subsequent period $t \geqslant 0$ after any sequence of experiments and realized signals.

The mapping g allows to characterize the behavior and the outcomes of both agents in terms of the posterior of just one agent. We pick the buyer's belief as a state variable and study how this belief influences the agents' behavior and payoffs.[6]

1.4.2 Posterior Drifts

Describing the agents' equilibrium behavior requires understanding not only how beliefs differ after each signal, but also how the agents disagree about the likelihood of the realization of such posterior beliefs. Assume that, at some period $t > 0$, the agents have beliefs (μ_b^t, μ_s^t) with $\mu_s^t = g(\mu_b^t)$. Observe, after the realization of experiment π^t, the agents may not only disagree about the posterior beliefs $(\mu_s^{t-1} \neq \mu_b^{t-1})$ at period $t-1$ but also about the likelihood of the realization of each pair $(\mu_b^{t-1}, \mu_s^{t-1})$. That is, the seller views certain pairs of posterior belief as more probable than the buyer does.

---

[6] The analysis is analogous if the analyst chooses the seller's belief as the state variable.

To describe this disagreement, fix a buyer's belief μ_b^t and define $\rho(\cdot \mid \mu_b^t) : \Delta\Theta \to \mathbb{R}$ by

$$\rho(\mu_b^{t-1} \mid \mu_b^t) := \frac{r \cdot \mu_b^{t-1}}{r \cdot \mu_b^t},$$

where r is the vector of likelihood ratios given by the priors (ν_s, ν_b). The following lemma shows that the mapping $\rho(\cdot \mid \mu_b^t)$ captures the agents' disagreement about their posteriors at period $t-1$, provided that the buyer's belief at period t is μ_b^t.

Lemma 1 *Assume that at period t the agents have prior beliefs (μ_b^t, μ_s^t) with $\mu_s^t = g(\mu_b^t)$. If π^t is an experiment such that for each agent $i \in \{s, b\}$ induces a posterior spread $\tau_i \in \mathrm{PS}[\mu_i^t]$, then for each posterior $\mu_b^{t-1} \in \Delta\Theta$,*

$$\tau_s(g(\mu_b^{t-1})) = \tau_b(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t).$$

Lemma 1 describes the agents' disagreement about the likelihood of their posterior beliefs. To provide some intuition, fix a belief μ_b^t and write $H_{\rho>1}(\mu_b^t) := \{\mu_b^{t-1} \in \Delta\Theta \mid \rho(\mu_b^{t-1} \mid \mu_b^t) > 1\}$ for the posterior beliefs that the seller deems more likely as compared to the buyer. Fix an experiment π^t, and let $\tau_b \in \mathrm{PS}[\mu_b^t]$ be the buyer's posterior spread it induces. Notice, the function $\rho(\mu_b^{t-1} \mid \mu_b^t)$ is linear in μ_b^{t-1}. Moreover, the set $H_{\rho>1}(\mu_b^t)$ is an open half-space containing μ_s^t. Figure 1.5 illustrates this open half-space as a shaded area of the simplex. Intuitively, if an experiment induces a posterior spread $\tau_b \in \mathrm{PS}[\mu_b^t]$ such that $\mathrm{Supp}\,(\tau_b) \cap H_{\rho>1}(\mu_b^t) \neq \varnothing$, then the seller expects the average buyer's posterior to "drift" towards the region $H_{\rho>1}(\mu_b^t)$. This has an important implication: while the buyer's posterior belief satisfies Bayesian plausibility from the buyer's perspective, it not satisfied from the seller's perspective. The next sections will show how the posterior drift affects equilibrium payoffs and behavior.

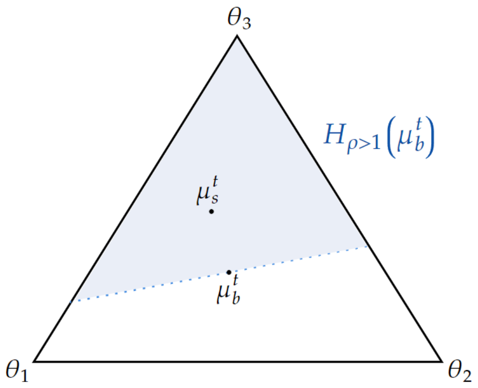


Figure 1.5. Illustration of the set $H_{\rho>1}(\mu_b^t)$.

1.4.3 Dynamic Programming: Last Period

We proceed by backward induction to characterize payoffs and behavior in the sequential game. We start by finding the experiment that the seller offers at period $t = 1$ after any history.

The buyer's value of each posterior belief is key in characterizing the equilibrium payoffs. With this in mind, write $\mu_b^0 \in \Delta\Theta$ for the buyer's posterior belief at period $t = 0$ and write

$$U(\mu_b^0) = \max_{a \in A} \mathbb{E}_{\mu_b^0}[u(a, \theta)]$$

for the buyer's expected payoff when he has posterior belief μ_b^0.

Lemma 2 *The function $U(\cdot)$ is continuous and convex.*

Following Blackwell (1951, 1953), Lemma 2 implies that the buyer is weakly better off under any posterior spread and, hence, is weakly better off by obtaining more information. Assume that the buyer has belief μ_b^1 at period $t = 1$. Notice, the buyer's willingness to pay for a posterior spread $\tau_b \in \mathrm{PS}[\mu_b^1]$ is non-negative and is given by $\mathbb{E}_{\tau_b}[U(\mu_b^0)] - U(\mu_b^1)$. Therefore,

$$V^1(\mu_b^1) := \sup_{\tau_b \in \mathrm{PS}[\mu_b^1]} \mathbb{E}_{\tau_b}[U(\mu_b^0)] - U(\mu_b^1)$$

is the seller's maximum revenue that can be achieved in a single period, provided that the buyer has belief $\mu_b^1 \in \Delta\Theta$. Moreover, since U is convex, the supremum defining V^1 is achieved by the posterior spread of a fully revealing experiment. Thus,

$$V^1(\mu_b^1) = \bar{U} \cdot \mu_b^1 - U(\mu_b^1),$$

where $\bar{U} = (U(\delta_\theta))_{\theta\in\Theta}$ denotes the vector of maximum utilities at each state. Note, Lemma 2 implies that V^1 is a continuous, non-negative, and concave mapping such that $V^1(\delta_\theta) = 0$ for each $\theta \in \Theta$. Moreover, because the buyer's decision problem is not trivial, $V^1(\mu_b^t) > 0$ for each $\mu_b^t \in \operatorname{int} \Delta\Theta$.

Notice, the agents' behavior in the last period is strategically equivalent to a setting with only one period in which the buyer has prior belief $\nu_b = \mu_b^1$. The following result characterizes behavior and payoffs for one-period protocols.

Theorem 1 *Assume there is only one period. There exists an SPE that satisfies the following:*

(i) The seller offers a fully revealing experiment,

(ii) the buyer accepts the offer with probability one,

(iii) the seller has expected payoff $V^1(\nu_b)$*, and*

(iv) the buyer has expected payoff $U(\nu_b)$*.*

Moreover, each SPE satisfies properties (ii)-(iv).[7]

Theorem 1 characterizes the unique equilibrium payoffs of the one-period game in terms of the model's primitives (A, u) and ν_b.[8] Observe, the seller's prior belief has no effect in the agents' payoffs. The proof follows from a standard argument used in ultimatum-style games. In each SPE, the seller offers the experiment that maximizes total surplus at the maximum price that the buyer is willing to accept. Observe, in each strategy profile where the buyer rejects with positive probability, the seller has a profitable deviation: offering a fully revealing experiment at a slightly lower price, which the buyer accepts with certainty.

---

[7] Notice, if there are some states $\theta, \theta' \in \Theta$ such that $\arg\max_{a\in A} u(a,\theta) = \arg\max_{a\in A} u(a,\theta')$, then there exists an optimal experiment that pools states θ and θ' into a single signal $m \in M$. Observe that, conditional on signal m, the buyer has no value from observing the true state. Therefore, condition (i) does not hold for some SPE.

[8] Indeed, the functions U and V^1 are defined by the decision problem (A, u).

1.4.4 Dynamic Programming: Previous Periods

This section characterizes equilibrium payoffs in an arbitrary number of periods by using a backward induction argument. Fix a period $t > 1$, a belief $\mu_t \in \Delta\Theta$, and a posterior spread $\tau_b \in \mathrm{PS}[\mu_t]$. Write $\tilde{\tau}_b \in \Delta(\Delta\Theta)$ for the distribution of beliefs such that $\tilde{\tau}_b(\mu_b^{t-1}) := \tau_b(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t)$. By Lemma 1, $\tilde{\tau}_b$ is a well defined probability measure that describes the seller's ex-ante beliefs about the buyer's posterior μ_b^{t-1} after observing some experiment π^t. So, $\tilde{\tau}_b$ is the buyer's posterior spread induced by τ_b under the seller's perspective.

We inductively define a sequence of real mappings $(V^t)_{t\in\mathbb{N}}$ defined on $\Delta\Theta$. Assume that $V^{t-1} : \Delta\Theta \to \mathbb{R}$ is well-defined and write

$$V^t(\mu_b^t) := \sup_{\tau_b \in \mathrm{PS}[\mu_b^t]} \left(\mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t) + \mathbb{E}_{\tilde{\tau}_b}[V^{t-1}(\mu_b^{t-1})]\right),$$

where $\tilde{\tau}_b$ is the posterior spread induced by τ_b under the seller's perspective.

We will show that the mapping V^t captures the seller's maximum total revenue that she can extract when there are $t > 1$ periods remaining and the buyer has belief $\mu_b^t \in \Delta\Theta$. (Recall that the seller has belief $\mu_s^t = g(\mu_b^t)$.) To see this, assume that the seller offers an experiment that induces a posterior spread $\tau_b \in \mathrm{PS}[\mu_b^t]$. The first component, $\mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t)$, captures the *seller's present revenue:* the buyer's willingness to pay for an experiment inducing a posterior spread τ_b. (This relies on anticipating that the buyer's continuation value for the next period $t-1$ is $U(\mu_b^{t-1})$.) The second component, $\mathbb{E}_{\tilde{\tau}_b}[V^{t-1}(\mu_b^{t-1})]$, captures the *seller's future revenue*: the sum of expected transfers from optimally selling information in the remaining $t-1$ periods. Importantly, while the expectation of the first component is based on the buyer's posterior spread τ_b, the expectation of the second component is based on $\tilde{\tau}_b$—which captures the buyer's posterior spread from the seller's perspective. The difference between $\tilde{\tau}_b$ and τ_b is the key driver of our results.

Notice that in principle, the supremum defining V^t may not be attained at some period t. We show that this is not the case. Moreover, we show that the optimization problem associated to $V^t(\mu_b^t)$ can be solved using the concave envelope (Aumann, Maschler, and Stearns, 1995; Kamenica and Gentzkow, 2011b). To do so, fix $\mu_b^t \in \Delta\Theta$ and define the auxiliary mapping $\Lambda^t(\cdot \mid \mu_b^t) : \Delta\Theta \to \mathbb{R}$ as

$$\Lambda^t(\mu_b^{t-1} \mid \mu_b^t) := U(\mu_b^{t-1}) + V^{t-1}(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t).$$

The following Lemma describes the value $V^t(\mu_b^t)$ as a standard concavification problem for the objective $\Lambda^t(\cdot \mid \mu_b^t)$.

Lemma 3 *For each $t > 1$ and each $\mu_b^t \in \Delta\Theta$,*

$$V^t(\mu_b^t) = \sup_{\tau_b \in \mathrm{PS}[\mu_b^t]} \mathbb{E}_{\tau_b}[\Lambda^t(\mu_b^{t-1} \mid \mu_b^t)] - U(\mu_b^t).$$

Moreover,

(i) For each $\mu_b^t \in \Delta\Theta$, the supremum defining $V^t(\mu_b^t)$ is achieved for some posterior spread $\tau_b \in \mathrm{PS}[\mu_b^t]$ that has at most $|\Theta|$ elements in its support.

(ii) The mapping $V^t(\cdot)$ is continuous.

(iii) The mapping $V^t(\cdot)$ satisfies $V^{t+1}(\cdot) \geqslant V^t(\cdot)$ and $V^t(\delta_\theta) = 0$ for each $\theta \in \Theta$.

Lemma 3 states that $V^t(\mu_b^t)$ is the value of a well-defined finite-dimensional Bayesian persuasion problem. Thus, the supremum defining $V^t(\mu_b^t)$ can be found by computing the concave envelope of $\Lambda^t(\cdot \mid \mu_b^t)$ evaluated at the belief $\mu_b^{t-1} = \mu_b^t$. Moreover, part (i) shows that the supremum is achieved by an experiment with at most $|\Theta|$ signals. So, provided that $|M| \geqslant |\Theta|$, there is some experiment $\pi^t : \Theta \to \Delta M$ that induces the optimal posterior spread. Part (ii) shows that V^t is continuous and therefore bounded. Part (iii) shows that the seller weakly benefits from having extra periods and that the future revenue is zero if the buyer becomes certain about the state.

Theorem 2 *Assume there are T periods. There exists an SPE that satisfies the following:*

(i) On the equilibrium path the buyer accepts each offer with probability one,

(ii) the seller has expected payoff $V^T(\nu_b)$, and

(iii) the buyer has expected payoff $U(\nu_b)$.

Moreover, each SPE satisfies these properties.

Theorem 2 characterizes the unique equilibrium payoffs of the T-period game in terms of the agent's prior beliefs. The proof follows an inductive argument reminiscent of those used in ultimatum-style games. Fix a history that induces buyer's belief μ_b^t at period t. By backward induction, in each SPE the agents anticipate that the buyer's continuation value is given by $U(\mu_b^{t-1})$. So, the seller takes this buyer's continuation payoff as given and offers an experiment that maximizes current and future payments described by the expression defining $V^t(\mu_b^t)$. Overall, this results in an expected payoff $V^T(\nu_b)$ for the seller and an expected payoff of $U(\nu_b)$ for the buyer. Notice, as in standard ultimatum-style games, the buyer accepts all offers on the equilibrium path. If, with positive probability, the buyer does not accept the offer, then the seller would have incentives to deviate to a lower price which the buyer accepts with probability one.

Observe that each history $h_s^t \in \mathcal{H}_s$ induces an information-selling game of t periods. Hence, the results of Theorem 2 extend to all such induced subgames, even those outside the equilibrium path. So, each history $h_s^t \in \mathcal{H}_s$ in which the buyer's initial belief is μ_b^t, the buyer accepts the equilibrium seller's offer with probability one, and the seller's and buyer's continuation expected payoffs are $V^t(\mu_b^t)$ and $U(\mu_b^t)$, respectively.

1.4.5 Example Revisited

We now apply the dynamic programming approach to analyze the example of Section 2.2. Recall that in this example the agents' priors are $(\nu_b(\overline{\theta}), \nu_s(\overline{\theta})) = (0.9, 0.5)$.

We first characterize the case in which the agents have two periods to trade information. In the first period $t = 2$, the seller seeks an experiment that, in expectation, maximizes the expected value of the objective

$$\Lambda^2(\cdot \mid \nu_b) := U(\cdot) + V^1(\cdot)\rho(\cdot \mid \nu_b).$$

For each $x \in [0, 1]$, write μ_x for the belief such that $\mu_x(\overline{\theta}) = x$. Observe, under these prior beliefs the distortion mapping satisfies $\rho(\mu_x \mid \nu_b) > 1$ if and only if $0.9 > x$. So, the seller believes that all the buyer's posteriors μ_x that satisfy $0.9 > x$ are more likely relative to the buyer. Hence, the seller can tailor an experiment that drives the average buyer's posterior towards posteriors μ_x that satisfy $0.9 > x$. Figure 1.6 plots the objective $\Lambda^2(\cdot \mid \nu_b)$ (in blue) and its concave envelope (in red dashed lines).

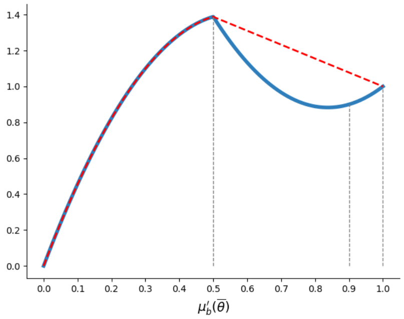


Figure 1.6. Mapping $\Lambda^2(\cdot \mid \nu_b)$ and its concave envelope.

Notice, the posterior spread $\tau \in \mathrm{PS}[\nu_b]$ that maximizes the objective $\mathbb{E}_\tau[\Lambda^2(\mu_b^1 \mid \nu_b)]$ is such that $\mathrm{Supp}\,(\tau) = \{\mu_{0.5}, \mu_1\}$. Moreover, the maximum price that the seller can achieve for the associated experiment is $p^1 = \mathbb{E}_\tau[U(\mu_b^1)] - U(\nu_b) = 0$. (Recall that $U(\cdot)$ is linear from $\mu_{0.5}$ to μ_1.) Therefore, provided there are two periods, the free-sampling selling scheme described in Section 2.2 maximizes the seller's revenue and provides an expected revenue of $V^2(\nu_b) = \frac{5}{18}$.

We can apply the dynamic programming approach to analyze the example in a setting with with three periods. The optimal selling scheme has the following features: In the first period, the seller provides a free sample experiment that induces a posterior spread $\tau_b^3 \in \mathrm{PS}[\nu_b]$ such that $\mathrm{Supp}\,(\tau_b^3) = \{\mu_{0.5}, \mu_1\}$. If the posterior $\mu_{0.5}$ is realized in the second period, then the seller offers an experiment inducing a posterior spread $\tau_b^2 \in \mathrm{PS}[\mu_{0.5}]$ such that $\mathrm{Supp}\,(\tau_b^2) = \{\mu_x, \mu_1\}$ for some $x \approx 0.3$. If the posterior μ_x is realized in the last period, the seller finally offers an experiment that fully reveals the state of the world. Notice, if at some period the posterior μ_1 is realized, then the buyer does not purchase further information.

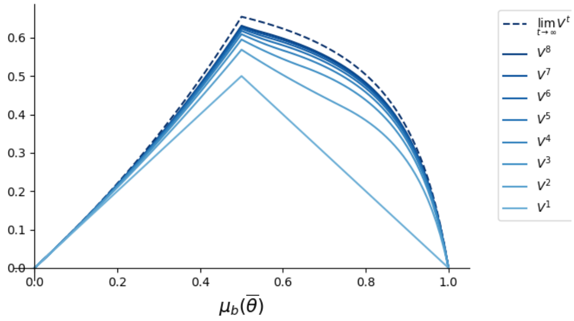


Figure 1.7. Mappings V^t in terms of $\mu_b(\overline{\theta})$, given that priors are $(\nu_b, \nu_s) = (0.9, 0.5)$.

The example can be easily extended to a sequential game with an arbitrary finite number of periods. By applying the concave envelope approach on the associated objective Λ^t, one can iteratively compute the value functions $(V^t)_{t\in\mathbb{N}}$ and the optimal selling scheme for an arbitrary number of periods. Figure 1.7 illustrates the mappings $V^t(\cdot)$ for $t = 1\ldots,9$. The asymptotic limit $\lim_{t\to\infty} V^t(\cdot)$ is plotted in dashed lines. There are two important observations. First, at each period $t > 0$ and each belief $\mu_b \in \operatorname{int}\Delta\Theta$, $V^{t+1}(\mu_b) > V^t(\mu_b)$. So, provided that there is information left to offer, the seller strictly benefits from having more time to interact with the buyer. Second, the value functions $(V^t)_{t\in\mathbb{N}}$ are bounded and converge to a continuous function. Section 1.5 shows that these two properties are universal and emerge in all environments with prior disagreement, regardless of the buyer's decision problem (Θ, A, u) or the agents' prior beliefs (ν_b, ν_s).

1.5 Main Results

Theorem 2 describes the seller's payoffs of each T-period game in terms of the mapping V^T. This section explores the properties of the mappings $(V^t)_{t\in\mathbb{N}}$ as a way to describe the payoffs and behavior of the sequential game with an arbitrary number of periods.

To provide the results it will be useful to introduce some notation. Write $\mathcal{C}$ for the space of real continuous functions defined in $\Delta\Theta$ and write

$$\mathcal{F} = \{V \in \mathcal{C} : \ V \text{ is non-negative and } V(\delta_\theta) = 0 \text{ for each } \theta \in \Theta\}.$$

Note that Lemma 3 and 2 imply that $V^t \in \mathcal{F}$ for each $t \in \mathbb{N}$. The key idea is to use the recursive nature of the mappings $(V^t)_{t\in\mathbb{N}}$ as a way to identify the regions where the value functions strictly increase with respect to t. With this in mind, define the **Bellman operator** $\phi : \mathcal{F} \to \mathcal{F}$ as follows:

$$\phi(V)(\mu_b) := \sup_{\tau\in\mathrm{PS}[\mu_b]} \mathbb{E}_\tau\big[U(\mu_b') + V(\mu_b')\rho(\mu_b' \mid \mu_b)\big] - U(\mu_b).$$

Observe, since $\phi(V) \in \mathcal{F}$ for each $V \in \mathcal{F}$, the Bellman operator is well-defined. (See Lemma 9 in the appendix). The Bellman operator identifies the mappings $(V^t)_{t\in\mathbb{N}}$ in the sense that

$V^{t+1} = \phi^t(V^1)$ for each $t \in \mathbb{N}$. Consequently, features of the agents' payoffs and behavior are derived from the monotonicity properties of ϕ. (Lemma 10 in the appendix describes these properties.)

1.5.1 The Impact of Time

Theorem 2 and Lemma 3 together show that the seller's subjective expected revenue weakly increases with the length of the interaction. The example in Section 2.2 shows that the seller strictly benefits from having additional periods to trade. In this section, we prove that this phenomenon is not an isolated case but a universal property satisfied in all environments with prior disagreement.

Write $\mathcal{D}^+ := \big\{\mu_b \in \Delta\Theta : \mu_b \neq g(\mu_b) \text{ and } V^1(\mu_b) > 0\big\}$ for the set of buyer's beliefs where the agents disagree and the buyer has a positive value for information.[9] The following result identifies two key properties of the set $\mathcal{D}^+$.

Lemma 4

(i) If $\nu_s = \nu_b$, then $\mathcal{D}^+ = \varnothing$.

(ii) If $\nu_s \neq \nu_b$, then $\operatorname{int} \Delta\Theta \subseteq \mathcal{D}^+$.

Lemma 4 follows from two observations. First, given that the decision problem is not trivial, at any interior belief, the buyer strictly values information. Second, if the agents disagree at the priors (ν_s, ν_b), then they disagree at any interior posterior belief. Consequently, there are two possibilities, either the agents share a common prior and thus $\mathcal{D}^+$ is empty, or the agents agree to disagree and $\mathcal{D}^+$ contains all beliefs with full support. Our first main result shows that $\mathcal{D}^+$ characterizes the region of beliefs at which the value functions $(V^t)_{t\in\mathbb{N}}$ strictly increase in t.

Theorem 3 *The following holds for each $t \in \mathbb{N}$:*

$$\mathcal{D}^+ = \big\{\mu_b \in \Delta\Theta : V^{t+1}(\mu_b) > V^t(\mu_b)\big\}.$$

---

[9] Notice,the set $\mathcal{D}^+$ depend on both, the decision problem (Θ, A, u) and the prior beliefs (ν_b, ν_s).

The proof of Theorem 3 follows from an inductive argument. The first step uses a geometric approach to show that the convex envelope defining $V^2(\mu_b)$ is strictly above the value $V^1(\mu_b)$. This relies on the existence of some belief $\mu_b' \in \mathcal{D}^+$ that the seller deems relatively more likely, (that is, $\rho(\mu_b' \mid \mu_b) > 1$). The second step uses the monotonic properties of the Banach operator ϕ. (See Lemma 10.) Since $V^3(\cdot)$ is defined in terms of $V^2(\cdot)$, and $V^2(\cdot)$ is defined in terms of $V^1(\cdot)$, it follows that $V^2(\cdot) > V^1(\cdot)$ in $\mathcal{D}^+$ implies that $V^3(\cdot) > V^2(\cdot)$ in $\mathcal{D}^+$. An inductive argument shows the result holds for all $t \in \mathbb{N}$.

Theorem 3 characterizes the seller's equilibrium payoffs for each belief μ_b, including the exogenous prior belief ν_b. Moreover, it provides a sharp difference between the case of common prior and the case of prior disagreement. Notice, under a common prior $\mathcal{D}^+ = \varnothing$. Consequently, $V^{T+1}(\nu_b) = V^T(\nu_b) = \ldots = V^1(\nu_b)$, implying that no selling strategy surpasses an experiment that fully reveals the state in the first period. By contrast, under prior disagreement $\nu_b \in \operatorname{int}\Delta\Theta \subseteq \mathcal{D}^+$. As a result, $V^{T+1}(\nu_b) > V^T(\nu_b) > \ldots > V^1(\nu_b) > 0$, showing that the seller's expected revenue strictly increases in the total number of periods, T.

Theorem 3 not only characterizes the equilibrium payoffs but also identifies the seller's behavior at any history in the game. Note, since fully revealing the state at any stage yields a payoff of $V^1(\mu_b^t)$, it is strictly suboptimal to fully reveal the state whenever the current belief is $\mu_b^t \in \mathcal{D}^+$ and $t > 1$. This implies the following corollary.

Corollary 1 *Assume that $\nu_s \neq \nu_b$ and $T > 1$. At each SPE the first experiment offered by the seller is not fully revealing. Moreover, in each history at period t that induces a buyer's belief $\mu_b^t \in \mathcal{D}^+$, the seller does not offer a fully revealing experiment.*

Intuitively, Theorem 3 and Corollary 1 follow from the interplay between selling information in the current period versus future periods. On the one hand, the sequential framework allows the seller to steer the buyer's beliefs toward paths in which the seller expects larger future payments. On the other hand, this drift comes with the cost of depleting the 'stock' of information available for future interactions. So, revealing too much information diminishes the available information for future sales, and withholding too much information decreases the posterior drift. Consequently, the optimal revenue is not achieved at extremes—neither through full disclosure nor complete withholding—but rather through an experiment that partially reveals information.

1.5.2 *Asymptotic Payoffs*

This section identifies the seller's subjective expected revenue in environments in which the agents interact in an asymptotically large number of periods. Typically, In dynamic settings, long-run outcomes are characterized by the convergence properties of Banach contractions. However, in our setting, the Bellman operator ϕ is not a Banach contraction, as it admits multiple fixed points. Despite this, we show that the set of fixed points of ϕ provide a bound on the seller's asymptotic revenue.

We now construct a key fixed point of ϕ. Write $\theta^* \in \arg\max_{\theta \in \Theta} r(\theta)$ for a state that maximizes the likelihood ratio of the agents. We define $B : \Delta\Theta \to \mathbb{R}$ as $B(\mu_b) := V^1(\mu_b)\rho(\delta_{\theta*} \mid \mu_b)$.[10]

Lemma 5 *The following statements hold:*

(i) The mapping B is a fixed point of ϕ.

(ii) For each $\mu_b \in \Delta\Theta$, $B(\mu_b) \geqslant V^1(\mu_b)$.

Lemma 5 shows that B fixed point of ϕ that dominates the value function V^1. Moreover, due to the monotonicity properties of ϕ, the fact that B dominates V^1 implies that B dominates all the mappings $(V^t)_{t \in \mathbb{N}}$. (See Lemma 10.) Figure 1.8 illustrates the mapping B (in red) and the value functions $(V_t)_{t \in \mathbb{N}}$ (in blue) for the example of Section 2.2.

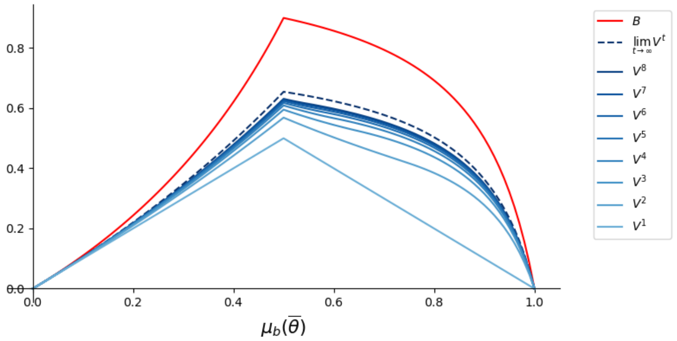


Figure 1.8. Mappings $(V^t)_{t \in \mathbb{N}}$ and B in terms of $\mu_b(\overline{\theta})$, given that priors are $(\nu_b, \nu_s) = (0.9, 0.5)$.

---

[10] Notice, under common prior, $\rho(\delta_{\theta*} \mid \mu_b) = 1$ for each $\mu_b \in \Delta\Theta$. Under prior disagreement $\rho(\delta_{\theta*} \mid \mu_b) > 1$ for each $\mu_b \in \operatorname{int} \Delta\Theta$. (See Lemma 11.)

Fix a belief $\mu_b \in \Delta\Theta$. Since the sequence $V^t(\mu_b)$ is monotone and bounded by $B(\mu_b)$, its limit exists. The following result shows that the point-wise limit of the sequence $(V^t)_{t\in\mathbb{N}}$ defines a well-behaved mapping.

Theorem 4 *There exists a mapping $V^\infty : \Delta\Theta \to \mathbb{R}$ so that, for each $\mu_b \in \Delta\Theta$,*

$$V^\infty(\mu_b) = \lim_{t\to\infty} V^t(\mu_b).$$

Moreover, $V^\infty \in \mathcal{F}$ and V^∞ is a fixed point of ϕ.

Theorem 4 states that the seller's asymptotic revenue is captured by V^∞, a continuous (and hence bounded) mapping. This result implies that the seller's marginal revenue with respect to t sharply decreases. Intuitively, the optimal strategy reveals most of the information in the early periods, depleting the stock of information available for the final extra periods. Consequently, when t is large, an additional period generates almost no extra value for the seller.

1.6 Discussion

1.6.1 Subgame Perfect Equilibrium

As described in Section 1.3, the agents in this game have no private information. Moreover, since prior beliefs are transparent and all signals are public, the agents' posterior beliefs remain transparent after any sequence of experiments.

We impose the restriction that agents "cannot signal what they do not know." That is, the actions of the co-player do not convey information about the state θ. At each history, the agents' beliefs about the state depend solely on the signals selected by chance. More specifically, given a stream of experiments $\boldsymbol{\pi} = (\pi^t)_{t\in\mathcal{T}}$ purchased by the buyer at periods $\mathcal{T} \subseteq \{1, 2, \ldots, T\}$ the conditional distribution of the stream of signals $\boldsymbol{m} = (m^t)_{t\in\mathcal{T}}$ is given by

$$\mathbb{P}_{\boldsymbol{\pi}}[\boldsymbol{m} \mid \theta] := \prod_{t\in\mathcal{T}} \pi^t(m^t \mid \theta).$$

Therefore, since i has prior $\nu_i \in \operatorname{int} \Delta\Theta$, i's posterior beliefs at period $\hat{t} < \min(\mathcal{T})$ are given by

$$\mu_i^{\hat{t}}(\theta) = \frac{\nu_i(\theta)\mathbb{P}_{\boldsymbol{\pi}}[\boldsymbol{m} \mid \theta]}{\sum\limits_{\theta' \in \Theta} \nu_i(\theta')\mathbb{P}_{\boldsymbol{\pi}}[\boldsymbol{m} \mid \theta']}. \tag{1.1}$$

Absent any experiment, the belief of each agent i remains fixed at the prior ν_i, even after a deviation of the co-player.

In this game, the agents' posterior beliefs are transparent after any history. Hence, subgame perfect equilibrium is equivalent to strong perfect Bayesian equilibrium under the following requirements: First, at each history that implements some experiments, the agents' beliefs are described by Equation (1.1). Second, at each history in which the buyer does not purchase any experiment, the belief of agent i equals its prior ν_i.

1.6.2 Discounting the Future

This paper investigates the effects of belief disagreement in information markets. Our main finding reveals that the seller benefits from longer interactions. To isolate the effects of prior disagreement, we abstract from other factors that could influence our dynamic setting, such as time discounting.

Incorporating a discounting parameter $\delta \in (0, 1)$ makes deferring the payments less attractive for the seller. As a result, the seller may refrain from exploiting trading opportunities in the long run and instead, opts to disclose more information in early stages. At one extreme, when the seller is highly impatient ($\delta \approx 0$), the optimal strategy entails disclosing all information within a single period. At the opposite extreme, with a sufficiently patient seller ($\delta \approx 1$), our characterization closely approximates the equilibrium payoffs of the agents.

1.6.3 Costly Experiments

This paper assumes that the seller faces no cost for executing experiments. This benchmark covers multiple economic interactions in which the marginal cost per experiment is negligible. For instance, software firms incur zero marginal costs for running antivirus tests.

We anticipate that adding costs to the experiments will impact the experiments that the seller offers in equilibrium. If the seller faces a small fixed cost per experiment, the seller will reveal more information (in comparison with no-cost environment) with the goal of decreasing the expected number of experiments executed. If the cost per experiment is sufficiently big, then the seller will opt to offer a fully-revealing experiment in the first period.

Chapter 2

Language Constraints in Multi-Sender Communication Games

This chapter studies optimal communication under language constraints. The model features multiple Senders and a single Receiver, in which the only barrier to communication is a limitation imposed by language: Senders are identically informed, all players have aligned incentives, the state space is infinite, but the Senders' message spaces are each finite. Thus, while players prefer full information revelation, their language does not allow for it. Efficient communication has two important properties. First, Senders use language heterogeneously. Second, Senders assign non-convex meaning to individual messages, while ensuring that the meaning of message profiles remain convex.

2.1 Introduction

In many economic settings, a decision maker lacks expertise. Instead, other agents may have the expertise that the decision maker desires. A prominent case is where those experts convey information to the decision maker via cheap-talk messages. Typically, this is studied in settings in which the experts have a rich language—in particular, a language rich enough to verbalize all relevant information, should they desire to do so. But, in practice, there may be important language constraints.

Language constraints are pervasive in everyday communication. They naturally arise when experts interact with decision makers under time pressure. Because technical jargon can require lengthy explanations, experts often restrict their language to words that have everyday meaning. For instance, lawyers may avoid complex legal jargon, doctors may avoid medical terminology, and technical advisors may avoid communicating complex findings.

In other cases, language constraints stem from institutional requirements. Organizations often mandate that experts condense detailed technical information into coarse categories when communicating with the public. In the U.S., certain public schools must report performance using a simplified rating scale (instead of providing details on test scores). In the EU, appliance manufacturers are required to provide information on electrical consumption, by using energy labels that rely on a coarse rating system. Similarly, in several Latin American countries, food companies must print coarse warning symbols on packaging to alert consumers about potentially harmful ingredients.

In these situations, experts are constrained to use a sparse language. They cannot use a language that is rich enough to convey every detail to the decision maker. In light of these and similar examples, the literature has recognized the importance of studying such language constraints (Crémer et al., 2007; Jäger et al., 2011; Blume and Board, 2013; Sobel, 2016; Aybas and Turkel, 2024).

This paper examines how experts best communicate under language constraints in a particularly tractable setting—one in which it should be easy to communicate. Specifically, it looks at a setting where the experts and the decision maker have aligned incentives. Additionally, each expert has the same information. So, absent language constraints, the experts should be able to communicate all relevant information. Moreover, the presence of one vs. multiple experts should not impact what information is conveyed. Yet, with language constraints, it does: The presence of one vs. multiple experts impacts not only what information is conveyed but also how it is conveyed.

To understand why, suppose there is a continuum of states. Each expert, or Sender, knows the state, but only has access to a finite set of (cheap-talk) messages. So, when a Sender conveys information, it must be clustered into coarse categories. That is, full-information revelation is not possible.

The main results concern the meaning of a message versus the meaning of a message profile. Both the meaning of a message and the meaning of a message profile are endogenously determined by equilibrium. For instance, suppose there are two Senders, S1 and S2, who can each report either *low* or *high*. The meaning of the message *low* for S1 is the set of states at which (in equilibrium) S1 says *low*. Similarly, the meaning of the message profile (*low*, *high*) is the set of states at which (in equilibrium) S1 says *low* and S2 says *high*. Notice the meaning of the message *low* may be different for S1 versus S2. But, it is the meaning of the message profile that determines the Receiver's best response.

What is the structure of the meaning of messages, when communication is efficient? Proposition 3 shows that, in any efficient equilibrium, message profiles essentially have convex meaning. Theorem 5, however, shows that the same conclusion cannot be drawn for the meaning of an individual Sender's message. In particular, when the state space is multi-dimensional and the choice set is sufficiently large, efficient communication typically requires that each Sender's messages have non-convex meaning.

To understand the interest in convex meanings, note that we typically think that states have a natural order—one that meaningfully conveys information about payoffs. For instance, suppose the set of states $[0, 1]$ represents the severity of a medical condition. A natural order specifies that higher states are associated with more severe medical conditions. So, if the agents want to choose the same treatment t at both states θ and θ', then presumably they should also want to choose treatment t at any convex combination of θ and θ'. In this sense, we would expect convex meaning. And, indeed, in any efficient equilibrium, message profiles do have convex meaning. Since the meaning of message profiles determines the Receiver's best response, the Receiver would choose treatment t at any convex combination of θ and θ'.

But, importantly, to achieve efficiency, the Senders may need to both use language heterogeneously and use language in a way that lacks convex meaning. To see this, suppose each of two (doctor) Senders can only report *low* or *high*. So, there are four message profiles that the (patient) Receiver can observe. Assume that an efficient equilibrium induces four possible actions—that is, each message profile induces a different equilibrium treatment. This cannot involve Senders using language homogeneously: If it did, the Receiver would only observe two message profiles and could not choose four different equilibrium actions. It also cannot involve messages of each individual Sender having convex meaning. If it did, each Sender Si would have a cutoff $\hat{\theta}_i$ so that Si reports *low* below $\hat{\theta}_i$ and reports *high* above $\hat{\theta}_i$. In that case, the Receiver would only observe (at most) three message profiles and, again, could not choose four different equilibrium actions.

Notice, the Receiver would observe four message profiles if one Sender adopts a language with convex meaning while the other does not. To see this, suppose an efficient equilibrium requires inducing different actions when the state belongs to $[0, \frac{1}{4})$, $[\frac{1}{4}, \frac{1}{2})$, $[\frac{1}{2}, \frac{3}{4})$, and $[\frac{3}{4}, 1]$. The Receiver can distinguish these four regions if: (i) S1 reports *low* when the state is in $[0, \frac{1}{2})$ and reports *high* otherwise and (ii) S2 reports *low* when the state is in $[\frac{1}{4}, \frac{3}{4})$ and reports *high* otherwise. This desire to communicate efficiently leads S2 to use language in a way that lacks convex meaning.

In this example, the desire to communicate efficiently requires that at least one Sender use language in a way that lacks convex meaning. The other Sender may well use language so that it has convex meaning. However, this is an artifact of the assumption that the state space is unidimensional. When the state space is multidimensional, the desire to communicate efficiently will typically require that all Senders use language in a way that lacks convex meaning. See Section 2.2 for an example.

At a broader level, the paper identifies two important and distinct patterns of language use in efficient equilibria: Senders use language heterogeneously, and their languages lack a convex meaning. This is despite the fact that a payoff assumption (a single-crossing property) favors convexity. These patterns of language use suggest that obtaining efficiency with multiple Senders requires solving a delicate coordination problem among them.

The findings of this paper have implications for institutional design. When an institution constrains experts to provide coarse reports to a decision maker, what reporting standards should be adopted? The results suggest that experts should be allowed to use different reporting standards, and should have the flexibility to issue non-convex reports.

While the paper focuses on the multi-Sender case, the model and analysis admit an alternate interpretation.[1] Consider a single Sender endowed with a language that consists of finitely many words. The Sender can use those words to form sentences—formally, sequences of words of finite length—and thereby enrich her language. The Receiver observes the sentence and takes an action.

This setting differs from the cheap-talk literature, where a message represents the full description of the information conveyed. Because there are typically no language constraints,

---

[1] See Section 2.5.1 for a detailed discussion.

it is without loss to focus on the case where a message represents a complete narrative. But, when there are language constraints, the Sender may want to piece together multiple messages to construct a narrative. This raises the question: How should the Sender structure sentences to communicate efficiently?

The results of the multi-Sender model can address this question. In particular, a model with K Senders is equivalent to a model with one Sender who can convey sentences with K words. Under this reinterpretation, the main results yield two implications. First, efficient word use requires context-dependent meaning: The meaning of each word depends on both its position in a sentence and the surrounding words. Second, although complete sentences must have convex meaning, incomplete sentences typically exhibit non-convex meaning. These insights suggest that efficiently constructing narratives from a finite set of words requires care.

To understand the policy implications, return to the front-of-package nutrition labels example. These policies mandate food companies to print a sequence of ingredient-related labels on packaging; each label indicates that a harmful feature of the product exceeds a certain threshold. (Examples of such features include calories, sugar, sodium, and saturated fats.) In practice, the labeling requires context-independence and convex meaning. The results in this paper suggest that communication can be improved by relaxing these requirements.

This paper falls within a growing literature that uses a game theoretic analysis to raise questions about language and language use (Rubinstein, 1996; Blume, 2000; Glazer and Rubinstein, 2001; Blume, 2004; Crémer et al., 2007; Lipman, 2009; Jäger et al., 2011; Blume and Board, 2013; Sobel, 2016; Heumann, 2020; Suzuki, 2023; Aybas and Turkel, 2024; Bauch, 2024; Blume, 2024; Dilmé, 2024; Lipman, 2025). It studies the meaning in language in cheap talk settings (Crawford and Sobel, 1982a; Green and Stokey, 2007).

The paper shows a trade-off between efficiency and messages having convex meaning. Convexity is not a new property. In fact, the linguistics literature has viewed convexity of meaning as a structural design feature of natural languages. See Gärdenfors (2000); Jäger (2007); Jäger et al. (2010); Gärdenfors (2014); Kirby (2017); Carr et al. (2017). Intuitively, it is easier to learn a language that exhibits convexity of meaning. As a matter of fact, consider a language with this property, and suppose that a speaker hears that two different colors are called ‘red’. The speaker would then infer that any convex combination of these two colors is also called ‘red’. This shows that convexity of meaning allows speakers to generalize from a few observations to a continuum of cases.

Here, convex meaning or the lack thereof, is an output of equilibrium. The result on message profiles having convex meaning is familiar from the literature on cheap talk. (This insight goes back to Crawford and Sobel (1982a).) In models with a unidimensional state space, convexity arises from preferences that favor interval choice, such as Milgrom and Shannon’s (1994) single-crossing property or Kartik et al.’s (2024) single-crossing differences. Analogously, in the multidimensional case, convexity results from preferences that favor convex choice such as Grandmont’s (1978) intermediate preferences, Jäger et al.’s (2011) convex loss function, Sobel’s (2016) convex choice, Saint-Paul’s (2017) generalized single-crossing condition, Kartik and Kleiner’s (2024) directionally single-crossing differences, or the strict-single crossing (SSC) property introduced in this article.

The literature has pointed to other reasons why messages may not have convex meaning. For instance, in Krishna and Morgan (2004), a lack of convexity arises because of dynamics. In Antic et al. (2023), a lack of convexity arises from the experts’ desire to communicate in the presence of an outside observer who can veto the experts’ choices.

The closest paper to this one is Jäger et al. (2011). They consider a model with a single Sender whose language is constrained. As a result, they cannot address the important tradeoffs that arise in the multiple-Sender case. In particular, their results cannot speak to whether Senders should use language homo- vs. heterogeneously. (That is, their results cannot speak to whether the meaning of the message *low* should be the same or different for different Senders.) Likewise, they cannot study a trade-off between efficiency and convex meaning: With one Sender, in any efficient equilibrium, each message must have convex meaning. (This follows from the fact that, with one Sender, a message profile is a message.)

This paper shows that efficiency may require that Senders use language heterogeneously. This stands in contrast to Crémer et al. (2007) and Sobel (2016). In those papers, the Receiver has a language constraint. As a result of the Receiver’s language constraint, efficient use of language requires that the Senders use language homogeneously.

The remainder of the paper is organized as follows. The next section illustrates the main results in a simple example. Section 3 introduces the model. Section 4 presents the results of the paper. Finally, Section 5 discusses and concludes the paper. All proofs are collected in the appendices.

2.2 Example

Assume that a single Receiver interacts with two Senders. The state space is $\Theta = \operatorname{Conv}\{(0,0),(0,1),(1,0)\} \subseteq \mathbb{R}^2$, and the common prior is uniform over Θ. The game starts with the realization of the state. Both Senders observe the state, but the Receiver does not. After learning the state, Senders simultaneously send messages to the Receiver. The Receiver then chooses an action from the choice set A with $\Theta \subseteq A \subseteq \mathbb{R}^2$. All players have identical preferences given by the utility function $u(a,\theta) = -\|a-\theta\|^2$. Note, the utility is decreasing in the distance between the action and the state. Hence, all player would like the Receiver to correctly guess the state.

Rich language First consider the benchmark case of a rich language. That is, suppose that each Sender has access to a message space large enough to label every state of the world. The simplest case occurs when $M_1 = M_2 = \Theta$.

There are many equilibria in this game. In particular, there are uninformative (or babbling) equilibria. One such equilibrium involves each Sender uniformly randomizing over her message space; upon observing any message profile, the Receiver takes the ex-ante optimal action $a^* = (\frac{1}{3}, \frac{1}{3})$. Because the Receiver's strategy is irresponsive to message profiles, both Senders are indifferent among their messages. Moreover, given the Senders' strategy profile, no message profile reveals information about the state. As a result, the ex-ante optimal action remains a best response to every message profile.

Note, however, an uninformative equilibrium is inefficient because it does not maximize the ex-ante expected utility. There is an efficient fully revealing equilibrium satisfying the following properties: If the state is θ, then each Sender i sends message $m_i = \theta$. Upon observing the message profile (m_1, m_2), the Receiver chooses action $a = m_1$ if $m_1 = m_2$, and $a^* = (\frac{1}{3}, \frac{1}{3})$ otherwise. This fully revealing equilibrium is efficient because the Receiver chooses the optimal action at every state of the world.

Importantly, any efficient equilibrium must be fully revealing. Moreover, the efficient equilibrium outcome is robust to the number of Senders: Independent of the number of Senders, the Receiver accurately learns the state of the world and chooses the *ex post* optimal action.

Sparse Language Suppose that there is a sparse language. That is, suppose that no Sender has access to a message space large enough to label every state of the world. The sparsest case occurs when $M_1 = M_2 = \{\ell, h\}$.

As in the rich language case, this game has many equilibria. In particular, there are uninformative equilibria. However, no uninformative equilibrium is efficient even relative to the constrained message spaces. Moreover, an efficient equilibrium cannot be fully revealing because there are not enough messages to separate all states of the world. This raises the question: How do efficient equilibria use language?

To answer this question, let us first look at an auxiliary decision problem. Note that the Receiver can observe at most four message profiles: $(\ell, \ell), (\ell, h), (h, \ell)$, and (h, h). Therefore, through communication, he can learn that the state belongs to one of K regions, with $K \leqslant 4$. Each region $P_k \subseteq \Theta$ will be associated with an action a_k. With this in mind, fix a profile of K actions, $(a_k : k \leqslant K)$. The first step asks how the Receiver should partition the state space into K regions, if he wants to use these actions optimally. Thus, the first step will associate each profile of K actions with an optimal partition of the state space. The second step will address which action profile and associated partition is best for the Receiver.

First, fix a profile of actions $(a_k : k \leqslant K)$. The Receiver will assign state θ to region P_k if the distance between θ and a_k is smaller than the distance between θ and any other a_j. This implies that, if θ and θ' belong to region P_k, then any convex combination of θ and θ' must also be in region P_k: since both θ and θ' are closest to a_k , so is any convex combination of θ and θ'. Thus, the Receiver should divide the state space into convex regions.

Second, the Receiver should partition the state space into four regions. To see why, consider a profile of three actions (a_1, a_2, a_3) associated with a partition $\mathcal{P} = \{P_1, P_2, P_3\}$.

Pick a state $\theta' \notin \{a_1, a_2, a_3\}$. Let $\varepsilon > 0$ denote the minimum distance between state θ' and each a_k. Define B as the open ball centered at θ' with radius ε. Now, consider a new partition $\mathcal{P}' = \{P'_1, P'_2, P'_3, P'_4\}$ constructed as follows: $P'_k = P_k \backslash B$ for each $k \leqslant 3$, and $P'_4 = B$. For $k \leqslant 3$, associate each region P'_k with the action a_k; associate P'_4 with action $a_4 = \theta'$. Notice, partition $\mathcal{P}'$ (and the associated action profile) gives the Receiver a strictly higher expected utility than partition $\mathcal{P}$, so the Receiver should partition the state space into four regions.

Third, the optimal partition involves four regions of approximately the same mass. This arises from the fact that the Receiver wants to choose a profile of actions that minimizes the variance within regions. Thus, the Receiver wants to choose a profile of actions whose associated regions have approximately the same probability. Because the prior is uniform, this implies that the regions have approximately the same mass.

Figure 2.1 depicts the unique solution to the Receiver's decision problem. The actions associated with each region are $a_{\ell\ell} = (0.17, 0.17)$, $a_{\ell h} = (0.13, 0.65)$, $a_{h\ell} = (0.65, 0.13)$, and $a_{hh} = (0.39, 0.39)$.

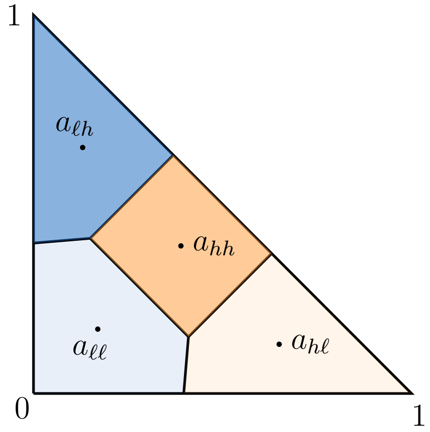
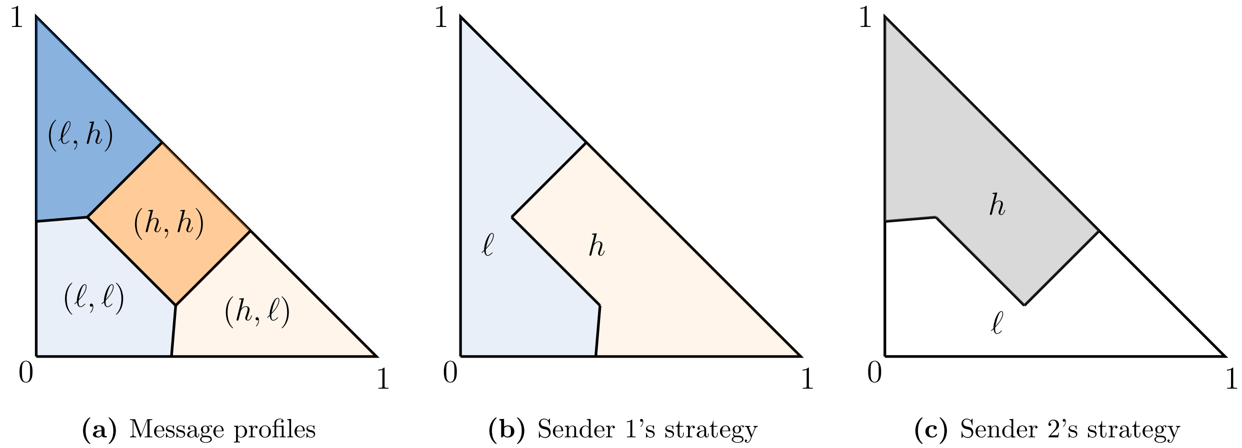


Figure 2.1. Optimal partition

The solution to the auxiliary decision problem directly determines the efficient equilibrium outcome. Since all players have aligned incentives, the Receiver's utility maximization in the decision problem ensures that the Senders' utilities are maximized. To construct an equilibrium that induces the optimal partition in Figure 2.1, recall that the Receiver selects an action based on the observed message profile. Therefore, it suffices to associate each region with a specific message profile, as Figure 2.2a illustrates. Figures 2.2b and 2.2c decompose Figure 2.2a into the strategies chosen by each Sender.



(a) Message profiles **(b)** Sender 1's strategy **(c)** Sender 2's strategy

Figure 2.2. Message profiles and Senders' strategies

Think of the *meaning of message* m_i for Sender i as the set of states in which Sender i transmits message m_i. Similarly, think of the *meaning of message profile* (m_1, m_2) as the set of states in which Sender 1 and Sender 2 transmit messages m_1 and m_2, respectively.

The equilibrium Senders' strategies in Figures 2.2b and 2.2c exhibit three features. First, Senders assign heterogeneous meaning to each message. That is, the meaning of message ℓ differs between the two Senders. Second, the meaning of each message is not convex. That is, there exist states θ and θ' at which Sender i sends message ℓ, even though Sender i does not send the same message at the state $\alpha\theta + (1 - \alpha)\theta'$ for some $\alpha \in (0, 1)$. Third, the meaning of each message profile is convex. That is, if Senders transmit message profile (m_1, m_2) at two states θ and θ', then they transmit the same message profile at every state of the form $\alpha\theta + (1 - \alpha)\theta'$ for $\alpha \in (0, 1)$.

These features are properties of each efficient equilibrium. Indeed, if the two Senders assigned the same meaning to some message, then the Receiver would observe at most two message profiles: (ℓ, ℓ) and (h, h). Therefore, he would identify at most two regions in the state space. However, no partition with less than four regions can be optimal as we previously argued. Hence, Senders must assign heterogeneous meaning to each message.

Similarly, Senders must assign non-convex meaning to each message. To see it, note that the meaning of message m_i for Sender i corresponds to the union of the meaning of the message profiles in which Sender i transmits message m_i. That is, the meaning of message m_i consists of the union of the meanings of two message profiles: (m_i, ℓ) and (m_i, h). Since the meaning of each message profile corresponds to a region in Figure 2.1, and the union of any two such regions is non-convex, it follows that the meaning of message m_i must also be non-convex. Note, this is different from the one-dimensional case, where some Senders' messages were convex.

Finally, the meaning of each message profile must be convex. As a matter of fact, the meaning of each message profile must correspond to a region in Figure 2.1. Since each such region is convex, so is the meaning of each message profile.

This analysis highlights, under a sparse language, efficient communication requires coordination among Senders and a specific structure on language use. In particular, it requires Senders to assign non-convex heterogeneous meaning to messages, while ensuring that message profiles carry convex meaning. The remainder of the paper examines the generality of these results.

2.3 Model

Consider a cheap-talk model with a finite set of Senders I ($|I| \geqslant 2$) and one Receiver. Each Sender (she) observes a payoff-relevant state, but the Receiver (he) does not. The Senders then simultaneously send messages to the Receiver. The Receiver observes the profile of messages sent and chooses an action.

Let $\Theta \subseteq \mathbb{R}^n$ be the set of states, where $n \geqslant 2$. It is assumed that Θ is a compact convex set with non-empty interior. The state is drawn from a full-support absolutely continuous distribution F, with density f. Sender i's set of messages is a finite set M_i with $|M_i| \geqslant 2$. Note, the Senders' language is sparse since $|\Theta| > |\times_{i \in I} M_i|$. The set of the Receiver's actions is a compact metric set A.

The Senders and the Receiver have identical preferences represented by a utility function $u : A \times \Theta \to \mathbb{R}$. The utility function satisfies a *strict single-crossing* (SSC) property and a *non-triviality* condition.

Definition 1 *The utility function* $u : A \times \Theta \to \mathbb{R}$ *satisfies the* strict single-crossing (SSC) property *if, for each* $a, a' \in A$ *and* $\theta, \theta' \in \Theta$*, the following holds: If* $u(a,\theta) \geqslant u(a',\theta)$ *and* $u(a,\theta') > u(a',\theta')$*, then*

$$u(a, \alpha\theta + (1-\alpha)\theta') > u(a', \alpha\theta + (1-\alpha)\theta'),$$

for each $\alpha \in (0,1)$*.*

The SSC property requires the following: If players weakly prefer action a over a' at one state and strictly prefer a over a' at another, then they must strictly prefer a over a' at any non-degenerate convex combination of these two states. So, if players rank two actions the same way at two distinct states, the ranking cannot reverse at any intermediate state.

Definition 2 *The utility function* $u : A \times \Theta \to \mathbb{R}$ *satisfies the* non-triviality condition *if, for each* $a, a' \in A$*, the following holds: If* $u(a,\theta) = u(a',\theta)$ *for each* $\theta \in \Theta$*, then* $a = a'$*.*

The non-triviality condition requires that the players are not indifferent between any distinct pair of actions across all states. Throughout the paper, a utility function $u : A \times \Theta \to \mathbb{R}$ that satisfies both the SSC property and the non-triviality condition will be referred to as an *SSC-utility function.*

A (pure) strategy for Sender i is a measurable mapping $\sigma_i : \Theta \to M_i$. So, σ_i specifies the message Sender i sends contingent on the state. A (pure) strategy for the Receiver is a mapping $\rho : M \to A$, where $M = \bigtimes_{i \in I} M_i$. So, ρ specifies the action the Receiver selects contingent on the observed message profile. It is convenient to define the joint Senders' strategy profile, i.e., $\sigma : \Theta \to M$, where $\sigma(\theta) = (\sigma_i(\theta))_{i \in I}$. Likewise, let $\sigma_{-i} : \Theta \to M_{-i}$, where $M_{-i} = \bigtimes_{j \in I, j \neq i} M_j$, denote the vector of all Senders' strategies but i's. That is, $\sigma_{-i}(\theta) = (\sigma_j(\theta))_{j \in I, j \neq i}$.

Write

$$u(a,\mu) = \int_{\Theta} u(a,\theta) d\mu$$

for the Receiver's expected utility of action a when his belief about the state of the world is $\mu \in \Delta\Theta$. The solution concept is efficient (pure-strategy) perfect Bayesian equilibrium.

Definition 3 *A strategy profile* (σ, ρ) *is a* perfect Bayesian equilibrium (PBE) *if there exists some belief system* $\mu : M \to \Delta\Theta$ *so that the following holds:*

(i) For each $\theta \in \Theta$ *and every* $i \in I$, $\sigma_i(\theta) \in \arg\max_{m_i \in M_i} u(\rho(m_i, \sigma_{-i}(\theta)), \theta)$.

(ii) For each $m \in M$, $\rho(m) \in \arg\max_{a \in A} u(a, \mu(m))$.

(iii) For each $(\theta, m) \in \Theta \times M$, $\mu(\theta \mid m) \int \mathbb{1}_{\sigma^{-1}(\{m\})}(\theta')dF(\theta') = \mathbb{1}_{\sigma^{-1}(\{m\})}(\theta)f(\theta)$.

A strategy profile (σ, ρ) is a PBE if it satisfies three conditions. Condition (i) says that each Sender will choose a message that maximizes her payoff, given that the other Senders transmit messages according to σ_{-i}, and that the Receiver selects the action that ρ specifies. Condition (ii) says that the Receiver will choose an action that maximizes his expected payoff given the belief induced by the Senders' message profile. Condition (iii) says that the Receiver's belief about the state conditional upon observing a message profile must be derived from the prior and the Senders' strategies via Bayes' rule.

For a given strategy profile (σ, ρ), let $v(\sigma, \rho) \in \mathbb{R}$ denote the induced ex-ante expected payoff. Formally,

$$v(\sigma, \rho) = \int u((\rho \circ \sigma)(\theta), \theta)dF(\theta).$$

The next definition formalizes the notion of efficiency.

Definition 4 *Let* (σ, ρ) *be a PBE. Say* (σ, ρ) *is an* efficient PBE *if* $v(\sigma, \rho) \geqslant v(\sigma', \rho')$, *for each PBE* (σ', ρ').

A PBE (σ, ρ) is efficient if (σ, ρ) induces the maximum ex-ante expected payoff across all PBEs. The class of efficient PBE is appealing for three reasons.[2] First, it is focal because if players could coordinate on a specific PBE, they would choose an efficient one. Second, analyzing this class provides a valuable benchmark, as efficient PBE payoffs are the highest among the PBE payoffs of the communication game. This is similar to papers that study the upper bound on the Sender's gains from communication, as in the Bayesian persuasion model (Kamenica and Gentzkow, 2011a). Finally, efficient PBE payoffs can be attained

---

[2] Typically, the common-interest communication literature pays special attention to the set of efficient equilibria. See, for example, Crémer et al. (2007); Lipman (2025); Jäger et al. (2011); Blume and Board (2013); Sobel (2016).

through pure strategies (See Theorem 1 in Lipman (2025)). This implies that excluding mixed strategies is without loss of generality.

The next Proposition ensures that this equilibrium class is non-empty provided that the utility function is continuous.

Proposition 2 *Assume $u : A \times \Theta \to \mathbb{R}$ is continuous. There exists an efficient PBE.*

The proof of Proposition 2 follows a standard argument. The compactness of the state space, message sets, and action set ensures that the set of strategy profiles is compact. By continuity of the utility function, there exists a strategy profile that maximizes the ex-ante payoff. This strategy profile constitutes an efficient equilibrium: since preferences are aligned, no player has an incentive to deviate.[3]

For the ease of exposition, I focus on a particular class of efficient PBE. Call a PBE (σ, ρ) *non-redundant* if for every $m, m' \in \sigma(\Theta)$, $\rho(m) = \rho'(m)$ implies that $m = m'$. In other words, a PBE is non-redundant if every pair of distinct message profiles sent along the equilibrium path of play induce different responses from the Receiver.

Requiring non-redundancy is without loss of generality since every efficient PBE outcome can be achieved by a non-redundant PBE. Intuitively, any redundant PBE can be transformed into a non-redundant PBE by merging messages that induce the same action. Henceforth, an efficient pure-strategy non-redundant PBE will be referred to simply as an efficient PBE.

2.4 Results

The main results of the paper concern the structure of meaning in an efficient PBE. In the model, two objects carry meaning: messages and message profiles. The following definition formalizes these notions.

Definition 5 *Let (σ, ρ) be a strategy profile.*

---

[3] This argument does not rely on the SSC property. Moreover, a similar reasoning applies to any game with compact strategy sets. The key observation is that the set of equilibrium payoffs is compact since equilibrium conditions are defined by weak inequalities. This stronger result is not stated in the paper, as doing so would require introducing additional notation.

(i) The meaning of message $m_i \in M_i$ *of Sender* i *is* $\sigma_i^{-1}(\{m_i\}) = \{\theta \in \Theta \mid \sigma_i(\theta) = m_i\}$.

(ii) The meaning of message profile $m \in M$ *is* $\sigma^{-1}(\{m\}) = \bigcap_{i \in I} \sigma_i^{-1}(\{\mathrm{proj}_i(m)\})$.

The meaning of message m_i of Sender i is the set of states in which Sender i sends m_i. The meaning of a message profile $m = (m_i)_{i \in I}$ is the set of states in which each Sender i sends their respective message m_i. In other words, the meaning of a message or a message profile corresponds to the set of states in which it is used. The next result establishes that efficiency requires message profiles to have almost surely convex meaning.[4]

Proposition 3 *Assume* $u : A \times \Theta \to \mathbb{R}$ *is an SSC-utility function. In an efficient PBE, the meaning of each message profile is almost surely convex.*

The proof of Proposition 3 follows from the SSC property and the *non-triviality* condition. To see this, first, note that in an efficient equilibrium, the ex-ante expected utility is maximized subject to the language constraints. Therefore, each equilibrium action is associated with the set of states where it is strictly optimal. The SSC property ensures that this set is convex.

It remains to examine the states associated with each equilibrium action, where the action is only weakly optimal. Below, we will see that the set of such states has measure zero, so it can be ignored. As a result, each equilibrium action is associated with an almost surely convex set of states. Finally, because the Receiver best responds to message profiles, each message profile must pool an almost surely convex set of states.

Note, the states at which an equilibrium action is only weakly optimal are precisely those where there is another equilibrium action that is also optimal. The SSC property and the non-triviality condition together imply that the set of states at which players are indifferent between two actions forms a low-dimensional affine subspace of the state space. Before providing the intuition behind this statement, observe that this implies that indifference is probabilistically negligible, and therefore, so is the set of weakly optimal states.

---

[4] Formally, a set $C \subseteq \Theta$ is *almost surely convex* if there exists a convex set $C' \subseteq \Theta$ such that their symmetric difference, $C \triangle C'$, has Lebesgue measure zero. Conversely, a set C fails to be almost surely convex if, for every convex set C', the symmetric difference $C \triangle C'$ has positive measure.

The underlying intuition for why this set forms an affine subspace has two steps. First, the SSC property ensures the affine structure. To see this, let θ and θ' be two distinct states, and let a and a' be two distinct actions such that players are indifferent between a and a' at both states. Suppose, for contradiction, that there exists $t \in \mathbb{R}$ such that players are not indifferent between a and a' at the state $t\theta' + (1-t)\theta$. For simplicity, assume $t > 1$ and that players strictly prefer a over a' at this state.[5] Observe that θ' is a convex combination of θ and $t\theta' + (1-t)\theta$. Since players strictly prefer a over a' at $t\theta' + (1-t)\theta$, and are indifferent between a and a' at θ, the SSC property implies that players must strictly prefer a at θ', which is a contradiction.

Hence, the set of states at which players are indifferent between two different actions is either empty, a singleton, a line, a hyperplane, or the entire state space. Second, the *non-triviality* condition rules out the last possibility, ensuring that this affine subspace has strictly lower dimension. Consequently, indifference occurs on a negligible set and does not affect the analysis.

Proposition 3 prompts the question of whether each Sender's message must have almost surely convex meaning in an efficient equilibrium, when players have an SSC-utility function. The example in Section 2.2 illustrates that this is not necessarily the case: All individual messages can lack convex meaning even though message profiles do have convex meaning. The following theorem examines how prevalent this phenomenon is.

Theorem 5 *Let $|A| \geqslant |M|$. There exists an SSC-utility function so that, in each efficient PBE, the following holds: For each Sender i, the meaning of any message m_i fails to be almost surely convex.*

Theorem 5 establishes that even in environments that inherently favor convexity, non-convex meaning is essential for efficient communication. More precisely, there exists an SSC-utility function for which the efficient use of a sparse language requires each Sender to adopt a language that lacks convex meaning. Notably, Proposition 3 and Theorem 5 together reveal that efficient communication relies on a subtle form of coordination among Senders.

---

[5] See Lemma 15 in Appendix A for the remaining cases.

Specifically, while each Sender's message has non-convex meaning, the aggregate meaning of message profiles remains convex.

The proof of Theorem 5 focuses on constructing an SSC-utility function that induces a unique efficient PBE outcome (up to a null set). The construction has a structure that resembles the example in Section 2.2: The idea is to partition the state space into $|M|$ regions. Figure 2.3 describes how the partition is constructed for the case $|M| = 6$. Notice, there is a regular polygon with $|M| - 1 = 5$ sides. The $|M| - 1 = 5$ other regions are constructed by drawing rays from the center of the polygon through each of its vertices and extending them to the boundary of the state space. This forms a partition of the state space into $|M| = 6$ convex regions.

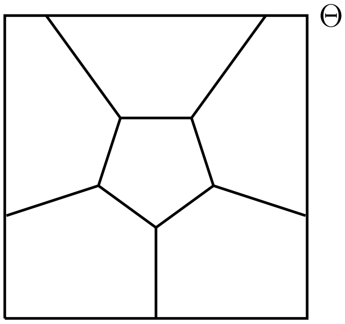


Figure 2.3. Induced optimal partition Theorem 5 (case $|M| = 6$)

The proof constructs an SSC-utility function so that each region is associated with a distinct action that is optimal for that region. Thus, there is an efficient PBE with the following property: Each region P_k is associated with a different message profile. Conditional upon observing the message profile associated with P_k, the Receiver chooses the action associated with P_k. (In fact, this property holds for each efficient PBE, up to a null set.)

Under the construction, each region corresponds to the meaning of some message profile. So, each message profile has convex meaning. However, the meaning of an individual message is the union of the meanings of all message profiles in which the message appears. Refer to Figure 2.3. The union of any two or more regions is not convex, provided the union is not all the state space Θ. Since each individual message is associated with at least two regions and is not associated with all regions, the meaning of any individual message is non-convex.

Note, Theorem 5 does not apply to every SSC-utility function, but only to certain ones. Example 1 provides an SSC-utility function and an efficient equilibrium in which one of the Senders assigns convex meaning to all of her messages.

Example 1 *Assume that a single Receiver interacts with two Senders. The state space is* $\Theta = \operatorname{Conv}\{(0,0),(0,1),(1,0)\} \subseteq \mathbb{R}^2$*, and the common prior is uniform over* Θ*. The Senders message sets are* $M_1 = M_2 = \{\ell, h\}$*. The Receiver's action set is* A *with* $\Theta \subseteq A \subseteq \mathbb{R}^2$*. All players have identical preferences, represented by the SSC-utility function*

$$u(a,\theta) = -\sqrt{(a-\theta)^T W (a-\theta)},$$

where

$$W = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 0.5 \end{pmatrix}.$$

That is, utility decreases with the weighted Euclidean distance between the chosen action and the true state, with weight determined by the positive definite matrix W*. Intuitively, players want the Receiver to correctly identify the state but place twice as much importance on accuracy in the first state component. Moreover, given an incorrect guess, players prefer the Receiver's errors to be positively correlated—meaning they would rather have the Receiver consistently overestimate or underestimate both state components rather than overestimating one and underestimating the other.*

Note, as in the example of Section 2.2, the Receiver can observe at most four message profiles. Therefore, through communication, he can learn that the state belongs to one of K *regions, with* $K \leqslant 4$*. After learning the region that the true state belongs to, he chooses an action. If the Receiver were able to divide the state space into* K *regions, how should he do so? Figure 2.4 depicts the unique solution to his decision problem. The actions corresponding to each region are* $a_{\ell,\ell} = (0.28.0.29)$*,* $a_{\ell,h} = (0.13, 0.62)$*,* $a_{h,\ell} = (0.86, 0.06)$*, and* $a_{h,h} = (0.57, 0.19)$*.*

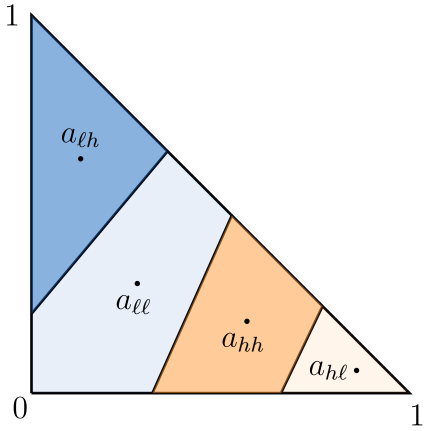
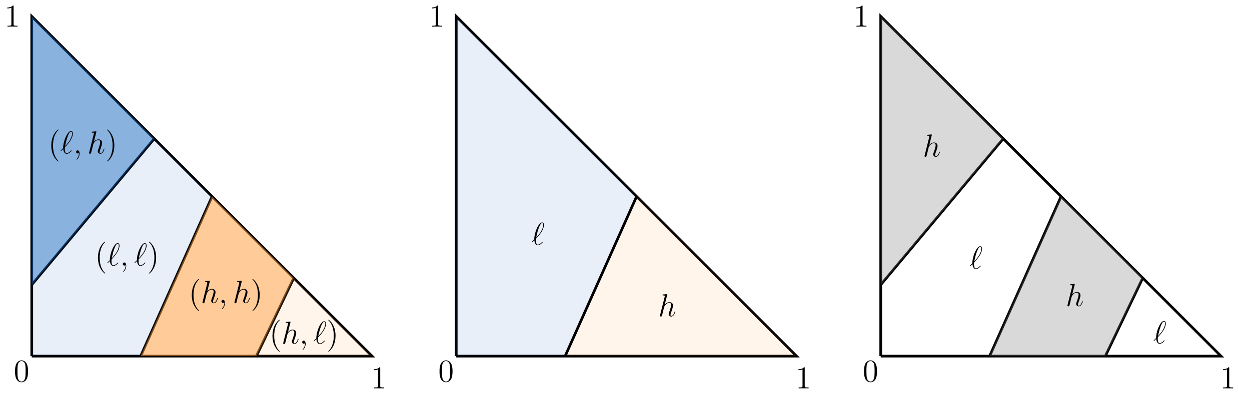


Figure 2.4. Optimal partition with weighted Euclidean preferences

In order to construct an equilibrium that induces the optimal partition in Figure 2.4, recall that the Receiver selects an action based on the observed message profile. Therefore, it suffices to associate each region with a specific message profile, as Figure 2.5a illustrates. Figures 2.5b and 2.5c decompose Figure 2.5a into the strategies chosen by each Sender.



(a) Message profiles **(b)** Sender 1's strategy **(c)** Sender 2's strategy

Figure 2.5. Message profiles and Senders' strategies with weighted Euclidean preferences

Observe, Sender 1 assigns convex meaning to each of her messages. Hence, the conclusion of Theorem 5 fails when preferences are given by the SSC-utility function $u(a, \theta) = -\sqrt{(a-\theta)^T W (a-\theta)}$.

In Example 1, Sender 1's messages have convex meaning but Sender 2's message do not. This raises the question of whether some Sender must use a strategy that involves non-convex meaning. Appendix C shows that this need not be the case. In particular, it uses the method of proof for Theorem 5 to show that there exists an SSC-utility function and an efficient equilibrium thereof, so that all Senders use strategies that involve convex meaning.

As a final remark, observe that Proposition 3 and Theorem 5 together provide conditions under which efficient communication requires Senders to use language in a heterogeneous manner. Without such heterogeneity, the non-convex meaning of individual messages could not combine to yield the convex meaning of message profiles.

2.5 Discussion and conclusion

2.5.1 A single Sender interpretation

The model and analysis presented in this paper admit an alternative interpretation to the multi-Sender setting. Consider a single Sender and a single Receiver. The Sender observes a payoff-relevant state, but the Receiver does not. Upon observing the state, the Sender sends a message to the Receiver. The Receiver observes the message and selects an action.

The sets Θ and A, along with the common prior and the utility function, satisfy the assumptions introduced in the model section (Section 2.3). However, the set of messages available to the Sender is now $M = \bigtimes_{i\in I} M_i$, where each M_i is a finite set and I is a finite index set. For ease of exposition, refer to the typical element m of M as a *sentence*. Thus, the Sender communicates by constructing sentences of length $|I|$. Given a sentence m, refer to $\operatorname{proj}_i(m)$ as the *i-th word* in sentence m. Let $W = \bigcup_{i\in I} M_i$ denote the set of all words that may appear in a sentence.

For each $i \in I$, define extended message sets $\tilde{M}_i = M_i \cup \{\varnothing\}$. Let $\tilde{M} = [\bigtimes_{i\in I} \tilde{M}_i] \setminus [M \cup \{(\varnothing)_{i\in I}\}]$ denote the set of *partial sentences*, i.e., sentences in which some but not all positions are unspecified (denoted by $\varnothing$). For each word $w \in W$, define $[w] = \{m \in M \mid \exists i \in I, \operatorname{proj}_i(m) = w\}$. That is, $[w]$ is the set of sentences in which the word w appears. For each partial sentence $\tilde{m}$, define $[\tilde{m}] = \{m \in M \mid \forall i \in I,\ \operatorname{proj}_i(\tilde{m}) \in \{\varnothing, \operatorname{proj}_i(m)\}\}$. In other words, $[\tilde{m}]$ is the set of sentences that are consistent with $\tilde{m}$, that is, sentences that match the partial sentence in every specified position.

The following definition formalizes the meaning of a sentence, a word, and a partial sentence under a given strategy profile.

Definition 6 *Let (σ, ρ) be a strategy profile.*

(i) The meaning of sentence $m \in M$ *is* $\sigma^{-1}(\{m\}) = \{\theta \in \Theta \mid \sigma(\theta) = m\}$.

(ii) The meaning of the word $w \in W$ *is* $\bigcup_{m \in [w]} \sigma^{-1}(m)$.

(iii) The meaning of the partial sentence $\tilde{m} \in \tilde{M}$ *is* $\bigcup_{m \in [\tilde{m}]} \sigma^{-1}(\{m\})$.

The meaning of a sentence is the set of states in which the Sender uses it. The meaning of a word aggregates the meaning of all sentences in which it appears. The meaning of a partial sentence corresponds to the meaning of all consistent sentences.[6]

Proposition 3 directly translates into this alternative framework.

Corollary 2 *Assume $u : A \times \Theta \to \mathbb{R}$ is an SSC-utility function. In an efficient PBE, the meaning of each sentence is almost surely convex.*

While Theorem 5 does not translate directly into this framework, its proof implies that the union of the meanings of any proper subset of sentences fails to be almost surely convex. This remark has two noteworthy implications. First, the meaning of a word w is context-dependent: it depends on the word's position within a sentence, and the other words in the sentence. This observation relates to pragmatics, the study of how context shapes the meaning of utterances. The second implication is formalized in the following result.

Corollary 3 *Let $|A| \geqslant |M|$. There exists an SSC-utility function so that, in each efficient PBE, the meaning of any word and any partial sentence fails to be almost surely convex.*

Corollary 3 speaks to semantics, the study of meaning. It shows that non-convexity of meaning is essential for efficient communication. This is the case even in very simple languages: those with only two words and sentences of length two.

To illustrate, suppose a Sender uses a language with $M_1 = \{e, m\}$ and $M_2 = \{\ell, h\}$, where e, m, ℓ, and h stand for *extreme*, *moderate*, *low*, and *high*, respectively. Let the state space and the action set be the unit interval. Assume the Sender and Receiver's utility is

---

[6] These definitions align with Frege's context principle. It states that the meaning of an expression is determined by the meanings of all complex expressions in which it appears. See Szabó (2024) for an exposition.

$u(a, \theta) = -(a - \theta)^2$, which is an SSC-utility function. In an efficient PBE, the state space is partitioned into four intervals: $[0, \frac{1}{4})$, $[\frac{1}{4}, \frac{1}{2})$, $[\frac{1}{2}, \frac{3}{4})$, and $[\frac{3}{4}, 1]$. A natural language might associate the first interval with the sentence (e, ℓ), the second with (m, ℓ), the third with (m, h), and the fourth with (e, h). In this equilibrium, the meaning of the word *extreme* is $[0, \frac{1}{4}) \cup [\frac{3}{4}, 1]$, which is not convex.

This example (and more broadly, Corollary 3) challenges the idea, often emphasized in the linguistics literature, that convexity of meaning is a feature of natural languages.

2.5.2 The SSC Property

This paper builds on a key assumption: the SSC property. This condition can be interpreted as a convexity-based consistency requirement on preferences: if players rank two actions identically at two distinct states, their ranking must remain unchanged at any intermediate state. As shown in Proposition 3, the SSC property guarantees that message profiles have convex meaning.

Convexity of message profiles is a necessary condition for individual messages to have convex meaning, as the meaning of a message profile corresponds to the intersection of the meanings of its component messages. If message profiles fail to exhibit convexity, then individual messages cannot be convex either.

Therefore, without the SSC property, neither message profiles nor individual messages exhibit convex meaning. In the single-Sender setting, this implies that neither sentences nor words can be assigned convex meanings. The SSC property thus provides the most favorable environment for convexity of meaning. Yet, Theorem 5 reveals that non-convex meaning is essential for achieving efficient communication.

2.5.3 Concluding remarks

This paper studies how to communicate efficiently in the presence of language constraints. To this end, it considers a cheap-talk model with multiple Senders and a single Receiver. In the absence of language constraints, achieving efficient communication is straightforward: each Sender can truthfully report the state to the Receiver, who then selects the optimal action in every state of the world.

By contrast, under language constraints, efficient communication exhibits two important properties. First, Senders use language heterogeneously. Second, Senders typically assign non-convex meaning to individual messages, while ensuring that the meaning of message profiles remain convex. Thus, efficient communication hinges on a specific and coordinated structure of language use.

The model and analysis also admit an alternative interpretation to the multi-Sender setting. Specifically, a cheap-talk game with a single Sender and a single Receiver, in which the Sender communicates using sentences of finite length. At interim stages, the Sender conveys partial information through partial sentences, which typically have non-convex meaning. However, the meaning of complete sentences is convex. These findings underscore the subtlety involved in structuring sentences efficiently.

Chapter 3

Skill Acquisition under Strategic Advice

This chapter examines strategic communication in environments where a decision-maker can acquire a limited skill set before facing a decision problem. Using a Sender–Receiver model, the analysis focuses on situations in which the Receiver can prepare for at most K actions. Within the class of interval equilibria, the paper shows that the informativeness of the Sender's strategy decreases as the Receiver's skill capacity increases. Despite this tension, a higher Receiver's skill capacity may improve overall welfare, as the Receiver's ability to act effectively in more contingencies may outweigh the informational loss.

3.1 Introduction

Many decision problems require a specific set of skills to be effectively addressed. For example, an investor must possess financial literacy to make optimal investment choices, and a firm needs a market expert to implement effective marketing strategies. Similarly, a law student must acquire legal knowledge to pass the bar examination. In such cases, decision-makers can typically train to develop the necessary skills before encountering the decision problem. However, it is often infeasible to be fully prepared for every possible variation the problem may take, as there may be a large number of scenarios and only limited resources available to prepare for each of them.

In light of this, decision-makers may find it helpful to seek advice from experts. For example, an investor might consult a financial expert to gain insight into future market trends, a firm may turn to a consulting agency to identify the most suitable candidate profile for a market analyst position, and a law student could seek guidance from professors to better prepare for the bar examination. In such cases, expert advice can help guide the decision-maker toward acquiring the most relevant skill set.

This paper examines how an expert communicates with a decision maker via cheap-talk messages. Importantly, the decision maker has the opportunity to acquire a skill set before facing the decision problem. To analyze this interaction, the paper introduces a simple Sender–Receiver model that captures the key features illustrated in the motivating examples.

There is a decision problem. The Receiver's optimal course of action depends on an underlying state of the world. The Sender, however, has a different objective: She is biased in favor of higher actions. The Sender and the Receiver also differ in both the timing at which they learn the state and in their roles: The Sender has an early informational access, while the Receiver holds the decision-making authority. The interaction unfolds as follows. First, the state of the world is realized. The Sender observes the state and sends a message to the Receiver. Upon observing the message, the Receiver faces two sequential decisions. First, he selects a skill set to acquire. A skill set corresponds to a set of actions he will be able to perform. Second, upon choosing a skill set, the Receiver learns the true state and selects an action from within the chosen skill set.

To reflect that it is often infeasible to prepare for all contingencies, the model assumes that there are infinitely many possible states, but the Receiver can only train for a finite number of actions. Formally, the Receiver may choose a skill set consisting of at most K actions, for some $K \in \mathbb{N}$. Refer to K as the Receiver's skill capacity.

The paper focuses on a particular class of equilibria, referred to as *interval equilibria*. An equilibrium is an interval equilibrium if the set of states for which the Sender transmits a given message forms an interval. An interval equilibrium is a *n-step equilibrium* if the Sender transmits n different messages. Within this class, the paper shows a negative relationship between the Receiver's skill capacity and the quality of communication. More precisely, the maximum number of steps an interval equilibrium can have decreases with the Receiver's skill capacity. In other words, the Sender is able to transmit less information to a Receiver with a higher skill capacity.

The intuition behind this result is as follows. A Receiver with greater skill capacity is better prepared to act across a wider range of states and therefore gains less from obtaining precise information. Given the divergence in preferences between the Sender and the Receiver, this reduced informational value makes it more difficult for the Sender to credibly transmit detailed information when the Receiver has a high skill capacity.

Section 3.3 illustrates this relationship through a simple example. It also shows that, although a higher skill capacity hampers communication, the overall effect on welfare can be positive. In particular, a greater skill capacity may be ex ante beneficial for both the Sender and the Receiver, even if less information is transmitted. Intuitively, the Receiver's skill capacity and the information provided by the Sender are complementary in addressing the decision problem: both allow the Receiver to act more effectively across a wider range of contingencies. Thus, a higher skill capacity affects welfare in two opposing ways: it improves welfare by enabling the Receiver to handle more contingencies, but it also reduces welfare due to the loss of information. Section 3.3 illustrates that the former positive effect can outweigh the latter negative one.

This paper contributes to the communication literature in which players may have some form of commitment power (Kolotilin et al., 2013; Deimen and Szalay, 2019; Blume et al., 2022; Lipnowski et al., 2022; Blume and Deimen, 2024). It specifically examines strategic advice for skill acquisition in a cheap talk setting (Crawford and Sobel, 1982b; Green and Stokey, 2007).

The closest paper to this one is Kolotilin et al. (2013), which also studies a Sender–Receiver model in which the Receiver commits to a set of actions. The key difference lies in the timing of this commitment: while they analyze the case where the Receiver commits before any interaction with the Sender, this paper considers a setting where the Receiver commits after observing the Sender's message. This distinction leads to different predictions. In their model, the Receiver faces no constraint on the number of actions he can commit to, and they show that committing to a finite set is optimal. In contrast, the findings in this paper suggest the opposite: the Receiver prefers to commit to as many actions as possible. Moreover, in their setting, commitment facilitates communication, whereas in this paper, greater commitment power hampers communication.

The remainder of the paper is organized as follows. The next section introduces the model. Section 3 illustrates the main result in a simple example. Section 4 presents the results of the paper. Finally, Section 5 discusses and concludes the paper. All proofs are collected in the appendices.

3.2 Model

There are two agents: one called the Sender (S, she) and the second called the Receiver (R, he). The agents have state-dependent preferences over actions. Write $\Theta = [0, 1]$ for the set of states. It is common knowledge amongst the agents that the state is drawn from a uniform distribution. Write F for the CDF of the uniform distribution and f for the associated PDF. The set of actions is $A = \mathbb{R}$. The Sender's utility function is $u_S(a, \theta) = -(a - \theta - b)^2$, where $b > 0$, and the Receiver's utility function is $u_R(a, \theta) = -(a - \theta)^2$. So, the Receiver would like to match the state, but the Sender is biased to choose higher actions.

The agents differ both in the timing of when they learn the true state and in the decision-making authority. First, Nature draws a state θ from Θ. The Sender learns the state but the Receiver does not. The Sender then chooses a message m from a set of messages M with $|M| = |\Theta|$. The Receiver observes the message and chooses a skill set to acquire. The set of skill sets is some $\mathcal{A} \subseteq 2^A$, to be described below. After acquiring a skill set, the Receiver learns the true state and chooses an action a. Note, the Receiver can only choose action a if he acquired the skills for a, i.e., if he chose some $\alpha \in \mathcal{A}$ where $a \in \alpha$.

Observe that the Receiver makes two decisions, based on different information. First, there is a *skill acquisition stage.* At that stage, the Receiver only observes a message sent by the Sender (about the state). The Receiver chooses a skill set, i.e., a set of actions that the Receiver chooses to learn about. The Receiver cannot obtain all skills. Thus, there is some finite $K \geqslant 1$ so that

$$\mathcal{A} = \{\alpha \subseteq A : \alpha \neq \varnothing \text{ and } |\alpha| \leqslant K\}.$$

That is, the Receiver must choose some skill ($\alpha \neq \varnothing$) but the skill set must have K or fewer actions. After the Receiver chooses the skill set, he learns the true state and chooses an action from his skill set. This is the *implementation stage.*

A (pure) strategy for the Sender is a measurable map $\sigma : \Theta \to M$. Refer to σ as a *communication rule.* A (pure) strategy for the Receiver consists of a pair (ρ, d), where

$\rho : M \to \mathcal{A}$ and $d : M \times \mathcal{A} \times \Theta \to A$, where $d(m, \alpha, \theta) \in \alpha$. Refer to ρ as a *skill acquisition rule* and refer to d as an *implementation rule*. Observe that, the communication rule, the skill acquisition rule, and the implementation rule are assumed to be pure.

Write

$$u_R(a, \mu) = \int_\Theta u_R(a, \theta) \mathrm{d}\mu,$$

for the Receiver's expected utility of action a when his belief about the state is $\mu \in \Delta\Theta$.

The solution concept is (pure) interval perfect Bayesian equilibrium.

Definition 7 *A strategy profile* (σ, ρ, d) *is a* perfect Bayesian equilibrium *(*PBE*) if there exists some belief system* $\mu : M \to \Delta\Theta$ *so that the following holds:*

(i) For each $\theta \in \Theta$, $\sigma(\theta) \in \arg\max_{m \in M} u_S(d(m, \rho(m), \theta), \theta)$.

(ii) For each $m \in M$, $\rho(m) \in \arg\max_{\alpha \in \mathcal{A}} u_R(d(m, \alpha, \mu(m)), \mu(m))$.

(iii) For each $(m, \alpha, \theta) \in M \times \mathcal{A} \times \Theta$, $d(m, \alpha, \theta) \in \arg\max_{a \in \alpha} u_R(a, \theta)$.

(iv) For each $(\theta, m) \in \Theta \times M$, $\mu(\theta \mid m) \int_\Theta \mathbb{1}_{\sigma^{-1}(\{m\})}(\theta') \mathrm{d}F(\theta') = \sigma(m \mid \theta) f(\theta)$.

So, the strategy profile (σ, ρ, d) is a PBE if it satisfies four conditions. Condition (i) says that the Sender chooses a message that maximizes her payoffs, given that the Receiver will choose a skill set according to the skill acquisition rule ρ and will then choose an action according to the implementation rule d evaluated at the true state. Condition (ii)-(iii) say that the Receiver will choose an implementation rule that maximizes his payoffs given the known information and, anticipating that, will choose a skill acquisition rule to maximize his expected payoff given the belief induced by the Sender's message. Condition (iv) says that the Receiver's belief about the state conditional upon observing a message m must be derived from the prior and the Sender's strategy via Bayes' rule.

The paper will focus on a particular class of equilibria, called interval equilibria. Note, $\sigma^{-1}(\{m\}) = \{\theta \in \Theta : \sigma(\theta) = m\}$ is the set of states at which the Sender sends message m. A special case will be a communication rule where each $\sigma^{-1}(\{m\})$ is either an interval or empty. In this case, say that σ is an *interval communication rule*. A strategy profile (σ, ρ, d)

is an *interval strategy profile* if σ is an interval communication rule. Likewise, a PBE (σ, ρ, d) is an *interval PBE* if σ is an interval communication rule.

Given a strategy profile (σ, ρ, d), write ψ for the induced outcome, i.e., $\psi : \Theta \to A$ defined by

$$\psi(\theta) = d(\sigma(\theta), \rho(\sigma(\theta)), \theta).$$

The next proposition simplifies the identification of interval PBE outcomes.

Proposition 4 *An interval strategy profile (σ, ρ, d) induces an interval PBE outcome if and only if*

(i) For each $\theta, \theta' \in \Theta$, $u_S(d(\sigma(\theta), \rho(\sigma(\theta)), \theta), \theta) \geqslant u_S(d(\sigma(\theta'), \rho(\sigma(\theta')), \theta), \theta)$.

(ii) For each $m \in \sigma(\Theta)$, $\rho(m) \in \arg\max_{\alpha \in \mathcal{A}} \int_{\sigma^{-1}(m)} u_R(d(m, \alpha, \theta), \theta) d\theta$.

(iii) For each $(m, \alpha, \theta) \in \sigma(\Theta) \times \mathcal{A} \times \Theta$, $d(m, \alpha, \theta) \in \arg\max_{a \in \alpha} u_R(a, \theta)$.

Proposition 4 states that an interval strategy profile induces an interval PBE outcome if it satisfies three conditions. Condition (i) says that the communication rule is incentive compatible for the Sender given the Receiver's strategy. That is, the Sender does not have incentives to misrepresent the state given the communication rule and the Receiver's strategy. Conditions (ii)-(iii) say that the Receiver chooses an implementation rule that maximizes her payoffs given the known information, and anticipating that, chooses a skill acquisition rule that maximizes his expected payoff given the information the on-path messages provide.

Proposition 4 shows that off-path behavior is irrelevant from the perspective of identifying equilibria. To understand the result, consider an interval strategy profile (σ, ρ, d) satisfying conditions (i)-(iii). We can construct an equilibrium that does not change the communication rule but does change the Receiver's strategy. The new Receiver's strategy is so that the skill acquisition rule prescribes some on-path skill set upon observing any off-path message, and the implementation rule prescribes some on-path action upon observing an off-path message. This can be done by having the Receiver update differently conditional upon observing any off-path message. This new belief is chosen so that an on-path skill set constitutes a best response to any off-path message.

3.3 An Example

To see how the Receiver's skill capacity impacts communication, this section compares the interval equilibria of a game in which the Receiver's skill capacity is $K = 1$ with another in which $K = 2$. To that end, assume the preference bias parameter is $b = \frac{1}{16}$. Refer to the game with $K = 1$ as the *low skill capacity* game, and to the game with $K = 2$ as the *high skill capacity* game.

3.3.1 Low skill capacity game

Assume the Receiver's capacity is $K = 1$. Suppose the Receiver chose the skill set $\alpha = \{a\}$ in response to message m. Then, upon observing the state, he will select the unique action a in α. I now proceed to examine the Receiver's skill acquisition behavior. In light of Proposition 4, it suffices to examine the skill sets the Receiver would acquire when the state is uniformly drawn from an interval. More precisely, the Receiver's problem at this stage is:

$$\max_{a \in A} \frac{1}{\overline{\theta} - \underline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} u_R(a, \theta)\, d\theta.$$

The unique solution to this problem is $a^* = E[\theta \mid (\underline{\theta}, \overline{\theta})] = \frac{\underline{\theta}+\overline{\theta}}{2}$.

Let us now analyze the Sender's equilibrium communication rules. To determine the number of steps an interval PBE may have, I examine the relationship between the lengths of two adjacent intervals in the partition $\sigma^{-1}(M)$. Suppose the Sender sends messages m and m' when the state belongs to the intervals (θ_0, θ_1) and (θ_1, θ_2), respectively. Write x and x' for the lengths of these intervals, i.e., $x = \theta_1 - \theta_0$ and $x' = \theta_2 - \theta_1$.

Since the Sender anticipates the Receiver's behavior, she knows that upon observing message m, the Receiver will acquire the skill set $\alpha = \{a\} = \{\theta_1 - \frac{x}{2}\}$ and will then choose action $a = \theta_1 - \frac{x}{2}$. Similarly, upon observing message m', the Receiver will acquire the skill set $\alpha' = \{a'\} = \{\theta_1 + \frac{x'}{2}\}$ and will then choose action $a' = \theta_1 + \frac{x'}{2}$. Therefore, the Sender must be indifferent between these two actions when the state is θ_1.

Note, the Sender prefers actions closer to $\theta_1 + \frac{1}{16}$ at the state θ_1. Thus, actions a and a' must be equidistant from $\theta_1 + \frac{1}{16}$, implying: $\theta_1 + \frac{1}{16} - a = a' - \theta_1 - \frac{1}{16}$. Solving for x' in terms of x, we obtain $x' = x + \frac{1}{4}$. In words, the length of the next interval increases by $\frac{1}{4}$.

As a result, an interval PBE has at most three steps. To see why, suppose the Sender could send at least four distinct messages. Then, the measure of the set of states in which these messages are used must be at least $\frac{1}{4} + \frac{2}{4} + \frac{3}{4} = \frac{3}{2}$. However, this is impossible, as the total measure of the state space is 1. Thus, no minimal interval PBE can sustain more than three messages.

Note that a 3-step equilibrium is characterized by two thresholds θ_1, θ_2 with $0 < \theta_1 < \theta_2 < 1$. The Sender conveys message m when the state is below θ_1, message m' when the state belongs to the interval (θ_1, θ_2), and message m'' when the state is above θ_2. Given the previous analysis, it must hold that $\theta_1 + (\theta_1 + \frac{1}{4}) + (\theta_1 + \frac{1}{2}) = 1$. Solving for θ_1, we obtain $\theta_1 = \frac{1}{12}$. Consequently, $\theta_2 = \frac{5}{12}$. Upon observing message m, the Receiver acquires the skill set $\alpha = \{\frac{1}{24}\}$; upon observing message m', the Receiver acquires the skill set $\alpha' = \{\frac{1}{4}\}$; and upon observing message m'', the Receiver acquires the skill set $\alpha'' = \{\frac{17}{24}\}$. The next figure illustrates the 3-step equilibrium outcome.

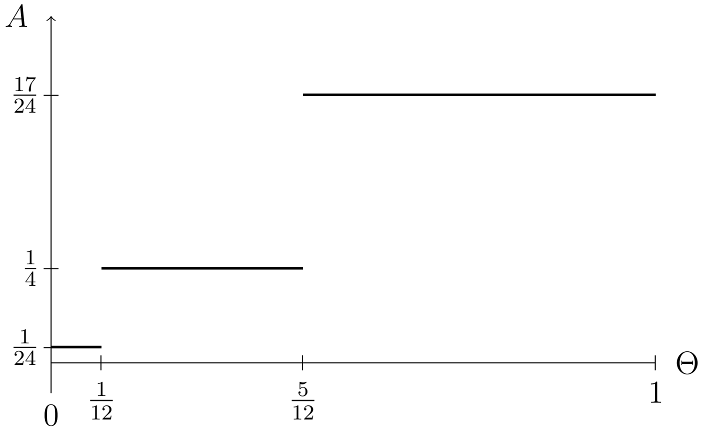


Figure 3.1. Maximum-step equilibrium outcome (Case $K = 1$)

3.3.2 High skill capacity game

Assume the Receiver's skill capacity is $K = 2$. For ease of exposition, I will focus on skills $\alpha \in \mathcal{A}$ with $|\alpha| = 2$ throughout this section.[1]

First, consider the Receiver's implementation rule. Suppose the Receiver chose the skill set $\alpha = \{a_1, a_2\}$ in response to message m. Upon observing the state, the Receiver will choose the action in α that is closest to the state. That is, the Receiver selects a_1 if the state is below $\frac{a_1+a_2}{2}$, and a_2 otherwise.

Now, I proceed to examine the Receiver's skill acquisition behavior. In light of Proposition 4, it suffices to examine the skill sets the Receiver would acquire assuming that the state is uniformly drawn from an interval. More precisely, the Receiver's problem at this stage is:

$$\max_{a_1,a_2} \frac{1}{\overline{\theta}-\underline{\theta}} \int_{\underline{\theta}}^{\frac{a_1+a_2}{2}} u_R(a_1,\theta)\,d\theta + \frac{1}{\overline{\theta}-\underline{\theta}} \int_{\frac{a_1+a_2}{2}}^{\overline{\theta}} u_R(a_2,\theta)\,d\theta.$$

The unique solution to the Receiver's problem is:

$$a_1^* = \frac{3\underline{\theta}+\overline{\theta}}{4} \quad \text{and} \quad a_2^* = \frac{\underline{\theta}+3\overline{\theta}}{4}.$$

This solution has two important features. First, the cutoff that determines the state at which the Receiver switches from action a_1^* to a_2^* is $\frac{\underline{\theta}+\overline{\theta}}{2}$. That is, the Receiver partitions the interval $(\underline{\theta}, \overline{\theta})$ into two equal subintervals, each associated with one of the two actions. Second, each a_j^* is equal to the optimal action conditional on the corresponding subinterval.

Let us now analyze the Sender's equilibrium communication rules. As in the low skill capacity game, I examine the relationship between the lengths of two adjacent intervals in the partition $\sigma^{-1}(M)$ to determine the number of steps that an interval PBE may have. Suppose the Sender sends messages m and m' when the state belongs to the intervals (θ_0, θ_1) and (θ_1, θ_2), respectively. Let x and x' denote the lengths of these intervals, i.e., $x = \theta_1 - \theta_0$ and $x' = \theta_2 - \theta_1$.

---

[1] Lemma 7 shows that the Receiver exhausts his skill capacity in every interval PBE. Thus, this restriction does not affect the analysis.

Since the Sender anticipates the Receiver's behavior, she knows that upon observing the messages m and m', the Receiver will acquire the skill sets $\alpha = \{a_1, a_2\} = \{\theta_0 + \frac{1}{4}x, \theta_1 - \frac{1}{4}x\}$ and $\alpha' = \{a'_1, a'_2\} = \{\theta_1 + \frac{1}{4}x', \theta_2 - \frac{1}{4}x'\}$, respectively. Note, when the state is close to θ_1, the Receiver will choose action a_2 given (m, α), and will choose action a'_1 given (m', α'). Therefore, the Sender must be indifferent between these two actions when the state is θ_1.

Note, the Sender prefers actions closer to $\theta_1 + \frac{1}{16}$ at the state θ_1. Thus, actions a_2 and a'_1 must be equidistant from $\theta_1 + \frac{1}{16}$, implying: $\theta_1 + \frac{1}{16} - a_2 = a'_1 - \theta_1 - \frac{1}{16}$. Solving for x' in terms of x, we obtain $x' = x + \frac{1}{2}$. In words, the length of the next interval increases by $\frac{1}{2}$.

As a result, an interval PBE can have at most two steps. To see why, suppose the Sender could send at least three distinct messages. Then, the measure of the set of states in which these messages are used must be at least $\frac{1}{2} + 1 = \frac{3}{2}$. However, this is impossible, as the total measure of the state space is 1. Thus, no interval PBE can sustain more than two messages.

Note that in a 2-step equilibrium, the Sender conveys message m when the state is below a threshold θ_1, and message m' otherwise. Based on the previous analysis, it must hold that $\theta_1 + (\theta_1 + \frac{1}{2}) = 1$. Solving for θ_1, we obtain $\theta_1 = \frac{1}{4}$. Upon observing message m, the Receiver acquires the skill set $\alpha = \{\frac{1}{16}, \frac{3}{16}\}$, while upon observing message m', the Receiver acquires the skill set $\alpha' = \{\frac{7}{16}, \frac{13}{16}\}$. Given skill set α, the Receiver selects action $\frac{1}{16}$ when $\theta < \frac{1}{8}$, and $\frac{3}{16}$ otherwise. Given skill set α', the Receiver chooses action $\frac{7}{16}$ when $\theta < \frac{5}{8}$, and $\frac{13}{16}$ otherwise. The next figure illustrates the 2-step equilibrium outcome.

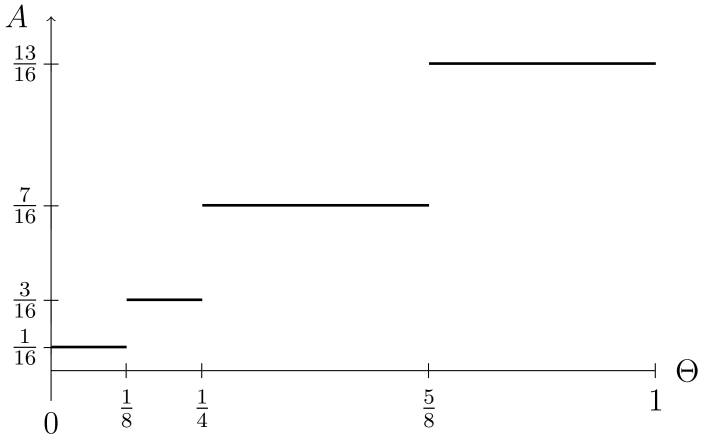


Figure 3.2. Maximum-step equilibrium outcome (Case $K = 2$)

3.3.3 Comparison

Observe that the maximum number of steps in an interval equilibrium is higher in the low skill capacity game than in the high skill capacity game. This illustrates a negative relationship between the Receiver's skill capacity and the quality of communication.

However, despite this reduction in communication quality, the ex-ante expected utilities of both the Sender and the Receiver improve as the skill capacity increases. In the low skill capacity game, the ex-ante expected utilities of the Sender and the Receiver are approximately -0.023 and -0.019, respectively. In the high skill capacity game, these expected utilities increase to -0.013 and -0.009, representing roughly a 50% improvement. Hence, although higher skill capacity reduces the Sender's ability to transmit information, it simultaneously improves welfare by enabling the Receiver to implement actions that are better tailored to the realized state. This example suggests that the positive welfare effect of higher skill capacity may dominate the negative welfare effect of reduced information transmission.

3.4 Equilibrium Analysis

3.4.1 Receiver's equilibrium behavior

The Receiver makes two sequential decisions, each based on different information. First, upon observing the message sent by the Sender, he chooses a skill set. Second, after observing the realized state, he selects an action from the chosen skill set. In equilibrium, the Receiver anticipates how he will select actions from skill sets when deciding which skill set to acquire. Hence, it is necessary to first characterize the Receiver's equilibrium implementation rules.

To understand what an equilibrium implementation rule entails, suppose the Receiver chose the skill set α in response to message m, and then observes the realized state θ. The Receiver's best response is to choose the action $a \in \alpha$ that is closest to θ. To formalize this, define the set $\Theta(a \mid \alpha) \subseteq \Theta$ for each $a \in A$ and $\alpha \in \mathcal{A}$ as follows:

$$\Theta(a \mid \alpha) = \bigcap_{a' \in \alpha, a' \neq a} \{\theta \in \Theta \mid u_R(a, \theta) \geqslant u_R(a', \theta)\}$$

In words, $\Theta(a \mid \alpha)$ is the set of states for which action a yields at least as high a payoff as any other action in α. Consequently, if the Receiver selects action $a \in \alpha$ given the skill set α and the state θ, then necessarily $\theta \in \Theta(a \mid \alpha)$.

Lemma 6 *In a PBE, (σ, ρ, d), the Receiver's implementation rule satisfies the following: For each $(m, \alpha, \theta) \in M \times \mathcal{A} \times \Theta$ and $a \in \alpha$,*

(i) If $d(m, \alpha, \theta) = a$, then $\theta \in \Theta(a \mid \alpha)$.

(ii) If $\theta \in \operatorname{int} \Theta(a \mid \alpha)$, then $d(m, \alpha, \theta) = a$.

Lemma 6 characterizes the Receiver's equilibrium implementation rule. Item(i) states that $\theta \in \Theta(a \mid \alpha)$ is a necessary condition for choosing action a given message m, skill set α, and state θ. However, this condition alone is not sufficient, as indifference may arise when θ is equidistant to two different actions a and a' in α, making it optimal to choose either. By contrast, Item (ii) states that when $\theta \in \operatorname{int} \Theta(a \mid \alpha)$, action a is strictly preferred over every other action in α, and hence uniquely optimal given m, α, and θ.

Now, we proceed to analyze the Receiver's skill acquisition behavior. The Receiver chooses which skill set to acquire after observing the message transmitted by the Sender. By Proposition 4, it suffices to examine which skill sets the Receiver would choose upon learning that the state belongs to a particular interval. Moreover, the Receiver anticipates that he will subsequently select an action optimally from the chosen skill set.

Assuming the state is uniformly drawn from the interval $(\underline{\theta}, \overline{\theta})$, the Receiver's problem is

$$\max_{\alpha \in \mathcal{A}} \frac{1}{\overline{\theta} - \underline{\theta}} \sum_{a \in \alpha} \int_{\Theta(a|\alpha) \cap (\overline{\theta}, \underline{\theta})} u_R(a, \theta) d\theta$$

In solving this problem, the Receiver balances the following trade-off: (i) Each $\Theta(a \mid \alpha) \cap (\underline{\theta}, \overline{\theta})$ should have approximately equal size because the state is uniformly distributed; and (ii) each action $a \in \alpha$ should approximate the expected utility-maximizing action conditional on $\Theta(a \mid \alpha)$. That is, the Receiver aims to set a as close as possible to $E[\theta \mid \Theta(a \mid \alpha)]$. The next lemma characterizes the equilibrium skill acquisition rules.

Lemma 7 *In an interval PBE* (σ, ρ, d)*, the Receiver's skill acquisition rule satisfies the following: For each* $m \in \sigma(\Theta)$*,*

$$\rho(m) = \left\{ \frac{2j-1}{2K}(\overline{\theta} - \underline{\theta}) + \underline{\theta} \,\middle|\, j = 1, \ldots, K \right\}$$

where $\underline{\theta} = \inf \sigma^{-1}(m)$ *and* $\overline{\theta} = \sup \sigma^{-1}(m)$*.*

Lemma 7 characterizes the Receiver's skill acquisition rule in a minimal interval PBE and yields two insights. First, the Receiver always exhausts his skill capacity. To see this, consider a skill set α with $|\alpha| < K$. Fix an action $a \in \Theta \backslash \alpha$, and define $\alpha' = \alpha \cup \{a\}$. Note that, at each state, the Receiver obtains a weakly larger utility given skill set α' than with α. Moreover, for each state $\theta \in \operatorname{int} \Theta(a \mid \alpha')$, he obtains strictly higher utility with α' than with α. Since $\operatorname{int} \Theta(a \mid \alpha')$ has positive measure (because $a \notin \alpha$), it follows that α cannot be prescribed by an equilibrium skill acquisition rule.

Second, the Receiver's choice of skill sets balances the trade-off outlined above precisely: For each $a \in \rho(m)$, $\Theta(a \mid \rho(m))$ has measure $\frac{\overline{\theta} - \underline{\theta}}{K}$, and a coincides with the expected state conditional on $\Theta(a \mid \rho(m))$.

3.4.2 Sender's equilibrium behavior

The Sender chooses a message after observing the state. Note that, in equilibrium, the Sender anticipates the Receiver's behavior. First, she anticipates which skill set the Receiver will acquire in response to the message (see Lemma 7). Second, she anticipates which action from the skill set the Receiver will implement upon learning the state (see Lemma 6).

In light of Proposition 4, the interval strategy profile (σ, ρ, d) induces an interval equilibrium outcome if the Receiver's strategy (ρ, d) satisfies Lemmas 6–7, and the communication rule σ is incentive compatible for the Sender. The following notation is useful in characterizing incentive-compatible communication rules. Say that the interval communication rule σ has n-steps if $|\sigma(\Theta)| = n$. That is, the number of steps represents the number of messages prescribed by σ. In addition, for a communication rule with n-steps, let

$$\sigma^{-1}(M) = \big\{ (\theta_0, \theta_1), (\theta_1, \theta_2), \ldots, (\theta_{n-1}, \theta_n) \big\}$$

where $0 = \theta_0 < \theta_1 < \ldots < \theta_{n-1} < \theta_n = 1$. Define $I_j = (\theta_{j-1}, \theta_j)$ and $|I_j| = \theta_j - \theta_{j-1}$ for each $j = 1, \ldots, n$.

The next proposition characterizes the interval strategy profiles that induce an interval PBE outcome.

Proposition 5 *The interval strategy profile (σ, ρ, d) induces an interval PBE outcome if and only if $|I_{j+1}| = |I_j| + 4bK$ for each $j = 1, \ldots, n-1$, and (ρ, d) satisfies Lemmas 6–7.*

Proposition 5 states that an interval communication rule σ is incentive compatible if and only if the sizes of the intervals characterizing σ satisfy a monotonic property. Precisely, the size of each subsequent interval increases by $4bK$. To understand this result, consider an interval PBE (σ, ρ, d). Suppose the Sender sends messages m and m' when the state belongs to I_j and I_{j+1}, respectively. Then the Sender must be indifferent between $\max \rho(m) = \theta_j - \frac{|I_j|}{2K}$ and $\min \rho(m') = \theta_j + \frac{|I_{j+1}|}{2K}$ at the boundary state θ_j. This means that:

$$\theta_j + b - \left(\theta_j - \frac{|I_j|}{2K}\right) = \left(\theta_j + \frac{|I_{j+1}|}{2K}\right) - \theta_j - b.$$

Solving yields:

$$|I_{j+1}| = |I_j| + 4bK.$$

Corollary 4 *There exists an n-step interval PBE (σ, ρ, d) if and only if $(n-1)n < \frac{1}{2bK}$*

Corollary 4 reveals that the maximum number of steps that can be supported in equilibrium is decreasing in the bias b and the Receiver's skill capacity K. Whereas the dependence on b is well-known from Crawford and Sobel (1982b), the dependence on K is a new insight in the literature. Thus, Corollary 4 can be interpreted as revealing a negative relationship between the quality of communication and the Receiver's skill capacity.

3.5 Conclusions

This paper studies strategic communication in environments where a decision-maker can acquire a skill set before facing a decision problem. Importantly, the Receiver faces an economic constraint: He may choose a skill set consisting of at most K actions, for some $K \in$

$\mathbb{N}$. The main result shows that as the Receiver's skill capacity increases, the informativeness of the Sender's strategy decreases. This occurs because a more flexible Receiver benefits less from precise communication, making it harder for a biased Sender to credibly convey useful information.

Despite this tension, a simple example in the paper illustrates that a higher skill capacity can improve overall welfare. The Receiver's ability to prepare for a broader range of contingencies can outweigh the informational loss, leading to better outcomes for both parties ex ante. These findings reveal a fundamental trade-off between skill and information acquisition: while better training enhances a decision-maker's capability, it may simultaneously diminish the effectiveness of expert advice. An important direction for future research is to explore how general this welfare effect is when the Receiver's constraint is relaxed across different communication environments.

Appendix A

Chapter 1

A.1 Proofs of Section 2.2

Proof of Proposition 1. The statement directly follows from Theorems 2 and 3. ■

A.2 Proofs of Section 1.4

Proof of Lemma 1. Assume that at the beginning of period t, the agents' beliefs are (μ_b^t, μ_s^t) with $\mu_s^t = g(\mu_b^t)$. Fix an experiment π^t. Let $\tau_i \in \mathrm{PS}[\mu_i]$ be the associated agent i's posterior spread. Write $M[\mu_i^{t-1} \mid \mu_i^t] \subseteq M$ for the set of messages that induce posterior μ_i^{t-1} on i given that i has prior μ_i^t. Then, notice that, for each i

$$\tau_i(\mu_i^{t-1}) = \sum_{m \in M[\mu_i^{t-1}|\mu_i^t]} \sum_{\theta \in \Theta} \pi^t(m \mid \theta)\mu_i^t(\theta).$$

Moreover, for each posterior belief μ_b^{t-1}, $M[g(\mu_b^{t-1})|\mu_b^t] = M[\mu_s^{t-1}|\mu_s^t]$. Notice, Bayes rule states that for each $m \in M[\mu_i^{t-1} \mid \mu_i^t]$,

$$\pi^t(m \mid \theta)\mu_i^t(\theta) = \mu_i^{t-1}(\theta) \left(\sum_{\theta \in \Theta} \pi^t(m \mid \theta)\mu_i^t(\theta) \right)$$

In addition, recall that

$$\mu_s^{t-1}(\theta) = g(\mu_b^{t-1})(\theta) = \frac{r(\theta)\mu_b^{t-1}(\theta)}{r \cdot \mu_b^{t-1}}. \tag{A.1}$$

Notice, since $\operatorname{Supp}\mu_s^t = \operatorname{Supp}\mu_b^t$, it follows that

$$\begin{aligned}
\tau_s(\mu_s^{t-1}) &= \sum_{m \in M[\mu_s^{t-1}|\mu_s^t]} \sum_{\theta \in \operatorname{Supp}\mu_s^t} \pi^t(m \mid \theta)\mu_s^t(\theta) \\
&= \sum_{m \in M[\mu_s^{t-1}|\mu_s^t]} \sum_{\theta \in \operatorname{Supp}\mu_s^t} \left(\frac{\mu_b^{t-1}(\theta)}{\mu_b^t(\theta)} \sum_{\theta' \in \Theta} \pi^t(m \mid \theta')\mu_b^t(\theta') \right) \mu_s^t(\theta) \\
&= \sum_{\theta \in \operatorname{Supp}\mu_s^t} \frac{\mu_b^{t-1}(\theta)\mu_s^t(\theta)}{\mu_b^t(\theta)} \left(\sum_{m \in M[\mu_s^{t-1}|\mu_s^t]} \sum_{\theta' \in \Theta} \pi^t(m \mid \theta')\mu_b^t(\theta') \right) \\
&= \sum_{\theta \in \operatorname{Supp}\mu_s^t} \frac{\mu_b^{t-1}(\theta)\mu_s^t(\theta)}{\mu_b^t(\theta)} \tau_b(\mu_b^{t-1}) \\
&= \sum_{\theta \in \operatorname{Supp}\mu_s^t} \frac{r(\theta)\mu_b^{t-1}(\theta)}{r \cdot \mu_b^t} \tau_b(\mu_b^{t-1}) \\
&= \frac{r \cdot \mu_b^{t-1}}{r \cdot \mu_b^t} \tau_b(\mu_b^{t-1}),
\end{aligned}$$

where the fourth equality follows from Equation (A.1). ■

Proof of Lemma 2. Fix an action $a \in A$. Notice that the function $u_a : \Delta\Theta \to \mathbb{R}$ defined by $u_a(\mu) = \sum_{\theta \in \Theta} u(a, \theta)\mu(\theta)$ is linear, which implies that it is convex and continuous. Hence, observe that $U(\mu) = \max_{a \in A} u_a(\mu)$, so U is convex and continuous. ■

Proof of Theorem 1. First, we show existence. Let $E^1 = (\pi^1, p^1)$ be a priced experiment that fully reveals the state at price $p^1 = V^1(\nu_b) = \overline{U} \cdot \nu_b - U(\nu_b)$. Write σ for the seller's strategy that selects E^1 at the root of the game. Write $c : \mathcal{E} \to \Delta\{\text{accept}, \text{reject}\}$ for the buyer's strategy such that satisfies the following: for each priced experiment $E' = (\pi', p') \in \mathcal{E}$,

$$c(E')(\text{accept}) = \begin{cases} 1 & \text{if } p' \leqslant \mathbb{E}_{\tau'}[U(\mu_b^0)] - U(\nu_b) \\ 0 & \text{otherwise,} \end{cases}$$

where $\tau' \in \mathrm{PS}[\nu_b]$ is the posterior spread induced by π'. Finally, write α for the strategy profile such that prescribes an optimal action given the buyer's beliefs. That is, for each history $h^0 \in \mathcal{H}^0$,

$$\operatorname{Supp}(\alpha(h^0)) \subseteq \arg\max_{a \in A} \mathbb{E}_{\mu_b^0}[u(a,\theta)], \tag{A.2}$$

where μ_b^0 is the buyer's belief at history h^0. Notice, by construction, under the profile $(\sigma, (c, \alpha))$ the buyer and the seller have no profitable deviation at any history, so it is an SPE. Moreover, note that $(\sigma, (c, \alpha))$ satisfies properties (i)-(iv).

Now, we show that each SPE satisfies properties (ii)-(iv). Fix an SPE $(\sigma, (c, \alpha))$. Notice, each history $h^0 \in \mathcal{H}^0$ must satisfy Equation (A.2). Hence, after any history, the buyer's value from accepting an experiment $E = (\pi, p) \in \mathcal{E}$ is $\mathbb{E}_\tau[U(\mu_b^0)] - U(\nu_b)$, where $\tau \in \mathrm{PS}[\nu_b]$ is the posterior spread induced by π. Consequently, it must be that $c(E) = 1$ if $p < \mathbb{E}_\tau[U(\mu_b^0)] - U(\nu_b)$, and $c(E) = 0$ if $p > \mathbb{E}_\tau[U(\mu_b^0)] - U(\nu_b)$.

In the case where $p = \mathbb{E}_\tau[U(\mu_b^0)] - U(\nu_b)$, the buyer is indifferent between accepting and rejecting, so any randomization $c(E) \in \Delta\{\text{accept}, \text{reject}\}$ is optimal. We will show that the buyer must accept at least one fully revealing experiment at a price $V^1(\nu_b)$ with probability one. Suppose, by way of contradiction, that he rejects with positive probability all such priced experiments. Consequently, the seller cannot achieve an expected revenue of $V^1(\nu_b)$. However, he can achieve any strictly lower payoff since the buyer would accept with probability one any fully revealing experiment at price p, for any $p \in \mathbb{R}_+$ with $p < V^1(\nu_b)$. Therefore, the seller has no optimal choice, which contradicts that the strategy profile $(\sigma, (c, \alpha))$ is a SPE.

As a result, the buyer must accept at least one fully revealing experiment at a price $V^1(\nu_b)$ with probability one. Since such a priced experiment yields the highest possible expected revenue, then σ must prescribe choosing one of such experiments. In conclusion, any SPE satisfies properties (ii)-(iv). ■

Lemma 8 *Fix a mapping $f : \Delta\Theta \times \Delta\Theta \to \mathbb{R}$ and let $F : \Delta\Theta \to \mathbb{R}$ be defined by $F(\mu) = \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_{\mu' \sim \tau}[f(\mu', \mu)]$. If f is continuous, then F is continuous.*

Proof. Fix $\mu \in \Delta\Theta$. We show that F is continuous at μ. Fix a sequence $(\mu_k)_{k \in \mathbb{N}}$ such that $\mu_k \in \Delta\Theta$ and $\lim \mu_k = \mu$. We show $\lim_{k \to \infty} F(\mu_k) = F(\mu)$. We divide the proof into two steps. Step one shows that $\limsup_{k\to\infty} V(\mu_k) \leqslant V(\mu)$ and step two shows that $\liminf_{k\to\infty} V(\mu_k) \leqslant V(\mu)$.

Step 1. Notice, since f is continuous, there is some $\tau \in \mathrm{PS}[\mu]$ such that $\mathbb{E}_\tau[f(\mu', \mu)] = F(\mu)$ (See Kamenica and Gentzkow (2011b).) Moreover, there exist an affine mapping $L : \Delta\Theta \to \mathbb{R}$ such that

(i) $L(\mu) = \mathbb{E}_\tau[f(\mu', \mu)] = F(\mu)$.

(ii) $L(\mu') \geqslant f(\mu', \mu)$ for each $\mu' \in \Delta(\Theta)$.

Notice, since the sets $\Delta\Theta$ and $\Delta\Theta \times \Delta\Theta$ is compact, the mappings L and f are uniformly continuous. Hence, there is some $\delta > 0$ such that $||\mu - \mu_k||_\infty < \delta$ implies that for each $\mu' \in \Delta\Theta$, $|f(\mu', \mu_k) - f(\mu', \mu)| < \frac{\varepsilon}{2}$ and $|L(\mu_k) - L(\mu)| < \frac{\varepsilon}{2}$. So,

$$\begin{aligned} F(\mu_k) &= \sup_{\tau' \in \mathrm{PS}[\mu_k]} \mathbb{E}_{\mu' \sim \tau'}[f(\mu', \mu_k)] \\ &\leqslant \sup_{\tau' \in \mathrm{PS}[\mu_k]} \mathbb{E}_{\mu' \sim \tau'}[L(\mu') + \tfrac{\varepsilon}{2}] \\ &= L(\mu_k) + \tfrac{\varepsilon}{2} \\ &< L(\mu) + \varepsilon \\ &= F(\mu) + \varepsilon. \end{aligned}$$

Note, since $\varepsilon > 0$ is arbitrary and $\lim_{k\to\infty} \mu_k = \mu$, it follows that $\limsup_{k\to\infty} F(\mu_k) \leqslant F(\mu)$.

Step 2. Write $\tau \in \mathrm{PS}[\mu]$ for the posterior spread that satisfies $F(\mu) = \mathbb{E}_{\mu' \sim \tau}[f(\mu', \mu)]$. Note, by Kamenica and Gentzkow (2011b), there is some finite message space M with $|M| \leqslant |\Theta|$ and some experiment $\pi : \Theta \to \Delta(M)$ such that π induces τ. Let $\tau_k \in \mathrm{PS}[\mu_k]$ be the posterior spread induced by π under prior belief μ_k.

For each $m \in M$ and each prior belief μ' write $\mathbb{P}_\pi[m \mid \mu']$ for the probability of m under prior belief $\mu' \in \Delta\Theta$. Likewise, write $\mathcal{P}_m(\mu') \in \Delta\Theta$ for the posterior belief induced by a message $m \in M$ and prior belief $\mu' \in \Delta\Theta$. Notice that $\mathcal{P}_m(\mu')$ and $\mathbb{P}_\pi[m \mid \mu']$ are continuous at μ. Hence, for each $m \in M$, $\lim_{k\to\infty} \mathbb{P}_\pi[m \mid \mu_k] = \mathbb{P}_\pi[m \mid \mu]$ and $\lim_{k\to\infty} \mathcal{P}_m(\mu_k) = \mathcal{P}_m(\mu)$.

Moreover, since f is continuous,

$$\lim_{k\to\infty} \mathbb{P}_\pi[m \mid \mu_k] f(\mathcal{P}_m(\mu_k), \mu_k) = \mathbb{P}_\pi[m \mid \mu] f(\mathcal{P}_m(\mu), \mu).$$

Therefore,

$$\begin{aligned} \lim_{k\to\infty} \mathbb{E}_{\mu'\sim\tau_k}[f(\mu', \mu_k)] &= \lim_{k\to\infty} \sum_{\mu'\in \mathrm{Supp}\,(\tau_k)} \tau_k(\mu') f(\mu', \mu_k) \\ &= \lim_{k\to\infty} \sum_{m\in M} \mathbb{P}_\pi[m \mid \mu_k] f(\mathcal{P}_m(\mu_k), \mu_k) \\ &= \sum_{m\in M} \mathbb{P}_\pi[m \mid \mu] f(\mathcal{P}_m(\mu), \mu) \\ &= F(\mu). \end{aligned}$$

Finally, notice that for each $k \in \mathbb{N}$,

$$F(\mu_k) = \sup_{\tau'\in \mathrm{PS}[\mu_k]} \mathbb{E}_{\mu'\sim\tau'}[f(\mu', \mu_k)] \geqslant \mathbb{E}_{\mu'\sim\tau_k}[f(\mu', \mu_k)]$$

Thus, it follows that $\liminf_{k\to\infty} F(\mu_k) \geqslant F(\mu)$, as desired. ■

Proof of Lemma 3. Fix $t > 1$ and $\mu_b^t \in \Delta\Theta$. Lemma 1 ensures that for any posterior spread $\tau_b \in \mathrm{PS}[\mu_b^t]$ the associated seller's beliefs about the buyer's posterior μ_b^{t-1} are given by $\tilde{\tau}_b(\mu_b^{t-1}) = \tau_b(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t)$. As a result,

$$\mathbb{E}_{\tilde{\tau}_b}[V^{t-1}(\mu_b^{t-1})] = \sum_{\mu_b^{t-1}\in \mathrm{Supp}\,\tilde{\tau}_b} V^1(\mu_b^{t-1})\tau(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t) = \mathbb{E}_{\tau_b}[V^1(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t)].$$

Therefore,

$$\begin{aligned} \mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t) + \mathbb{E}_{\tilde{\tau}_b}[V^{t-1}(\mu_b^{t-1})] &= \mathbb{E}_{\tau_b}[U(\mu_b^{t-1}) + V^1(\mu_b^{t-1})\rho(\mu_b^{t-1} \mid \mu_b^t)] - U(\mu_b^t) \\ &= \mathbb{E}_{\tau_b}[\Lambda^t(\mu_b^{t-1} \mid \mu_b^t)] - U(\mu_b^t). \end{aligned}$$

From here we conclude that

$$V^t(\mu_b^t) = \sup_{\tau_b \in \mathrm{PS}[\mu_b^t]} \mathbb{E}_{\tau_b}[\Lambda^t(\mu_b^{t-1} \mid \mu_b^t)] - U(\mu_b^t).$$

In addition, notice that condition (i) follows from Proposition 9 in the working paper version of Kamenica and Gentzkow (2011b). As for condition (ii), observe that Lemmas 8 and 2 ensure that $\sup_{\tau_b \in \mathrm{PS}[\mu_b^t]} \mathbb{E}_{\tau_b}[\Lambda^t(\mu_b^{t-1} \mid \mu_b^t)]$ and $U(\mu_b^t)$ are continuous on μ_b^t. Hence, $V^t(\mu_b^t)$ is a continuous mapping.

Finally, to prove condition (iii), we first show that $V^{t+1} \geqslant V^t$ by arguing that a seller with $t+1$ periods can offer a non-informative experiment in period $t+1$ and then behave optimally from period t on. Second, we show that $V^t(\delta_\theta) = 0$ for each $t \in \mathbb{N}$ and each $\theta \in \Theta$ through an inductive argument.

First, fix $\mu \in \Delta\Theta$ and $t \geqslant 1$. Notice that $\Lambda^{t+1}(\mu \mid \mu) = U(\mu) + V^t(\mu)\rho(\mu \mid \mu) = U(\mu) + V^t(\mu)$. In addition, observe that $\delta_\mu \in \mathrm{PS}[\mu]$. As a result,

$$\begin{aligned} V^{t+1}(\mu) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[\Lambda^{t+1}(\mu' \mid \mu)] - U(\mu) \\ &\geqslant \mathbb{E}_{\delta_\mu}[\Lambda^{t+1}(\mu' \mid \mu)] - U(\mu) \\ &= \Lambda^{t+1}(\mu \mid \mu) - U(\mu) \\ &= V^t(\mu). \end{aligned}$$

Second, fix $\theta \in \Theta$. We will show that $V^1(\delta_\theta) = 0$ for each $t \in \mathbb{N}$. We proceed by induction. Notice that $V^1(\delta_\theta) = \overline{U} \cdot \delta_\theta - U(\delta_\theta) = 0$. Now, fix $t \geqslant 1$. Assume that $V^t(\delta_\theta) = 0$. Since δ_θ is an extreme point of $\Delta\Theta$, it cannot be written as a non-trivial convex combination of elements of $\Delta\Theta$. Hence, $\mathrm{PS}[\delta_\theta] = \{\delta_\theta\}$. As a result,

$$\begin{aligned} V^{t+1}(\delta_\theta) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[\Lambda^{t+1}(\mu' \mid \mu)] - U(\mu) \\ &= \Lambda^t(\delta_\theta \mid \delta_\theta)] - U(\delta_\theta) \\ &= V^t(\delta_\theta) \\ &= 0. \end{aligned}$$

In conclusion, $V^t(\delta_\theta) = 0$ for each $t \in \mathbb{N}$. ■

Proof of Theorem 2. We divide the proof into two steps. Step 1 shows existence of a SPE that satisfies conditions (i)-(iii). Step 2 shows that each SPE satisfies conditions (i)-(iii).

Step 1. We proceed by induction on T. Notice that the base case ($T = 1$) is shown in Theorem 1.

Now we show the inductive step. Fix $T \geqslant 1$. Assume that, for every pair of agents' beliefs (μ_b^T, μ_s^T) with $\mu_s^T = g(\mu_b^T)$ at the beginning of period T, there exists an SPE $(\sigma', (c', \alpha'))$ as described in Theorem 2 in the game with T periods. We will use these strategies to construct the SPE in the game of $T + 1$ periods. For any seller's (non-initial) history at period $T + 1 > t \geqslant 1$, $h_s^t = \{(E^{t'}, c^{t'}, m^{t'}(c^{t'}))\}_{t'=T+1}^{t+1}$, write $\hat{h}_s^t = \{(E^{t'}, c^{t'}, m^{t'}(c^{t'}))\}_{t'=T}^{t+1}$ for the associated pruned seller's history, which describes the history of play in the subgame that starts after $\{\varnothing, (E^{T+1}, c^{T+1}, m^{T+1}(c^{T+1})\}$. Similarly, for any buyer's history h_b^t, write $\hat{h}_b^t$ for the associated pruned buyer's history that describes the history of play in the subgame that starts after $\{\varnothing, (E^{T+1}, c^{T+1}, m^{T+1}(c^{T+1}), E_T\}$. Last, for every trading history h^0, write $\hat{h}^0$ for the associated pruned trading history that describes the history of play in the subgame that starts after $\{\varnothing, (E^{T+1}, c^{T+1}, m^{T+1}(c^{T+1})\}$.

Now, let $\tilde{\pi}^{T+1}$ be the experiment inducing an optimal posterior spread $\tilde{\tau}$ such that (1) $|\text{Supp}\,(\tilde{\tau})| \leqslant |\Theta|$ and (2) $V^{t+1}(\nu_b) = \mathbb{E}_\tau[\Lambda^{t+1}(\mu' \mid \nu_b)] - U(\nu_b)$. Set $\tilde{p}^{T+1} = \mathbb{E}_{\tilde{\tau}}[U(\mu_b^T)] - U(\nu_b)$, and $\tilde{E}^{T+1} = (\tilde{\pi}^{T+1}, \tilde{p}^{T+1})$. Consider the seller's strategy given by:

$$\sigma(h_s^t) = \begin{cases} \tilde{E}^{T+1} & \text{if } t = T + 1 \\ \sigma'(\hat{h}_s^t) & \text{if } t \leqslant T. \end{cases}$$

The buyer's acceptance rule defined by:

$$c(h_b^t)(\text{accept}) = \begin{cases} 1 & \text{if } h_b^t = \{(\pi^{T+1\prime}, p^{T+1\prime})\} \text{ and } p^{T+1\prime} \leqslant \mathbb{E}_{\tau'}[U(\mu_b^T)] - U(\nu_b) \\ 0 & \text{if } h_b^t = \{(\pi^{T+1\prime}, p^{T+1\prime})\} \text{ and } p^{T+1\prime} > \mathbb{E}_{\tau'}[U(\mu_b^T)] - U(\nu_b) \\ c'(\hat{h}_b^t) & \text{if } t \leqslant T. \end{cases}$$

where $c'(\hat{h}_b^t)$ is the acceptance rule in the SPE associated with the initial buyer's belief induced by the subgame that starts after $\{\varnothing, (E^{T+1}, c^{T+1}, m^{T+1}(c^{T+1})\}$. Likewise, the buyer's final action rule given by $\alpha(h^0) = \alpha'(\hat{h}^0)$ where $\alpha'(\hat{h}^0)$ is the final choice rule in the SPE associated to the initial buyer's belief induced by the subgame that starts after $\{\varnothing, (E^{T+1}, c^{T+1}, m^{T+1}(c^{T+1})\}$.

To show that $(\sigma, (c, \alpha))$ is an SPE, we appeal to the one-shot deviation principle. Since $(\sigma', (c', \alpha'))$ is an SPE for any game of T periods, then no single-deviation is optimal at any history after period T.

Now, a buyer's history at period $T+1$ is characterized by an initial priced experiment (π^{T+1}, p^{T+1}) offered. In the case that $p^{T+1} \leqslant \mathbb{E}_\tau[U(\mu_b^T)] - U(\nu_b)$, a buyer's deviation would imply a positive probability of rejection, which implies that the buyer's continuation payoffs will be a convex combination of $U(\nu_b)$ and $\mathbb{E}_\tau[U(\mu_b^T)] - p^{T+1}$, which will be weakly smaller than the payoff from sticking to c, $\mathbb{E}_\tau[U(\mu_b^T)] - p^{T+1}$. In the case that $p^{T+1} > \mathbb{E}_\tau[U(\mu_b^T)] - U(\nu_b)$, a buyer's deviation would imply a positive probability of accepting the experiment, which implies that the buyer's continuation payoffs will be a convex combination between $U(\nu_b)$ and $\mathbb{E}_\tau[U(\mu_b^T)] - p^{T+1}$, which will be weakly smaller than the payoff from sticking to c, $U(\nu_b)$.

Finally, a one-shot deviation for the seller at the root of the game would imply choosing other priced experiment than $\tilde{E}^{T+1}$. A deviation to the priced experiment (π^{T+1}, p^{T+1}) would imply a seller's expected revenue of either $p^{T+1} + \mathbb{E}_{\tilde{\tau}}[V^T(\mu_b^T)] \leqslant \mathbb{E}_\tau[U(\mu_b^T)] - U(\nu_b) + \mathbb{E}_{\tilde{\tau}}[V^T(\mu_b^T)]$ or $V^T(\nu_b^T)$, which, either way, is lower than $V^{T+1}(\nu_b) = \sup_{\tau_b \in \mathrm{PS}[\nu_b]} \mathbb{E}_{\tau_b}[U(\mu_b^T)] - U(\nu_b) + \mathbb{E}_{\tilde{\tau}_b}[V^T(\mu_b^T)]$. Hence, the seller has no incentives to deviate once at the root and then conform back to σ.

Step 2. Fix a SPE $(\sigma, (c, \alpha))$. To see that $(\sigma, (c, \alpha))$ satisfies condition (i), first notice that in any SPE, it must be that $\operatorname{Supp} \alpha(h^0) \subseteq \arg\max_{a \in A} \mathbb{E}_{\mu_b^0}[u(a, \theta)] = \arg\max_{a \in A} U(\mu_0^b)$ where μ_b^0 is the buyer's belief at history h^0. Therefore, at any trading history $h_b^t = \{h_s^t, E^t\} \in \mathcal{H}_b$ in which the buyer's initial belief is μ_b^t, the buyer's value of accepting the priced experiment $E^t = (\pi^t, p^t)$ is $\mathbb{E}_\tau[U(\mu_b^{t-1})] - U(\mu_b^t)$, where τ is the posterior spread induced by π^t. As a result, it must be that

$$c(h_b^t)(\text{accept}) = \begin{cases} 1 & \text{if } p^t < \mathbb{E}_\tau[U(\mu_b^{t-1})] - U(\mu_b^t) \\ 0 & \text{if } p^t > \mathbb{E}_\tau[U(\mu_b^{t-1})] - U(\mu_b^t) \end{cases}.$$

In the case that $p^t = \mathbb{E}_\tau[U(\mu_b^{t-1})] - U(\mu_b^t)$, the receiver is indifferent between accepting and rejecting the offer, so any randomization is optimal. We will show that the buyer must accept with probability one at least one experiment inducing the posterior spread τ_b that defines $V^t(\mu_b^t)$ at a price $p^t = \mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t)$. Suppose, by contradiction, that the buyer rejects all such priced experiments with positive probability. Hence, the seller cannot achieve an expected revenue stream of $V^t(\mu_b^t)$. However, he can achieve any strictly lower payoff stream since the buyer would accept with probability one any priced experiment inducing posterior spread τ_b at price p, for any $p \in \mathbb{R}$ with $p < \mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t)$. Therefore, the seller has no optimal choice, which contradicts that the strategy profile $(\sigma, (c, \alpha))$ is a SPE.

As a result, the buyer must accept at least one such priced experiment with probability one. Since this experiment yields the highest possible expected revenue stream, then σ must prescribe choosing one of such experiments. In conclusion, any SPE satisfies property (i). Notice that this implies that the seller's expected revenue stream at any given history h_s^t is given by $V^t(\mu_b^t)$ where μ_b^t is the buyer's initial belief at period t in such a history. In particular, the seller's ex-ante expected revenue is $V^T(\nu_b)$, so property (ii) holds. In addition, observe that at any on-path buyer's history h_b^t with initial buyer's belief μ_b^t, we have that the buyer accepts the seller's offered experiment at a price $p^t = \mathbb{E}_{\tau_b}[U(\mu_b^{t-1})] - U(\mu_b^t)$. This implies that the buyer's continuation payoffs at such a history are $U(\mu_b^t)$. In particular, the ex-ante buyer's expected payoff is $U(\nu_b)$. In other words, any SPE satisfies property (iii). ■

A.3 Proofs of Section 1.5

Lemma 9 *Assume that* $V \in \mathcal{F}$. *Then, the mapping* $\phi(V) : \Delta\Theta \to \mathbb{R}$ *is continuous and satisfies* $\phi(V)(\delta_\theta) = 0$ *for each* $\theta \in \Theta$.

Proof. Notice that $V \in \mathcal{F}$ implies V is continuous. Hence, the mapping defined by $f(\mu', \mu) = U(\mu') + V(\mu')\rho(\mu' \mid \mu)$ is continuous. Thus, the mapping $\phi(V)(\cdot) : \Delta\Theta \to \mathbb{R}$ is continuous. (See Lemma 8.)

Fix $\theta \in \Theta$. Notice $\tau \in \mathrm{PS}[\delta_\theta]$ if and only if $\mathrm{Supp}\,(\tau) = \{\delta_\theta\}$. Hence

$$\phi(V)(\delta_\theta) = \mathbb{E}_\tau[U(\mu') + V(\mu')\rho(\mu' \mid \delta_\theta)] = U(\delta_\theta) + V(\delta_\theta) - U(\delta_\theta) = 0.$$

Therefore, $\phi(V) \in \mathcal{F}$. ■

Lemma 10 *Fix a pair of mappings $V, W \in \mathcal{F}$ such that $W(\mu) \geqslant V(\mu)$ for each $\mu \in \Delta\Theta$, and write $\mathcal{S} = \{\mu \in \Delta\Theta : V(\mu) = W(\mu)\}$. The following properties hold:*

(i) For each $\mu \in \Delta\Theta$, $\phi(V)(\mu) \geqslant V(\mu)$.

(ii) For each $\mu \in \Delta\Theta$, $\phi(W)(\mu) \geqslant \phi(V)(\mu)$.

(iii) If $\phi(W)(\mu) = \phi(V)(\mu)$, then there is some $\tau \in \mathrm{PS}[\mu]$ with $\mathrm{Supp}\,(\tau) \subseteq \mathcal{S}$ such that

$$\phi(V)(\mu) = \mathbb{E}_\tau[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] - U(\mu) = \phi(W)(\mu) \tag{A.3}$$

Proof. We first show part (i). Fix $\mu \in \Delta\Theta$ and $V \in \mathcal{F}$. Since $\delta_\mu \in \mathrm{PS}[\mu]$, then

$$\begin{aligned}\phi(V)(\mu) &\geqslant \mathbb{E}_{\delta_\mu}[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] - U(\mu)\\ &= U(\mu) + V(\mu)\rho(\mu \mid \mu) - U(\mu)\\ &= V(\mu).\end{aligned}$$

We now show part (ii). Fix $\mu \in \Delta\Theta$ and notice that

$$\begin{aligned}\phi(W)(\mu) &= \sup_{\tau' \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau'}[U(\mu') + W(\mu')\rho(\mu' \mid \mu)] - U(\mu)\\ &\geqslant \sup_{\tau' \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau'}[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] - U(\mu)\\ &= \phi(V)(\mu).\end{aligned}$$

Now we show part (iii). Assume that $\phi(W)(\mu) = \phi(V)(\mu)$. Notice, since $\phi(V)$ is continuous, then $U(\mu') + \phi(V)(\mu')\rho(\mu' \mid \mu)$ is continuous in μ'. Thus, there is some $\tau \in \mathrm{PS}[\mu]$ with finite support such that

$$\phi(V)(\mu) = \max_{\tau' \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau'}[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] = \mathbb{E}_{\tau}[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] \tag{A.4}$$

Thus,

$$\begin{aligned}\phi(V)(\mu) &= \mathbb{E}_{\tau}[U(\mu') + V(\mu')\rho(\mu' \mid \mu)] \\ &\leqslant \mathbb{E}_{\tau}[U(\mu') + W(\mu')\rho(\mu' \mid \mu)] \\ &= \max_{\tau' \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau'}[U(\mu') + W(\mu')\rho(\mu' \mid \mu)] \\ &\leqslant \phi(W)(\mu) \\ &= \phi(V)(\mu),\end{aligned}$$

which implies Equation (A.3). Moreover, if $\phi(V)(\mu') > V(\mu')$ for some $\mu' \in \operatorname{Supp}(\tau)$, then we obtain $\phi^2(V)(\mu) > \phi(V)(\mu)$, a contradiction. Thus, $\phi(V)(\mu') = V(\mu')$ for each $\mu' \in \operatorname{Supp}\tau$, as desired. ■

Lemma 11 *Fix $\mu \in \Delta\Theta$. The following statements are equivalent:*

(i) $g(\mu) = \mu$.

(ii) $r(\theta) = r(\theta')$ *for each* $\theta, \theta' \in \operatorname{Supp}\mu$.

(iii) $\rho(\mu' \mid \mu) = 1$ *for each* $\mu' \in \Delta\Theta$ *with* $\operatorname{Supp}\mu' \subseteq \operatorname{Supp}\mu$.

(iv) $g(\mu') = \mu'$ *for all* $\mu' \in \Delta\Theta$ *with* $\operatorname{Supp}\mu' \subseteq \operatorname{Supp}\mu$.

Proof. Fix $\mu \in \Delta\Theta$. We first prove that condition (i) is equivalent to condition (ii). Assume that $g(\mu) = \mu$. Let $\theta \in \operatorname{Supp}\mu$, then $g(\mu)(\theta) = \frac{r(\theta)\mu(\theta)}{r \cdot \mu} = \mu(\theta) > 0$. This implies that $r(\theta) = r \cdot \mu$, which is independent of θ. Conversely, assume that for each $\theta \in \operatorname{Supp}\mu$, $r(\theta) = c$ is constant. Then, $g(\mu)(\theta) = \frac{r(\theta)\mu(\theta)}{r \cdot \mu} = \frac{c\mu(\theta)}{c} = \mu(\theta)$.

Now, we prove that condition (ii) implies conditions (iv) and (iii). Assume that $r(\theta) = c$ is constant for all $\theta \in \operatorname{Supp}\mu$. Fix $\mu' \in \Delta\operatorname{Supp}\mu$. Then, $\operatorname{Supp}\mu' \subset \operatorname{Supp}\mu$. Therefore, for each $\theta \in \operatorname{Supp}\mu' \subset \operatorname{Supp}\mu$, we conclude that $r(\theta) = c$. Hence, $g(\mu') = \mu'$ given the equivalence between conditions (i) and (ii). Moreover, $\rho(\mu' \mid \mu) = \frac{r \cdot \mu'}{r \cdot \mu} = \frac{c}{c} = 1$.

Next, we show that condition (iii) implies condition (ii). Assume that $\rho(\mu' \mid \mu) = 1$ for each $\mu' \in \Delta\Theta$ with $\operatorname{Supp}\mu' \subset \operatorname{Supp}\mu$. Then, for any $\theta \in \operatorname{Supp}\mu$, we have that $1 = \rho(\delta_\theta \mid \mu) = \frac{r(\theta)}{r \cdot \mu}$. Thus, $r(\theta) = r \cdot \mu$ is constant as it is independent of θ.

Finally, notice that condition (iv) implies condition (i) since $\operatorname{Supp}\mu \subset \operatorname{Supp}\mu$, so $g(\mu) = \mu$.

■

Lemma 12 *Fix $\mu \in \Delta\Theta$. The following statements are equivalent:*

(i) $V^1(\mu) = 0$.

(ii) $\arg\max_{a \in A} \mathbb{E}_\mu[u(a, \theta)] \subseteq \arg\max_{a \in A} u(a, \theta)$ *for each* $\theta \in \operatorname{Supp}\mu$.

(iii) $V^1(\mu') = 0$ *for every* $\mu' \in \Delta\Theta$ *with* $\operatorname{Supp}\mu' \subseteq \operatorname{Supp}\mu$.

Proof. Fix $\mu \in \Delta\Theta$. We first prove that parts (i) and (ii) are equivalent. First, suppose that $V^1(\mu) = \overline{U} \cdot \mu - U(\mu) = 0$. Fix $\theta \in \operatorname{Supp}\mu$ and $a' \in A$. Assume, by contradiction, that $a' \in \arg\max_{a \in A} \mathbb{E}_\mu[u(a, \theta)]$ but $a' \notin \arg\max_{a \in A} u(a, \theta)$ for some $\theta \in \operatorname{Supp}\mu$. Hence, $U(\delta_\theta) > U(\mu)$. Since $U(\delta_{\theta'}) \geqslant U(\mu)$, then $\overline{U} \cdot \mu > U(\mu)$, a contradiction.

Conversely, assume that $\arg\max_{a \in A} \mathbb{E}_\mu[u(a, \theta)] \subset \arg\max_{a \in A} u(a, \theta)$ for each $\theta \in \operatorname{Supp}\mu$. Hence, $U(\mu) = U(\delta_\theta)$ for all $\theta \in \operatorname{Supp}\mu$. Finally, $\overline{U} \cdot \mu = U(\mu)$ which is to say $V^1(\mu) = 0$.

Now, we prove that part (ii) implies part (iii). Assume that for each $\theta \in \operatorname{Supp}\mu$ the following is satisfied: $\arg\max_{a \in A} \mathbb{E}_\mu[u(a, \theta)] \subset \arg\max_{a \in A} u(a, \theta)$. Fix $\mu' \in \Delta\Theta$ with $\operatorname{Supp}\mu' \subset \operatorname{Supp}\mu$, and $a' \in \arg\max_{a \in A} \mathbb{E}_\mu[u(a, \theta)]$. Therefore, $u(a', \theta) = U(\delta_\theta)$ for each $\theta \in \operatorname{Supp}\mu$. As a result, $\overline{U} \cdot \mu' \geqslant U(\mu') \geqslant \mathbb{E}_{\mu'}[u(a', \theta)] = \mathbb{E}_{\mu'}[U(\delta_\theta)] = \overline{U} \cdot \mu'$. In conclusion, $V^1(\mu') = 0$.

Last, we prove that part (iii) implies part (i). Note, since $\operatorname{Supp}\mu' \subseteq \operatorname{Supp}\mu$, then $V^1(\mu) = 0$. ■

Proof of Lemma 4. First assume that $\nu_s = \nu_b$. This implies that $g(\mu_b) = \mu_b$ for each μ_b. Thus $\mathcal{D}^+ = \varnothing$. Now assume that $\nu_s = g(\nu_b) \neq \nu_b$ and fix $\mu_b \in \operatorname{int}\Delta\theta$. Notice, by Lemma 11, $g(\mu_b) \neq \mu_b$ for each $\mu_b \in \operatorname{int}\Delta\Theta$. Notice, since the buyer's decision problem is not trivial, there are some $\theta, \theta' \in \Theta$ such that $\arg\max_{a \in A} u(a, \theta) \cap \arg\max_{a \in A} u(a, \theta') = \varnothing$. Since $\operatorname{Supp}(\mu_b) = \Theta$, it follows that $V^1(\mu_b) > 0$. (See Lemma 12.) Therefore, $\operatorname{int}\Theta \subseteq \mathcal{D}^+$.

■

Lemma 13 *Fix $\mu \in \Delta\Theta$. Then, $\mu \in \mathcal{D}^+$ if and only if $V^2(\mu) > V^1(\mu)$.*

Proof. Let $\mu \in \Delta\Theta$. First, assume that $\mu \in \mathcal{D}^+$, which means that $\mu \neq g(\mu)$ and $V^1(\mu) > 0$. We first show that there is some $\tau \in \mathrm{PS}[\mu]$ such that $\mathbb{E}_\tau[\Lambda^2(\mu' \mid \mu)] > \mathbb{E}_\tau[\overline{U} \cdot \mu']$. To show this, we will show the following:

(i) There is some $\mu' \in \operatorname{int}(\Delta\operatorname{Supp}\mu)$, such that $\Lambda^2(\mu' \mid \mu) > \overline{U} \cdot \mu'$.

(ii) For each $\theta \in \Theta$ it follows that $\Lambda^2(\delta_\theta \mid \mu) = \overline{U} \cdot \delta_\theta$.

So, if $\tau \in \mathrm{PS}[\mu]$ is such that $\operatorname{Supp}\tau = \{\mu'\} \cup \{\delta_\theta : \theta \in \Theta\}$, then

$$\begin{aligned} V^2(\mu) &= \sup_{\tau' \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau'}[\Lambda^2(\mu' \mid \mu)] - U(\mu) \\ &\geqslant \mathbb{E}_\tau[\Lambda^2(\mu' \mid \mu)] - U(\mu) \\ &> \overline{U} \cdot \mu - U(\mu) \\ &= V^1(\mu), \end{aligned}$$

as desired.

To show condition (i), notice that $\Delta\operatorname{Supp}\mu \not\subset H_{\rho=1}(\mu)$. (See Lemma 11.) Moreover, for every $\theta \in \Theta$ for which $g(\mu)(\theta) > \mu(\theta)$, it holds that $\delta_\theta \in H_{\rho>1}$. Hence, by linearity of ρ , it follows that $H_{\rho>1} \cap \Delta\operatorname{Supp}\mu$ is non-empty and open relative to $\Delta\operatorname{Supp}\mu$. Fix $\mu' \in H_{\rho>1} \cap \Delta\operatorname{Supp}\mu$. Notice, that $V^1(\mu') > 0$. (See Lemma 12.) Thus,

$$\Lambda^2(\mu' \mid \mu) = U(\mu') + V^1(\mu')\rho(\mu' \mid \mu) > U(\mu') + V^1(\mu') = \overline{U} \cdot \mu'.$$

To show condition (ii), fix $\theta \in \Theta$ and notice that

$$\Lambda^2(\delta_\theta \mid \mu) = U(\delta_\theta) + V^1(\delta_\theta)\rho(\delta_\theta \mid \mu) = U(\delta_\theta) = \overline{U} \cdot \delta_\theta.$$

Now we show the converse. Assume, by contrapositive, that $\mu \notin \mathcal{D}^+$. So, either $V^1(\mu) = 0$ or $g(\mu) = \mu$. First consider the case $g(\mu) = \mu$. Lemma 11 shows that for for each $\tau \in \mathrm{PS}[\mu]$

and $\mu' \in \operatorname{Supp}\tau$, it follows that $\rho(\mu' \mid \mu) = 1$. Hence,

$$\begin{aligned} V^2(\mu) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[U(\mu') + V^1(\mu')\rho(\mu' \mid \mu)] - U(\mu) \\ &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[U(\mu') + V^1(\mu)] - U(\mu) \\ &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[\overline{U} \cdot \mu'] - U(\mu_b) \\ &= V^1(\mu). \end{aligned}$$

Consider the case $V^1(\mu) = 0$. Then, Lemma 12 shows that, for each $\tau \in \mathrm{PS}[\mu]$ and each $\mu' \in \operatorname{Supp}\tau$ it follows that $V^1(\mu') = 0$. Consequently,

$$\begin{aligned} V^2(\mu) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[U(\mu') + V^1(\mu')\rho(\mu' \mid \mu)] - U(\mu) \\ &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau[U(\mu')] - U(\mu) \\ &= V^1(\mu). \end{aligned}$$

Therefore, $\mu \notin \mathcal{D}^+$ implies $V^2(\mu) = V^1(\mu)$ as desired. ■

Proof of Theorem 3. We proceed by induction on $t \in \mathbb{N}$. The base case $t = 1$ follows directly from Lemma 13. Fix $t > 1$ and assume that for each $t \geqslant t' \geqslant 1$.

(i) $V^{t'+1}(\mu') = V^{t'}(\mu')$ for each $\mu' \notin \mathcal{D}^+$, and

(ii) $V^{t'+1}(\mu') > V^{t'}(\mu')$ for each $\mu' \in \mathcal{D}^+$.

We will show that these two statements hold for $t + 1$. Notice that $V^{t+2}(\mu') \geqslant V^{t+1}(\mu')$ for each $\mu' \in \Delta\Theta$ (See Lemma 3). Thus, it suffices to show that that $V^{t+2}(\mu') > V^{t+1}(\mu')$ if and only if $\mu' \in \mathcal{D}^+$.

First consider the case $\mu \notin \mathcal{D}^+$. Note that $\operatorname{Supp}(\tau) \cap \mathcal{D}^+ = \varnothing$ for each $\tau \in \mathrm{PS}[\mu]$. (See Lemmata 11 and 12). Therefore, for each $\tau \in \mathrm{PS}[\mu]$ and each $\mu' \in \operatorname{Supp}(\tau)$, $V^{t+1}(\mu') =$

$V^t(\mu')$. This implies that

$$\begin{aligned} V^{t+2}(\mu) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau\big[U(\mu') + V^{t+1}(\mu')\rho(\mu' \mid \mu)\big] - U(\mu) \\ &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau\big[U(\mu') + V^{t}(\mu')\rho(\mu' \mid \mu)\big] - U(\mu) \\ &= V^{t+1}(\mu), \end{aligned}$$

as desired.

Now consider the case $\mu \in \mathcal{D}^+$. We show that $V^{t+2}(\mu) > V^{t+1}(\mu)$ by contradiction. Suppose $\phi(V^{t+1})(\mu) = V^{t+2}(\mu) = V^{t+1}(\mu) = \phi(V^t)(\mu)$. Then, by Lemma 10, there is some $\tau \in \mathrm{PS}[\mu]$ such that

$$V^{t+1}(\mu) = \mathbb{E}_\tau\big[U(\mu') + V^t(\mu')\rho(\mu' \mid \mu)\big] - U(\mu) = V^{t+2}(\mu),$$

and $\mathrm{Supp}\,(\tau) \subseteq \big\{\mu' : V^{t+1}(\mu') = V^t(\mu')\big\}$. Moreover, by conditions (i) and (ii), it follows that

$$\mathrm{Supp}\,(\tau) \subseteq \Delta\Theta \backslash \mathcal{D}^+ = \big\{\mu' : V^t(\mu') = V^{t-1}(\mu')\big\}.$$

Therefore,

$$\begin{aligned} V^{t+1}(\mu) &= \mathbb{E}_\tau[U(\mu') + V^t(\mu')\rho(\mu' \mid \mu)] - U(\mu) \\ &= \mathbb{E}_\tau[U(\mu') + V^{t-1}(\mu')\rho(\mu' \mid \mu)] - U(\mu) \\ &\leqslant \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_\tau\big[U(\mu') + V^{t-1}(\mu')\rho(\mu' \mid \mu)] - U(\mu) \\ &= V^t(\mu). \end{aligned}$$

Which contradicts condition (ii). Therefore, we conclude that $V^{t+2}(\mu) > V^{t+1}(\mu)$, as desired.

■

Lemma 14 *Let* $\hat{\theta} \in \arg\max_{\theta \in \Theta}\{r(\theta)\}$. *Then,* $\rho(\delta_{\hat{\theta}} \mid \mu_b) \geqslant 1$ *for each* $\mu_b \in \Delta(\Theta)$.

Proof. Note that, by definition of ρ, for each $\mu_b \in \Delta\Theta$, $\rho(\delta_{\hat{\theta}} \mid \mu_b)\rho(\mu_b \mid \delta_{\hat{\theta}}) = 1$. Hence, it suffices to show that $\rho(\mu_b \mid \delta_{\hat{\theta}}) \leqslant 1$ for each $\mu_b \in \Delta\Theta$. Notice, for each $\theta \in \Theta$, $\rho(\delta_\theta \mid \delta_{\hat{\theta}}) =$

$\frac{r(\theta)}{r(\hat{\theta})} \leqslant 1$. Thus, since $\rho(\mu_b \mid \delta_{\hat{\theta}})$ is linear in μ_b, if follows that $\rho(\mu_b \mid \delta_{\hat{\theta}}) \leqslant 1$ for each $\mu_b \in \Delta\Theta$. ■

Proof of Lemma 5. We first show part (i). Fix $\mu_b \in \Delta(\Theta)$. First we prove that for the mapping

$$\Lambda_B(\mu_b' \mid \mu_b) := U(\mu_b') + B(\mu_b') \cdot \rho(\mu_b' \mid \mu_b)$$

is weakly concave in μ_b'. To show this, note that

$$\begin{aligned}
\Lambda_B(\mu_b' \mid \mu_b) &= U(\mu_b') + B(\mu_b') \cdot \rho(\mu_b' \mid \mu_b) \\
&= U(\mu_b') + V^1(\mu_b') \cdot \rho(\delta_{\hat{\theta}} \mid \mu_b') \cdot \rho(\mu_b' \mid \mu_b) \\
&= U(\mu_b') + V^1(\mu_b') \cdot \rho(\delta_{\hat{\theta}} \mid \mu_b) \\
&= U(\mu_b') + \left(\lambda \sum_{\theta} \mu_b'(\theta) \cdot U(\delta_\theta) - U(\mu_b')\right) \cdot \rho(\delta_{\hat{\theta}} \mid \mu_b) \\
&= -U(\mu_b')\left(\rho(\delta_{\hat{\theta}} \mid \mu_b) - 1\right) + \rho(\delta_{\hat{\theta}} \mid \mu_b) \sum_{\theta \in \Theta} \mu_b'(\theta) \cdot U(\delta_\theta).
\end{aligned}$$

In addition, note that $\rho(\delta_{\hat{\theta}} \mid \mu_b) \geqslant 1$. (See Lemma 14.) Hence, for each $\mu_b \in \Delta(\Theta)$, the mapping $\Lambda_B(\cdot \mid \mu_B)$ is the sum of a weakly concave function and a linear function. Hence, it is weakly concave. Thus, for each μ_b,

$$\sup_{\tau \in \mathrm{PS}[\mu_b]} \mathbb{E}_\tau[\Lambda_B(\mu_b' \mid \mu_b)] = \Lambda_B(\mu_b \mid \mu_b).$$

We now show that B is a fixed point of ϕ. Fix $\mu_b \in \Delta\Theta$ and note that

$$\begin{aligned}
\phi(B)(\mu_b) &= \sup_{\tau \in \mathrm{PS}[\mu_b]} \mathbb{E}_\tau[\Lambda_B(\mu_b' \mid \mu_b)] - U(\mu_b) \\
&= \Lambda_B(\mu_b \mid \mu_b) - U(\mu_b) \\
&= B(\mu_b) \cdot \rho(\mu_b \mid \mu_b) \\
&= B(\mu_b),
\end{aligned}$$

where the third equality equation follows from definition of Λ_B. This shows that B is a fixed point of ϕ.

We now show part (ii). Fix $\mu_b \in \Delta(\Theta)$ and notice that $\rho(\delta_{\hat{\theta}} \mid \mu_b) \geqslant 1$. (See Lemma 14). Hence, $B(\mu_b) = V^1(\mu_b) \cdot \rho(\delta_{\hat{\theta}} \mid \mu_b) \geqslant V^1(\mu_b)$, as desired. ■

Proof of Theorem 4. First we show that $V^t(\mu) \leqslant B(\mu)$ for each $t \in \mathbb{N}$ and each belief $\mu \in \Delta(\Theta)$. We proceed by induction. Notice that $V^1(\cdot) \leqslant B(\cdot)$ (See Lemma 5). Now, assume that $V^t(\cdot) \leqslant B(\cdot)$ for $k \geqslant 1$. Therefore, by monotonicity of ϕ, for each $\mu \in \Delta\Theta$, it follows that

$$V^{k+1}(\mu) = \phi(V^t)(\mu) \leqslant \phi(B)(\mu) = B(\mu). \tag{A.5}$$

(See Lemma 10.)

This implies that $(V^t(\mu))_{t\in\mathbb{N}}$ is a bounded increasing sequence and hence it has a limit. Write $V^\infty(\mu) := \lim_{t\to\infty} V^t(\mu)$.

We first will show that V^∞ satisfies the following: for each $\mu_b \in \Delta\Theta$,

$$V^\infty(\mu) = \sup_{\tau\in \mathrm{PS}[\mu]} \mathbb{E}_\tau[U(\mu') + V^\infty(\mu') \cdot \rho(\mu' \mid \mu)] - U(\mu). \tag{A.6}$$

To show this, first fix $\tau \in \mathrm{PS}[\mu]$ Notice, since $V^1 \leqslant V^\infty$, it follows that

$$\mathbb{E}_\tau[U(\mu') + V^t(\mu') \cdot \rho(\mu' \mid \mu)] \leqslant \mathbb{E}_\tau[U(\mu') + V^\infty(\mu') \cdot \rho(\mu' \mid \mu)]. \tag{A.7}$$

In addition, notice that for each $t \in \mathbb{N}$, the mapping $U(\mu') + V^t(\mu') \cdot \rho(\mu' \mid \mu)$ is bounded by the continuous mapping $U(\mu') + B(\mu') \cdot \rho(\mu' \mid \mu)$. (See Equation A.5.) Then, by the Dominated Convergence Theorem,

$$\mathbb{E}_\tau[U(\mu') + \lim_{t\to\infty} V^t(\mu') \cdot \rho(\mu' \mid \mu)] = \lim_{t\to\infty} \mathbb{E}_\tau[U(\mu') + V^t(\mu') \cdot \rho(\mu' \mid \mu)] \tag{A.8}$$

Finally, notice that, by definition of V^{t+1}.

$$\mathbb{E}_\tau[U(\mu') + V^t(\mu') \cdot \rho(\mu' \mid \mu)] - U(\mu) \leqslant V^{t+1}(\mu) \tag{A.9}$$

Therefore,

$$\begin{aligned}
V^{\infty}(\mu) &= \lim_{t\to\infty} V^{t+1}(\mu) \\
&= \lim_{t\to\infty} \sup_{\tau\in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[U(\mu') + V^{t}(\mu')\cdot \rho(\mu' \mid \mu)] - U(\mu) \\
&\leqslant \sup_{\tau\in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[U(\mu') + V^{\infty}(\mu')\cdot \rho(\mu' \mid \mu)] - U(\mu) \\
&= \sup_{\tau\in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[U(\mu') + \lim_{t\to\infty} V^{t}(\mu')\cdot \rho(\mu' \mid \mu)] - U(\mu) \\
&= \sup_{\tau\in \mathrm{PS}[\mu]} \lim_{t\to\infty} \mathbb{E}_{\tau}[U(\mu') + V^{t}(\mu')\cdot \rho(\mu' \mid \mu)] - U(\mu) \\
&\leqslant \sup_{\tau\in \mathrm{PS}[\mu]} \lim_{t\to\infty} V^{t+1}(\mu) \\
&= V^{\infty}(\mu),
\end{aligned}$$

where the first inequality follows from Equation(A.7), the fourth equality follows from Equation (A.8), and the last inequality follows from Equation (A.9). This shows Equation (A.6).

Notice that $V^{\infty}(\delta_{\theta}) = 0$ for each $\theta \in \Theta$. Thus, to show that $V^{\infty} \in \mathcal{F}$ is suffices to show that V^{∞} is continuous.

Fix $\mu \in \Delta\Theta$ and let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence such that $\mu_k \in \Delta\Theta$ and $\lim \mu_k = \mu$. We show $\lim V^{\infty}(\mu_k) = V^{\infty}(\mu)$. We divide the proof into two steps. Step one shows that $\limsup_{k\to\infty} V^{\infty}(\mu_k) \leqslant V(\mu)$ and step two shows that $\liminf_{k\to\infty} V^{\infty}(\mu_k) \leqslant V^{\infty}(\mu)$.

Step 1. Let $f : \Delta\Theta \times \Delta\Theta \to \mathbb{R}$ given by $f(\mu', \mu) = U(\mu') + V^{\infty}(\mu')\rho(\mu' \mid \mu)$. Notice, by Equation (A.6)

$$V^{\infty}(\mu) = \sup_{\tau\in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[f(\mu', \mu)] - U(\mu)$$

Notice that f is bounded since $U(\mu') + V^{\infty}(\mu')\rho(\mu' \mid \mu) \leqslant U(\mu') + B(\mu')\rho(\mu' \mid \mu)$. Thus, there exist an affine mapping $L : \Delta\Theta \to \mathbb{R}$ such that

(i) $L(\mu') \geqslant f(\mu', \mu)$ for each $\mu' \in \Delta(\Theta)$.

(ii) $L(\mu) = \sup_{\tau} \mathbb{E}_{\tau}[f(\mu', \mu)]$.

Let $M > 0$ be a bound of $V^{\infty}(\cdot)$. Since the set $\Delta\Theta \times \Delta\Theta$ is compact, the mapping ρ is uniformly continuous. Hence, there is some $\delta > 0$ such that $||\mu - \mu_k||_{\infty} < \delta$ implies that for each $\mu' \in \Delta\Theta$,

$$\rho(\mu' \mid \mu_k) < \rho(\mu' \mid \mu) + \tfrac{\varepsilon}{2M},$$

Then, for each $\mu' \in \Delta\Theta$,

$$V^{\infty}(\mu')\rho(\mu' \mid \mu_k) < V^{\infty}(\mu')\rho(\mu' \mid \mu) + \tfrac{\varepsilon}{2}.$$

Thus,

$$\begin{aligned} f(\mu', \mu_k) &= U(\mu') + V^{\infty}(\mu')\rho(\mu' \mid \mu_k) \\ &< U(\mu') + V^{\infty}(\mu')\rho(\mu' \mid \mu) + \tfrac{\epsilon}{2} \\ &= f(\mu', \mu) + \tfrac{\epsilon}{2} \\ &\leqslant L(\mu') + \tfrac{\epsilon}{2}. \end{aligned}$$

Consequently,

$$\begin{aligned} V^{\infty}(\mu_k) &= \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[f(\mu', \mu_k)] - U(\mu_k) \\ &\leqslant \sup_{\tau \in \mathrm{PS}[\mu]} \mathbb{E}_{\tau}[L(\mu') + \tfrac{\varepsilon}{2}] - U(\mu) + \tfrac{\varepsilon}{2} \\ &= L(\mu) - U(\mu) + \varepsilon \\ &= \sup_{\tau} \mathbb{E}_{\tau}[f(\mu', \mu)] - U(\mu) + \varepsilon \\ &= V^{\infty}(\mu) + \varepsilon. \end{aligned}$$

This implies that $\limsup_{k\to\infty} V^{\infty}(\mu_k) \leqslant V^{\infty}(\mu) + \varepsilon$. Moreover, since $\varepsilon > 0$ is arbitrary it follows that $\limsup_{k\to\infty} V^{\infty}(\mu_k) \leqslant V^{\infty}(\mu)$.

Step 2. Fix $\varepsilon > 0$. Notice, there is some $K \in \mathbb{N}$ such that $V^{K}(\mu) > V^{\infty}(\mu) + \frac{\varepsilon}{2}$. Moreover, there is some $\delta > 0$ such that $||\mu_k - \mu||_{\infty} < \delta$ implies $V^{K}(\mu_k) > V^{K}(\mu) + \frac{\varepsilon}{2}$. Therefore, if $||\mu_k - \mu||_{\infty} < \delta$, then

$$V^{\infty}(\mu_k) \geqslant V^{K}(\mu_k) \geqslant V^{K}(\mu) - \tfrac{\varepsilon}{2} \geqslant V^{\infty}(\mu) - \epsilon.$$

Therefore, $\liminf_{k\to\infty} V^{\infty}(\mu_k) \geqslant V^{\infty}(\mu) - \varepsilon$. Moreover, since the $\varepsilon >$ is arbitrary, it follows that $\liminf_{k\to\infty} V^{\infty}(\mu_k) \geqslant V^{\infty}(\mu)$. ■

Appendix B

Chapter 2

B.1 Proofs of Propositions

Proof of Proposition 2. Let S denote the set of strategy profiles. Since Θ, M, and A are compact spaces, S is also compact. Moreover, since u is continuous, so is v (see Theorem 1 in Milgrom and Weber (1985)). Hence, by the Weierstrass theorem, there exists a strategy profile $(\sigma, \rho) \in S$ that maximizes v.

Now, consider a Senders' strategy profile $\sigma' : \Theta \to M$ defined such that $\sigma'(\theta) = m$ only if

$$u(\rho(m), \theta) \geqslant u(\rho(m'), \theta) \quad \text{for all } m' \in M.$$

That is, σ' maps each state to a message that induces an optimal action in $\rho(M)$. Consequently, for each $\theta \in \Theta$, we have

$$u((\rho \circ \sigma')(\theta), \theta) \geqslant u((\rho \circ \sigma)(\theta), \theta).$$

This implies that $v(\sigma', \rho) \geqslant v(\sigma, \rho)$, meaning that (σ', ρ) is also a maximizer of v.

Furthermore, observe that (σ', ρ) constitutes a Bayes-Nash equilibrium. Since the Senders induce an optimal action in $\rho(M)$ for each state, no unilateral deviation is profitable. Similarly, the Receiver attains the highest possible expected utility, ensuring he has no incentive to deviate.

Since every off-path response can be replaced by an on-path response, every Bayes-Nash equilibrium outcome can be induced by a PBE. In conclusion, an efficient PBE exists.

■

Corollary 5 *If (σ, ρ) is an efficient PBE, then $v(\sigma, \rho) \geqslant v(\sigma', \rho')$ for each strategy profile (σ', ρ').*

The next definition and lemmas are instrumental in the proof of Proposition 3.

Definition 8 *Let $V \subseteq W \subseteq \mathbb{R}^n$. The set V is* affine relative to W *if, for each $x, y \in V$ and $t \in \mathbb{R}$,*

$$tx + (1-t)y \in W \text{ implies } tx + (1-t)y \in V.$$

A set V is affine relative to another set W if, for any two points in V, the line segment connecting them that lies in W is contained in V. In other words, $V \cap W$ is an affine subspace of W. More explicitly, $V \cap W$ is either empty, a singleton, a line, a hyperplane, or equal to $V \cap W$.

Lemma 15 *Let $a, a' \in A$. The set $\{\theta \in \operatorname{int}\Theta \mid u(a,\theta) = u(a',\theta)\}$ is an affine set relative to $\operatorname{int}\Theta$.*

Proof. Take $\theta, \theta' \in \operatorname{int}\Theta$ and $t \in \mathbb{R}$. Let $\theta^t = t\theta' + (1-t)\theta$. Suppose, for contradiction, that $u(a,\theta) = u(a',\theta)$ and $u(a,\theta') = u(a',\theta')$, but $u(a,\theta^t) \neq u(a',\theta^t)$, despite $\theta^t \in \operatorname{int}\Theta$. Without loss of generality, assume that $u(a,\theta^t) > u(a',\theta^t)$.

Consider the following cases:

(i) If $t \in \{0,1\}$, then θ^t is either θ or θ', which contradicts the assumption that $u(a,\theta^t) \neq u(a',\theta^t)$.

(ii) If $t \in (0,1)$, then θ^t is a convex combination of θ and θ'. Since $\theta \in \operatorname{int}\Theta$, we can choose $\hat{\theta} \in \operatorname{int}\Theta$ such that θ is a convex combination of $\hat{\theta}$ and θ^t. Then, the SSC property implies that

$$u(a,\hat{\theta}) < u(a',\hat{\theta}),$$

since otherwise $u(a,\theta) > u(a',\theta)$, contradicting the initial assumption. However, since θ^t is also a convex combination of $\hat{\theta}$ and θ', the SSC property then implies

$$u(a,\theta^t) < u(a',\theta^t),$$

which contradicts our assumption that $u(a,\theta^t) > u(a',\theta^t)$.

(iii) If $t \notin [0,1]$, then either θ is a convex combination of θ' and θ^t, or vice versa. Assume, without loss of generality, that θ' is a convex combination of θ and θ^t. Since $u(a,\theta) = u(a',\theta)$ and $u(a,\theta^t) > u(a',\theta^t)$, the SSC property implies

$$u(a,\theta') > u(a',\theta'),$$

which contradicts our assumption that $u(a,\theta') = u(a',\theta')$.

Therefore, it must be that $u(a,\theta^t) = u(a',\theta^t)$, completing the proof. ■

Lemma 15 establishes that the SSC property implies that the set of states where two actions yield the same utility is affine relative to the interior of the state space. Therefore, the conjunction of the SSC property and the *non-triviality* condition ensures that such indifference occurs on an affine set of dimension strictly lower than n, and is thus negligible. The next corollary formalizes this.

Corollary 6 *Let $u : A \times \Theta \to \mathbb{R}$ be an SSC-utility function. Then, for every pair of distinct actions $a, a' \in A$, the set*

$$\{\theta \in \Theta \mid u(a,\theta) = u(a',\theta)\}$$

has measure zero.

Proof. Let $\Theta_\sim = \{\theta \in \Theta \mid u(a,\theta) = u(a',\theta)\}$. Observe that

$$\Theta_\sim \subseteq (\Theta_\sim \cap \operatorname{int}\Theta) \cup \partial\Theta.$$

By Lemma 15, the set $\Theta_\sim \cap \operatorname{int}\Theta$ is contained in an affine subset of $\operatorname{int}\Theta$. Due to the non-triviality condition, this affine set must have dimension strictly less than n, and hence has Lebesgue measure zero in $\mathbb{R}^n$.

Moreover, since Θ is convex, its boundary $\partial\Theta$ also has Lebesgue measure zero (see Theorem 1 in Lang (1986)).

Therefore, $\Theta_\sim$ also has measure zero. ■

Define sets $\Theta(a \mid \hat{A}) = \bigcap_{a' \in \hat{A}, a' \neq a} \{\theta \in \Theta \mid u(a,\theta) > u(a',\theta)\}$ and $\overline{\Theta}(a \mid \hat{A}) = \bigcap_{a' \in \hat{A}, a' \neq a} \{\theta \in \Theta \mid u(a,\theta) \geqslant u(a',\theta)\}$, for each $\hat{A} \subseteq A$ and $a \in \hat{A}$. In other words, $\Theta(a \mid \hat{A})$ is the set of states in which action $a \in \hat{A}$ is uniquely optimal among all actions in $\hat{A}$. Similarly, $\overline{\Theta}(a \mid \hat{A})$ corresponds to the set of states in which action $a \in \hat{A}$ is optimal among all actions in $\hat{A}$.

Lemma 16 *For each finite $\hat{A} \subseteq A$ and $a \in \hat{A}$, it holds that $\Theta(a \mid \hat{A}) = \overline{\Theta}(a \mid \hat{A})$ almost surely.*

Proof. Observe that

$$\Theta(a \mid \hat{A}) \subseteq \overline{\Theta}(a \mid \hat{A}) \subseteq \Theta(a \mid \hat{A}) \cup \left(\bigcup_{\substack{a,a' \in \hat{A} \\ a \neq a'}} \{\theta \in \Theta \mid u(a,\theta) = u(a',\theta)\} \right).$$

By Corollary 6, the union on the right-hand side has measure zero, as it is a finite union of null sets. This completes the proof. ■

Proof of Proposition 3. Let (σ, ρ) be a non-redundant efficient PBE. By Corollary 5, we have $v(\sigma, \rho) \geqslant v(\sigma', \rho')$ for every strategy profile (σ', ρ'). Note that, given the Receiver's strategy ρ, state-wise optimization requires Senders to associated each state θ with a message m satisfying $u(\rho(m), \theta) \geqslant u(a, \theta)$ for all $a \in \rho(M)$. Equivalently, the Senders' strategy profile $\hat{\sigma} : \Theta \to M$ is state-wise optimal given ρ if

$$\Theta(\rho(m) \mid \rho(M)) \subseteq \hat{\sigma}^{-1}(\{m\}) \subseteq \overline{\Theta}(\rho(m) \mid \rho(M)).$$

Thus, for each $m \in M$, it follows that $\sigma^{-1}(\{m\}) = \hat{\sigma}^{-1}(\{m\})$ almost surely. Moreover, by Lemma 16, we have $\hat{\sigma}^{-1}(\{m\}) = \Theta(\rho(m) \mid \rho(M))$ almost surely. Consequently, $\sigma^{-1}(\{m\}) = \Theta(\rho(m) \mid \rho(M))$ almost surely. Finally, observe that $\Theta(\rho(m) \mid \rho(M))$ is a convex set, as it is the intersection of convex sets. This implies that $\sigma^{-1}(\{m\})$ is almost surely convex. ■

B.2 Proof of Theorem 5

This appendix contains the proof of Theorem 5. The proof proceeds in three steps. The first two steps establish the result for a bi-dimensional state space, while the final step extends the argument to any multidimensional state space.

Before detailing the proof, I outline the first two steps and explain why they suffice to establish Theorem 5 in the bi-dimensional case:

- **Step 1: Constructing preference relations**

  This step defines a preference relation over actions for each state. For any pair of distinct actions, the preference relation satisfies two properties: (i) the set of states in which players are indifferent between the two actions is either empty or a hyperplane, and (ii) one action is weakly preferred to the other either throughout the entire state space or within a closed half-space. This ensures that any utility representation is an SSC-utility function.

  In addition, the preference relations are such that there are exactly $|M|$ ex-post optimal actions. That is, there exists a set of $|M|$ actions such that, at every state, one of these actions is optimal. Note that the set of states in which each of these ex-post optimal actions is uniquely optimal is convex, due to the properties outlined above. However, the set of states where two or more—but fewer than $|M|$—of these actions are simultaneously optimal is, in general, not convex.

  This ensures that, if a utility representation exists, then in any efficient PBE, the meaning of message profiles is almost surely convex, whereas the meaning of each individual Sender's message is generally not.

- **Step 2: Constructing a utility representation**

  This step constructs a utility function that represents the preference relations defined in the previous step. As previously outlined, any utility representation is an SSC-utility function, and in any efficient PBE, the meaning of each Sender's message is not almost surely convex. Note, this completes the proof.

Now, I move on to the details of each step.

B.1. Step 1: Preference relations

Assume, without loss of generality, that $0 \in \operatorname{int}\Theta$. Let $\varepsilon > 0$ be such that $B_\varepsilon(0) \subset \operatorname{int}\Theta$, and pick an arbitrary $\theta_0 \in B_\varepsilon(0)\backslash\{0\}$. Let $r : \mathbb{R}^2 \to \mathbb{R}^2$ denote the clockwise rotation about 0 by an angle of $\frac{\pi}{|M|-1}$ radians. For each $q \in \mathbb{Z}$, let r^q denote the q-fold composition of r with itself (where r^{-1} represents the counterclockwise rotation by the same angle, i.e., the inverse of r).

Let $J' = \{0, 1, \ldots, |M| - 2\}$. Define $\theta_j = r^{2j}(\theta_0)$ for each $j \in J'$. Define the function $g : J' \times J' \to J'$ by

$$g(j,k) = \begin{cases} k - (j+1) & \text{if } j < k, \\ |M| - 2 + k - j & \text{if } j > k. \end{cases}$$

Note that $g(j, k)$ counts the number of elements between j and the smallest $k' > j$ such that $k' \equiv k \pmod{|M| - 1}$.

For each distinct pair $j, k \in J'$, define the (closed) half-space

$$H_{j,k} = \{\theta \in \Theta \mid r^{g(j,k)}(\theta_j) \cdot \theta \geqslant 0\}.$$

Note, $r^{g(j,k)}(\theta_j)$ bisects the angle between θ_j and θ_k. Let $s : \mathbb{R}^2 \to \mathbb{R}^2$ denote the clockwise rotation about 0 by an angle of $\frac{\pi}{2}$ radians. In addition, for each $j \in J'$, define

$$H_{j,|M|-1} = \{\theta \in \Theta \mid (0.5\theta_j + 0.5r^{|M|-3}(\theta_j)) \cdot \theta \geqslant (0.5\theta_j + 0.5r^{|M|-3}(\theta_j)) \cdot s(\theta_j)\},$$ [1]

and $H_{|M|-1,j} = r^{|M|-1}(H_{j,|M|-1})$.

Let $J = \{0, 1, \ldots, |M| - 2, |M| - 1\}$. The previous constructions define a collection of half-spaces $\{H_{j,k}\}_{j,k \in J; j \neq k}$. Figure B.1 illustrates this construction for the case $|M| = 6$.

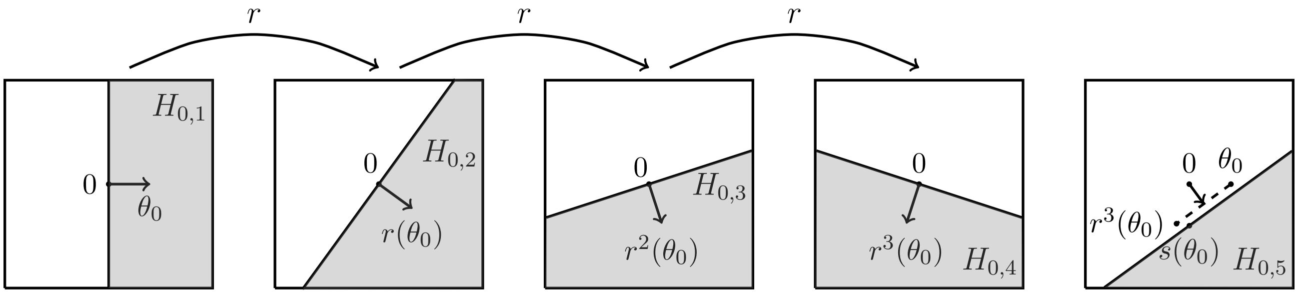


Figure B.1. Half-spaces $H_{0,j}$ for $j \in J$ with $j > 0$. (Case $|M| = 6$)

---

[1] Note that $H_{j,|M|-1}$ is well-defined since $r^{|M|-3}(\theta_j) \neq -\theta_j$. Indeed, $r^{|M|-3}(\theta_j)$ corresponds to θ_j rotated about 0 by $\frac{|M|-3}{|M|-1}\pi$ radians. Thus, $r^{|M|-3}(\theta_j)$ and $-\theta_j$ have different directions.

Now, pick $|M|$ distinct actions, and let $\hat{A} = \{a_0, a_1, \ldots, a_{|M|-1}\}$ denote the set of such actions. For each $\theta \in \Theta$, define a preference relation $\succeq_\theta$ on $\hat{A}$ as follows

$$a_j \succeq_\theta a_k \quad \text{if and only if} \quad a_j = a_k \quad \text{or} \quad (a_j \neq a_k \text{ and } \theta \in H_{j,k}).$$

The next lemmas are useful to prove that each preference relation is complete and transitive.

Lemma 17 *For each $j, k \in J'$ distinct, $H_{k,j} = r^{|M|-1}(H_{j,k})$.*

Proof. Assume, without loss of generality, that $j < k$. First, observe that

$$\theta_k = r^{2k}(\theta_0) = r^{2(k-j)+2j}(\theta_0) = r^{2(k-j)}(r^{2j}(\theta_0)) = r^{2(k-j)}(\theta_j).$$

In addition, note that

$$g(k, j) = |M| - 2 + j - k = |M| - 3 - g(j, k).$$

Therefore,

$$\begin{aligned} r^{g(k,j)}(\theta_k) &= r^{|M|-3-g(j,k)}(r^{2(k-j)}(\theta_j)) \\ &= r^{|M|-1+k-(j+1)}(\theta_j) \\ &= r^{|M|-1}(r^{g(j,k)}(\theta_j)). \end{aligned}$$

Thus, the direction of $H_{k,j}$ is obtained by rotating the direction of $H_{j,k}$ by π radians around the origin. In other words, $H_{k,j} = r^{|M|-1}(H_{j,k})$. ■

Lemma 18 *For each $j, k, l \in J'$ distinct, $H_{j,k} \cap H_{k,l} \subseteq H_{j,l}$.*

Proof. Let $d_{j,k}$, $d_{k,l}$, and $d_{j,l}$ denote the directions of $H_{j,k}$, $H_{k,l}$, and $H_{j,l}$, respectively. It suffices to show that there exist $\lambda_1, \lambda_2 \geqslant 0$ such that

$$d_{j,l} = \lambda_1 d_{j,k} + \lambda_2 d_{k,l}.$$

To see this, assume $\theta \in H_{j,k} \cap H_{k,l}$. That is, $d_{j,k} \cdot \theta \geqslant 0$ and $d_{k,l} \cdot \theta \geqslant 0$. Then,

$$(\lambda_1 d_{j,k} + \lambda_2 d_{k,l}) \cdot \theta \geqslant 0$$

for any $\lambda_1, \lambda_2 \geqslant 0$. Figure B.2 illustrates the relevant cases in which $d_{j,l} = \lambda_1 d_{j,k} + \lambda_2 d_{k,l}$ holds.

(a) Clockwise angle between $d_{j,k}$ and $d_{k,l}$ is smaller than π

(b) Clockwise angle between $d_{j,k}$ and $d_{k,l}$ is larger than π

Figure B.2. $d_{j,l} = \lambda_1 d_{j,k} + \lambda_2 d_{k,l}$ for some $\lambda_1, \lambda_2 \geqslant 0$

Note that $H_{j,j+1} = r^{-g(j,k)}(H_{j,k})$ since $H_{j,l} = r^{g(j,l)}(H_{j,j+1})$. Therefore,

$$H_{j,l} = r^{g(j,l)}(H_{j,j+1}) = r^{g(j,l)-g(j,k)}(H_{j,k}),$$

and so $d_{j,l}$ is obtained by rotating $d_{j,k}$ clockwise $g(j,l) - g(j,k)$ times. Considering only positive clockwise rotations and discounting full rotations, we obtain

$$g(j,l) - g(j,k) = \begin{cases} l - k & \text{if } l > k > j \text{ or } j > l > k, \\ 2(|M| - 1) + (l - k) & \text{if } k > l > j \text{ or } j > k > l, \\ (|M| - 1) + (l - k) & \text{otherwise.} \end{cases}$$

Similarly, $H_{k,l} = r^{g(k,l)}(H_{k,k+1})$ and $H_{k,k+1} = r^{-g(k,j)}(H_{k,j})$. Also, Lemma 17 implies $H_{k,j} = r^{|M|-1}(H_{j,k})$. Hence,

$$H_{k,l} = r^{g(k,l)}(H_{k,k+1}) = r^{g(k,l)-g(k,j)}(H_{k,j}) = r^{|M|-1+g(k,l)-g(k,j)}(H_{j,k}),$$

so $d_{k,l}$ is obtained by rotating $d_{j,k}$ clockwise $|M|-1+g(k,l)-g(k,j)$ times. Again, accounting for only positive clockwise rotations and discounting full rotations, we get

$$|M|-1+g(k,l)-g(k,j) = \begin{cases} l-j & \text{if } l > k > j, \\ 2(|M|-1)+(l-j) & \text{if } j > k > l, \\ (|M|-1)+(l-j) & \text{otherwise.} \end{cases}$$

Therefore, it suffices to show the following implications:

- If $|M|-1+g(k,l)-g(k,j) < |M|-1$, then $|M|-1+g(k,l)-g(k,j) > g(j,l)-g(j,k)$.
- If $|M|-1+g(k,l)-g(k,j) > |M|-1$, then $|M|-1+g(k,l)-g(k,j) < g(j,l)-g(j,k)$.

We proceed by cases:

Case $j < k < l$: $|M|-1+g(k,l)-g(k,j) = l-j < |M|-1$, so we must show $l-j > l-k$, which simplifies to $j < k$, true by assumption.

Case $j < l < k$: $|M|-1+g(k,l)-g(k,j) = |M|-1+(l-j) > |M|-1$, and we must show $|M|-1+(l-j) < 2(|M|-1)+(l-k)$, which reduces to $k-j < |M|-1$, true since $k-j \leqslant |M|-2$.

Case $k < j < l$: $|M|-1+(l-j) > |M|-1$, and we must show $|M|-1+(l-j) < (|M|-1)+(l-k)$, which reduces to $k < j$, true by assumption.

Case $k < l < j$: $|M|-1+(l-j) < |M|-1$, and we must show $|M|-1+(l-j) > 2(|M|-1)+(l-k)$, which simplifies to $j-k < |M|-1$, true since $j-k \leqslant |M|-2$.

Case $l < j < k$: $|M|-1+(l-j) < |M|-1$, and we must show $j < k$, which holds by assumption.

Case $l < k < j$: $|M|-1+g(k,l)-g(k,j) = 2(|M|-1)+(l-j) > |M|-1$, and we must show $2(|M|-1)+(l-j) < (|M|-1)+(l-k)$, which simplifies to $k < j$, true by assumption.

This completes the proof. ■

Lemma 19 *$H_{j,k} \cap H_{k,l} \subseteq H_{j,l}$, for each $j,k,l \in J$ distinct with one of them being $|M|-1$.*

Proof. Assume, without loss of generality, that $l = |M|-1$. We will establish that $H_{j,k} \cap H_{k,|M|-1} \subseteq H_{j,|M|-1}$. First, observe that $\{\theta_l \mid l \in J'\}$ is a collection of $|M|-1$ equally spaced directions relative to θ_0. Therefore, the set $\{s(\theta_l) \mid l \in J'\}$ consists of $|M|-1$ equally spaced vectors. These vectors determine the vertices of a regular polygon with $|M|-1$ sides, which we denote by $P_{|M|-1}$.

Now consider $d_j := \frac{1}{2}\theta_j + \frac{1}{2}r^{|M|-3}(\theta_j)$, which bisects the angle between θ_j and $r^{|M|-3}(\theta_j)$. Since θ_j and $r^{|M|-3}(\theta_j)$ are symmetric around d_j, their orthogonal directions $\pm s(\theta_j)$ and $\pm s^{-1}(r^{|M|-3}(\theta_j))$ are symmetric around the direction orthogonal to d_j. Hence, one of the two pairs of opposite orthogonal vectors — either $\{s(\theta_j),\, s^{-1}(r^{|M|-3}(\theta_j))\}$ or $\{s^{-1}(\theta_j),\, s(r^{|M|-3}(\theta_j))\}$ — must lie along the same line. By definition, $s(\theta_j)$ lies on the boundary of $H_{j,|M|-1}$, and thus $s^{-1}(r^{|M|-3}(\theta_j))$ must lie on this boundary as well.

Next, note that the angle between θ_j and $r^{|M|-3}(\theta_j)$ is $\frac{|M|-3}{|M|-1}\pi$, so the angle between $s(\theta_j)$ and $s^{-1}(r^{|M|-3}(\theta_j))$ is $\pi + \frac{|M|-3}{|M|-1}\pi = 2\pi - \frac{2\pi}{|M|-1} = -\frac{2\pi}{|M|-1} \mod 2\pi$. Hence,

$$s^{-1}(r^{|M|-3}(\theta_j)) = r^{-2}(s(\theta_j)) = s(\theta_{j-1}).$$

Therefore, the side of the polygon $P_{|M|-1}$ determined by the consecutive vertices $s(\theta_j)$ and $s(\theta_{j-1})$ lies on the boundary of $H_{j,|M|-1}$.

On the other hand, observe that $H_{k,|M|-1} = r^{2(k-j)}(H_{j,|M|-1})$. Thus, $H_{j,|M|-1}$ and $H_{k,|M|-1}$ are symmetric with respect to the line defined by the boundary of $H_{j,k}$, which is orthogonal to $r^{g(j,k)}(\theta_j)$ — the angle bisector of θ_j and θ_k. In other words, the boundary of H_j, k serves as the axis of reflection between $H_{j,|M|-1}$ and $H_{k,|M|-1}$.

Since $H_{j,k}$ is the half-space on one side of this reflection axis, the intersection $H_{j,k} \cap H_{k,|M|-1}$ is the portion of $H_{k,|M|-1}$ that lies on the same side of the axis as $H_{j,|M|-1}$. But because the two regions are symmetric and $H_{j,k}$ selects the side containing $H_{j,|M|-1}$, this intersection must lie within $H_{j,|M|-1}$. That is, $H_{j,k} \cap H_{k,|M|-1} \subseteq H_{j,|M|-1}$. ■

Proposition 6 *For each $\theta \in \Theta$, the preference relation $\succeq_\theta$ is complete and transitive.*

Proof. Let $\theta \in \Theta$ and $a_j, a_k, a_l \in \hat{A}$.

(i) If $j = k$, then $a_j \succeq_\theta a_k$.

If $j \neq k$ and $j, k < |M|-1$, then Lemma 17 ensures that $H_{j,k} \cup H_{k,j} = \Theta$. Consequently, $\theta \in H_{j,k}$ or $\theta \in H_{k,j}$. In other words, $a_j \succeq_\theta a_k$ or $a_k \succeq_\theta a_j$.

If $j \neq k$ and either $j = |M|-1$ or $k = |M|-1$, then $H_{j,k} \cup H_{k,j} = \Theta$ by construction. Consequently, $a_j \succeq_\theta a_k$ or $a_k \succeq_\theta a_j$.

In conclusion, $\succeq_\theta$ is complete.

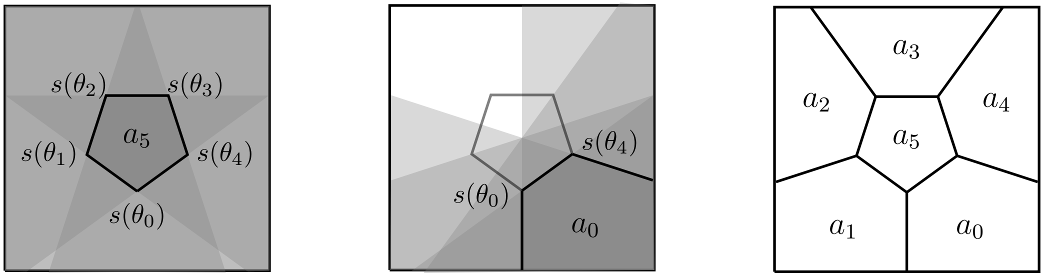
(ii) Assume that $a_j \succeq_\theta a_k$ and $a_k \succeq_\theta a_l$. Note, if two of the indices are equal, then $a_j \succeq_\theta a_l$. If j, k, l are distinct, then $\theta \in H_{j,k} \cap H_{k,l}$. Note, Lemmas 18 and 19 ensure that $H_{j,k} \cap H_{k,l} \subset H_{j,l}$. Therefore, $\theta \in H_{j,l}$, which means that $a_j \succeq_\theta a_l$. In conclusion, $\succeq_\theta$ is transitive.

■

Remark 1 *Fix* $a_j \in \hat{A}$. *Observe, action* a_j *is optimal at state* θ *if and only if* θ *belongs to* $\bigcap_{k \neq j, k \in J} H_{j,k}$. *Consider the following cases:*

(i) Assume $j = |M|-1$. *As Figure B.3a illustrates,* $\bigcap_{j \in J'} H_{|M|-1,j}$ *corresponds to the polygon* $P_{|M|-1}$ *and all states within.*

(ii) Assume $j \in J'$. *As Figure B.3b illustrates,* $\bigcap_{k \neq j, k \in J} H_{j,k}$ *corresponds to the region outside of the polygon* $P_{|M|-1}$ *limited by the rays connecting the origin with* $s(\theta_{j-1})$ *and the origin with* $s(\theta_j)$.



(a) $\bigcap_{j<5} H_{5,j}$ **(b)** $\bigcap_{j>0} H_{0,j}$ **(c)** Optimal Partition

Figure B.3. Set of states in which action a_j is optimal. (Case $|M| = 6$)

Figure B.3c illustrates that the previous construction the state space into $|M|$ *almost surely convex regions, with the following property: The union of any subset of the partition is almost surely convex only if the subset contains zero, one, or all* $|M|$ *regions. To see this, consider two cases. Suppose, first, that the union includes the regular polygon. Since there are* $|M|-1$

outer regions, the union must omit at least one of them. In particular, there will be two adjacent outer regions such that only one is included. As a result, the union corresponds to a polygon that has an internal angle greater than π radians, so it fails to be almost surely convex.[2]

Now suppose the polygon is not included in the union. In this case, the union consists solely of a subset of the outer regions. If all outer regions are included, the resulting set fails to be almost surely convex. If only some outer regions are included, the union is either disconnected or forms a polygon with at least one internal angle exceeding π. In either case, the union fails to be almost surely convex.

B.2. Step 2: A utility representation

Fix $\theta \in \Theta$. Since the preference relation $\succeq_\theta$ on $\hat{A}$ is complete and transitive, and $\hat{A}$ is a finite set, it admits a utility representation. Let $\hat{u}_\theta : \hat{A} \to \mathbb{R}$ be such a utility function, normalized so that $\min_{\hat{a}\in\hat{A}} \hat{u}_\theta(\hat{a}) = 0$.

Now, observe that A is either finite or countable because it is a compact metric space. Hence, the set difference $A\backslash\hat{A}$ is at most countable, and we may index its elements by $\Lambda \subseteq (0,1)$; that is, $A\backslash\hat{A} = \{a_\lambda\}_{\lambda\in\Lambda}$. Define an extension of $\hat{u}_\theta$ to all of A by

$$u_\theta(a) = \begin{cases} \hat{u}_\theta(a) & \text{if } a \in \hat{A}, \\ -\lambda & \text{if } a = a_\lambda \text{ for some } \lambda \in \Lambda. \end{cases}$$

By construction, $\arg\max_{a\in A} u_\theta(a) = \arg\max_{\hat{a}\in\hat{A}} \hat{u}_\theta(\hat{a})$.

Now define $u : A \times \Theta \to \mathbb{R}$ by $u(a,\theta) = u_\theta(a)$. Observe that for each $\theta \in \Theta$, an action a^* maximizes $u(\cdot,\theta)$ if and only if $a^* \in \hat{A}$ and $a^* \succeq_\theta \hat{a}$ for all $\hat{a} \in \hat{A}$. Thus, for any measurable function $g : \Theta \to A$ such that $u(g(\theta),\theta) \geqslant u(g'(\theta),\theta)$ for all measurable $g' : \Theta \to A$, it must be that $g(\Theta) = \hat{A}$ and, for each $j \in J$, $g^{-1}(a_j)$ is almost surely equal to $\bigcap_{k\neq j,\, k\in J} H_{j,k}$.

Since $|M| = |\hat{A}|$, there exists a PBE (σ^*,ρ^*) with $\rho^* \circ \sigma^* = g$, so (σ^*,ρ^*) is efficient. Moreover, any efficient PBE (σ,ρ) must satisfy $\rho \circ \sigma = g$ almost surely. That is, g is the unique outcome of any efficient PBE.

---

[2] Formally, there are open balls $B_\delta(\theta)$ and $B_\delta(\theta')$ almost surely contained in the union and $\alpha \in (0,1)$ such that $B_\delta(\alpha\theta' + (1-\alpha)\theta)$ is almost surely contained in the complement of the union.

B.3. Step 3: Extension to higher dimensions

The argument naturally extends to higher dimensions by selecting θ_0 in $B_\varepsilon(0)$ and restricting attention to the X_1X_2-plane. With this modification, the convex regions depicted in Figure B.3 become prisms. Formally, assume that $\Theta \subseteq \mathbb{R}^n$ for $n \geqslant 3$. Pick θ_0 in the intersection of $B_\varepsilon(0)\backslash\{0\}$ and the X_1X_2-plane, and follow the same construction as before.

Then, each set S defined previously becomes $[S \times \mathbb{R}^{n-2}] \cap \Theta$. Note that $[S \times \mathbb{R}^{n-2}] \cap \Theta$ is almost surely convex if and only if S is almost surely convex. Thus, the proof carries over.

B.3 Scope of Theorem 5

Let $\bar{\theta} = (0.3, 0.3)$. Let $r : \mathbb{R}^2 \to \mathbb{R}^2$ denote the clockwise rotation about $\bar{\theta}$ by an angle of $\frac{\pi}{4}$ radians. For each $q \in \mathbb{Z}$, let r^q denote the q-fold composition of r with itself (with r^{-1} denoting the inverse of r).

Let $J = \{0, 1, 2, 3\}$. Define the function $g : J \times J \to J$ by

$$g(j,k) = \begin{cases} k - (j+1) & \text{if } j < k, \\ 3 + k - j & \text{if } j > k. \end{cases}$$

Note that $g(j,k)$ counts the number of elements between j and the smallest $k' > j$ such that $k' \equiv k \pmod 4$.

Let $\eta_{1,0} = (1, 0)$. Define $\eta_{j,k} = r^{g(j,k)+2j}(\eta_{0,1})$ for each distinct $j, k \in J$. Finally, define $H_{j,k} = \{\theta \in \Theta \mid \eta_{j,k} \cdot (\theta - \bar{\theta}) \geqslant 0\}$.

Now, pick distinct actions $a_0, a_1, a_2, a_3 \in A$, with $a_0 = (0, 0)$. Let $\hat{A} = \{a_0, a_1, a_2, a_3\}$. For each $\theta \in \Theta$, define the following relation on $\hat{A}$:

$$a_j \succeq_\theta a_k \text{ if and only if } j = k \text{ or } \theta \in H_{j,k}.$$

Proposition 7 *For each $\theta \in \Theta$, the relation $\succeq_\theta$ is complete and transitive.*

Proof. Fix $\theta \in \Theta$ and, let $a_j, a_k, a_l \in \hat{A}$.

(i) Note that if $j = k$, then trivially $a_j \succeq_\theta a_k$. Now assume $j \neq k$, and without loss of generality, suppose $j < k$. Observe that

$$H_{k,j} = r^{g(k,j)+2k}(H_{0,1}) \quad \text{and} \quad H_{0,1} = r^{-g(j,k)-2j}(H_{j,k}).$$

Substituting the second equation into the first gives:

$$H_{k,j} = r^{g(k,j)-g(j,k)+2(k-j)}(H_{j,k}).$$

Since $g(k,j) - g(j,k) = 4 + 2(j-k)$, it follows that:

$$H_{k,j} = r^4(H_{j,k}),$$

which implies that $H_{k,j}$ is a π-radian rotation of $H_{j,k}$ about $\bar{\theta}$. As a result, $H_{j,k} \cup H_{k,j} = \Theta$. Consequently, for any $\theta \in \Theta$, either $a_j \succeq_\theta a_k$ or $a_k \succeq_\theta a_j$.

(ii) Assume $a_j \succeq_\theta a_k$ and $a_k \succeq_\theta a_l$. Note, if at least two actions are equal, then trivially $a_j \succeq_\theta a_k$. Now assume all actions are different, and without loss of generality, suppose $j < k < l$. That is, we know that $\theta \in H_{j,k} \cap H_{k,l}$. We want to show that $\theta \in H_{j,l}$. It suffices to show that $\eta_{j,l} = \lambda_1 \eta_{j,k} + \lambda_2 \eta_{k,l}$, for some $\lambda_1, \lambda_2 \geqslant 0$.

Since

$$H_{j,l} = r^{g(j,l)}(H_{j,j+1}) = r^{g(j,l)-g(j,k)}(H_{j,k}),$$

and so $\eta_{j,l}$ is obtained by rotating $\eta_{j,k}$ clockwise $g(j,l) - g(j,k)$ times. Considering only positive clockwise rotations and discounting full rotations, we obtain $g(j,l) - g(j,k) = l - k$.

Similarly,

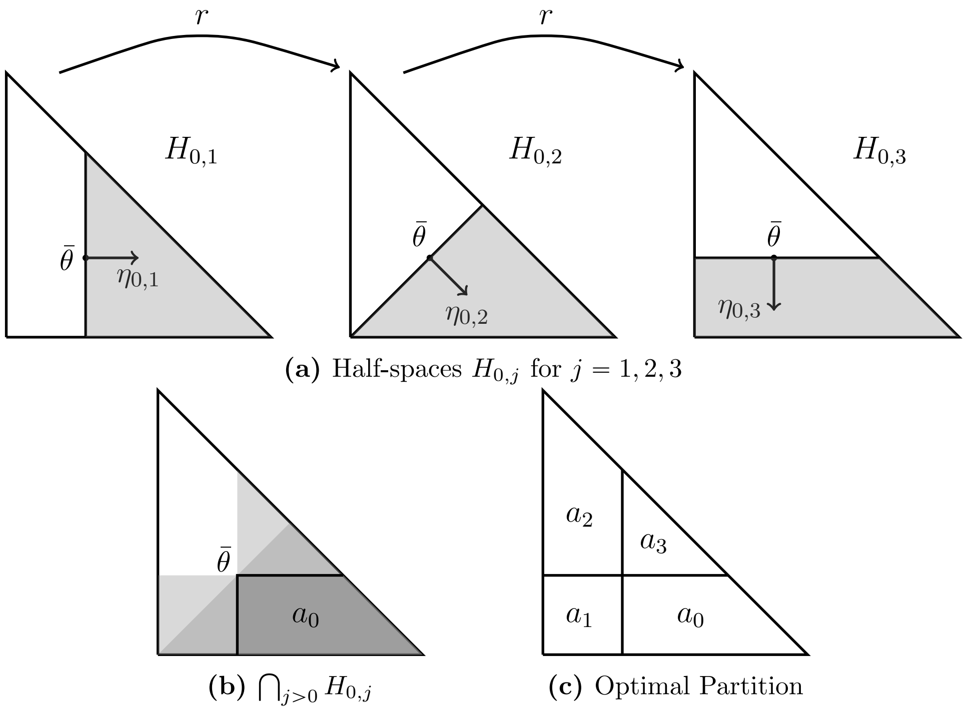
$$H_{k,l} = r^{g(k,l)}(H_{k,k+1}) = r^{g(k,l)-g(k,j)}(H_{k,j}) = r^{3+g(k,l)-g(k,j)}(H_{j,k}),$$

so $\eta_{k,l}$ is obtained by rotating $\eta_{j,k}$ clockwise $|M| - 1 + g(k,l) - g(k,j)$ times. Again, accounting for only positive clockwise rotations and discounting full rotations, we get

$3 + g(k,l) - g(k,j) = l - j$. Therefore, it suffices to show that $l - j > l - k$, which simplifies to $j < k$ and is true by assumption. This completes the proof.

■

Remark 2 *Figure B.4a illustrates the closed half-spaces* $H_{0,j}$ *for each* $j \in \{1,2,3\}$*, representing the set of states in which action* a_0 *is preferred to action* a_j*. Figure B.4b shows the set of states in which action* a_0 *is optimal. Finally, Figure B.4c depicts the partition of the state space induced by these preferences.*



(a) Half-spaces $H_{0,j}$ for $j = 1,2,3$

(b) $\bigcap_{j>0} H_{0,j}$ **(c)** Optimal Partition

Figure B.4. Constructed preferences

By Proposition 7, the preference relation $\succeq_\theta$ on $\hat{A}$ admits a utility representation for each $\theta \in \Theta$. Let $\hat{u}_\theta : \hat{A} \to \mathbb{R}$ be such a representation, normalized so that $\min_{\hat{a} \in \hat{A}} \hat{u}_\theta(\hat{a}) = 0$ for each θ. Extend $\hat{u}_\theta$ to all of A by defining

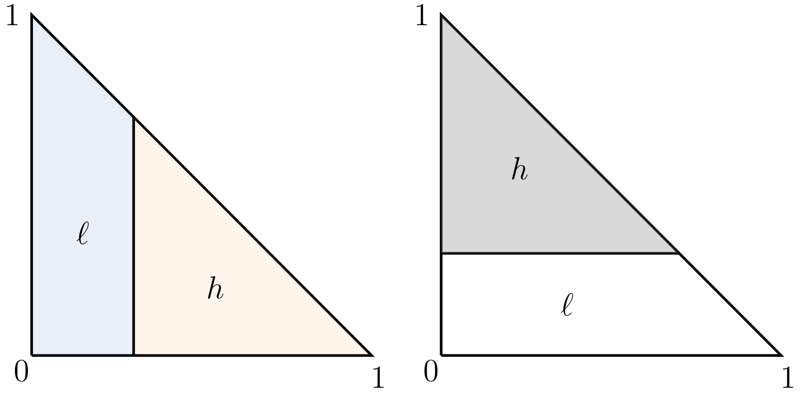
$$u_\theta(a) = \begin{cases} \hat{u}_\theta(a) & \text{if } a \in \hat{A}, \\ -a_1 - a_2 & \text{otherwise.} \end{cases}$$

By construction, $\arg\max_{a \in A} u_\theta(a) = \arg\max_{\hat{a} \in \hat{A}} \hat{u}_\theta(\hat{a})$.

Now define $u : A \times \Theta \to \mathbb{R}$ by $u(a, \theta) = u_\theta(a)$. Observe that for each $\theta \in \Theta$, an action a^* maximizes $u(\cdot, \theta)$ if and only if $a^* \in \hat{A}$ and $a^* \geq_\theta \hat{a}$ for all $\hat{a} \in \hat{A}$. Thus, for any measurable function $g : \Theta \to A$ such that $u(g(\theta), \theta) \geqslant u(g'(\theta), \theta)$ for all measurable $g' : \Theta \to A$, it must be that $g(\Theta) = \hat{A}$ and, for each $j \in \{0, 1, 2, 3\}$, $g^{-1}(a_j)$ is almost surely equal to $\bigcap_{k \neq j} H_{j,k}$.

Since $|M| = |\hat{A}|$, there exists a PBE (σ^*, ρ^*) such that $\rho^* \circ \sigma^* = g$, so (σ^*, ρ^*) is efficient. Moreover, any efficient PBE (σ, ρ) must satisfy $\rho \circ \sigma = g$ almost surely. That is, g is the unique outcome of any efficient PBE.

Finally, observe that there exists an efficient PBE in which Sender 1's and Sender 2's strategies are given by Figures B.5a and B.5b.



(a) Sender 1's strategy **(b)** Sender 2's strategy

Figure B.5. Senders strategy profile

Appendix C

Chapter 3

Proof of Proposition 4. First, let (σ, ρ, d) be an interval strategy profile that induces an interval PBE outcome. That is, there exists an interval PBE (σ', ρ', d') such that $d(\sigma(\theta), \rho(\sigma(\theta)), \theta) = d'(\sigma'(\theta), \rho'(\sigma'(\theta)), \theta)$ for each $\theta \in \Theta$.

Therefore, the players do not have incentives to deviate on-path under (σ', ρ', d'), which implies that the same holds for (σ, ρ, d). That is, (σ, ρ, d) satisfies conditions (i)-(iii).

Now, let (σ, ρ, d) be an interval strategy profile satisfying conditions (i)-(iii). Pick $m' \in \sigma(\Theta)$. Consider the interval strategy profile (σ, ρ', d') defined by

$$\rho'(m) = \begin{cases} \rho(m) & \text{if } m \in \sigma(\Theta) \\ \rho(m') & \text{else} \end{cases},$$

and

$$d'(m, \alpha, \theta) = \begin{cases} d(m, \alpha, \theta) & \text{if } m \in \sigma(\Theta) \\ d(m', \alpha, \theta) & \text{else} \end{cases}.$$

Note, (σ, ρ, d) and (σ, ρ', d') are outcome equivalent because they may only differ at off-path histories. In addition, (σ, ρ', d') is a PBE with the belief system $\mu : \Theta \to M$ defined by

$$\mu(m) = \begin{cases} U[\sigma^{-1}(m)] & \text{if } m \in \sigma(\Theta) \\ U[\sigma^{-1}(m')] & \text{else} \end{cases}.$$

In conclusion, (σ, ρ, d) induces an interval PBE outcome. ■

Proof of Lemma 6. Let (σ, ρ, d) be a PBE. Fix $(m, \alpha, \theta) \in M \times \mathcal{A} \times \Theta$ and $a \in \alpha$.

(i) Assume that $d(m, \alpha, \theta) = a$, so $a \in \arg\max_{a' \in \alpha} u_R(a', \theta)$. That is, a is at least as close to θ than any other action in α. Equivalently, θ is closer to a relative to any other action in α, i.e., $\theta \in \Theta(a \mid \alpha)$.

(ii) Assume that $\theta \in \operatorname{int} \Theta(a \mid \alpha)$, so θ is strictly closer to action a relative to any other action in α. That is, $u_R(a, \theta) > u_R(a', \theta)$ for all $a' \in \alpha$, $a' \neq a$. Consequently, $d(m, \alpha, \theta) = a$.

■

The following notation will be useful for the exposition. Given a skill set $\alpha = \{a_1, a_2, \ldots, a_k\}$, define $\overline{a}_j = \frac{a_j + a_{j+1}}{2}$ for $j = 1, \ldots, k-1$. So, $\Theta(a_1 \mid \alpha) = [0, \overline{a}_1] \cap \Theta$, $\Theta(a_k \mid \alpha) = [\overline{a}_{k-1}, 1] \cap \Theta$, and $\Theta(a_j \mid \alpha) = [\overline{a}_j, \overline{a}_{j+1}] \cap \Theta$ for $j = 2, \ldots, k-1$. In addition, refer to the following system of linear equations as (FOC):

$$\begin{cases} 2a_1 & = \overline{a}_1 + \underline{\theta} \\ 2a_j & = \overline{a}_j + \overline{a}_{j-1} \text{ for } j = 2, \ldots, k-1 \\ 2a_k & = \overline{\theta} + \overline{a}_{k-1} \end{cases} \tag{FOC}$$

where $\underline{\theta}, \overline{\theta} \in \Theta$ with $\underline{\theta} < \overline{\theta}$.

Lemma 20 *Let $\alpha \in \mathcal{A}$ be a skill set. If α satisfies (FOC), then $\overline{\theta} - \overline{a}_{k-1} = \overline{a}_j - \overline{a}_{j-1} = \overline{a}_1 - \underline{\theta}$ for all $j = 2, \ldots, k-1$.*

Proof. By induction on j. (FOC) implies that $2a_1 = \overline{a}_1 + \underline{\theta}$ and $2a_2 = \overline{a}_2 + \overline{a}_1$. Adding these two equalities up yields $4\overline{a}_1 = 2\overline{a}_1 + \overline{a}_2 + \underline{\theta}$. Rearranging gives $\overline{a}_2 - \overline{a}_1 = \overline{a}_1 - \underline{\theta}$.

Now, assume that $\overline{a}_j - \overline{a}_{j-1} = \overline{a}_1 - \underline{\theta}$ for some $2 \leqslant j \leqslant k-2$. Then, $2a_j + 2a_{j+1} = \overline{a}_j + \overline{a}_{j-1} + \overline{a}_{j+1} + \overline{a}_j$ follows from (FOC). Rearranging, we obtain $\overline{a}_{j+1} - \overline{a}_j = \overline{a}_j - \overline{a}_{j-1} = \overline{a}_1 - \underline{\theta}$, where the last equality holds by the inductive hypothesis. In conclusion, $\overline{a}_j - \overline{a}_{j-1} = \overline{a}_1 - \underline{\theta}$ for all $j = 2, \ldots, k-1$

Finally, observe that $2a_k + 2a_{k-1} = \overline{\theta} + \overline{a}_{k-1} + \overline{a}_{k-1} + \overline{a}_{k-2}$ follows from (FOC). Rearranging, we obtain $\overline{\theta} - \overline{a}_{k-1} = \overline{a}_{k-1} - \overline{a}_{k-2} = \overline{a}_1 - \underline{\theta}$, where the last inequality holds by the inductive argument. This completes the proof. ■

Lemma 21 *Let $\alpha \in \mathcal{A}$ be a skill set. If α satisfies* (FOC)*, then $\overline{a}_j = \frac{j}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta}$ for all $j = 1, \ldots, k-1$.*

Proof. By induction on j. First, note that

$$\begin{aligned}\overline{\theta} - \underline{\theta} &= \underline{\theta} - \overline{a}_{k-1} + \sum_{j=2}^{k-1}(\overline{a}_j - \overline{a}_{j-1}) + \overline{a}_1 - \underline{\theta} \\ &= k(\overline{a}_1 - \underline{\theta})\end{aligned}$$

where the last inequality follows from Lemma 20. Solving, we obtain $\overline{a}_1 = \frac{1}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta}$.

Now, assume that $\overline{a}_j = \frac{j}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta}$ for some $j < k - 1$. Then,

$$\begin{aligned}\overline{a}_{j+1} &= \overline{a}_j + \overline{a}_1 - \underline{\theta} \\ &= \left(\frac{j}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta}\right) + \left(\frac{1}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta}\right) - \underline{\theta} \\ &= \frac{j+1}{k}(\overline{\theta} - \underline{\theta}) + \underline{\theta},\end{aligned}$$

where the first equality follows from Lemma 20, and the second one follows from the base step and the inductive hypothesis. ■

Proof of Lemma 7. Let (σ, ρ, d) be a minimal interval PBE, and $m \in \sigma(\Theta)$. Let $\underline{\theta} = \inf \sigma^{-1}(m)$ and $\overline{\theta} = \sup \sigma^{-1}(m)$. By definition, $\rho(m)$ solves

$$\min_{\alpha \in \mathcal{A}} \sum_{a \in \alpha} \int_{\Theta(a|\alpha) \cap (\overline{\theta}, \underline{\theta})} (a - \theta)^2 d\theta \tag{C.1}$$

Let $\alpha^* = \{a_j^* \mid j = 1, \ldots, k\}$ be a solution to (C.1). Observe that α^* must satisfy that (i) $|\alpha^*| = K$, and (ii) $\alpha^* \subseteq [\underline{\theta}, \overline{\theta}]$. To see this, first, consider a skill set α with $|\alpha| < K$. Fix an action $a \in \Theta \backslash \alpha$, and define $\alpha' = \alpha \cup \{a\}$. Note, the Receiver obtains a weakly larger expected utility given skill set α' relative to α. Moreover, for every state $\theta \in \operatorname{int} \Theta(a \mid \alpha')$, he obtains a strictly higher utility with α' than with α. Since $\operatorname{int} \Theta(a \mid \alpha')$ has positive measure (because $a \notin \alpha$), it follows that α is not a solution to (C.1).

Second, consider a skill set α such that $\alpha^* \nsubseteq [\underline{\theta}, \overline{\theta}]$. Suppose, without loss of generality, that there is some $a \in \alpha$ with $a < \underline{\theta}$. Define $\alpha' = (\alpha \backslash \{a\}) \cup \{\underline{\theta}\}$. Note that the Receiver obtains a strictly higher expected utility from α' than form α. Hence, α cannot be a solution to (C.1).

Consequently, the optimization problem (C.1) is equivalent to

$$\begin{aligned}
&\arg\min_{\alpha\in\mathcal{A}} \int_{\underline{\theta}}^{\overline{a}_1}(a_1-\theta)^2\mathrm{d}\theta + \sum_{j=2}^{K-1}\int_{\overline{a}_{j-1}}^{\overline{a}_j}(a_j-\theta)^2\mathrm{d}\theta + \int_{\overline{a}_{K-1}}^{\overline{\theta}}(a_K-\theta)^2\mathrm{d}\theta \\
&= \arg\min_{\alpha\in\mathcal{A}} a_1^2(\overline{a}_1-\underline{\theta}) - a_1(\overline{a}_1^2-\underline{\theta}^2) + a_K^2(\overline{\theta}-\overline{a}_{K-1}) - a_K(\overline{\theta}^2-\overline{a}_{K-1}^2) \\
&+ \sum_{j=2}^{K-1}[a_j^2(\overline{a}_j-\overline{a}_{j-1}) - a_j(\overline{a}_j^2-\overline{a}_{j-1}^2)]
\end{aligned}$$

where $\overline{a}_j = \frac{a_j+a_{j+1}}{2}$ for $j=1,\dots,K-1$.

Hence, α^* solves the first-order conditions:

$$\begin{cases}
(\overline{a}_1-\underline{\theta})(2a_1-\overline{a}_1-\underline{\theta}) & = 0 \\
(\overline{a}_j-\overline{a}_{j-1})(2a_j-\overline{a}_j-\overline{a}_{j-1}) & = 0 \text{ for } j=2,\dots,K-1 \\
(\overline{\theta}-\overline{a}_{K-1})(2a_K-\overline{\theta}-\overline{a}_{K-1}) & = 0
\end{cases}$$

Let α' be a solution to the first-order conditions. Note, if $\overline{a}'_1 = \underline{\theta}$, then either $a'_1 = a'_2 = \underline{\theta}$ or $a'_1 < \underline{\theta} < a'_2$. In any case, α' is not a solution to the Receiver's problem. Similarly, α' is not a solution if $\overline{\theta} = \overline{a}'_{K-1}$. Last, α' is not a solution if $\overline{a}'_j = \overline{a}'_{j-1}$ for some $j=2,\dots,K-1$ because that implies $a_{j+1} = a_{j-1}$.

Consequently, α^* is a solution to (FOC). Lemma 21 implies that $a_j = \frac{1}{2}(\overline{a}_j + \overline{a}_{j-1})$ for all $j=1,\dots,K$, where $\overline{a}_0 = \underline{\theta}$ and $\overline{a}_K = \overline{\theta}$. By Lemma 21, we conclude that

$$\begin{aligned}
a_j &= \frac{1}{2}(\overline{a}_j + \overline{a}_{j-1}) \\
&= \frac{1}{2}\left(\frac{j}{K}(\overline{\theta}-\underline{\theta}) + \underline{\theta} + \frac{j-1}{K}(\overline{\theta}-\underline{\theta}) + \underline{\theta}\right) \\
&= \frac{2j-1}{2K}(\overline{\theta}-\underline{\theta}) + \underline{\theta},
\end{aligned}$$

for all $j=1,\dots,K$. This concludes the proof. ■

Lemma 22 *In an interval PBE* (σ,ρ,d), $|I_{j+1}| = |I_j| + 4bK$ *for each* $j=1,\dots,n-1$.

Proof of Lemma 22. Let (σ, ρ, d) be an interval PBE. Pick arbitrary $I_{j+1}, I_j \in \sigma^{-1}(M)$. Suppose the Sender sends messages m and m' when the state belongs to I_j and I_{j+1}, respectively. Therefore, the Sender must be indifferent between $\max \rho(m) = \theta_j - \frac{|I_j|}{2K}$ and $\min \rho(m') = \theta_j + \frac{|I_{j+1}|}{2K}$ at the state θ_j. That is, $\theta_j + b$ must be equidistant from $\theta_j - \frac{|I_j|}{2K}$ and $\theta_j + \frac{|I_{j+1}|}{2K}$; i.e., $\theta_j + b - \left(\theta_j - \frac{|I_j|}{2K}\right) = \left(\theta_j + \frac{|I_{j+1}|}{2K}\right) - \theta_j - b$. Solving yields $|I_{j+1}| = |I_j| + 4bK$. ■

Corollary 7 *In an interval PBE,* $(n-1)n < \frac{1}{2bK}$.

Proof of Corollary 7. Let (σ, ρ, d) be an interval PBE. Suppose the initial interval is $[0, x)$ for some $x \in (0,1)$. Therefore, the total length of all n intervals in $\sigma^{-1}(M)$ satisfies

$$\begin{aligned} 1 &= nx + 4bK \sum_{j=1}^{n-1} j \\ &= nx + 4bK \frac{(n-1)n}{2}. \end{aligned}$$

Consequently, it must be that $2bK(n-1)n < 1$. This concludes the proof. ■

Lemma 23 *Let* σ *be an interval communication rule with intervals* $\sigma^{-1}(M) = \{I_1, I_2, \ldots, I_n\}$. *If* $|I_{q+1}| = |I_q| + 4bK$ *for each* $q = 1, \ldots, n-1$, *then for each* $m, m' \in \sigma(\Theta)$ *with* $\sigma^{-1}(m) = I_j$, $\sigma^{-1}(m') = I_l$, *and* $l > j$, *there exists* $\hat{\theta} \in [\theta_j, \theta_{l-1}]$ *such that the Sender is indifferent between* $\max \rho(m)$ *and* $\min \rho(m')$ *at the state* $\hat{\theta}$.

Proof of Lemma 23. Let σ be an interval communication rule with intervals $\sigma^{-1}(M) = \{I_1, I_2, \ldots, I_n\}$. Assume that $|I_{j+1}| = |I_j| + 4bK$ for each $j = 1, \ldots, n-1$.

Let $I_j = (\theta_j, \theta_{j+1})$ and $I_{j+l} = (\theta_{j+l}, \theta_{j+l+1})$ be two intervals in $\sigma^{-1}(M)$ with $l \geqslant 1$. Write m and m' for the messages such that $\sigma^{-1}(m) = I_j$ and $\sigma^{-1}(m') = I_{j+l}$. First, note that the Sender is indifferent between $\max \rho(m)$ and $\min \rho(m')$ at state

$$\hat{\theta} = \frac{\theta_{j+1} + \theta_{j+l}}{2} + \frac{|I_{j+l}| - |I_j|}{4K} - b \in [\theta_{j+1}, \theta_{j+l}].$$

Indeed,

(i) By assumption, $\frac{|I_{j+l}|-|I_j|}{4K} \geqslant \frac{4bK}{4K} = b$. Therefore,

$$\begin{aligned}\hat{\theta} &= \frac{\theta_{j+1}+\theta_{j+l}}{2} + \frac{|I_{j+l}|-|I_j|}{4K} - b \\ &\geqslant \frac{\theta_{j+1}+\theta_{j+1}}{2} \\ &= \theta_{j+1}.\end{aligned}$$

(ii) Note, $\hat{\theta} \leqslant \theta_{j+l}$ is equivalent to

$$\frac{|I_{j+l}|-|I_j|}{4K} \leqslant \frac{\theta_{j+l}-\theta_{j+l+1}}{2} + b = \frac{\sum_{m=j+1}^{j+l-1}|I_m|}{2} + b.$$

We show that this inequality holds by induction on the number of intervals between I_{j+l} and I_j: $l-1 \geqslant 0$.

- Suppose there are no intervals in between I_{j+l} and I_j. That is, suppose $l = 1$. Then, by assumption $\frac{|I_{j+l}|-|I_j|}{4K} = \frac{4Kb}{4K} = b$, and $\frac{\sum_{m=j+1}^{j+l-1}|I_m|}{2} + b = b$. Thus, the inequality holds.
- Suppose that the inequality holds when there $l-1 \geqslant 0$ intervals in between. That is, suppose that

  $$\frac{|I_{j+l}|-|I_j|}{4K} \leqslant \frac{\sum_{m=j+1}^{j+l-1}|I_m|}{2} + b.$$

  Now, observe that

  $$\begin{aligned}\frac{|I_{j+l+1}|-|I_j|}{4K} &= \frac{|I_{j+l+1}|-|I_{j+l}|}{4K} + \frac{|I_{j+l}|-|I_j|}{4K} \\ &\leqslant b + \frac{\sum_{m=j+1}^{j+l-1}|I_m|}{2} + b \\ &\leqslant \frac{|I_{j+l}|}{2} + \frac{\sum_{m=j+1}^{j+l-1}|I_m|}{2} + b \\ &= \frac{\sum_{m=j+1}^{j+l}|I_m|}{2} + b,\end{aligned}$$

  where the second inequality follows from the inductive hypothesis, and the third follows from the fact that $\frac{|I_{j+l}|}{2} \geqslant \frac{4bK}{2} \geqslant b$.

■

Let σ be an interval communication rule. Suppose $\sigma^{-1}(m) = I_j = (\theta_j, \theta_{j+1})$ for some $m \in M$. Write $\rho(m) = \{a_1^j, a_2^j, \ldots, a_K^j\}$, and $c_0^j = \theta_j$, $c_K^j = \theta_{j+1}$, and $c_l^j = \frac{a_l^j + a_{l+1}^j}{2}$ for $l = 1, \ldots, K-1$. Notice, $\Theta(a_l^j \mid \rho(m)) \cap [\theta_j, \theta_{j+1}] = [c_{l-1}^j, c_l^j]$ for each $l = 1, \ldots, K$.

Lemma 24 *Let σ be an interval communication rule with $\sigma^{-1}(m) = I_j = (\theta_j, \theta_{j+1})$ for some $m \in M$. Let $a > \max \rho(m) = a_K^j$. If $u_S(a_K^j, c_K^j) \geqslant u_S(a, c_K^j)$, then $u_S(a_l^j, \theta) \geqslant u_S(a, \theta)$ for each $\theta \in [c_{l-1}^j, c_l^j]$, for every $a_l^j \in \rho(m)$.*

Proof of Lemma 24. Assume that $u_S(a_K^j, c_K^j) \geqslant u_S(a, c_K^j)$. Since u_S satisfies increasing differences, $u_S(a_K^j, \theta) \geqslant u_S(a, \theta)$ for each $\theta \in [c_{K-1}^j, c_K^j]$.

Now, assume that $u_S(a_l^j, \theta) \geqslant u_S(a, \theta)$ for each $\theta \in [c_{l-1}^j, c_l^j]$ for some $l \leqslant K$. In particular, $u_S(a_l^j, c_l^j) - u_S(a, c_l^j) \geqslant 0$ is equivalent to $a > a_{l+1}^j + 2b$. Consequently, $a > a_l^j + 2b$, which is equivalent to $u_S(a_{l-1}^j, c_{l-1}^j) - u_S(a, c_{l-1}^j) \geqslant 0$. Since u_S satisfies increasing differences, $u_S(a_{l-1}^j, \theta) \geqslant u_S(a, \theta)$ for each $\theta \in [c_{l-2}^j, c_{l-1}^j]$. ■

Lemma 25 *Let σ be an interval communication rule with $\sigma^{-1}(m) = I_j = (\theta_j, \theta_{j+1})$ for some $m \in M$. If $a < \min \rho(m) = a_1^j$, then $u_S(a_l^j, \theta) \geqslant u_S(a, \theta)$ for each $\theta \in [c_{l-1}^j, c_l^j]$, for every $l \geqslant 2$.*

Proof of Lemma 25. Notice, $u_S(a_l^j, c_{l-1}^j) - u_S(a, c_{l-1}^j) = (a - a_l^j)(a + a_l^j - 2c_{l-1}^j - 2b) = (a - a_l^j)(a - a_{l-1}^j - 2b)$. Since $a \leqslant a_l^j$ and $a \leqslant a_{l-1}^j + 2b$, $u_S(a_l^j, c_{l-1}^j) - u_S(a, c_{l-1}^j) \geqslant 0$. ■

Proof of Proposition 5. The direct implication follows from Lemma 22.

Now, consider an interval strategy profile (σ, ρ, d) such that $|I_{j+1}| = |I_j| + 4bK$ for each $j = 1, \ldots, n-1$, and (ρ, d) satisfies Lemmas 6-7.

It remains to show that σ is incentive compatible. Fix $\theta \in \Theta$. Assume, without loss of generality, that $\theta \in I_j$. We want to show that the Sender prefers to reveal that the state belongs to I_j rather than to any other interval in $\sigma^{-1}(M)$.

Let I_l be an arbitrary interval in $\sigma^{-1}(M)$ with $l \neq j$. Consider the following cases:

- Assume $j < l$. By Lemma 23, there exists $\hat{\theta} \in [\theta_{j+1}, \theta_l]$ such that the Sender is indifferent between a_K^j and a_1^l. Since u_S has increasing differences, $u_S(a_K^j, \theta_{j+1}) \geqslant u_S(a_1^l, \theta_{j+1})$. Consequently, Lemma 24 implies that $u_S(a_q^j, \theta) \geqslant u_S(a_1^l, \theta)$ for each $\theta \in$

$[c^j_{l-1}, c^j_l]$, for every $q = 1, \ldots, K$. Since $I_j \subseteq \operatorname{int}\Theta(a^l_1 \mid \{a^l_1, \ldots, a^l_K\})$, the Sender prefers to reveal that the state belongs to I_j rather than to I_l.

- Assume $l < j$. By Lemma 23, there exists $\hat{\theta} \in [\theta_{l+1}, \theta_j]$ such that the Sender is indifferent between a^l_K and a^j_1. Since u_S has increasing differences, $u_S(a^j_1, \theta) \geqslant u_S(a^l_K, \theta)$ for each $\theta \in [\theta_j, c^j_1]$. In addition, since $a^l_K < a^j_1$, Lemma 25 guarantees that $u_S(a^j_q, \theta) \geqslant u_S(a^l_K, \theta)$ for each $\theta \in [c^j_{q-1}, c^j_q]$, for every $q \geqslant 2$. Finally, since $I_j \subseteq \operatorname{int}\Theta(a^l_K \mid \{a^l_1, \ldots, a^l_K\})$, we conclude that the Sender prefers to reveal that the state belongs to I_j rather than to I_l.

This concludes the proof. ■

References

A. Acharya and J. Ortner. Progressive learning. *Econometrica*, 85(6):1965–1990, 2017.

A. R. Admati and P. Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

S. N. Ali, N. Haghpanah, X. Lin, and R. Siegel. How to sell hard information. *The Quarterly Journal of Economics*, 137(1):619–678, 2022.

R. Alonso and O. Câmara. Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165:672–706, 2016.

N. Antic, A. Chakraborty, and R. Harbaugh. Subversive conversations. *Working paper*, 2023.

K. J. Arrow. *Information and economic behavior*, volume 28. Federation of Swedish Industries Stockholm, 1973.

R. Aumann. Agreeing to disagree. *The annals of statistics*, 4(6):1236–1239, 1976.

R. J. Aumann, M. Maschler, and R. E. Stearns. *Repeated games with incomplete information.* MIT press, 1995.

Y. Aybas and E. Turkel. Persuasion with coarse communication. *Working paper*, 2024.

G. Bauch. Effects of noise on the grammar of languages. *Working paper*, 2024.

D. Bergemann and A. Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11:85–107, 2019.

D. Bergemann, A. Bonatti, and A. Smolin. The design and price of information. *American Economic Review*, 108(1):1–48, 2018.

D. Bergemann, A. Bonatti, and T. Gan. The economics of social data. *The RAND Journal of Economics*, 53(2):263–296, 2022.

H. Bester and R. Strausz. Contracting with imperfect commitment and the revelation principle: the single agent case. *Econometrica*, 69(4):1077–1098, 2001.

J. Bizzotto, J. Rüdiger, and A. Vigier. Dynamic persuasion with outside information. *American Economic Journal: Microeconomics*, 13(1):179–194, 2021.

D. Blackwell. Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, volume 2, pages 93–103. University of California Press, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

A. Blume. Coordination and learning with a partial language. *Journal of economic theory*, 95(1):1–36, 2000.

A. Blume. A learning-efficiency explanation of structure in language. *Theory and decision*, 57(3):265–285, 2004.

A. Blume. Meaning in communication games. *Working paper*, 2024.

A. Blume and O. Board. Language barriers. *Econometrica*, 81(2):781–812, 2013.

A. Blume and I. Deimen. Strategic information transmission in the employment relationship. *Working paper*, 2024.

A. Blume, I. Deimen, and S. Inoue. Incomplete contracts versus communication. *Journal of economic theory*, 205:105544, 2022.

J. W. Carr, K. Smith, H. Cornish, and S. Kirby. The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive science*, 41(4):892–923, 2017.

Y.-K. Che and N. Kartik. Opinions as incentives. *Journal of Political Economy*, 117(5): 815–860, 2009.

Y.-K. Che, K. Kim, and K. Mierendorff. Keeping the listener engaged: a dynamic model of bayesian persuasion. *Journal of Political Economy*, 131(7):000–000, 2023.

V. Crawford and J. Sobel. Strategic information transmission. *Econometrica*, 50(6): 1431–1451, 1982a. ISSN 0012-9682.

V. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982b.

J. Crémer, L. Garicano, and A. Prat. Language and the theory of the firm. *The Quarterly Journal of Economics*, 122(1):373–407, 2007. ISSN 0033-5533.

I. Deimen and D. Szalay. Delegated expertise, authority, and communication. *The American Economic Review*, 109(4):1349, 2019. ISSN 0002-8282. doi: 10.1257/aer.20161109.

F. Dilmé. Precision vs complexity: Efficient communication and optimal codes. *Working paper*, 2024.

L. Doval and V. Skreta. Mechanism design with limited commitment. *Econometrica*, 90(4): 1463–1500, 2022.

K. Drakopoulos and A. Makhdoumi. Providing data samples for free. *Management Science*, 69(6):3536–3560, 2023.

J. Ely, A. Frankel, and E. Kamenica. Suspense and surprise. *Journal of Political Economy*, 123(1):215–260, 2015.

J. C. Ely. Beeps. *American Economic Review*, 107(1):31–53, 2017.

J. C. Ely and M. Szydlowski. Moving the goalposts. *Journal of Political Economy*, 128(2): 468–506, 2020.

M. Escudé and L. Sinander. Slow persuasion. *Theoretical Economics*, 18(1):129–162, 2023.

J. Glazer and A. Rubinstein. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior*, 36(2):158–173, 2001.

J.-M. Grandmont. Intermediate preferences and the majotiry rule. *Econometrica*, 46(2): 317–330, 1978.

J. Green and N. Stokey. A two-person game of information transmission. *Journal of Economic Theory*, 135:90–104, 2007.

M. D. Grubb. Selling to overconfident consumers. *American Economic Review*, 99(5):1770–1807, 2009.

P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. The MIT Press, 2000.

P. Gärdenfors. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press, 2014.

J. Y. Halpern. Characterizing the common prior assumption. *Journal of Economic Theory*, 106(2):316–355, 2002.

J. C. Harsanyi. Games with incomplete information played by "bayesian" players part ii. bayesian equilibrium points. *Management Science*, 14(5):320–334, 1968.

T. Heumann. On the cardinality of the message space in sender–receiver games. *Journal of mathematical economics*, 90:109–118, 2020.

J. Hörner and A. Skrzypacz. Selling information. *Journal of Political Economy*, 124(6): 1515–1562, 2016.

S. Ichihashi. The economics of data externalities. *Journal of Economic Theory*, 196:105316, 2021.

G. Jäger. The evolution of convex categories. *Linguist and Philos*, 30:551–564, 2007.

G. Jäger, T. de Jager, M. Aloni, K. Schulz, and H. Bastiaanse. *Natural Color Categories Are Convex Sets*, page 11–20. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 3642142869. doi: 10.1007/978-3-642-14287-1_2.

G. Jäger, L. P. Metzger, and F. Riedel. Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior*, 73(2):517, 2011.

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

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

N. Kartik and A. Kleiner. Convex choice. *Working paper*, 2024.

N. Kartik, S. Lee, and D. Rappoport. Single-crossing differences in convex environments. *The Review of Economic Studies*, 91(5):2981 – 3012, 2024.

S. Kirby. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, 24(1):118–137, 2017.

A. Kolotilin, H. Li, and W. Li. Optimal limited authority for principal. *Journal of Economic Theory*, 148:2344–2383, 2013.

V. Krishna and J. Morgan. The art of conversation: eliciting information from experts through multi-stage communication. *Journal of Economic Theory*, 117(2):147–179, 2004.

V. Krishna and J. Morgan. Contracting for information under imperfect commitment. *The RAND Journal of Economics*, 39(4):905–925, 2008.

R. Lang. A note on the measurability of convex sets. *Archiv der Mathematik*, 47:90–92, 1986.

B. Lipman. Why is Language Vague? *Working paper*, 2009.

B. Lipman. Why is Language Vague? *International Journal of Game Theory*, 54(8), 2025.

E. Lipnowski, D. Ravid, and D. Shishkin. Persuasion via weak institutions. *Journal of Political Economy*, 130:2705–2730, 2022.

J. Miao. *Economic dynamics in discrete time.* MIT press, 2020.

P. Milgrom and C. Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, 1994.

P. R. Milgrom and R. J. Weber. Distributional strategies for games with incomplete information. *Mathematics of operations research*, 10(4):619–632, 1985.

S. Morris. The common prior assumption in economic theory. *Economics & Philosophy*, 11 (2):227–253, 1995.

L. Rayo and I. Segal. Optimal information disclosure. *Journal of political Economy*, 118(5): 949–987, 2010.

J. Renault, E. Solan, and N. Vieille. Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349, 2017.

A. Rubinstein. Why are certain properties of binary relations relatively more common in natural language? *Econometrica*, 64(2):343–355, 1996.

G. Saint-Paul. A "quantized" approach to rational innatention. *European Economic Review*, 100:50–71, 2017.

L. J. Savage. *The foundations of statistics.* Courier Corporation, 1972.

J. Sobel. Broad terms and organizational codes. *Working paper*, 2016.

N. L. Stokey. *Recursive methods in economic dynamics.* Harvard University Press, 1989.

T. Suzuki. Endogenous ambiguity and rational miscommunication. *Journal of Economic Theory*, 211:1056–1086, 2023.

Z. G. Szabó. Compositionality. In E. N. Zalta and U. Nodelman, editors, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Fall 2024 edition, 2024.

S. Zheng and Y. Chen. Optimal advertising for information products. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 888–906, 2021.

W. Zhong. Optimal dynamic information acquisition. *Econometrica*, 90(4):1537–1582, 2022.