THREE ESSAYS ON INFORMATION TRANSMISSION GAMES AND BELIEFS IN PERFECT INFORMATION GAMES

by

Yun Wang

B.A., Peking University, 2008

M.A., University of Pittsburgh, 2010

Submitted to the Graduate Faculty of

the Department of Economics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF ECONOMICS

This dissertation was presented

by

Yun Wang

It was defended on

June 21st, 2013

and approved by

Dr. Andreas Blume, University of Arizona, Economics

Dr. John Duffy, University of Pittsburgh, Economics

Dr. Sourav Bhattacharya, University of Pittsburgh, Economics

Dr. Isa Hafalir, Carnegie Mellon University, Tepper School of Business

Dr. Luca Rigotti, University of Pittsburgh, Economics

Dr. Roee Teper, University of Pittsburgh, Economics

Dissertation Advisors: Dr. Andreas Blume, University of Arizona, Economics,

Dr. John Duffy, University of Pittsburgh, Economics

THREE ESSAYS ON INFORMATION TRANSMISSION GAMES AND BELIEFS IN PERFECT INFORMATION GAMES

Yun Wang, PhD

University of Pittsburgh, 2013

This dissertation consists of three chapters. The first two chapters study strategic information transmission. The third chapter studies beliefs in perfect information games.

The first chapter examines strategic information transmission with interacting decision-makers. I analyze a cheap-talk game between an informed agent and two uninformed decision-makers who coordinate their actions. I compare public communication with private communication. I find that the agent responds to the decision-makers' coordination by providing less precise information. Conditions that support a full information revelation equilibrium in private communication also support the same type of equilibrium in public communication; but the reverse is not true.

The second chapter investigates the role of persuasion mechanisms in collective decision-making. A persuasion mechanism consists of a family of conditional distributions over the underlying state space. A biased, perfectly informed sender adopts a persuasion mechanism to provide a group of uninformed receivers with signals about the unknown state of the world. I compare public persuasion with private persuasion. I find that the sender can always reach the convex upper bound of the set of expected payoffs under public persuasion, regardless of the number of signals or the signals' correlation structure. The sender is worse off under private persuasion. Moreover, I show that private persuasion is always more informative than its public counterpart. As a result, the receivers make better decisions under private persuasion.

The third chapter experimentally explores people's beliefs behind the failure of backward induction in the centipede games. I elicit players' beliefs about opponents' strategies and 1st-order beliefs. I find that subjects maximize their monetary payoffs according to their stated beliefs less frequently in the Baseline Centipede treatment where an efficient non-equilibrium outcome exists; they do so more frequently in the Constant Sum treatment where the efficiency property

is removed. Moreover, subjects believe their opponents' maximizing behavior and expect their opponents to hold the same belief less frequently in the Baseline Centipede treatment and more frequently in the Constant Sum treatment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My immense gratitude to my primary advisor, Andreas Blume, for his dedication, guidance and support throughout the five years of my graduate study. I am also grateful to John Duffy, who shares his deep insights and astounding breadth of knowledge in many discussions. I thank other committee members, Sourav Bhattacharya, Isa Hafalir, Luca Rigotti, and Roee Teper for their long-time encouragement, help and guidance. I am deeply indebted to all my committee members.

I thank Professor Pierre Liang, Thomas Rawski, and Alistair Wilson, who have generously shared their insights during every conversation, discussion, and seminar presentation. I also thank my peers at the University of Pittsburgh for their help, comments, and suggestions. Especial gratitude to my senior classmate Zhuomin Liu, whose perfectionism in research has consistently influenced me over the years.

Last but not least, I express my endless gratitude and love to my father, Qingxin Wang, and mother, Meiqi Shi, for inspiring my intellectual curiosity from the earliest age.

1.0 GOING PUBLIC OR GOING PRIVATE? CHEAP-TALK WITH INTERACTING DECISION-MAKERS

1.1 INTRODUCTION

This paper studies strategic communication with interacting decision-makers. In many economic settings, decision-makers get informed from a certain information source before taking actions that affect each other's payoffs. Morris and Shin (2002, 2007) investigate a situation where each decision-maker wants an action close to both the underlying state and the average of all others' choices. The authors examine the role of public information that comes from a non-strategic social planner. Yet in reality the information source that the decision-makers consult is often biased. Farrell and Gibbons (1989) consider a game of information transmission between a self-interested agent and two uninformed decision-makers. The authors compare the conditions that support an informative equilibrium under public communication versus that under private communication. But their framework omits the strategic complementarity between decision-makers' actions. In this paper we intend to bring the two approaches together.

We analyze an agent's cheap talk communication with interacting decision-makers (hereafter "principals"). There are two major departures from Morris and Shin (2002, 2007): first, we allow heterogeneity for the principals' desire to adapt to the underlying state and for their coordination motives; whereas Morris and Shin consider only homogeneity among the principals. Second, we investigate the information transmission from a self-interested agent, who privately observes the underlying state of the world. The agent can either send cheap talk messages to both principals publicly, or to each principal privately, but not both. While Morris and Shin (2002) assume that the public information is from a benevolent planner who maximizes the social welfare and the private information is from another un-biased information source. We compare public communication with private communication. We show, in the latter environment, that the agent responds to the principals' strategic complementary actions by providing less precise information. For any partition equilibrium with the same number of partition elements, the agent faces tighter incentive compatibility constraints under private communication. Moreover, Morris and Shin (2002) show that the principals strictly benefit from public information only when no private information is of high precision. We find that both the agent and the principals are better off under public communication. This is true even when the equilibrium under each type of communication have the same level of information precision.

Our finding regarding the agent's information revelation incentives appears to be similar to Farrell and Gibbons (1989). They show that whenever there exists an equilibrium in which the agent communicates informatively under private communication, there is an equilibrium under public communication in which the agent does the same; but not vice versa. We also find that the agent's incentive compatibility constraints are more restrictive under private communication than that under public communication. Nevertheless, the underlying reason of our result is different from that of Farrell and Gibbons (1989). In our model with strategic complementarity in principals' actions, under public communication the principals take actions based on the same posterior belief. This is because public information induces common belief. Here public messages from the agent serve as a coordination device for the principals and enhance both principals' welfare. On the other side, in our model private communication still does not allow the agent to tell different lies to different principals. The principals' actions are not completely uncorrelated even though the agent communicates with each principal separately. Any change in the agent's "talking" strategy with one principal will also affect the other principal's best response. For the agent's incentive compatibility in any informative equilibrium, all possible deviations and the consequent changes in the principals' induced actions need to be excluded. Therefore, the agent is more tightly constrained under private communication.

Meanwhile, we rank the agent's and the principals' welfare over the two types of communication for equilibria with the same number of partition elements. We show, by a numerical example, that both the agent and the principals are better off under public communication. The agent's ex-ante expected utility under public communication weakly dominates that under private communication; while for both principals' the dominance relation is strict. Moreover, the gap between the agent's expected utilities under the two types of communication shrinks as one principal's coordination motive grows. On the other hand, the gap between the principals' expected utilities may either shrink or expand depending on that principal's relative weight on the strategic complementary term in his utility.

We also show, under both types of communication, that equilibria take the partition form as characterized by Crawford and Sobel (1982). The agent partitions the continuous state space into finite number of elements and sends only one message representing the partition element in which the true underlying state lies. For any given equilibria, the higher the number of partition elements, the more precise information the agent's messages convey. We show, under public communication, that the maximum number of partition elements decreases monotonically with the principals' biases, i.e. the larger the biases, the smaller the incentives for the agent to provide more precise information. Moreover, it changes non-monotonically with the principals' coordination motives. The pattern of the change depends on principals' relative bias in preferences. In other words, as the coordination need of a principal whose interest is less misaligned with the agent grows larger, the effectiveness of communication reduces. Intuitively, coordinating with somebody else who stands further from the agent' ideal position undermines the agent's well-being.

Additionally, we investigate possible communication improvement compared to the cheap talk game without the principals' strategic complementarity in actions (Goltsman and Pavlov (2011)). We say communication improves if in our model there exists a least informative equilibrium supported by a range of the parameters which does not support any informative equilibrium in Goltsman and Pavlov (2011). We find that communication improvement exists under public communication but not under private communication. Moreover, communication improvement under public communication only happens when the principals attach different weights on the strategic complementary terms in their actions.

The remainder of the paper is organized as follows. Section 1.2 introduces the model, strategies, and equilibrium concept. Section 1.3 provides the conditions that support full revealing equilibrium under public and private communication, respectively. Section 1.4 proves the existence of partition equilibrium and derives the maximum size of partition equilibria under public and private communication, respectively. This section also compares the agent's incentives to partially reveal information under both types of communication. Section 1.5 compares the agent's welfare and the principals' welfare under both types of communication, respectively. We show, by a numerical example, that both the agent and the principals are better off under public communication. Section 1.6 reviews the related literature on strategic information transmission and on organization design. Section 1.7 concludes.

1.2 THE MODEL

Suppose there are three players, one agent(S, "she") and two principals (R_1 and R_2, "he"). A random variable, t, denotes the underlying state of the world. As the game starts, nature determines the realization of $t \in T = [0, 1]$. The prior probability distribution $t \sim F$ over T is commonly known to all participants. The agent observes the state of the world, t, while the principals do not. Each receiver i takes an action $a_i \in A_i = \mathbf{R}, i = 1, 2$ simultaneously and independently. The receivers' actions and the realized state of the world determine all players' utilities:

- The agent's utility:
$$U^S(a_1, a_2, t) = -\ell_1(|a_1 - t|) - \ell_2(|a_2 - t|)$$
where $\ell_i \in \mathcal{C}^2, \ell_i'(\cdot) \geq 0, \ell_i''(\cdot) > 0, \forall i = 1, 2$
- Each principal's utility:
$$U_i^R(a_1, a_2, b_i, t) = -L_i(|a_i - t - b_i|) - \Lambda_i(|a_1 - a_2|)$$
where $L_i \in \mathcal{C}^2, L_i'(\cdot) \geq 0, L_i''(\cdot) > 0, \Lambda_i'(\cdot) > 0, \Lambda_i''(\cdot) > 0, b_i \in \mathbf{R}, \forall i = 1, 2$

The utility functions indicate that the players' interests are not perfectly aligned. In every state t, the agent's objective is to induce a_1 and a_2 that exactly coincide with the current state. But the two principals have two-fold concerns: on one hand, each principal has an ideal action, $t + b_1$ and $t + b_2$ respectively; on the other hand, each receiver cares about the distance between the induced actions. The closer the two actions, a_1 and a_2, the smaller the losses for both principals. We call this effect the "coordination concern" and model it with a twice differentiable concave function $\Lambda_i(\cdot)$[1]. We allow the $\Lambda(\cdot)$ function to be different to capture heterogenous coordination concerns across different principals.

A special case of this model is the uniform-quadratic form which are widely employed by the literature: assuming $t \sim U[0, 1]$, $\ell_i(\cdot), L_i(\cdot)$ functions to be of the quadratic form, and $\Lambda_i(|a_1 - a_2|) = \lambda_i \cdot (a_1 - a_2)^2, \lambda_i \in R_+, \forall i = 1, 2$. A simple algebra shows that attaching strictly positive λ_i to the quadratic utility loss is equivalent to attaching weights $1/(1 + \lambda_i)$ and $\lambda_i/(1 + \lambda_i)$ to principal i's own preferred action and the coordination concern, respectively.

[1]Here, by "coordination" we refer to *tacit* coordination, namely, coordination without explicit conversation between the two principals. There are some work investigating coordination with the decision-makers' explicit communication, a topic of which we do not discuss in this paper.

1.2.1 Public Communication

Under public communication, the agent sends costless and nonverifiable messages $m \in M$ to the principals. The agent is only allowed to send a message publicly, i.e. she is bound to send messages that will be publicly observed by both principals.

1.2.1.1 Strategies and Weak Perfect Bayesian Nash Equilibrium

We now specify each player's strategy and the principals' beliefs upon receiving the agent's signal, under the public communication protocol:

- The agent's strategy, $\tau : T \rightarrow \Delta M$, is a mapping from the state space to the set of distributions over the message space. The message she finally chooses can be observed by both principals publicly.
- Each principal's strategy, $\alpha_i : M \rightarrow A_i, i = 1, 2$, is a mapping from the message space to his action space. Potentially a principal could mix over the action space A_i, but he has a unique best response to each belief. Therefore, we replace ΔA_i with A_i. Upon observing the agent's message, each principal chooses his action separately, which determines the utilities of all players.
- Each principal's belief, $\mu_i : M \rightarrow \Delta T, i = 1, 2$, is a mapping from the message space to the set of distributions over the state space. Upon receiving the agent's message, each principal updates his belief about the true state t.

We adopt Weak Perfect Bayesian Nash Equilibrium as the equilibrium concept throughout the rest of the paper. Such equilibrium requires that the agent's strategy maximizes her ex-ante expected utility in every possible state given each principal's strategy, that each principal's strategy maximizes his expected utility given the agent's revelation strategy and the other principal's best response, and that each principal makes inferences about the realized value of t based on observed messages via the Bayes' Rule wherever possible, i.e. at every information set along every on-equilibrium paths with $\forall m \in M$ within the support of $\tau(\cdot)$:

- Equilibrium Strategy Profile $(\tau^*, \alpha_1^*, \alpha_2^*)$
- For all $t \in [0, 1]$, define $\tau(m|t)$ as the agent's talking strategy given t. $\int_0^1 \tau(m|t)dm = 1$ and

$$m \in \arg\max_{m'} U^S(a_1^*(m'), a_2^*(m'); t, b_1, b_2, \lambda_1, \lambda_2)$$

- For all $m \in M$ and $i = 1, 2$, let $a_i, i = 1, 2$ denote the pure strategy according to α_i

$$a_i^*(m) = \arg\max_{a'} \int_0^1 U^{R_i}(a_1', a_2'; t, b_1, b_2, \lambda_1, \lambda_2)\mu_i^*(t|m)dt$$

- Each receiver's belief $\mu_i^*(t|m)$ is derived from Bayes' Rule whenever possible: $\mu_i^*(t|m) = \frac{\tau^*(m|t)f(t)}{\int_0^1 \tau^*(m|t)f(t)dt}$

1.2.2 Private Communication

Under private communication, the agent can send messages to each principal separately and privately. Each principal observes only the messages that are designated for him, but not the ones for his counterpart.

1.2.2.1 Strategies and Weak Perfect Bayesian Nash Equilibrium

Again, we specify each player's strategy and the principals' beliefs upon receiving the agent's signals, under the private communication protocol:

- The agent's strategy, $\tau : T \to \Delta(M \times M)$, is a mapping from the state space to the set of distributions over the message space. Denote $\tau(\cdot) = (\tau_1(\cdot), \tau_2(\cdot))$. Privately informed about t, she chooses two messages and communicates with each principal separately.
- Principal i's strategy, $\alpha_i : M \to A_i, i = 1, 2$, is a mapping from the message space to the set of probability distributions over his action space. Each principal has a unique best response to each belief. Therefore, we can replace ΔA_i with A_i.
- Principal i's belief, $\mu_i : M \to \Delta T, i = 1, 2$, is a mapping from the message space to the set of distribution over the state space. Upon observing the agent's message, principal i updates his expectation about the true state variable t.

Weak Perfect Bayesian Nash Equilibrium requires that the agent's strategy maximizes her expected utility in every possible state given each principal's strategies, that each principal's strategy maximizes his expected utility given the agent's strategy and the other principal's best response, and that each principal makes inference about the realized value of t based on observed messages via Bayes' Rule wherever possible, i.e. at every information set along every on-equilibrium paths with $\forall m_i \in M, i = 1, 2$ within the support of $\tau(\cdot)$:

- Equilibrium Strategy Profile $(\tau_1^*, \tau_2^*, \alpha_1^*, \alpha_2^*)$

- For all $t \in [0,1]$, define $\tau_i^*(m_i|t), i = 1,2$ as the agent's talking strategies given t. $\int_0^1 \tau_i(m_i|t)dm_i = 1, i = 1,2$ and

$$m_1 \in \arg\max_{m'} U^S(a_1^*(m'), a_2^*(m'); t, b_1, b_2, \lambda_1, \lambda_2)$$
$$m_2 \in \arg\max_{m'} U^S(a_1^*(m'), a_2^*(m'); t, b_1, b_2, \lambda_1, \lambda_2)$$

- For all $m_i \in M$ and $i = 1,2$, let $a_i, i = 1,2$ denote the pure strategy according to α_i

$$a_i^*(m_i) = \arg\max_{a'} \int_0^1 U^{R_i}(a_1', a_2'; t, b_1, b_2, \lambda_1, \lambda_2)\mu_i^*(t|m_i)dt$$

- Each principal's belief $\mu_i^*(t|m)$ is derived from Bayes' Rule whenever possible: $\mu_i^*(t|m_i) = \frac{\tau_i^*(m_i|t)f(t)}{\int_0^1 \tau_i^*(m_i|t)f(t)dt}$

1.3 FULL INFORMATION REVELATION

In this section we derive conditions that support *separating equilibria* under public and private communication, respectively. In any such equilibrium, each agent type separates herself by sending distinct messages which can be interpreted as and only as from the true type; and each principals knows the agent's exact type and makes decision with complete information. Thus we call a separating equilibrium *full-revealing.* The following proposition establishes the conditions under which full-revealing can be supported under each type of communication. Moreover, it shows that the incentive compatibility constraints are tighter for the agent to reveal full information under private communication than that under public communication.

Proposition 1. *Conditions for full information revelation*

1. *In the cheap-talk game with interacting principals, full information revelation under private communication implies full revelation under public communication; but not vice versa.*
2. *When the common prior is the uniform distribution $t \sim U[0,1]$, the utility loss $\ell_i(\cdot), L_i(\cdot), \forall i = 1,2$ are quadratic functions, and the principal's utility loss from mis-coordination is $\Lambda_i(|a_1 - a_2|) = \lambda_i \cdot (a_1 - a_2)^2, \lambda_i \in R_+, \forall i = 1,2$*
   a. *Under public communication, a separating equilibrium in which every agent type reveals the true realization of t to both principals publicly, exists if and only if*

$$\frac{(1+2\lambda_2)b_1 + (1+2\lambda_1)b_2}{1+\lambda_1+\lambda_2} = 0$$

b. Under private communication, a separating equilibrium in which every agent type reveals the true realization of the state variable t to each principal separately, exists if and only if $b_1 = b_2 = 0$.

The proof of this proposition is included in the Appendix. The underlying intuition is two-fold. The first part is similar to the reason underlying Farrell and Gibbons' (1989) main result: under public communication the agent's messages are observed by the principals at the same time; while private communication allows the agent to send different messages to different principals. Thus, in the latter case the agent's strategy changes from *one mapping* to *a pair of mappings* from the state space to the message space. A separating equilibrium imposes stronger restrictions on the parameter space in order to exclude all possible profitable deviations, compared to the non-deviation restriction imposed in the public communication. These profitable deviations include lying to Principal 1, lying to Principal 2, and lying to both principals at the same time. As the incentive compatibility constraints become tighter, the range of the parameters that support the separating equilibrium shrinks to $b_1 = b_2 = 0$.

The second reason is from the interacting term in the principals' utility functions in our model: the principal's utility functions are *interdependent* through the coordination term. Suppose, under private communication, agent type t deviates from full-revealing with Principal 1 while remains full-revealing with Principal 2. The direct consequence is that Principal 1's best response changes. Moreover, there is an indirect consequence. Since both principals care about coordination, Principal 2's best response function will also change accordingly. Even though the agent sends messages to each principal separately, any change in the "talking" strategy with one principal will also affect the other principal's best response. Therefore, the agent is more tightly constrained under private communication.

Part 2.a of Proposition 1 provides the condition for full information revelation under public communication. In Goltsman and Pavlov (2011) a parallel condition requires $b_1 + b_2 = 0$, which means the "aggregate bias" equals zero. In other words, if there is only one "representative" principal whose bliss action is the average of the two principals', separating equilibrium exists only when this principal's interest is perfectly aligned with the agent. In our model, the truth-telling condition $\frac{(1+2\lambda_2)b_1+(1+2\lambda_1)b_2}{1+\lambda_1+\lambda_2} = 0$ can be rewritten as $b_1 + \frac{\lambda_2-\lambda_1}{1+\lambda_1+\lambda_2}b_1 = -(b_2 + \frac{\lambda_1-\lambda_2}{1+\lambda_1+\lambda_2}b_2)$. The parameter b_i can be interpreted as principal i's "absolute bias" and the term $\frac{\lambda_j-\lambda_i}{1+\lambda_1+\lambda_2}b_i$ as "relative bias." Notice that the latter is the absolute bias times the weighed differences in coordination needs λ_i.Therefore, part 2.a states that when one principal's relative bias offsets with the others', there exists separating equilibrium in public communication.

Part 2.b of Proposition 1 shows the tightened condition that support a separating equilibrium under private communication. Here we have two remarks. First, $b_1 = b_2 = 0$ means that the principals do not have bias any longer. Their ideal actions are to coincide with the current state and with their co-decision-maker. Therefore all players's interests are perfectly aligned and the message(s) become a device to coordinate actions. Second, $b_1 = b_2 = 0$ implies $(1 + 2\lambda_2)b_1 + (1 + 2\lambda_1)b_2 = 0$, the condition for the existence of separating equilibrium in the public communication, but not vice versa. Hence, if there is separating equilibrium under private communication protocol, there exists one in the public communication, but the converse is not true.

1.4 PARTIAL INFORMATION REVELATION

1.4.1 Partition Equilibria

In this section, we investigate non-full revealing equilibria. Section 1.3 indicates that full revelation can be supported under very restricted conditions. When these conditions are not satisfied, the principals' actions a_1 and a_2 will not coincide with each other, nor will they coincide with the true state, t, as the agent wishes. Thus the agent is better off by not always revealing the true state. In the most extreme case, the agent adopts the same strategy $\tau(\cdot)$ in all states; the principals interpret all messages in the same way and take actions that maximize the expected utilities based on the prior probability distribution $F(\cdot)$. This type of "babbling" equilibrium conveys no information regarding the true state. The principals' posterior beliefs after receiving any message remains the same as the common prior belief. We formally define:

Definition 1. *An equilibrium is* informative *if at least one principal's posterior belief is different from the common prior belief with positive probability.*

As shown in Crawford and Sobel (1982), in a one-sender one-receiver cheap talk game all equilibria take the partition form as long as there is misalignment between the agent's and the principals' interests. A babbling equilibrium, an equilibrium with the coarsest partition, is uninformative. In contrast, any partition equilibrium with at least two partition elements is informative. In such an equilibrium the principals' posterior belief(s) has incorporated the probability that the true state variable falls in one of the partition elements.

Next, we characterize all non-separating equilibria. Similar to Crawford and Sobel (1982), we demonstrate that all partial-revealing equilibria are of the partition form; the set of action pairs induced in equilibrium is finite. The agent adopts semi-pooling revelation strategy, partitioning the state space $[0,1]$ into finitely many elements, in which all agent types pool to send the same message.

Proposition 2. *Equilibrium Characterization*

1. *For all $t \in [0,1]$, if the action pair (a_1^*, a_2^*) that solves $\partial U^S(a_1,a_2,t)/\partial a_1 = 0$ and $\partial U^S(a_1,a_2,t)/\partial a_2 = 0$ does not coincide with the one (a_1', a_2') that solves $\partial U_i^R(a_1,a_2,b_i,t)/\partial a_i = 0, i = 1,2$, then there exists a vector ϵ such that if two action pairs $\vec{u} = (\hat{a_1}, \hat{a_2})$ and $\vec{v} = (\tilde{a_1}, \tilde{a_2})$ are induced in equilibrium, $||\vec{u} - \vec{v}|| > \epsilon$. Moreover, the set of action pairs induced in equilibrium is finite.*
2. *Under public communication there exists a positive integer N_M, such that for every $n \in \boldsymbol{N}$ and $1 \leq n \leq N_M$, there exists at least one equilibrium in which all agent types within $t \in [t_i, t_{i+1}]$ with $i = 0, ..., n-1$ send the same message. Denote the principals' action pair induced by all agent types within $[t_i, t_{i+1}]$ as $(a_1([t_i, t_{i+1}]), a_2([t_i, t_{i+1}]))$:*
   - *The agent type on each boundary, namely, $t_i, i = 1, ...n$ is indifferent between inducing the action pair $(a_1([t_i, t_{i+1}]), a_2([t_i, t_{i+1}]))$ and $(a_1([t_{i-1}, t_i]), a_2([t_{i-1}, t_i]))$ where $i = 1...n$*
   - *The principals' best response profile upon receiving each message is a pair of actions $(a_1([t_{i-1}, t_i]), a_2([t_{i-1}, t_i]))$*
   - *The first and last boundary agent types are $t_0 = 0$ and $t_n = 1$ respectively.*
   - *When the common prior is the uniform distribution $t \sim U[0,1]$, the utility loss $\ell_i(\cdot), L_i(\cdot), \forall i = 1,2$ are quadratic functions, and the principal's utility loss from mis-coordination is $\Lambda_i(|a_1 - a_2|) = \lambda_i \cdot (a_1 - a_2)^2, \lambda_i \in R_+, \forall i = 1,2$, the size of each partition element changes by $\frac{2(b_1+b_2+2b_1\lambda_2+2b_2\lambda_1)}{1+\lambda_1+\lambda_2}$*

The proof of this result is in the Appendix. Equilibria with the partition form can be interpreted as imprecise information revelation when all player rationally account others' best responses and consistently update their beliefs. Although the agent is allowed to send any cheap-talk messages, she will not lie *in equilibrium*; rather, the agent speaks in an imprecise manner: instead of revealing the exact value of the true state variable t, she sends out a message representing the interval where t lies. Moreover, the finer the partition, the more precise the agent's communication is. So it is natural to ask how finest the partition could be. We investigate this question under public communication.

The following corollary provides the upper bound of the number of partition elements N_M as a function of $b_1, b_2, \lambda_1, \lambda_2$, when the common prior is the uniform distribution and the players' utility functions all take the quadratic form:

Corollary 1. *The number of equilibrium partition elements for given parameter values* $(b_1, b_2, \lambda_1, \lambda_2)$ *is bounded by*

$$N_M = \langle \frac{1}{2} + \frac{[(b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2) \times (b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2 + 4(1 + \lambda_1 + \lambda_2))]^{\frac{1}{2}}}{2(b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2)} \rangle$$

where $\langle x \rangle$ *represents the largest integer below the value of* x*.*

The proof of this corollary is in the Appendix. N_M is a measure for the maximum degree of information precision given the parameters $(b_1, b_2, \lambda_1, \lambda_2)$. Examining the marginal effects with respect to the parameters, we have

- $\partial N_M / \partial b_i < 0, i = 1, 2$.
- $\partial N_M / \partial \lambda_i > 0$ if $b_i > b_j$; while $\partial N_M / \partial \lambda_i < 0$ if $b_i < b_j$, $i = 1, 2$

The first partial derivative indicates that the value of N_M decreases monotonically with the principals' biases. Given $\lambda_1, \lambda_2, b_j$, the higher one principal's bias is, the smaller the maximum number of partition elements. In contrast, the second partial derivative shows a non-monotonic relation between N_M and the principals' coordination concerns: if principal i has a relatively larger bias, an increase in i's coordination need leads to an increase in the value of N_M; on the contrary, if i's bias is relatively smaller, the higher i's coordination concern, the smaller the value of N_M. In other words, when comparing the information precision with respect to the principals' coordination concerns, the agent cares more about the *relative* bias rather than *absolute* biases. Principal i with a larger bias seems undesirable for the agent, yet if he has a greater need to coordinate with the other principal whose bias is smaller, the agent is willing to share more precise information publicly. On the other hand, if principal i is originally less biased compared with the other principal but develops greater need to coordinate with his highly-biased co-decision-maker, the precision of communication will decrease. In a nutshell, coordinating with somebody else who stands farther away from the agent' ideal position undermines the agent's well-being.

Under public communication, for given values of $(b_1, b_2, \lambda_1, \lambda_2)$, the number of partition elements is bounded by the N_M as derived above. A natural question is how close is the finest partition equilibrium to a separating equilibrium. The next result provides the condition under which the finest partition equilibrium is sufficiently close to a full-revealing equilibrium.

Proposition 3. *In the four-dimensional parameter space, when the values of $(b_1, b_2, \lambda_1, \lambda_2)$ get arbitrary close to the subspace $\frac{(1+2\lambda_2)b_1+(1+2\lambda_1)b_2}{1+\lambda_1+\lambda_2} = 0$, i.e. the condition for separating equilibrium, we have $N_M \to \infty$. That is, the equilibrium with N_M partition elements gets arbitrarily close to the separating equilibrium when the above expression approaches 0.*

Next, we investigate partial information revelation under private communication. From Section 1.3 we know that the conditions supporting a separating equilibrium under private communication also support a separating equilibrium under public communication; but the converse is not true. In other words, the incentive compatibility constraints are tighter for the agent to reveal full information under private communication than that under public communication. Here for partition equilibrium under private communication we establish a similar result.

Observation 1. *1. In the cheap talk game with interacting principals, given any partition equilibrium under public communication and any partition equilibrium with the same number of partition elements under private communication, the sender has more restrictive incentive compatibility constraints under private communication.*

2. Moreover, under private communication, compared to the cheap talk game with no interdependence between principals' actions, the sender has more restrictive incentive compatibility constraints in our model.

The proof of this observation is included in the Appendix. Similar to proposition 1, this observation results from two effects: the tightened incentive compatibility constraint under private communication and the interdependence between principals' actions. We shall illustrate the two effect by considering a two-step partition equilibrium between the agent and each principal, as shown in Figure 1.1. Suppose the agent adopts the following strategy: sending message m_1 if $t \in [0, t_1]$ and message n_1 if $t \in [t_1, 1]$ to Principal 1, while sending m_2 if $t \in [0, t_2]$ and n_2 if $t \in [t_2, 1]$ to Principal 2. Denote Principal 1's two induced actions $\hat{a_1}$ and $\tilde{a_1}$ and Principal 2' two induced actions $\hat{a_2}$ and $\tilde{a_2}$. Since the communication is private between the agent and one principal on each side, it is not necessarily $t_1 = t_2$. Without loss of generality, we examine the case when $t_1 \le t_2$.

The first part of this observation comes from the fact that more non–profitable–deviation restrictions needs to be satisfied under private communication than that under public communication. Specifically, this two-by-two partition equilibrium under private communication should exclude all possible deviations of the boundary types from (t_1, t_2) to (t_1', t_2), from (t_1, t_2) to (t_1, t_2'), and from (t_1, t_2) to (t_1', t_2'):

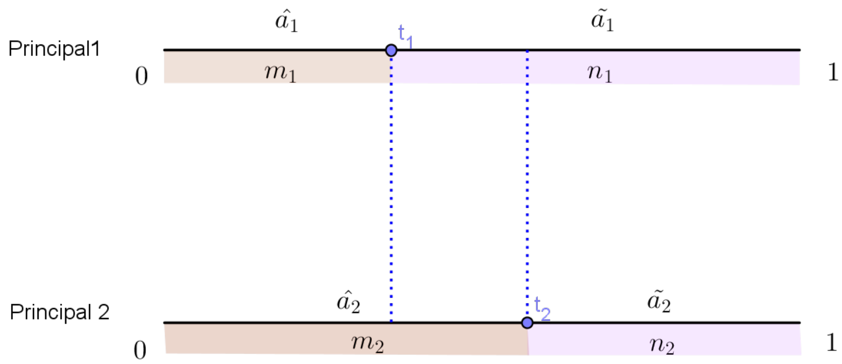


Figure 1.1: A Partition Equilibrium under Private Communication

1. Fix the boundary type t_2:

$$-\ell_1(|\hat{a_1} - t_1|) - \ell_2(|\hat{a_2} - t_1|) \geq -\ell_1(|\hat{a_1} - t_1'|) - \ell_2(|\hat{a_2} - t_1|), \forall t_1' > t_1$$

$$-\ell_1(|\tilde{a_1} - t_1|) - \ell_2(|\hat{a_2} - t_1|) \geq -\ell_1(|\tilde{a_1} - t_1'|) - \ell_2(|\hat{a_2} - t_1|), \forall t_1' < t_1$$

2. Fix the boundary type t_1:

$$-\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\hat{a_2} - t_2|) \geq -\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\hat{a_2} - t_2'|), \forall t_2' > t_2$$

$$-\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\tilde{a_2} - t_2|) \geq -\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\tilde{a_2} - t_2'|), \forall t_2' < t_2$$

3. For all $t_1' \in [0,1]$ and $t_2' \in [0,1]$

$$EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2) \geq EU^S_{\text{PRI}}(t_1', t_2'; b_1, b_2), \forall t_1', t_2' \in [0,1]$$

where $EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2) = \int_0^{t_1} -\ell_1(|\hat{a_1} - t|) - \ell_2(|\hat{a_2} - t|)dt + \int_{t_1}^{t_2} -\ell_1(|\tilde{a_1} - t|) - \ell_2(|\hat{a_2} - t|)dt + \int_{t_2}^{1} -\ell_1(|\tilde{a_1} - t|) - \ell_2(|\tilde{a_2} - t|)dt$

In contrast, under public communication there is only one boundary type $\bar{t}$ which satisfies

$$-\ell_1(|a_1 - \bar{t}|) - \ell_2(|a_2 - \bar{t}|) \geq -\ell_1(|a_1 - t'|) - \ell_2(|a_2 - t'|), \forall t' > \bar{t}$$

$$-\ell_1(|a_{11} - \bar{t}|) - \ell_2(|a_{22} - \bar{t}|) \geq -\ell_1(|a_{11} - t') - \ell_2(|a_{22} - t'|), \forall t' < \bar{t}$$

where (a_1, a_2) and (a_{11}, a_{22}) are the two action pairs induced in a two-step partition equilibrium under public communication. Thus the agent faces less restrictive incentive compatibility constraint under public communication. She is better disciplined under private communication: the principals' equilibrium actions reveal more information to each other than what the agent can control directly.

The second part of this observation results from the principals' interdependent actions. In the cheap talk model without such interaction terms (Goltsman and Pavlov (2011)), receiver i's induced actions are functions of only i's own related parameters t_i, b_i, i.e. $\hat{a_1} = \hat{a_1}(t_1, b_1), \tilde{a_1} = \tilde{a_1}(t_1, b_1)$ and $\hat{a_2} = \hat{a_2}(t_2, b_2), \tilde{a_2} = \tilde{a_2}(t_2, b_2)$. Thus a decomposition of incentive compatibility condition 3 above shows this condition is not binding as long as conditions 1 and 2 are satisfied. Nevertheless, this is not true in our model. The above condition 3 requires the agent's expected utility with boundary types t_1 and t_2 to be greater than her expected utility from deviations to any (t_1', t_2'). From the interacting term $\Lambda_i(|a_1 - a_2|)$, we know that the principals' induced actions are also interdependent: $\hat{a_1} = \hat{a_1}(t_1, t_2, b_1, b_2), \tilde{a_1} = \tilde{a_1}(t_1, t_2, b_1, b_2)$ and $\hat{a_2} = \hat{a_2}(t_1, t_2, b_1, b_2), \tilde{a_2} = \tilde{a_2}(t_1, t_2, b_1, b_2)$. Any type of deviation will result in a change in all the four induced actions. Thus condition 3 is not necessarily satisfied even when conditions 1 and 2 hold. Inductively, we can show that in any partition equilibrium with more than two steps, conditions 1 and 2 do not suffice for the agent's incentive compatibility. Therefore, we conclude that under private communication, the agent faces more restrictive constraints in equilibrium than she does in Goltsman and Pavlov (2011).

1.4.2 Comparative Statics

In this section we construct the least informative partition equilibrium under each type of communication. Assuming the state t is from a uniform distribution and the players' utility functions are quadratic, we derive the comparative statics of the equilibrium conditions with respect to parameters in our model.

Under public communication, informative communication requires at least two partition elements in the equilibrium. Denote the boundary type $\bar{t}$. In such an equilibrium, all agent types $t \in [0, \bar{t}]$ sends message m_1 while all $t \in [\bar{t}, 1]$ m_2. In the uniform-quadratic case, the boundary type equals[2]

$$\bar{t} = \frac{(1+2\lambda_2)(4b_1+1)+(1+2\lambda_1)(4b_2+1)}{4(1+\lambda_1+\lambda_2)}$$

It is easy to see that within a certain range of the parameters there exists partition equilibrium with at least two steps. No parameter values beyond this range support any non-babbling equilibrium. In a two-step equilibrium, as the parameter values increase to the boundary of this range, the value of $\bar{t}$ decreases to 0 or increases beyond 1. So we first investigate how $\bar{t}$ changes with the the bias parameters $b_i, i = 1, 2$. Taking partial derivative with respect to $|b_i|, i = 1, 2$, we have:

$$\frac{\partial \bar{t}}{\partial |b_i|} = \frac{1+2\lambda_j}{1+\lambda_i+\lambda_j} > 0, i, j = 1, 2$$

The larger the bias, the higher the value of $\bar{t}$. When $\bar{t}$ exceeds 1, the two-step equilibrium reduces to babbling equilibrium and informativeness diminishes.

Second, we examine the impact of the principals' interaction on the change of $\bar{t}$. Taking cross-effect with regard to the coordination parameters, $\lambda_i, i = 1, 2$, we have

$$\frac{\partial^2 \bar{t}}{\partial b_i \partial \lambda_i} = -\frac{1+2\lambda_j}{(1+\lambda_i+\lambda_j)^2} \leq 0, i, j = 1, 2$$

$$\frac{\partial^2 \bar{t}}{\partial b_i \partial \lambda_j} = \frac{2}{1+\lambda_i+\lambda_j} - \frac{1+2\lambda_j}{(1+\lambda_i+\lambda_j)^2} \geq 0, i, j = 1, 2$$

This result shows that an increase in different principals' coordination concerns exerts opposite effects on the marginal effect of $\frac{\partial \bar{t}}{\partial |b_i|}$. The higher principal i's coordination concern, the smaller the b_i's marginal effect on $\bar{t}$; on the contrary, the higher the other principal j's coordination concern, the larger the marginal effect. Since this marginal effect measures the speed of change of $\bar{t}$, a principal's greater coordination concern decreases the speed while the other principal's greater concern increases the speed.

Under private communication, the least informative equilibrium involves the agent's partitioning the state space into two steps in the private communication with Principal 1 while babbles in the

[2] It is easy to see that in such an equilibrium the principals' best response functions are

$$a_i = \frac{1}{1+\lambda_i}(\bar{t}/2+b_i) + \frac{\lambda_i}{1+\lambda_i}a_j, i = 3-j$$

upon receiving m_1 and

$$a_{ii} = \frac{1}{1+\lambda_i}((\bar{t}+1)/2+b_i) + \frac{\lambda_i}{1+\lambda_i}a_{jj}, i = 3-j$$

upon receiving m_2.The agent type $\bar{t}$ is indifferent between inducing action pairs (a_1, a_2) and (a_{11}, a_{22}). Solving the following equation, we get the indifferent type

$$-(a_1-\bar{t})^2 - (a_2-\bar{t})^2 = -(a_{11}-\bar{t})^2 - (a_{22}-\bar{t})^2$$

private communication with Principal 2. We are interested in the range of the parameters that supports a two-step partition equilibrium between the agent and one principal. Suppose the agent sends m_1 if $t \in [0, t_1]$ and n_1 if $t \in [t_1, 1]$ to Principal 1. It is easy to see that the boundary type t_1 is given by[3]:

$$t_1 = \frac{1 + (2 + 4b_2)\lambda_1^2 + \lambda_2 + 4b_1(1 + \lambda_1)(1 + \lambda_2) + \lambda_1(3 + 4b_2 + 2\lambda_2)}{2(1 + 2\lambda_1)(1 + \lambda_1 + \lambda_2)}$$

So we first investigate how the boundary type t_1 changes with the the bias parameters $b_i, i = 1, 2$. Taking partial derivative with respect to $|b_i|, i = 1, 2$, we have:

$$\frac{\partial t_1}{\partial |b_1|} = \frac{2(1 + \lambda_1)(1 + \lambda_2)}{(1 + 2\lambda_1)(1 + \lambda_1 + \lambda_2)} > 0$$

Second, we examine the impact of the principals' interaction on the change of t_1. Taking cross-effect with regard to the coordination parameters, λ_1, λ_2, we have

$$\frac{\partial^2 t_1}{\partial b_1 \partial \lambda_1} = -\frac{2(1 + \lambda_2)(2 + 4\lambda_1 + 2\lambda_1^2 + \lambda_2)}{(1 + 2\lambda_1)^2(1 + \lambda_1 + \lambda_2)^2} \leq 0$$

$$\frac{\partial^2 t_1}{\partial b_1 \partial \lambda_2} = \frac{2\lambda_1(1 + \lambda_1)}{(1 + 2\lambda_1)(1 + \lambda_1 + \lambda_2)^2} \geq 0$$

The interpretation is similar to the previous subsections: the higher principal Principal 1's coordination concern, the lower the marginal effect of the boundary type with respect to Principal 1's bias. On the contrary, the larger Principal 2's coordination concern, the greater the marginal effect.

In summary, under both public and private communication, the agent's incentive to reveal information changes non-monotonically in response to the principals' desire to coordinate. The

[3] In such an equilibrium R_1's best response functions are:

$$a_1 = \frac{1}{1 + \lambda_1}(t_1/2 + b_1) + \frac{\lambda_1}{1 + \lambda_1} a_2$$

$$a_{11} = \frac{1}{1 + \lambda_1}((t_1 + 1)/2 + b_1) + \frac{\lambda_1}{1 + \lambda_1} a_2$$

upon receiving m_1 and n_1 respectively. R_2's best response function is:

$$a_2 = \frac{1}{1 + \lambda_2}(1/2 + b_2) + \frac{\lambda_2}{1 + \lambda_2}(t_1 a_1 + (1 - t_1) a_{11})$$

Agent type t_1 is indifferent between inducing action pairs (a_1, a_2) and (a_{11}, a_2):

$$-(a_1 - t_1)^2 = -(a_{11} - t_1)^2$$

The boundary type t_1 is the solution to the following equation:

$$-(a_1 - t_1)^2 - (a_2 - t_1)^2 = -(a_{11} - t_1)^2 - (a_2 - t_1)^2$$

pattern of the change depends on the decision-makers' relative heterogeneous biases. In other words, as the coordination need of a principal whose interest is less misaligned with the agent grows larger, the effectiveness of communication reduces. Intuitively, coordinating with somebody else who stands further from the agent' ideal position undermines the agent's well-being.

1.4.3 Communication Improvement

In this section we present results on communication improvement. We focus on the two-step equilibrium under each type of communication when the state t is generated from a uniform distribution and the players' utility functions are quadratic. We compare our results to Goltsman and Pavlov (2011). If there exists a two-step equilibrium supported by a range of the parameters which does not support any informative equilibrium in Goltsman and Pavlov (2011), we say there is communication improvement in our model.

Notice that in Goltsman and Pavlov (2011) under public communication, when the parameters b_1, b_2 fall in the following range: $|(b_1 + b_2)/2| < 1/4$, there exists a two-step equilibrium. The following examples fix the value of coordination concerns, $\lambda_i, i = 1, 2$ and compare the range of b_i's to Goltsman and Pavlov (2011).

Example 1: $\lambda_1 = 1, \lambda_2 = 0$

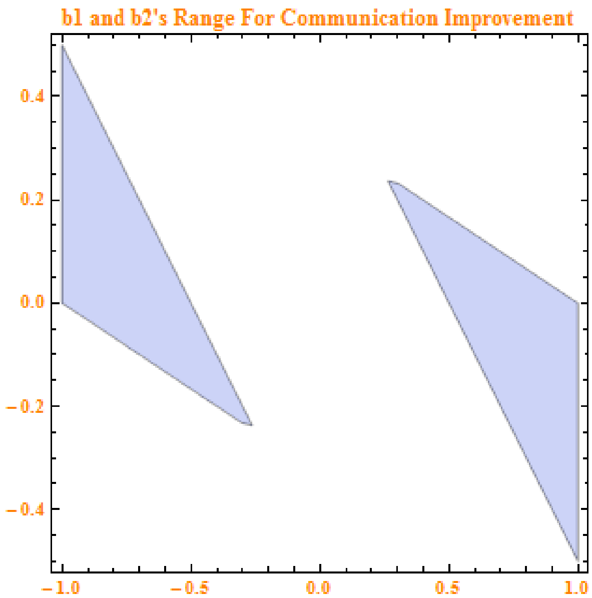
In this case, the boundary agent type is $\bar{t} = \frac{b_1+3b_2+1}{2}$. When the b_i's fall within the shaded range on Figure 1.2, we have $\bar{t} = \frac{b_1+3b_2+1}{2} \in (0, 1)$ while $|\frac{b_1+b_2}{2}| > 1/4$. This means under public communication there exists a two-step equilibrium in our model, but not in Goltsman and Pavlov (2011).

Example 2: $\lambda_1 = \lambda_2 = k$ where $k \in \mathbf{R}$ is any positive real number

In this case $\bar{t} = \frac{1+2(b_1+b_2)}{2}$. The inequalities $0 < \bar{t} = \frac{1+2(b_1+b_2)}{2} < 1$ and $|\frac{b_1+b_2}{2}| > 1/4$ yield no solution. There is *no* range of the bias parameters b_1, b_2 that supports informative communication in our model but only babbling equilibrium in the benchmark case. For all symmetric coordination concerns, communication improvement becomes impossible.

The examples show the impact of the coordination parameters on communication improvement:

Observation 2. *When the principals' coordination concerns are positive and different, communication improves in terms of the existence of informative communication under higher principals' biases. In contrast, when the two principals have symmetric coordination needs, such improvement no longer exists.*



Parameter values: $\lambda_1 = 1, \lambda_2 = 0$. The horizonal axis represents b_1 and the vertical axis b_2. The shaded areas represent the ranges such that $\bar{t} = \frac{b_1+3b_2+1}{2} \in (0,1)$ while $|\frac{b_1+b_2}{2}| > 1/4$.

Figure 1.2: Communication Improvement under Public Communication

Now we move to the private communication. Goltsman and Pavlov (2011), when the parameters b_1, b_2 fall in the following range: $|b_i| < 1/4, i = 1, 2$, there exists a two-step equilibrium under private communication. The following example shows that, at least under special values, communication improvement is not possible in our model.

Example 1: $\lambda_1 = k, \lambda_2 = 0$

We derive the two-by-two partition equilibrium as described in Figure 1.1. Assuming $b_1 = 0$ and the two boundary types $t_1 \leq t_2$. Informative equilibrium requires $0 < t_1 \leq t_2 < 1$. For all $k \in \mathbf{R}$, the range of b_2 that allows an informative equilibrium is equal to or smaller than $[0, 0.25]$. This means there is no communication improvement compared to Goltsman and Pavlov (2011) under private communication.

Observation 3. *Compared to the cheap talk game without the principals' interdependent actions, in our model there is communication improvement under public communication, while there is no such improvement under private communication.*

1.5 WELFARE COMPARISON

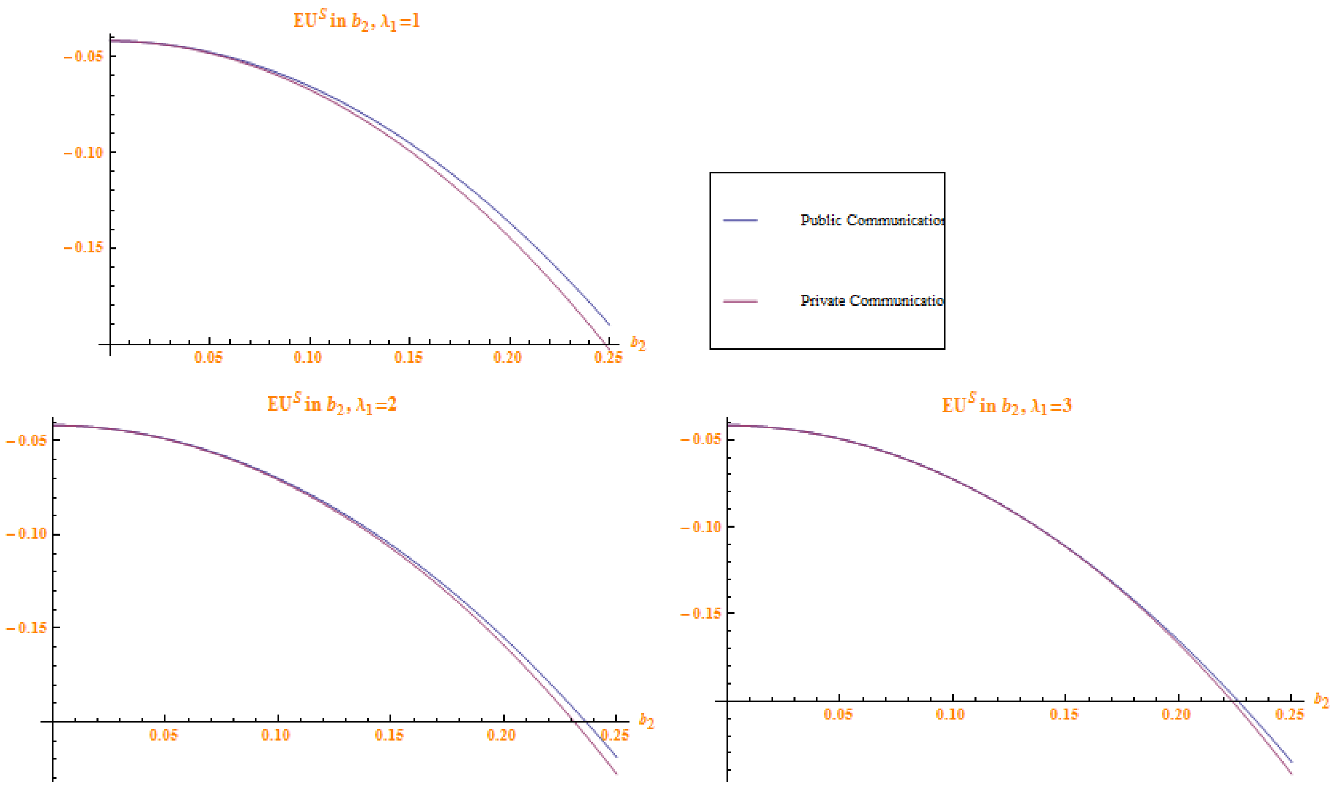
In this section we rank the players' welfare between public and private communication. There are two circumstances which will yield the same expected payoffs in both types of communication for each player: (1) separating equilibria in both types of communication, and (2) uninformative equilibria in both types of communication. In the following part, we skip these generic cases and provide comparison results for the more interesting partial revelation equilibria. Specifically, we compare a two-step partition equilibrium under public communication with a two-by-two step partition equilibrium under private communication with the boundary types $t_1 \leq t_2$. The equilibria under investigation are of the same number of partition elements. Thus our results show, even in the partition equilibrium with the same level of information precision, that the players are strictly better off under one type of communication.

We shall focus on the uniform-quadratic case and derive our comparison results for fixed parameter values. Namely, we look into the case when $\lambda_1 = k, \lambda_2 = 0, b_1 = 0, k = 1, 2, 3$[4]. In other words, principal 1's bias is zero; principal 2's utility does not depend on how well the coordination is. We derive the agent's and principals' utilities as functions of principal 2's bias, b_2.

Figure ?? draws the agent's expected utility functions as functions of b_2 under each type of communication, respectively. The blue line represents EU^S under public communication and the purple line the one under private communication. We have three observations: first, the agent's expected utility is decreasing in b_2, in either type of communication. Second, the agent's expected utility under public communication weakly dominates that under private communication. So the agent is better off under public communication. Third, the gap between the utilities from each type of communication shrinks as λ_1 increase. So the agent's marginal benefit from public communication decreases as principal 1's coordination concern becomes larger.

Figure ?? draws principal 1's expected utility functions as functions of b_2 under each type of communication, respectively. The blue line represents EU^{R1} under public communication and the purple line the one under private communication. We also have three observations here: first, principal 1's expected utility is decreasing in b_2, in either type of communication. Second, principal 1's expected utility under public communication is higher than that under private communication. Hence, principal 1 is also better off under public communication. Third, the gap between principal 1's utilities from each type of communication expands as λ_1 increase. This means principal 1's

[4] Also, we choose the parameter values to be comparable to Goltsman and Pavlov's (2011) comparison results.

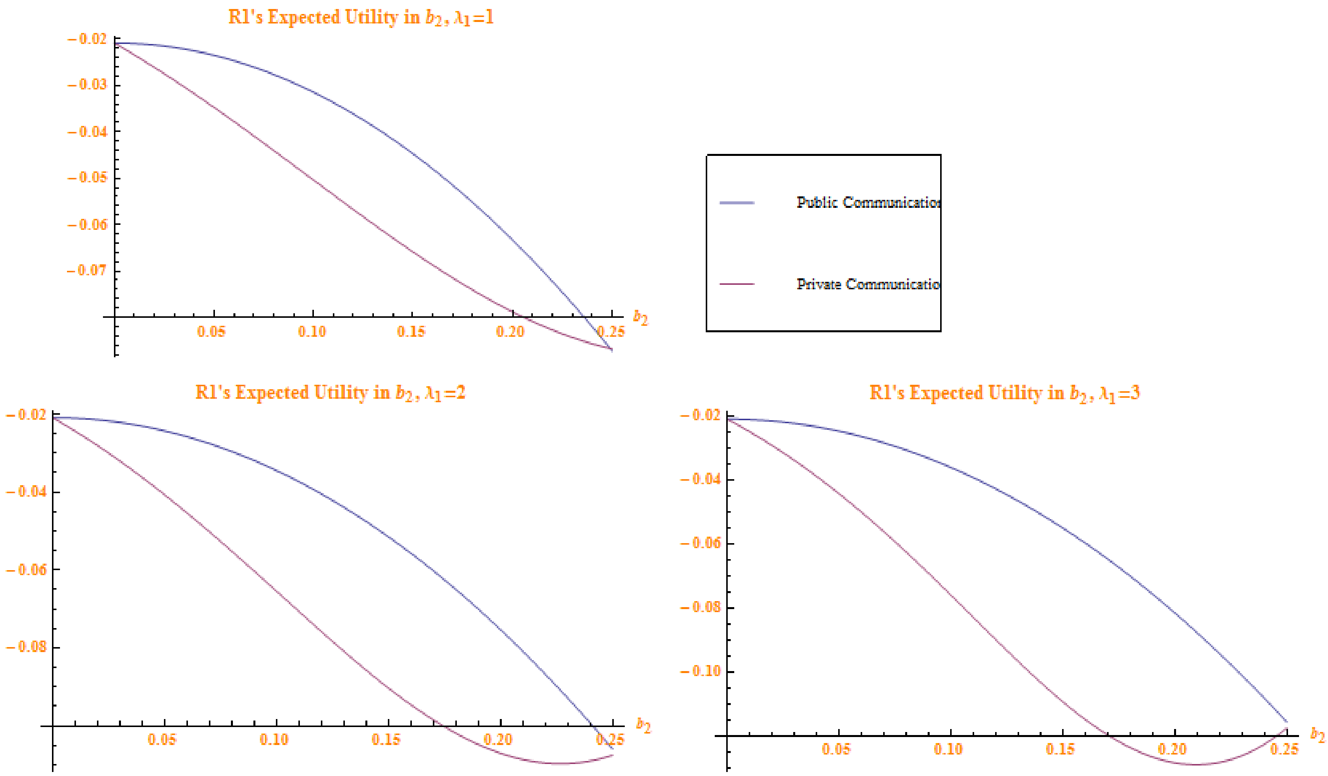


Figures: Agent's ex-ante expected utility as a function of b_2. EU^S_{PUB} (blue curve) weakly dominates EU^S_{PRI} (purple curve). As λ_1 increases, the difference between EU^S_{PUB} and EU^S_{PRI} shrinks. Parameter values: (left top) $\lambda_1 = 1, \lambda_2 = 0, b_1 = 0$; (left bottom) $\lambda_1 = 2, \lambda_2 = 0, b_1 = 0$; (right bottom) $\lambda_1 = 3, \lambda_2 = 0, b_1 = 0$.

Figure 1.3: Agent's Expected Utility in b_2

marginal benefit from public communication becomes larger as his coordination concerns grows.

Figure ?? draws principal 2's expected utility functions as functions of b_2 under each type of communication, respectively. The blue line represents EU^{R2} under public communication and the purple line the one under private communication. Again, there are three observations here: first, principal 2's expected utility is decreasing in b_2, in either type of communication. Second, principal 2's expected utility under public communication is higher than that under private communication. Thus principal 2 is also better off under public communication. Third, the gap between principal 2's utilities from each type of communication expands as λ_1 increase. This means principal 2's marginal benefit from public communication declines as his partner's coordination concerns rises.

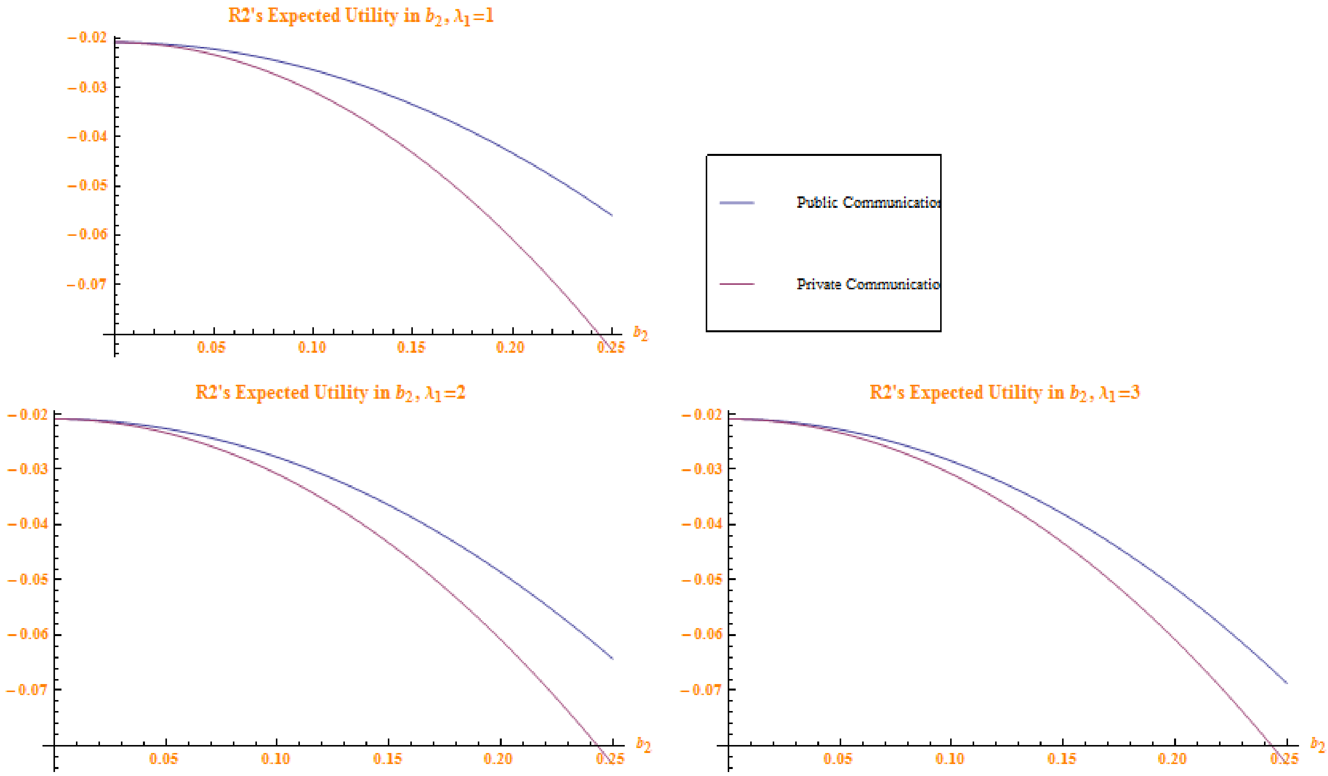


Figures: Principal 1's ex-ante expected utility as a function of b_2. EU^{R1}_{PUB} (blue curve) dominates EU^{R1}_{PRI} (purple curve). As λ_1 increases, the difference between EU^{R1}_{PUB} and EU^{R1}_{PRI} expands. Parameter values: (left top) $\lambda_1 = 1, \lambda_2 = 0, b_1 = 0$; (left bottom) $\lambda_1 = 2, \lambda_2 = 0, b_1 = 0$; (right bottom) $\lambda_1 = 3, \lambda_2 = 0, b_1 = 0$.

Figure 1.4: Principal 1's Expected Utility in b_2

1.6 RELATED LITERATURE

Out paper intends to link two strands of literature: the literature on strategic information transmission and the literature on organization design under incomplete information. The seminal cheap-talk model by Crawford and Sobel (1982) studies a better-informed agent sends a message to an uninformed principal, who takes an action that affects the welfare of both players. They show that as long as the agent's and principal's interests are not perfectly aligned, there is only finite number of actions induced in equilibrium. The agent partitions the type space into finite number of elements and all agents' types within one partition element pool to send one signal representing the element where the true type lies. The construction of partition equilibrium relies on indifference conditions of the agent's types on the boundaries of each partition element. This help us to find



Figures: Principal 2's ex-ante expected utility as a function of b_2. EU^{R2}_{PUB} (blue curve) dominates EU^{R2}_{PRI} (purple curve). As λ_1 increases, the difference between EU^{R2}_{PUB} and EU^{R2}_{PRI} shrinks. Parameter values: (left top) $\lambda_1 = 1, \lambda_2 = 0, b_1 = 0$; (left bottom) $\lambda_1 = 2, \lambda_2 = 0, b_1 = 0$; (right bottom) $\lambda_1 = 3, \lambda_2 = 0, b_1 = 0$.

Figure 1.5: Principal 2's Expected Utility in b_2

the key equilibrium condition in this model, too.

There are a few models that extend the CS model into a one-agent multiple-principal framework. Farrell and Gibbons (1989) study a two-principal framework with discrete state space and discrete action space. In their model the agent can send messages either publicly or privately but not both. They show that whenever there exists an equilibrium in which the sender communicates informatively under private communication, there is an equilibrium under public communication in which the sender does the same. But the converse is never true. Goltsman and Pavlov (2011) generalize Farrell and Gibbons (1989) framework into concave and differentiable payoff functions over a continuous type space and a continuous action space. They derive conditions under which public communication dominates private communication. They also show that the sender is willing to reveal more information if she combines the two messaging channels together. Koessler (2006) investigates the relation between public and private communication with certifiable messages. He shows that, under a type-dependent message space, restrictions on the principals' off-equilibrium beliefs exclude the circumstance when informative equilibrium exists in public but not private communication. He also shows that private communication strictly dominates public communication. There exists equilibrium with full information revelation in private communication when no public communication is possible. Nonetheless, all these work assumes no interactions between the principals' actions. The novelty of our model is that we allow the principals' utilities to be interdependent.

Related works addressing multi-principal communication issues include Newman and Sansing (1993) and Gigler (1994). Both investigate an incumbent firm's information disclosure to two audience groups: its competitor and the capital market/shareholders. There exists no private disclosure ever and all public disclosure equilibria are partition equilibria. Similarly, Ainsworth and Sened (1993) study an interest group's lobbying two separate groups: the congressmen and government officials. They compare the collective decisions with versus without a lobbyist in public communication. Koessler and Martimort (2008) study optimal communication mechanisms in a two-dimensional action space. When action on each dimension is controlled by a different principal and the coordination between two principals enters the agent's utility jointly, in private communication there is an asymmetric contractual externality between the two principals. They also show that public communication gives rise to mutual discipline or subversion effects according to the principals' biases. Johns (2007) examines a bureaucrat's public cheap-talk communication with two policymakers who then bargain for the final policy choice. He shows that the two policymakers are better off in the follow-up bargaining if they encounter a moderate bureaucrat instead of somebody else that biases towards one of them. Board and Dragu (2008) also study two-principal cheap-talk followed by an ultimatum bargaining between the principals for a final decision. They find that the additional principal with veto power might get lower welfare; nevertheless the original principal is better off than in the case without veto threat.

Another strand of literature includes works on players' coordination under incomplete information. Morris and Shin (2002) study the social value of public information provision. They investigate the setting where each decision maker, with a private noisy signal, aims at a decision appropriate to the underlying state and coordinating with the majority. They show that when none of the members holds any socially valuable private information, public information revelation always leads to welfare improvement. However, when every organization member receives a private signal besides the commonly observed message, greater provision of public information does not always enhance welfare. Morris and Shin (2007) emphasize the trade-off between information precision and common understanding among group members in designing the optimal communication mechanisms. In both papers the decision-maker's private information is exogenous; and the provision of public information is through a benevolent social planner who aims at maximizing the aggregated utility of all decision-makers. In our model the messages come from a biased agent whose preferred action does not completely coincide with the decision-makers' interests.

Alonso et. al.(2008) compare centralized communication and decentralized communication with coordination in a multi-division organization. They analyze a multi-agent single-principal cheap-talk setting. Each informed division observes its own type and sends a message to a head-quarter who chooses the final action (vertical communication). Alternatively, the two divisions can communicate with each other directly and then make two separate decisions (horizonal communication). A division is concerned not only about its own ideal action regarding the underlying state, but also the coordination with the other division. The authors show that higher coordination need improves horizontal communication but worsens vertical communication when the decision right is controlled by the uninformed principal. Kawamura (2010) considers the same mechanism with more than two informed agents. He points out that as the number of informed agents grows, information precision diminishes. The difference in our setting is that we consider a single-agent multi-principal setting. We aim at showing how the agent's incentive to reveal information changes in response to the principals' interactions.

Two other papers address the issue of cheap-talk with multiple decision-makers located on a network. Galeotti et. al. (2009) investigate a cheap-talk model in which each decision-maker's payoff depends on the cumulative distance between every other's action choice and his own bliss point. They show that information revelation to another player decreases when some opponents with highly misaligned interest also communicate with that player. Hagenbach and Koessler (2010) study information transmission within a network with coordinating decision-makers. Each decision-maker has his own preferred action plus the need for coordination, weighted inversely by the number of players. They show that larger groups tend to effectively share information even when there is no information flow in smaller groups. Different from our setting, every group member in their model observes a private noisy signal and is only allowed for public communication.

Our model highlights a few new features: first, compared to the existing multi-principal cheap talk literature, we incorporate the principals' coordination problem into the game situation. Second, we consider heterogeneous coordination concerns so as to capture the principals' different weights attached to their most preferred actions and coordination. Third, we analyze how the informed agent's information revelation incentives change in different types of communication.

1.7 CONCLUSION AND DISCUSSION

In this paper we investigate a cheap-talk communication setting with one informed agent and two coordinating uninformed principals. We show that under both the public communication and the private communication equilibria have the partition form characterized in Crawford and Sobel (1982). The agent partitions the continuous state space into finite number of elements and sends out only one message representing the partition element in which the true underlying state lies. The number of partition elements for given parameter values decreases monotonically with the principals' biases, but changes non-monotonically with the principals' coordination concerns, in a pattern that depends on principals' relative instead of the absolute biases. Moreover, we show that with the additional coordination concerns in the principals' utility functions, there exists communication improvement compared to Goltsman and Pavlov (2011). Informative communicative equilibrium can be supported with the principals' more extreme biases. In addition, we discuss the relation between public and private communications. Farrell and Gibbons (1989) and the benchmark model both emphasize that the presence of two principals mutually discipline the agent's incentives in information revelation, even if there is no such revelation privately. Here we find similar communication improvement in private asymmetric communication. The agent's signal becomes more informative if she sends out less precise signals to the other. In the most extreme case of such asymmetric revelation, the agent babbles with the principal whose bias is larger while communicates informatively with the principal whose bias is moderate. There exists such informative asymmetric equilibrium within the parameters' range where all equilibria in the public communication are uninformative.

There are two additional issues to mention. Firstly, we attach heterogeneous coordination concerns, λ_1 and λ_2, to the principals' utility functions. An alternative way to incorporate principals' different weights on the most preferred actions and on coordination is to attach $1-\alpha_i$ and α_i where $\alpha_i \in [0,1]$. Simple algebra shows that it is equivalent to our setting once we normalize $\frac{\lambda_i}{1+\lambda_i} = \alpha_i$.

Secondly, we do not add the coordinating term onto the agent's payoff function explicitly; yet the agent has to take the principals' coordination into account before delivering her message(s).

To be robust, we also check two possibilities: the agent derives utility solely from principals' coordination $U^S(a_1, a_2) = -(a_1 - a_2)^2$ and she also wants her own bliss point to be implemented $U^S(a_1, a_2, t) = -(a_1 - t)^2 - (a_2 - t)^2 - (a_1 - a_2)^2$. Notice the agent's utility does not depend on t in the former case, thus the agent aims at inducing two actions as close as possible regardless of the true state of the world. In equilibrium, the agent knows the best response of each principal on receiving the signals, thus the message transmitted to one principal is tailored towards the action of the other principal, unlike in the main model where message(s) is tailored to induce actions in exact coincidence with the underlying state. The essence of the problem does not change and the main results hold. In the latter case, the agent's utility function satisfies all conditions of Proposition 2 so similar results will follow, though the positions of the boundary types each partition equilibrium may be different.

The model can be extended to endogenous coordination concerns. By changing the two parameters λ_1, λ_2 to $\lambda_1(t), \lambda_2(t)$, the principals' desire for minimize the disparity between actions varies with each realization of the state variable t. In reality, there are circumstances in which principals value their favorite actions more when the underlying state is "moderate," while feel greater needs to coordinate when the underlying state becomes more extreme.

1.8 APPENDIX

1.8.1 Proofs

1.8.1.1 Proof of Proposition 1

Proof for Part 1

The existence of separating equilibrium requires every agent's type having incentive to reveal the true t. The principals knows the value of t, so the best response functions are

$$a_1 = h_1(a_2, b_2, t), a_2 = h_2(a_1, b_2, t)$$

Agent's type t's utility is $U^S(a_1, a_2, t) = -\ell_1(|h_1(a_2, b_2, t) - t|) - \ell_2(|h_2(a_1, b_2, t) - t|)$.

Under public communication, suppose the agent type t deviates to the mimic type $\tilde{t}$. Both principals observe the same messages from the agent and believe she is type $\tilde{t}$ with probability 1. Thus the agent type t's utility from such an deviation becomes $U^S(\tilde{a_1}, \tilde{a_2}, \tilde{t}) = -\ell_1(|h_1(\tilde{a_2}, b_2, \tilde{t}) - t|) - \ell_2(|h_2(\tilde{a_1}, b_2, \tilde{t}) - t|)$. The non-profitable deviation condition requires $U^S(a_1, a_2, t) \geq U^S(\tilde{a_1}, \tilde{a_2}, \tilde{t})$.

Under private communication, we need to check three non-profitable deviation conditions: the condition should exclude unilateral deviation with R_1, unilateral deviation with R_2, and bilateral deviation with R_1 and R_2. The first two cases are symmetric. Suppose the agent type t deviates to mimic type s with R_1 while remains truth-telling with R_2. The principals' best response functions become

$$a_1 = g_1(a_2, b_2, s), a_2 = g_2(a_1, b_2, t)$$

Thus non-profitable bilateral deviation requires

$$U^S(a_1, a_2, t) \geq U^S(a_1', a_2', t) = -\ell_1(|g_1(a_2, b_2, s) - t|) - \ell_2(|g_2(a_1, b_2, t) - t|)$$

The last case is to exclude the agent type t's bilateral deviation. Suppose that the agent type t deviates to mimic type s with R_1 and deviates to mimic type r with R_2. The principals' best response functions become

$$a_1 = f_1(a_2, b_2, s), a_2 = f_2(a_1, b_2, r)$$

Non-profitable deviation requires

$$U^S(a_1, a_2, t) \geq U^S(a_1'', a_2'', t) = -\ell_1(|f_1(a_2, b_2, s) - t|) - \ell_2(|f_2(a_1, b_2, r) - t|)$$

Therefore, the non-profitable deviation constraints are tighter under private communication.

Proof for Part $2.a$

In a separating equilibrium under public communication, every agent type t sends a distinct message and both principals interpret correctly on every observed message as the agent type from whom this message comes. Without loss of generality, let us assume the separating strategy for each $t \in [0, 1]$ is a report $m = t$. Upon receiving m, both principals believe it comes from type t.

Each principal solves the optimization problem:

$$\max_{a_i \in A} U^{R_i}(a_i, a_j, b_i, b_j, t) = -(a_i - (t + b_i))^2 - \lambda_i(a_1 - a_2)^2, i, j = 1, 2$$

From the interacting term in utility function, we get two best response functions, showing the interdependence of their actions:

$$a_1(a_2) = \frac{1}{1+\lambda_1}(t + b_1) + \frac{\lambda_1}{1+\lambda_1} a_2$$
$$a_2(a_1) = \frac{1}{1+\lambda_2}(t + b_2) + \frac{\lambda_2}{1+\lambda_2} a_1$$

Solving the best response functions yields:

$$a_1^t = \frac{(1+\lambda_2)(t+b_1)+\lambda_1(t+b_2)}{1+\lambda_1+\lambda_2}$$
$$a_2^t = \frac{(1+\lambda_1)(t+b_2)+\lambda_2(t+b_1)}{1+\lambda_1+\lambda_2}$$

Type t agent's utility is $U_t^S = -(a_1^t - t)^2 - (a_2^t - t)^2 = -(\frac{b_1+\lambda_2 b_1+\lambda_1 b_2}{1+\lambda_1+\lambda_2})^2 - (\frac{b_2+\lambda_1 b_2+\lambda_2 b_1}{1+\lambda_1+\lambda_2})^2$.To check the incentive compatibility, we need to prove that type t agent's expected utility from mimicking any other type would not exceed U_t^S.

Now suppose pretending to be another type $\tilde{t}$ is a profitable deviation for type t. Given the two principals' beliefs same as above, namely, they believe that the agent is type $\tilde{t}$ with probability 1. Therefore, the induced action pair change to:

$$a_1^{\tilde{t}} = \frac{(1+\lambda_2)(\tilde{t}+b_1)+\lambda_1(\tilde{t}+b_2)}{1+\lambda_1+\lambda_2}$$
$$a_2^{\tilde{t}} = \frac{(1+\lambda_1)(\tilde{t}+b_2)+\lambda_2(\tilde{t}+b_1)}{1+\lambda_1+\lambda_2}$$

The agent's utility change to $U_{\tilde{t}}^S = -(a_1^{\tilde{t}} - t)^2 - (a_2^{\tilde{t}} - t)^2$. Type t agent's incentive compatibility in separating equilibrium implies $U_t^S > U_{\tilde{t}}^S$ for all the parameter values $b_1, b_2, \lambda_1, \lambda_2$. The inequality holds if and only if $\frac{(1+2\lambda_2)b_1+(1+2\lambda_1)b_2}{1+\lambda_1+\lambda_2} = 0$. Thereby the conclusion follows.

Proof for Part 2.b

Suppose under private communication there exists separating equilibrium, in which every agent type t sends a distinct message to each principal separately and each principal interprets the message properly as the type from whom this message comes.

Each principal solves the optimization problem:

$$\max_{a_i \in A} U^{R_i}(a_i, a_j, b_i, b_j, t) = -(a_i - (t+b_i))^2 - \lambda_i(a_1 - a_2)^2, i, j = 1, 2$$

From the interacting term in utility function, we get two best response functions, showing the interdependence of their actions:

$$a_1(a_2) = \frac{1}{1+\lambda_1}(t+b_1) + \frac{\lambda_1}{1+\lambda_1}a_2$$
$$a_2(a_1) = \frac{1}{1+\lambda_2}(t+b_2) + \frac{\lambda_2}{1+\lambda_2}a_1$$

Solving the best response functions yields:

$$a_1^t = \frac{(1+\lambda_2)(t+b_1)+\lambda_1(t+b_2)}{1+\lambda_1+\lambda_2}$$
$$a_2^t = \frac{(1+\lambda_1)(t+b_2)+\lambda_2(t+b_1)}{1+\lambda_1+\lambda_2}$$

Type t agent's utility is $U_t^S = -(a_1^t - t)^2 - (a_2^t - t)^2 = -(\frac{b_1+\lambda_2 b_1+\lambda_1 b_2}{1+\lambda_1+\lambda_2})^2 - (\frac{b_2+\lambda_1 b_2+\lambda_2 b_1}{1+\lambda_1+\lambda_2})^2$.

Incentive comparability constraints for the agent require no-profitable deviation unilateral with $R1$, no-profitable deviation unilateral with $R2$, and no-profitable deviation with both principals bilaterally. We will exclude all of the three cases below.

Deviation 1: The type t agent pretends to be type s to $R1$ while truthfully reveals t to $R2$ given the belief pair $(\mu_1(t|\cdot), \mu_2(t|\cdot))$ such that each principal interprets the message(s) as truthful information revelation.

In this case the principals' best response functions become:

$$a_1(a_2) = \frac{1}{1+\lambda_1}(s+b_1) + \frac{\lambda_1}{1+\lambda_1}a_2$$
$$a_2(a_1) = \frac{1}{1+\lambda_2}(t+b_2) + \frac{\lambda_2}{1+\lambda_2}a_1$$

Solving the best response functions yields:

$$a_1^s = \frac{-s - b_1 - t\lambda_1 - b_2\lambda_1 - s\lambda_2 - b_1\lambda_2}{1+\lambda_1+\lambda_2}$$
$$a_2^t = \frac{-t - b_2 - t\lambda_1 - b_2\lambda_1 - s\lambda_2 - b_1\lambda_2}{1+\lambda_1+\lambda_2}$$

Denote agent type t's utility in this case as $U_{st}^S = -(a_1^s - t)^2 - (a_2^t - t)^2$. No-profitable deviation requires $U_{st}^S - U_t^S \leq 0$. This equality holds if and only if $-b_2(\lambda_1+\lambda_2+2\lambda_1\lambda_2) - b_1(1+2\lambda_2+2\lambda_2^2) = 0$.

Deviation 2: The type t agent pretends to be type r to $R2$ while truthfully reveals t to $R1$ given the belief pair $(\mu_1(t|\cdot), \mu_2(t|\cdot))$ such that each principal interprets the message(s) as truthful information revelation.

Similar to the above case, the principals' best response functions are:

$$a_1(a_2) = \frac{1}{1+\lambda_1}(t+b_1) + \frac{\lambda_1}{1+\lambda_1}a_2$$
$$a_2(a_1) = \frac{1}{1+\lambda_2}(r+b_2) + \frac{\lambda_2}{1+\lambda_2}a_1$$

All derivations are the same. Denote the agent type t's utility in this case as U_{rt}^S. No-profitable deviation requires $U_{rt}^S - U_t^S \leq 0$. This equality holds if and only if $-b_1(\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2) - b_2(1 + 2\lambda_1 + 2\lambda_1^2) = 0$.

Deviation 3: The type t agent pretends to be type s to $R1$, type r to $R2$ given the belief pair $(\mu_1(t|\cdot), \mu_2(t|\cdot))$ such that each principal interprets the message(s) as truthful information revelation.

The principals' best response functions become:

$$a_1(a_2) = \frac{1}{1+\lambda_1}(s+b_1) + \frac{\lambda_1}{1+\lambda_1}a_2$$
$$a_2(a_1) = \frac{1}{1+\lambda_2}(r+b_2) + \frac{\lambda_2}{1+\lambda_2}a_1$$

Denote the agent type t's utility in this case as U^S_{sr}. No-profitable deviation requires $U^S_{sr} - U^S_t \leq 0$.

$$U^S_{sr} - U^S_t = b_2^2(1+2\lambda_1+2\lambda_1^2) + b_1^2(1+2\lambda_2+2\lambda_2^2) + 2b_1b_2(\lambda_1+\lambda_2+2\lambda_1\lambda_2)$$
$$-[(r+b_2)(1+\lambda_1) + (s+b_1)\lambda_2 - t(1+\lambda_1+\lambda_2)]^2 - [(r+b_2)\lambda_1 + (s+b_1)(1+\lambda_2) - t(1+\lambda_1+\lambda_2)]^2$$

Thus we have the following incentive compatibility constraints to excludes all non-profitable deviation:

$$-b_2(\lambda_1+\lambda_2+2\lambda_1\lambda_2) - b_1(1+2\lambda_2+2\lambda_2^2) = 0$$
$$-b_1(\lambda_1+\lambda_2+2\lambda_1\lambda_2) - b_2(1+2\lambda_1+2\lambda_1^2) = 0$$
$$U^S_{sr} - U^S_t \leq 0$$

Since $\lambda_1, \lambda_2 > 0$, these conditions holds only when $b_1 = b_2 = 0$. Thus we conclude: if and only if $b_1 = b_2 = 0$ there exists separating equilibrium in private communication.

1.8.1.2 Proof of Proposition 2

The partition form of equilibria and finite actions induced in such equilibria result from a key property of the players' utility functions. So first we prove this critical *single-crossing* property for an extended two-dimension action space. It implies the agent's incentive compatibility, i.e. all agent types on one side of the indifferent type strictly prefer one induced action pair to the other.

We first prove the result when the common prior is the uniform distribution and the players' utility functions are of the quadratic form. In any equilibrium of public communication, upon receiving a message m, principal i solves

$$\max_{a_i} \int_{t\in[0,1]} -(a_i - (t+b_i))^2 - \lambda_i(a_1-a_2)^2 dF_m(t), i = 1, 2$$

Each $i's$ first-order condition is a reaction function $a_i(a_j)$. Parallelize the two f.o.c's and substitute variables, we get

$$(1+\lambda_1+\lambda_2)a_1 - b_1 = (1+\lambda_1+\lambda_2)a_2 - b_2 \tag{1.1}$$

which means "common message induces common beliefs" for both principals.

Now given any two pair of actions $(\hat{a_1}, \hat{a_2})$ and $(\tilde{a_1}, \tilde{a_2})$ where all the $a_i's$ are fixed numbers. Consider the agent's utility with different pairs, $\hat{U^S}$ and $\tilde{U^S}$. Single crossing condition requires $\hat{U^S} = \tilde{U^S}$ for exactly one type $t \in [0, 1]$. For all other types, the agent should have strict preference of one pair over the other.

$$
\begin{aligned}
\hat{U^S} - \tilde{U^S} &= [-(\hat{a_1} - t)^2 - (\hat{a_2} - t)^2] - [-(\tilde{a_1} - t)^2 - (\tilde{a_2} - t)^2] \\
&= (\hat{a_1} + \tilde{a_1} - 2t)(\tilde{a_1} - \hat{a_1}) + (\hat{a_2} + \tilde{a_2} - 2t)(\tilde{a_2} - \hat{a_2}) \\
&= (\tilde{a_1}^2 - \hat{a_1}^2) + (\tilde{a_2}^2 - \hat{a_2}^2) - 2t(\tilde{a_1} - \hat{a_1} + \tilde{a_2} - \hat{a_2})
\end{aligned}
$$

This is a linear function of t. Therefore, for all actions pairs $(\hat{a_1}, \hat{a_2})$ and $(\tilde{a_1}, \tilde{a_2})$ such that $\tilde{a_1} - \hat{a_1} + \tilde{a_2} - \hat{a_2} \neq 0$, the agent's indifference condition yields a unique solution to the "boundary" type $\bar{t}$:

$$
\bar{t} = \frac{(\tilde{a_1}^2 - \hat{a_1}^2) + (\tilde{a_2}^2 - \hat{a_2}^2)}{2(\tilde{a_1} - \hat{a_1} + \tilde{a_2} - \hat{a_2})}
$$

If $\tilde{a_1} - \hat{a_1} + \tilde{a_2} - \hat{a_2} = 0$, we have the following two cases:

Case 1: $(\tilde{a_1}^2 - \hat{a_1}^2) + (\tilde{a_2}^2 - \hat{a_2}^2) \neq 0$. This means all agent types $t \in [0, 1]$ has strict preference. Either $(\hat{a_1}, \hat{a_2}) \succ_S (\tilde{a_1}, \tilde{a_2})$ or $(\tilde{a_1}, \tilde{a_2}) \succ_S (\hat{a_1}, \hat{a_2})$ $\forall t \neq \bar{t}$. No "indifference" type in between implies violation of single-crossing condition.

Case 2: $(\tilde{a_1}^2 - \hat{a_1}^2) + (\tilde{a_2}^2 - \hat{a_2}^2) = 0$. In this case, $\hat{U^S} = \tilde{U^S}$ $\forall t \in [0, 1]$, the single-crossing condition does not hold either.

Next we shall prove the following claim:

Claim: Within the entire range of the bias parameters (b_1, b_2) such that $b_1 \neq b_2$, for any two different action pairs, neither case 1 or case 2 is possible given same belief for both principals and best response functions described in equation (1), thus the single-crossing condition is not violated.

For Case 1, in $\tilde{a_1} - \hat{a_1} + \tilde{a_2} - \hat{a_2} = 0$ we can write $\hat{a_2} = \tilde{a_1} - \hat{a_1} + \tilde{a_2}$ (2). Suppose there are two action pairs induced in equilibrium such that $\tilde{a_i} \neq \hat{a_i}, i = 1, 2$. From equation (1) we have:

$$
(1 + \lambda_1 + \lambda_2)\hat{a_1} - b_1 = (1 + \lambda_1 + \lambda_2)\hat{a_2} - b_2
$$

$$
(1 + \lambda_1 + \lambda_2)\tilde{a_1} - b_1 = (1 + \lambda_1 + \lambda_2)\tilde{a_2} - b_2
$$

Insert (2) and simplify the system of equations, we get:

$$
2(1 + \lambda_1 + \lambda_2)\hat{a_1} - b_1 = 2(1 + \lambda_1 + \lambda_2)\tilde{a_1} - b_1
$$

Which implies $\hat{a_1} = \tilde{a_1}$, a contradiction.

For Case 2, plug $\hat{a_2} = \tilde{a_1} - \hat{a_1} + \tilde{a_2}$ into $(\tilde{a_1}^2 - \hat{a_1}^2) + (\tilde{a_2}^2 - \hat{a_2}^2) = 0$, we get $\hat{a_1} = \tilde{a_2}$ and $\hat{a_2} = \tilde{a_1}$, i.e. the two pairs are symmetric about the 45°-line. On the other hand, if such action pairs $(\hat{a_1}, \hat{a_2})$ and $(\tilde{a_1}, \tilde{a_2})$ are induced in equilibrium upon observing any public message m, equation (1) yields:

$$(1 + \lambda_1 + \lambda_2)\hat{a_1} - b_1 = (1 + \lambda_1 + \lambda_2)\hat{a_2} - b_2$$
$$(1 + \lambda_1 + \lambda_2)\tilde{a_1} - b_1 = (1 + \lambda_1 + \lambda_2)\tilde{a_2} - b_2$$

Plug $\hat{a_1} = \tilde{a_2}$ and $\hat{a_2} = \tilde{a_1}$ into the above we get:

$$(1 + \lambda_1 + \lambda_2)\tilde{a_1} - b_1 = (1 + \lambda_1 + \lambda_2)\tilde{a_1} - b_2 + b_1 - b_2$$

The equation only holds when $b_1 = b_2$. Hence for all $b_1 \neq b_2$, such actions pairs cannot be induced at the same time in any equilibrium.

Next, we check the concavity of all three players' utility functions to guarantee there exists a unique solution for every utility maximization problem given each agent type t.

- $\partial U^S(a_1, a_2, t)/\partial a_1 = 0$ and $\partial U^S(a_1, a_2, t)/\partial a_2 = 0$ at point $a_1 = a_2 = t$
- The matrix $\begin{pmatrix} \frac{\partial^2 U^S}{\partial a_1^2} & \frac{\partial^2 U^S}{\partial a_1 \partial a_2} \\ \frac{\partial^2 U^S}{\partial a_1 \partial a_2} & \frac{\partial^2 U^S}{\partial a_2^2} \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$ is negative definite. Thus U^S is concave for every (a_1, a_2). The solution is the unique maximum point
- For R_i, $\partial U^{R_i}(a_1, a_2, t)/\partial a_i = 0$ yields $a_i = \frac{(1+\lambda_j)(t+b_i)+\lambda_i(t+b_j)}{1+\lambda_i+\lambda_j}, i, j = 1, 2$, so that $a_i = \frac{(1+\lambda_j)(t+b_i)+\lambda_i(t+b_j)}{1+\lambda_i+\lambda_j}$
- The matrix $\begin{pmatrix} \frac{\partial^2 U^{R_i}}{\partial a_1^2} & \frac{\partial^2 U^{R_i}}{\partial a_1 \partial a_2} \\ \frac{\partial^2 U^{R_i}}{\partial a_1 \partial a_2} & \frac{\partial^2 U^{R_i}}{\partial a_2^2} \end{pmatrix} = \begin{pmatrix} -2 - 2\lambda_i & 0 \\ 0 & -2 - 2\lambda_i \end{pmatrix}$ is negative definite since $\lambda_i > 0$. Thus U^{R_i} is concave for every (a_1, a_2). The solution is the unique maximum point.

As shown in Proposition 1, for each agent type $t \in [0, 1]$, the agent's optimal strategy is to induce an action pair (a_1, a_2) such that $a_1 = a_2 = t$. And for the principals, if the true agent type t were known, the actions $a_1^t = \frac{(1+\lambda_2)(t+b_1)+\lambda_1(t+b_2)}{1+\lambda_1+\lambda_2}$ and $a_2^t = \frac{(1+\lambda_1)(t+b_2)+\lambda_2(t+b_1)}{1+\lambda_1+\lambda_2}$ would maximize their utilities. Proposition 2 assumes $\forall b_1, b_2, \lambda_1, \lambda_2$, $(1 + \lambda_2)b_1 + \lambda_1 b_2 \neq 0$ and $(1 + \lambda_1)b_2 + \lambda_2 b_1 \neq 0$. Hence $t \neq \frac{(1+\lambda_2)(t+b_1)+\lambda_1(t+b_2)}{1+\lambda_1+\lambda_2} = a_1^t$ and $t \neq \frac{(1+\lambda_1)(t+b_2)+\lambda_2(t+b_1)}{1+\lambda_1+\lambda_2} = a_2^t$.

Suppose $\vec{u} = (\hat{a_1}, \hat{a_2})$ and $\vec{v} = (\tilde{a_1}, \tilde{a_2})$ are two pairs of actions induced in equilibrium. Then by the generalized single-crossing property, there exist t_u and t_v such that $U^S(\vec{u}, t_u) \geq U^S(\vec{v}, t_u)$ and $U^S(\vec{u}, t_v) \leq U^S(\vec{v}, t_v)$. Since U^S is continuous on a_1, a_2, there exists $\bar{t} \in [0, 1]$ s.t. $t_u < \bar{t} < t_v$

and $U^S(\vec{u},\bar{t}) = U^S(\vec{v},\bar{t})$. All agent types $t > \bar{t}$ prefer $\vec{v}$ to $\vec{u}$ and types $t < \bar{t}$ prefer $\vec{u}$ to $\vec{v}$. Hence $\hat{a_1} < \bar{t} < \tilde{a_1}$ and $\hat{a_2} < \bar{t} < \tilde{a_2}$. And for the agent type $\bar{t}$, the most preferred action pair is $a_1 = a_2 = \bar{t}$.

Now for the principals, their utilities are maximized at $a_1^{\bar{t}} = \frac{(1+\lambda_2)(\bar{t}+b_1)+\lambda_1(\bar{t}+b_2)}{1+\lambda_1+\lambda_2}$ and $a_2^{\bar{t}} = \frac{(1+\lambda_1)(\bar{t}+b_2)+\lambda_2(\bar{t}+b_1)}{1+\lambda_1+\lambda_2}$. As $t_u < \bar{t} < t_v$, we have $\hat{a_1} < a_1^{\bar{t}} < \tilde{a_1}$ and $\hat{a_2} < a_2^{\bar{t}} < \tilde{a_2}$.

From above we know $\bar{t} \neq a_1^{\bar{t}}$ and $\bar{t} \neq a_2^{\bar{t}}$ as long as $(1+\lambda_2)b_1+\lambda_1 b_2 \neq 0$ and $(1+\lambda_1)b_2+\lambda_2 b_1 \neq 0$. Specifically, $\forall \epsilon > 0$, $|a_1^{\bar{t}} - \bar{t}| > \epsilon/\sqrt{2}$ and $|a_2^{\bar{t}} - \bar{t}| > \epsilon/\sqrt{2}$. Therefore $|\tilde{a_1} - \hat{a_1}| > \epsilon/\sqrt{2}$ and $|\tilde{a_1} - \hat{a_1}| > \epsilon/\sqrt{2}$. It follows that

$$||\vec{u} - \vec{v}|| = \sqrt{(\tilde{a_1} - \hat{a_1})^2 + (\tilde{a_1} - \hat{a_1})^2} > \epsilon$$

Moreover, the agent's utility satisfies concavity and single-crossing condition, $U^S_{a_i t} > 0$ implies $G'(t) = U^S_t(v,t) - U^S_t(u,t) = \int_u^v U^S_{a_i t} > 0$. Therefore, $G(t) > 0$ for $t > \bar{t}$, while $G(t) < 0$ for $t < \bar{t}$. Therefore, the set of action pairs induced in equilibrium is finite.

For each boundary type t_i, the induced action pair $(a_1([t_{i-1},t_i]), a_2([t_{i-1},t_i]))$ increases with t_{i-1} and t_i in both dimensions. The generalized single-crossing property guarantees the agent's incentive capability:

$$U^S(a_1([t_i,t_{i+1}]), a_2([t_i,t_{i+1}]), t) = U^S(a_1([t_{i-1},t_i]), a_2([t_{i-1},t_i]), t)$$

Now consider the two nearby partition elements, $[t_{i-1},t_i]$ and $[t_i,t_{i+1}]$. Principals' best responses and agent type t_{i-1}, t_i, t_{i+1}'s indifference gives the following condition:

$$t_{i-1} + t_{i+1} = 2t_i - \frac{2(b_1 + b_2 + 2b_1\lambda_2 + 2b_2\lambda_1)}{1 + \lambda_1 + \lambda_2}$$

This implies the size of each partition elements decreases step by step, at the length $\frac{2(b_1+b_2+2b_1\lambda_2+2b_2\lambda_1)}{1+\lambda_1+\lambda_2}$.

1.8.1.3 Proof of Corollary 1

From Proposition 2, we know the size of each partition element d_n decreases with $\frac{2(b_1+b_2+2b_1\lambda_2+2b_2\lambda_1)}{1+\lambda_1+\lambda_2}$. Denote the size of the first partition element as d_1. We have $d_1 + d_2 + ... + d_N = 1$. This implies

$$N \times d_N + \frac{2(b_1 + b_2 + 2b_1\lambda_2 + 2b_2\lambda_1)}{1 + \lambda_1 + \lambda_2} \times \frac{N(N-1)}{2} = 1$$

N is solvable in terms of d_N and all parameters. Since $d_N > 0$,

$$N \leq N_M = \langle \frac{1}{2} + \frac{[(b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2) \times (b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2 + 4(1 + \lambda_1 + \lambda_2))]^{\frac{1}{2}}}{b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2} \rangle$$

Take partial derivatives with respect to b_1 and λ_1(the marginal effect with respect to b_2 and λ_2 are in the same spirit) we get the following results:

$$\frac{\partial N_M}{\partial b_1} = -\frac{2(1+\lambda_1+\lambda_2)(1+2\lambda_2)}{\Lambda[\Lambda(\Lambda+4(1+\lambda_1+\lambda_2))]^{\frac{1}{2}}} < 0$$

where $\Lambda = b_1 + b_2 + 2b_2\lambda_1 + 2b_1\lambda_2 > 0$

$$\frac{\partial N_M}{\partial \lambda_1} = \frac{2(b_1-b_2)(1+2\lambda_2)}{\Lambda[\Lambda(\Lambda+4(1+\lambda_1+\lambda_2))]^{\frac{1}{2}}}$$

Hence $\frac{\partial N_M}{\partial \lambda_1} < 0$ if $b_1 < b_2$ while > 0 $b_1 > b_2$.

1.8.1.4 Proof of Proposition 3

Consider the subspace $\frac{(1+2\lambda_2)b_1+(1+2\lambda_1)b_2}{1+\lambda_1+\lambda_2} = \epsilon$ where ϵ is arbitrarily small. As $\epsilon \to 0$, from Corollary 3 we have:

$$\begin{aligned} N_M &= \frac{1}{2} + \frac{[\epsilon(1+\lambda_1+\lambda_2)(\epsilon(1+\lambda_1+\lambda_2)+4(1+\lambda_1+\lambda_2))]^{\frac{1}{2}}}{2\epsilon(1+\lambda_1+\lambda_2)} \\ &= \frac{1}{2} + \frac{\sqrt{\epsilon^2+4\epsilon}}{2\epsilon} \end{aligned}$$

Applying L'Hôpital's rule we have:

$$\lim_{\epsilon \to 0} N_M = \frac{1}{2} + \lim_{\epsilon \to 0} \frac{2(\epsilon+2)}{\sqrt{\epsilon^2+4\epsilon}}$$

Therefore, as $\epsilon \to 0$, $N_M \to \infty$.

1.8.1.5 Proof of Observation 1

We prove this result by first considering a two-by-two partition equilibrium. As shown in the following figure, suppose the agent adopts the following strategy: sending message m_1 if $t \in [0, t_1]$ and message n_1 if $t \in [t_1, 1]$ to Principal 1, while sending m_2 if $t \in [0, t_2]$ and n_2 if $t \in [t_2, 1]$ to Principal 2. Denote Principal 1's two induced actions $\hat{a_1}$ and $\tilde{a_1}$ and Principal 2' two induced actions $\hat{a_2}$ and $\tilde{a_2}$. Since the communication is private between the agent and one principal on each side, it is not necessarily $t_1 = t_2$. Without loss of generality, we look into the case where $t_1 \leq t_2$. The agent's incentive compatibility constraints under private communication involve the following three conditions:

- For the boundary type t_1 (fix the boundary type t_2) to be indifferent between the induced action pairs $(\hat{a_1}, \hat{a_2})$ and $(\tilde{a_1}, \hat{a_2})$, t_1 solves the following equation

$$-\ell_1(|\hat{a_1} - t_1|) - \ell_2(|\hat{a_2} - t_1|) = -\ell_1(|\tilde{a_1} - t_1|) - \ell_2(|\hat{a_2} - t_1|) \quad (1.2)$$

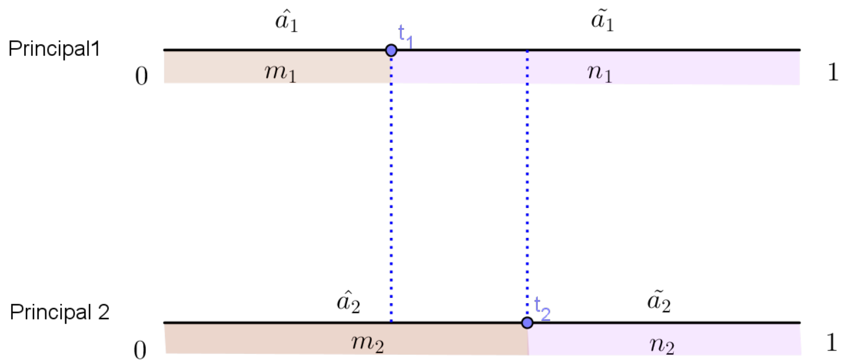


Figure 1.6: A Two-by-Two Partition Equilibrium under Private Communication

which implies $-\ell_1(|\hat{a_1}-t_1|) \geq -\ell_1(|\hat{a_1}-t_1'|), \forall t_1' > t_1$ and $-\ell_1(|\tilde{a_1}-t_1|) \geq -\ell_1(|\tilde{a_1}-t_1'|), \forall t_1' < t_1$.

- For the boundary type t_2 (fix the boundary type t_1) to be indifferent between the induced action pairs $(\tilde{a_1}, \hat{a_2})$ and $(\tilde{a_1}, \tilde{a_2})$

$$-\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\hat{a_2} - t_2|) = -\ell_1(|\tilde{a_1} - t_2|) - \ell_2(|\hat{a_2} - t_2|) \tag{1.3}$$

which implies $-\ell_2(|\hat{a_2}-t_2|) \geq -\ell_2(|\hat{a_2}-t_2'|), \forall t_2' > t_2$ and $-\ell_2(|\tilde{a_2}-t_2|) \geq -\ell_2(|\tilde{a_2}-t_2'|), \forall t_2' < t_2$

- For all $t_1' \in [0,1]$ and $t_2' \in [0,1]$, the agent's expected utility is the highest when she chooses the two boundary types as t_1 and t_2 and induces $\hat{a_1}$ and $\tilde{a_1}$ from Principal 1 and $\hat{a_2}$ and $\tilde{a_2}$ from Principal 2.

$$EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2) \geq EU^S_{\text{PRI}}(t_1', t_2'; b_1, b_2), \forall t_1', t_2' \in [0,1] \tag{1.4}$$

where $EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2) = \int_0^{t_1} -\ell_1(|\hat{a_1} - t|) - \ell_2(|\hat{a_2} - t|)dt + \int_{t_1}^{t_2} -\ell_1(|\tilde{a_1} - t|) - \ell_2(|\hat{a_2} - t|)dt + \int_{t_2}^{1} -\ell_1(|\tilde{a_1} - t|) - \ell_2(|\tilde{a_2} - t|)dt$

In contrast, under public communication in any two-step partition equilibrium, the agent have boundary type $\bar{t} \in [0,1]$ and principals have two induced pairs of actions (a_1, a_2) and (a_{11}, a_{22}) which solves:

$$\max_{a_i} \int_0^{\bar{t}} -L_i(|a_i - t - b_i|) - \Lambda_i(|a_1 - a_2|)dt$$
$$\max_{a_{ii}} \int_{\bar{t}}^{1} -L_i(|a_{ii} - t - b_i|) - \Lambda_i(|a_{11} - a_{22}|)dt$$

So the agent's incentive compatibility for boundary type $\bar{t}$ requires:

$$-\ell_1(|a_1 - \bar{t}|) - \ell_2(|a_2 - \bar{t}|) = -\ell_1(|a_{11} - \bar{t}|) - \ell_2(|a_{22} - \bar{t}|)$$

when this one condition holds, the global incentive compatibility constraint $EU^S_{\text{PUB}}(\bar{t}; b_1, b_2) \geq EU^S_{\text{PUB}}(t'; b_1, b_2), \forall t' \in [0,1]$ is always satisfied, as proved in the proof of proposition 2. So under public communication the incentive compatibility constraints for the agent are less tighter. Thus we prove the first part of this observation.

Now we move to the second part of the observation. In a cheap talk game without interdependent principals' actions (Goltsman and Pavlov (Goltsman and Pavlov 2011)), under private communication the principals' best response functions don't have interacting terms. Namely, $\hat{a_1} = \hat{a_1}(t_1, b_1), \tilde{a_1} = \tilde{a_1}(t_1, b_1)$ and $\hat{a_2} = \hat{a_2}(t_2, b_2), \tilde{a_2} = \tilde{a_2}(t_2, b_2)$. This helps us to reduce Equation 1.4 to

$$\begin{aligned}
EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2) &= \int_0^{t_1} -\ell_1(|\hat{a_1}(t_1, b_1) - t|)dt + \int_{t_1}^1 -\ell_1(|\tilde{a_1}(t_1, b_1) - t|)dt \\
&+ \int_0^{t_2} -\ell_2(|\hat{a_2}(t_2, b_2) - t|)dt + \int_{t_2}^1 -\ell_2(|\tilde{a_2}(t_2, b_2) - t|)dt \\
\geq EU^S_{\text{PRI}}(t'_1, t'_2; b_1, b_2) &= \int_0^{t'_1} -\ell_1(|\hat{a_1}(t'_1, b_1) - t|) + \int_{t'_1}^1 -\ell_1(|\tilde{a_1}(t'_1, b_1) - t|)dt \\
&+ \int_0^{t'_2} -\ell_2(|\hat{a_2}(t'_2, b_2) - t|)dt + \int_{t'_2}^1 -\ell_2(|\tilde{a_2}(t'_2, b_2) - t|)dt
\end{aligned}$$

as long as the boundary types t_1 and t_2 solves equations 1.2 and 1.3, the above inequality always holds. That is, in a cheap talk game without interacting principals, incentive compatibility conditions 1.2 and 1.3 implies condition 1.4. Therefore it suffices check the first two conditions for the private communication.

In our model with principals' interacting actions, however, checking only conditions 1.2 and 1.3 is not sufficient. To see this, notice that the principals' best response functions are interdependent:

$$\begin{aligned}
\hat{a_1} &= \max_{a_1} \int_0^{t_1} -L_1(|a_1 - t - b_1|) - \Lambda_1(|a_1 - a_2|)dt \\
\tilde{a_1} &= \max_{a_1} \int_{t_1}^1 -L_1(|a_1 - t - b_1|) - \Lambda_1(|a_1 - E(a_2)|)dt \\
\hat{a_2} &= \max_{a_2} \int_0^{t_2} -L_2(|a_2 - t - b_2|) - \Lambda_2(|E(a_1) - a_2|)dt \\
\tilde{a_2} &= \max_{a_2} \int_{t_2}^1 -L_2(|a_2 - t - b_2|) - \Lambda_2(|a_1 - a_2|)dt
\end{aligned}$$

where the values of a_2 (a_1) and $E(a_2)$ ($E(a_1)$) depend on t_2, b_2 (t_1, b_1). Therefore we have

$$\hat{a_1} = \hat{a_1}(t_1, t_2, b_1, b_2), \tilde{a_1} = \tilde{a_1}(t_1, t_2, b_1, b_2)$$

$$\hat{a_2} = \hat{a_2}(t_1, t_2, b_1, b_2), \tilde{a_2} = \tilde{a_2}(t_1, t_2, b_1, b_2)$$

Therefore the agent's expected utility from deviating to having boundary type t_1', t_2' is

$$EU^S_{\text{PRI}}(t_1', t_2'; b_1, b_2) = \int_0^{t_1'} -\ell_1(|\hat{a_1}(t_1', t_2', b_1, b_2) - t|) + \int_{t_1'}^1 -\ell_1(|\tilde{a_1}(t_1', t_2', b_1, b_2) - t|)dt$$
$$+ \int_0^{t_2'} -\ell_2(|\hat{a_2}(t_1', t_2', b_1, b_2) - t|)dt + \int_{t_2'}^1 -\ell_2(|\tilde{a_2}(t_1', t_2', b_1, b_2) - t|)dt$$

which is not necessarily less than or equal to $EU^S_{\text{PRI}}(t_1, t_2; b_1, b_2)$ merely under condition 1.2 and 1.3. In this case the incentive compatibility condition 1.4 is also binding. Therefore compared with Goltsman and Pavlov (2011), the agent faces tighter constraints under private communication in our model.

2.0 BAYESIAN PERSUASION WITH MULTIPLE RECEIVERS

2.1 INTRODUCTION

In many social or economic situations, a group of decision-makers receives advice from an informed agent. Examples include financial consulting, product advertising, and lobbying[1]. All these cases involve a biased agent: she wants the decision-makers to take the same action regardless of the state of the world. To attenuate the decision-makers' information disadvantage, in reality the agent's behavior is subject to certain restrictions. Not only is she prohibited from mis-reporting whatever information she has, but she also has to truthfully reveal the investigation process that leads her to discover the information. Kamenica and Gentzkow (2011) examine the case between one self-interested agent ("sender" thereafter) and one decision-maker ("receiver" thereafter). They show, despite her state-independent preference, that the sender can always be strictly better off than no persuasion by committing to a persuasion mechanism that provides noisy signals to the receivers. They also point out that this result does not easily extend to a multiple-receiver situation in which the receivers care about each other's decisions[2].

This paper investigates the role of persuasion mechanisms in multiple-receivers' collective decisions. We answer Kamenica and Gentzkow's open question: the persuasion environment matters to a great extent for the attainability of the sender's optimal payoffs. We compare public persuasion with private persuasion. In the former environment all receivers observe the sender's choice of the mechanism and the generated signals simultaneously. In the latter environment only the sender's mechanism is commonly known; each receiver gets a separate signal draw. We show, under public

[1] DellaVigna and Gentzkow's (2009) survey summarizes empirical studies on persuasion. Examples includes but not limited to persuading consumer of merchandise, persuading voters before elections, persuading donors to NPOs or charities, or persuading investors on financial markets. This survey also discusses persuader's incentives and roles, such as advertisers', financial analysts', and the Media's. Some persuasion channels are public while others are private; for instance, newspaper advertisements are to target all ages while propaganda through internet mainly attracts young citizens.

[2] See Kamenica and Gentzkow (2011) section 7.2.

persuasion, that the sender can always achieve the concave closure of the set of expected payoffs, no matter how many signals are drawn or what correlation structure they have. Under private persuasion the sender's payoff declines. From the sender's perspective a combination of both types of persuasion is also worse than a pure public one. In fact, any persuasion channel into which some private elements are introduced yields the sender less than her optimal payoff. For instance, a lobbyist might be better off if she holds a public meeting with all the legislators instead of approaching them privately. A salesman might benefit more if she makes a public disclosure about the product's quality to all consumers instead of persuading each of them separately[3].

We compare our results to Farrell and Gibbons (1989), the first discussion about public versus private information transmission. They show that whenever there exists an equilibrium in which the sender communicates informatively under private communication, there is an equilibrium under public communication in which the sender does the same. But the reverse is never true. We find, on the contrary, that private persuasion is always more informative than its public counterpart. If the signals are informative under public persuasion, the sender's optimal mechanism will generate more precise signals under private persuasion. Furthermore, Farrell and Gibbons (1989) does not have sharp welfare implications for the sender. We show, in contrast to their result, that the sender always gets a higher expected payoff under public persuasion. Moreover, this "higher payoff" is indeed the highest among all possible payoffs. It is also worth noting that private persuasion still yields the sender strictly positive payoff as compared to the one without persuasion, though the payoff level no longer reaches the upper bound of the set of all possible payoffs.

The key difference between two types of persuasion lies in their different impacts on receivers' beliefs. Signals from public persuasion lead to common belief about the unknown state of the world for all receivers; but signals from private persuasion do not. Thus similar to Kamenica and Gentzkow (2011), under public persuasion the sender can appropriately choose a state-dependent persuasion mechanism to manipulate these posterior beliefs. The mechanism generates noisy signals so that the posterior belief upon a "favorable" signal realization is just above the threshold doubt of a voting-pivot receiver[4]. Nevertheless, under private persuasion manipulating the receivers' posterior beliefs becomes more difficult for the sender. Each receiver forms his posterior belief not only according to his private signal, but also conditional on his vote being decisive. In such a

[3]One may argue that it increases the sender's cost to prepare different information packages for different people. However, we show that even *without* the additional cost in approaching the receivers separately, the sender still strictly prefers public persuasion.

[4]The receivers' votes are aggregated through a q-rule. As we shall show later, our results are robust to any q-rule, including majority, supermajority, and unanimity voting.

situation the sender's choice of the mechanism is subject to (1) the uncertainty of the distribution of all generated signals, and (2) belief interactions inherited in the receivers' strategic voting behavior.

The result that the sender benefits more from public persuasion is robust to the number of signals and the signals' correlation structure. The driving force underlying the welfare differences is not due to the number of trials. For example, the receivers may ask for N independent signal draws under public persuasion instead of only one. We show that the sender's payoff does not decrease to her payoff level under private persuasion, though the number of signals equals the one in a private environment. The reason is that the sender will adopt a different mechanism if she knows that multiple trials will be examined. Multiple trials are more likely to help the receivers to discover the true state, including the sender's least-favored state; sender's optimal mechanism thus becomes much less informative to offset the increased probability incurred by those unfavorable signals.

Finally, we demonstrate that the receivers' decision quality remains the same under public persuasion, regardless of the signals' structures. For instance, drawing N independent signals does not make the receivers better off. As discussed above, each signal becomes much less informative in this situation. This is because the receivers form the same posterior beliefs as long as the persuasion environment remains public. Thus the beliefs are under direct control of the sender by her appropriately choosing a signal-generating mechanism. In contrast, the sender's mechanism generates much more informative signals under private persuasion than under public persuasion. As a result, the receivers make better decisions under private persuasion.

The remainder of the paper is organized as follows. Section 2.2 introduces the model, strategies, and equilibrium concept. Section 2.3 compares the informativeness of the optimal mechanism under public persuasion with that under private persuasion, respectively. Section 2.4 presents the sender's and the receivers' welfare rankings over the two types of persuasion. Section 2.4.1.1 and 2.4.1.2 shows that the sender achieves the same level of expected payoffs under public persuasion, irrespective of the number of draws or the signals' correlation structure. And this level of expected payoffs are higher under public persuasion than that under private persuasion. Section 2.4.1.3 provides a stronger result that the sender does not only get "higher" expected payoffs under public persuasion; in fact the payoff under public persuasion serves as an upper bound for the sender's possible payoffs from any type of persuasion. Section 2.4.2 demonstrates that the receivers are better off under private persuasion; and drawing multiple signal draws under public persuasion fails to help them to reduce decision errors. Section 2.5 discusses two extensions of the model.

Section 2.6 provides detailed discussion on related literature, including literature on persuasion games and voting. Section 2.7 concludes.

2.2 MODEL

2.2.1 Setup

We analyze a Bayesian persuasion model with one sender ("she") and n receivers ("he"). Players' payoffs depend on the state of the world $t \in T = \{\alpha, \beta\}$, and and the receivers' collective decision. We assume the common prior probability distribution over T are: $\text{prob}(\alpha) = p, \text{prob}(\beta) = 1-p$, where $p \in [0,1]$. The collective decision is determined by voting, with alternatives denoted by $\{A, B\}$. Each receiver i casts a vote $v_i \in \{A, B\}$. Votes are aggregated by q-rule, which characterizes the minimum number of votes $m \in \{1, ..., n\}$ needed to implement alternative B[5]:

$$v(v_1, ..., v_n; m) = \begin{cases} B, & \text{if } |\{j : v_i = B\}| \geq m; \\ A, & \text{otherwise.} \end{cases}$$

Let $u^S(v,t)$ and $u_i^R(v,t), i = 1, ..., n$ denote the utility that the sender and each receiver derive from the implementation of the collective decision v in state t, respectively. We assume the sender's utility is state-independent:

$$u^S(B,t) > u^S(A,t), \forall t \in T$$

which means the sender always prefers alternative B to be implemented irrespective of the state t. The receivers, on the other hand, have state-dependent utilities.

$$u_i^R(B,\beta) > u_i^R(A,\beta), u_i^R(A,\alpha) > u_i^R(B,\alpha), \forall i = 1, ..., n$$

Receiver i prefers alternative A in state α and B in state β. And we allow heterogeneous preferences for different receivers, who might derive different levels of utility from each of the implemented alternatives, and assign different levels of utility loss to incorrect decisions. Namely, given the state t and final decision v, $u_j^R(v,t) \neq u_i^R(v,t)$, for $j \neq i$. Notice that all players' payoff structures are common knowledge; nevertheless, the receivers do not observe the true state of the world.

[5] We discuss all possible q-rules, where $q = \frac{m}{n}$, including simple majority ($m = \frac{n+1}{2}$), super-majority ($\frac{n+1}{2} < m < n$), and unanimity (m=n). Notice that under the q-rule, final decision is A if and only if at least $n - m + 1$ receivers vote for A. A receiver consider others' votes to be a tie when there are $m - 1$ votes for B and $n - m$ votes for A

The timing of the game is as follows. The sender sets up a signal-generating mechanism, which consists of a family of conditional distributions $\{\pi(\cdot|t)\}_{t\in T}$ over a space of signal realizations $S = \{a, b\}$[6]. The receivers get informed of $\{\pi(\cdot|t)\}_{t\in T}$. Then Nature determines the true state t, which is privately observed by the sender. The sender applies the conditional distribution in state t to generate noisy signal(s) and truthfully reveals the generated signal(s) to the receivers. Upon observing the signal(s), receivers cast votes simultaneously and the final decision is determined by a q-rule specified above. The timeline of the game is illustrated in Figure 2.1.

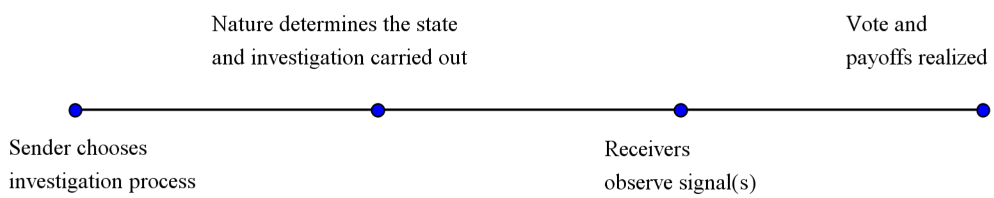


Figure 2.1: The Timeline of the Game

Note that different from traditional persuasion-game models, the sender does not have direct control over what the receivers might observe; instead, she tries to influence the receivers' decision by setting up a signal-generating mechanism. This can be interpreted as the revelation of the whole investigation process, such as conducting experiments, performing data analysis, or interviewing witnesses. Both the investigation process being employed and the evidence emerged are communicated to the decision-makers truthfully.

We compare the impacts of the sender's persuasion in two different institutions. Specifically, we are interested in the channel through which the generated signal realization(s) are transmitted, once the persuasion mechanism is established. We investigate two types of persuasion:

- Public persuasion: the sender's choice of $\{\pi(\cdot|t)\}_{t\in T}$ is commonly known to all receivers. The generated signal realization $s \in S$, is also public information. All receivers observe the same signal.
- Private persuasion: the sender's choice of $\{\pi(\cdot|t)\}_{t\in T}$ is commonly known to all receivers. The signal realizations, drawn from the mechanism independently, are observed by each receiver separately and privately. A receiver will observe either $s = a$ or $s = b$, but not both.

[6]Notice that the sender chooses this set of conditional distributions before she observes the true state. We generalize the signal realization space to $S = [0, 1]$ in Section 2.5.

2.2.2 Strategies and Equilibrium

This section specifies each player's strategy and the receivers' beliefs upon being informed of the signal-generating mechanism and observing the signal realization.

- The sender's strategy is to choose a family of conditional distributions $\{\pi(\cdot|t)\}_{t\in T}$ from $\mathcal{F} = \{\{\pi(\cdot|t)\}_{t\in T}|\pi : T \to \Delta S\}$, each element of which is a mapping from the state space to the simplex over the signal realization space.
- Each receiver's strategy, $\sigma_i : \mathcal{F} \times S \to \Delta\{A, B\}$, is a mapping from the Cartesian product of the collection of conditional distributions and the signal realization space to a simplex over voting alternatives[7].
- Each receiver's belief, $\mu_i : \mathcal{F} \times S \to \Delta T, i = 1, ..., n$, is a mapping from the Cartesian product of the collection of conditional distributions and the signal realization space to a simplex over the state space.

Similar to in Kamenica and Gentzkow (2011), we adopt *subgame perfect equilibrium* as the equilibrium notion here:

Definition 2. *A subgame perfect equilibrium of the multi-receiver Bayesian persuasion model is a strategy profile* $((\pi^*(\cdot|\alpha), \pi^*(\cdot|\beta)), \sigma_1^*, ..., \sigma_n^*)$ *such that*

- *Given* $(\sigma_1^*, ..., \sigma_n^*)$, $\forall t \in T$, $(\pi^*(\cdot|\alpha), \pi^*(\cdot|\beta))$ *is the maximizer of* $EU^S(v, t)$
- *Given the sender's choice of* $(\pi^*(\cdot|\alpha), \pi^*(\cdot|\beta))$,
  - *Each receiver's vote,* $\forall i \in \{1, ..., n\}, \forall s \in S$, $\sigma_i^*(\pi^*(\cdot|t), s) = \arg\max_{v_i} E_{\mu_i^*}(U_i^R(v, t))$, *where* v *is aggregated from* $v_i, i = 1, ..., n$ *via a q-rule*
  - *The receivers' posterior beliefs,* $\mu_i^*(t|s) = \frac{\pi(s|t)p(t)}{\sum_{t'\in T}\pi(s|t')p(t')}$, *where* $p(t)$ *denotes the common prior over* T

We assume that a receiver votes for B when he is indifferent between the two alternatives under the measure of his posterior belief μ_i. Thus with a binary signal realization space we can restrict our attention to pure strategy voting profile $(v_1^*, ..., v_n^*)$. More discussion about the voting behavior when space S is continuous is in Section 2.5.

[7] It is worth noting that the strategy space in our setting is different from the one(s) in cheap-talk games (Crawford and Sobel (1982), Green and Stokey (2007)). In the latter, the sender also chooses a family of signalling rules, which specifies a probability distribution over the message space in each state. The key difference is that in cheap-talk models the receiver(s) only observes the realized message(s), not the sender's choice of the family of signalling rules; whereas in our model the receivers observe both.

2.2.3 Preliminaries

This section provides preliminary analysis for the receivers' and the sender's problems, respectively. Two assumptions are specified before we proceed to the equilibrium characterization. One is on the receivers' collective preferences; the other is the *monotone likelihood ratio property* of the sender's signal-generating mechanism.

First, we simplify a receiver's problem under the measure of his posterior belief μ_i. Denote his posterior beliefs $\mu_i^\alpha, \mu_i^\beta$ for a given s induced by $\{\pi(\cdot|t)\}_{t\in T}$. This receiver's expected payoffs from voting for each of the alternatives are:

$$B : \mu_i^\beta \cdot u_i(B, \beta) + \mu_i^\alpha \cdot u_i(B, \alpha)$$
$$A : \mu_i^\beta \cdot u_i(A, \beta) + \mu_i^\alpha \cdot u_i(A, \alpha)$$

where $\mu_i^\beta + \mu_i^\alpha = 1$ given s. Thus this receiver votes for B if and only if

$$\mu_i^\beta \geq \frac{u_i(A, \alpha) - u_i(B, \alpha)}{u_i(A, \alpha) + u_i(B, \beta) - u_i(B, \alpha) - u_i(A, \beta)} \doteq q_i$$

We define, for receiver i, a *threshold doubt* q_i as a threshold value for the posterior belief above which this receiver will vote for B. We write the order statistics of the receivers' threshold doubts as $q_1 \leq ... \leq q_n$ and relabel the corresponding receivers as $R_1, ..., R_n$.

As assumed above, the sender prefers B to be implemented regardless of the state t. To demonstrate the sender's net benefit from either type of persuasion, we focus on the more interesting scenario in which the receivers' collective decision without the sender's persuasion is always $v = A$. It means that at least $n - m + 1$ receivers' threshold doubts are sufficiently high such that they would always vote for A based on their common prior. Formally, we assume:

Assumption 1. *Without the sender's persuasion, if the receivers were to cast a vote based on the common prior probability distribution over T, there will be less than m votes for option B, i.e.*

$$(1 - p) \cdot u_i(A, \beta) + p \cdot u_i(A, \alpha) \geq (1 - p) \cdot u_i(B, \beta) + p \cdot u_i(B, \alpha), i \in \{m, m + 1, ..., n\}$$

This is equivalent to a more convenient notation, $q_i \geq 1 - p$, for all $m \leq i \leq n$. We call A the receivers' *default* collective choice.

Next, we simplify the sender's problem. The sender derive utilities from each of the final decision $v \in \{A, B\}$ as:

$$EU^S = u^S(B, t) \cdot \text{Prob}(v = B) + u^S(A, t) \cdot \text{Prob}(v = A), \forall t \in T$$

where $\text{Prob}(v = A) = 1 - \text{Prob}(v = B)$ for each t. Thus the sender's problem reduces to choose $\{\pi(\cdot|t)\}_{t \in T}$ to maximize the probability that the voting outcome is B, since:

$$\max_{\{\pi(\cdot|t)\}_{t \in T}} EU^S = u^S(A,t) + (u^S(B,t) - u^S(A,t)) \cdot \text{Prob}(v = B), \forall t \in T \tag{2.1}$$

where all other terms besides $\text{Prob}(v = B)$ are constant. Note that the voting rule and Assumption 1 implies that this event occurs only when at least m receivers are persuaded to change their vote to the non-default alternative B.

The following assumption restricts our attention to a specific family of signal-generating mechanisms:

Assumption 2. *The signal-generating mechanism* $\{\pi(\cdot|t)\}_{t \in T}$ *satisfies* Monotone Likelihood Ratio Property *if*

$$\frac{\pi(s|t)}{\pi(s|t')} > \frac{\pi(s'|t)}{\pi(s'|t')}$$

for every $s > s'$ *and* $t > t'$[8].

With a binary state space $T = \{\alpha, \beta\}$, the sender's choice of a family of conditional distributions $\{\pi(\cdot|t)\}_{t \in T}$ is equivalent to finding densities $(\pi(s|\alpha), \pi(s|\beta))$. Assumption 2 implies the CDF $\Pi(\cdot|\beta)$ *first order stochastic dominates* $\Pi(s|\alpha)$, i.e. $\Pi(s|\beta) < \Pi(s|\alpha), \forall s \in S$[9]. Note that this assumption is not particularly restrictive: for any persuasion mechanism that does not satisfy MLRP, there exists another mechanism which satisfies the MLRP property and is outcome equivalent to the former.

Before proceeding to the main results of the paper, we have two remarks on the model setup. First, it does not matter whether the sender first chooses the investigation or Nature determines the true state, as long as the investigation process is truthfully revealed to the receivers and is verifiable ex-post. The sender's investigation specifies a family of conditional distributions over all possible states. Second, it is crucial that the investigation is reported to the receivers and is verifiable *ex-post*. The model reduces to a cheap-talk framework if the receivers only observe the generated signals but can never discover the signal-generating mechanism. In that case the sender would adopt a mechanism that would generate signal $s = b$ with probability 1 in both states and the receivers would rationally ignore any signal observation in equilibrium.

---

[8] The notion adopted here is *strictly monotone likelihood ratio* with the strict inequality held. Notice that a mechanism $\{\pi(\cdot|t)\}_{t \in T}$ with $\frac{\pi(s|t)}{\pi(s|t')} = \frac{\pi(s'|t)}{\pi(s'|t')}$ generates each signal $s \in S$ with the same probability in every state. It can be easily show that the receivers vote according to the prior probability distribution in this case.

[9] A brief proof is included in Appendix 2.8.1. And we will discuss the equilibrium characterization under this property more extensively in Section 2.5.

2.3 EQUILIBRIUM CHARACTERIZATION: PUBLIC VERSUS PRIVATE PERSUASION

In this section we first discuss the receivers' voting behavior upon observing signal realizations, taking the sender's persuasion mechanism $\{\pi(\cdot|t)\}_{t\in T}$ as fixed. Then we characterize the equilibrium of the persuasion game under public and private persuasion, respectively. Specifically, we compare the precision of signals generated by the sender's optimal mechanism under both types of persuasion.

2.3.1 Receivers' Voting Behavior

The receivers' voting behavior is described in terms of the threshold doubt q_i. Receiver i compares his posterior belief μ_i with q_i as shown in Section 2.2.3. We first define two terms that help us describe the voting strategies below.

Definition 3. *A receiver* votes sincerely *when he maximizes expected payoff conditional on his own signal observation only.*

Definition 4. *A receiver's vote is* informative *if his vote changes according to his own signal observation.*

Under Assumption 2, *informative voting* is equivalent to receiver i voting for B upon observing $s = b$ and voting for A upon observing $s = a$. We call a voting strategy *uninformative* if this receiver always votes for B (or A) regardless of the signal realization.

Lemma 1. *Under public persuasion, if Assumption 1 and 2 hold, then given any persuasion mechanism* $\{\pi(\cdot|t)\}_{t\in T}$ *the sender has chosen:*

- *Every receiver votes sincerely; and*
- *There exists a cutoff value* $\bar{q} \in [0, 1]$ *such that all receiver j's with threshold doubt* $q_j \in [\bar{q}, 1]$ *votes for A irrespective of his signal observation*

The proof of this lemma is included in the Appendix. The first statement is obvious. Under public persuasion all receivers have the same signal observation. So everybody casts vote conditional on his own observation. The second statement indicates that there are always a portion of the receivers whose threshold doubts are too large to be convinced by a signal realization $s = b$. So these receivers will always vote for A regardless of the signal observation generated by the persuasion mechanism. Note that if the number of these receivers exceeds $n - m$, the final decision will remain as the default A. Thus the sender has to appropriately choose the signal-generating mechanism so as to convince at least m receivers to vote informatively, as we shall show in the next section.

Next, we describe the receivers' voting behavior under private persuasion, when each of them observes independent signals draw separately. *Sincere voting* is no longer optimal. A rational receiver updates his posterior belief not only according to his own signal observation, but also according to the distribution of other receivers' observations. We define the following:

Definition 5. *A receiver* votes strategically *when he maximizes expected payoff conditional on his own observation and the event in which his vote is pivotal.*

Note that one's vote being pivotal is the only situation in which his vote will ever affect the voting outcome and his utility. Thus a receiver will infer the distribution of other receivers' observations from the event that his vote is pivotal and cast vote optimally. Denote $\gamma(k, r)$ the receiver's posterior belief μ_i^β when k out of r signals are $s = b$:

$$\gamma(k,r) = \frac{(1-p)(\pi(b|\beta))^k(1-\pi(b|\beta))^{r-k}}{p(\pi(b|\alpha))^k(1-\pi(b|\alpha)^{r-k}) + (1-p)(\pi(b|\beta))^k(1-\pi(b|\beta))^{r-k}}$$

We have the following result:

Lemma 2. *Under private persuasion, fix any persuasion mechanism $\{\pi(\cdot|t)\}_{t\in T}$ the sender has chosen. When each receiver votes strategically, there exists $0 \le q_k \le q_r \le 1$ such that*

- *A receiver i with threshold doubt $q_i \in [0, q_k]$ votes for B irrespective of his signal realization*
- *A receiver i with threshold doubt $q_i \in [q_k, q_r]$ votes informatively*
- *A receiver i with threshold doubt $q_i \in [q_r, 1]$ votes for A irrespective of his signal realization*

And the two cutoff values satisfy
$\gamma(m-k-1, r-k) \le q_k \le \gamma(m-k, r-k+1) \le q_r \le \gamma(m-k, r-k)$.

The proof of this result is included in the Appendix. Compared to Lemma 1, under private persuasion each receivers becomes more skeptical when casting his vote. This is because he updates his posterior belief upon the inferred distribution of signals from all informative-voters, not merely upon his own signal observation. Upon observing $s = a$, the event that one's vote being pivotal indicates some others' observing opposite signals $s = b$; thus μ_i^β, receiver i's posterior belief of the true state being β, increases. On the other hand, upon observing $s = b$, one's vote being decisive affects μ_i^β in a reversed manner: the presence of signals $s = a$ from some other receivers declines μ_i^β; thus this receiver i becomes more "skeptical" when casting a vote for B.

It is also worth noting that for receivers with heterogeneous preferences, strategic voting does not imply informative voting; nor vice versa. A receiver who votes uninformatively ignores his private signal rationally: his vote being decisive suggests that votes from the receivers who vote informatively constitute a tie. Upon observing either signal, a comparison between his updated posterior and threshold doubt might still lead to voting for one alternative irrespective of his own observation.

2.3.2 Informativeness of Sender's Optimal Mechanisms

2.3.2.1 Public Persuasion

Under public persuasion, all receivers observe the same signal realization. Since nobody has any private information, the distribution of their votes reveals no additional information regarding the true state t. Moreover, depending on the value of $\bar{q}$, only the receivers with threshold doubts $q_i \leq \bar{q}$ vote informatively. The final decision is B when at least m receivers do so upon observing a signal realization $s = b$. Therefore the sender's problem is to choose $(\pi(b|\alpha), \pi(b|\beta)) \in \mathcal{F}$ to maximize the probability of the voting outcome being $v = B$:

$$\max_{\pi(b|\alpha), \pi(b|\beta)} p \cdot \pi(b|\alpha) + (1-p) \cdot \pi(b|\beta) \tag{2.2}$$

subject to the receivers' voting behavior described in lemma 1. Then we have the following proposition:

Proposition 4. *Under public persuasion, the sender's optimal persuasion mechanism generates signal $s = b$ with positive probability in state α and generates signal $s = b$ with probability 1 in state β, i.e.*

$$\pi^*_{PUB}(b|\alpha) = \frac{(1-q_m) \cdot (1-p)}{q_m \cdot p}, \pi^*_{PUB}(b|\beta) = 1$$

which holds for all $m \leq n$. Moreover, there are at most m receivers who will vote informatively, i.e. $\bar{q} = q_m$.

The proof for this result is included in the Appendix. We make two remarks here. First, this proposition is true for all voting rules m, including majority rule $m = (n+1)/2$, super-majority rule $(n+1)/2 < m < n$, and the unanimity rule $m = n$. Second, under public persuasion there are $n - m$ receivers who has high threshold doubt voting uninformatively for A. This is because the sender only needs to convince m receivers to vote for B upon observing an $s = b$ observation.

Convincing more than m receivers is unnecessary; the sender's expected payoff will be lower if the sender produces more precise signals to convince an additional receiver.

2.3.2.2 Private Persuasion

Under private persuasion, the receivers observe independent draws of signal realizations separately. Compared to Section 2.3.2.1, both the receivers' voting behavior and the sender's problem change: as described in Section 2.3.1, the receivers take into account the distribution of other receivers' private signals as described. A receiver casts a vote according to his own signal observation as well as the probability of this vote being pivotal, the latter of which is the only event that his action might affect his utility. Thus each receiver updates posterior belief based on information beyond what his own private signal reveals. Consequently, the sender has to incorporate the receivers' strategic voting into her choice of the mechanism. Formally, the sender chooses $(\pi(b|\alpha), \pi(b|\beta)) \in \mathcal{F}$ to maximize:

$$\max_{\pi(b|\alpha),\pi(b|\beta)} p\cdot(\sum_{j=m-k}^{r-k}\binom{r-k}{j}(\pi(b|\alpha))^j(1-\pi(b|\alpha))^{r-k-j})+(1-p)\cdot(\sum_{j=m-k}^{r-k}\binom{r-k}{j}(\pi(b|\beta))^j(1-\pi(b|\beta))^{r-k-j}) \tag{2.3}$$

subject to the receivers' strategic voting behavior described in lemma 2. Note that the sender's problem differs from the one under public persuasion (Equation 2.2) in two aspects: (i) the probability of generating enough b signal realizations is written as a Binomial distribution function; and (ii) the maximization problem is constrained by the receiver's strategic-voting behavior. To better explain out result, we first discuss an intermediate case which only involves aspect (i) – we suppose the receivers vote sincerely. Sincere voting is not optimal for the receivers, but the result serves as a good benchmark as we compare public with private persuasion later on.

Now suppose each receiver observes independent draws of private signal realization and votes sincerely. The sender still chooses $(\pi(b|\alpha), \pi(b|\beta)) \in \mathcal{F}$ to maximize the probability of the voting outcome $v = B$. Nevertheless, she can no longer ensure that all receivers observe the same signal realization. For example, if the sender adopts a mechanism the same as in Section 2.3.2.1 such that exact m receivers will vote for B upon getting signal $s = b$, she will not always get the voting outcome as $v = B$ upon a b signal: some among these m receivers are very likely to observe an independent draw $s = a$ instead of $s = b$. Thus she will be better off by adopting another mechanism such that more than m receivers, i.e. receivers $\{R_1, ..., R_t\}$, are willing to vote informatively:

$$\max_{\pi(b|\alpha),\pi(b|\beta)} p\cdot(\sum_{j=m}^{t}\binom{t}{j}(\pi(b|\alpha))^j(1-\pi(b|\alpha))^{t-j})+(1-p)\cdot(\sum_{j=m}^{t}\binom{t}{j}(\pi(b|\beta))^j(1-\pi(b|\beta))^{t-j})$$

$$\text{subject to } \frac{y(1-p)}{xp+y(1-p)}\geq q_t$$

$$0\leq x\leq 1, 0\leq y\leq 1$$

Then we have the following observation[10]:

Observation 4. *Under private persuasion, when all receivers vote sincerely, the sender's optimal persuasion mechanism generates signals with probabilities:*

$$\pi_{SIN}(b|\alpha)=\frac{(1-q_t)\cdot(1-p)}{q_t\cdot p}, \pi_{SIN}(b|\beta)=1$$

where $\bar{q}=q_t\geq q_m$. *There are less than* $n-m$ *receivers who vote uninformatively for* A. *Moreover, this result holds for any voting rule* m.

The proof of this observation is in the Appendix. It says that under private persuasion the sender will choose the conditional distributions $\{\pi(\cdot|t)\}_{t\in T}$ to convince t receivers, where $m\leq t\leq n$, to vote for B upon observing a private signal $s=b$, if she knows that the receivers are going to vote sincerely. The intuition is that the sender does so to compensate the loss in probability when some receivers observe a signal $s=a$ instead of $s=b$, an event unfavorable to the sender. Moreover, the probability that a "wrong" signal is generated is lower under private persuasion than under public persuasion, i.e. $\pi_{\text{SIN}}(b|\alpha)\leq\pi^*_{\text{PUB}}(b|\alpha)$ since $q_t\geq q_m$. In other words, under private persuasion, the optimal mechanism generates signal with higher precision in the state where the sender's and the receivers' interests are mis-aligned.

Now we move to solve the sender's maximization problem with rational voting behavior under private persuasion. We have the following result:

Proposition 5. *Under private persuasion with any voting rules* $m\leq n$, *when all receivers vote strategically, the sender's optimal persuasion mechanism generates signal realizations with probability* $\pi^*_{PRI}(b|\alpha),\pi^*_{PRI}(b|\beta)$ *that satisfy:*

$$\pi^*_{PRI}(b|\alpha)<\pi_{SIN}(b|\alpha), \pi^*_{PRI}(b|\beta)\leq 1$$

The proof of this proposition is included in the Appendix. The first inequality shows that the probability that the optimal mechanism generates a "wrong" signal in state α is lower when

[10] As mentioned in section 2.3.1, sincere voting under private persuasion is not rational. Thus we write this result as an *Observation*, not a *Proposition*. And we do not attach $*$ to the sender's persuasion mechanism $\{\pi(\cdot|t)\}_{t\in T}$.

receivers behave strategically. This is a positive effect of the receivers' pivotal voting: inferring others' information from their votes makes a receiver more skeptical. Thus it forces the sender to reduce the probability of a "wrong" signal being generated. The second inequality, on the other hand, indicates that the optimal mechanism no longer generates signal $s = b$ with probability 1 even in the state with perfect interests alignment. This is from negative effect of the receivers' strategic interaction: some receivers are too skeptical to believe in their own signal observations. All in all, the inferred distribution of others' observations might not coincide with the real distribution of all signal realizations.

2.3.2.3 Comparison between the Two Institutions

This section compares the sender's optimal mechanism under public persuasion with the one under private information. We use the term *informativeness* to represent the probability that the signal representing the true state is generated in that state. Summarizing the results in Section 2.3.2.1 and 2.3.2.2, we have the following corollary:

Corollary 2. *For all voting rules $m \leq n$, in equilibrium the informativeness of the persuasion mechanism satisfies:*

$$\pi^*_{PRI}(a|\alpha) > \pi^*_{PUB}(a|\alpha)$$
$$\pi^*_{PRI}(b|\beta) \leq \pi^*_{PUB}(b|\beta) = 1$$

This result shows that, in state α where players' interests are mis-aligned, the informativeness of the sender's optimal mechanism is higher under private persuasion than that under public persuasion. On the other side, in state β, the mechanism generates a favorable signal $s = b$ with probability 1 under public persuasion, but with a probability less than 1 under private persuasion when the receivers voting strategically. In other words, the mechanism generate an unfavorable signal $s = a$ with positive probability even in the state where the players' interests are perfectly aligned.

These key differences result from differences in the sender's tradeoffs under each type of persuasion. Since less than m receivers would vote for alternative A based on the common prior, the sender adopts a mechanism such that more receivers will be convinced to vote informatively, i.e. vote for B instead of A upon observing a favorable signal $s = b$. Under public persuasion, the sender adjusts the precision of the signals so as to convince at least m receivers to vote for B upon observing $s = b$. But convincing more receivers with higher threshold doubts requires higher information precision in state α. This in turn increases the probability that the receivers discover the true state; thus the sender's benefit from the persuasion declines. We call this the **information precision effect**. Under private persuasion, the sender faces two additional tradeoffs: first, as discussed in Section 2.3.2.2, increasing the information precision and convincing more receivers with higher threshold doubts compensates the sender for the loss from the uncertainty associated with separate signal draws. The probability of observing at least m favorable signals out of a total of $m+j$ independent draws is always greater than the one of observing m such signals out of exactly m independent draws. This **probability increment effect** is irrelevant under public persuasion; yet it exists under private persuasion, no matter the receivers vote sincerely or strategically. Second, under private persuasion with strategic-voting receivers, as shown in lemma 2 and proposition 5, convincing more receivers to vote informatively provides them with additional information on the distribution of others' private signals. This makes each receiver more skeptical than if he were to vote sincerely. With one's own observation fixed, a receivers' posterior belief μ_i^β decreases when there is one more receiver who votes informatively, i.e. conditional on one's vote being pivotal, $\gamma(m, r) > \gamma(m, r+1)$. We call it **pivotal-voting effect**.Eventually it forces the sender to choose $\pi^*_{\text{PRI}}(b|\beta) \leq 1$.

It is also worth noting that the follow corollary holds in equilibrium under both public and private persuasion.

Corollary 3. *Under both public and private persuasion there is no equilibrium in which all receivers vote **uninformatively**.*

This corollary excludes two cases in any *subgame perfect equilibrium*: first, eliminating dominated strategies excludes the case in which all receivers ignore their observations and vote for B uninformatively. Second, nor is it possible that all receivers vote for A irrespective of signal observations in equilibrium. Under either type of persuasion, the sender optimally chooses the signal-generating mechanism such that at least a portion of the receivers cast their votes based on relevant information generated by the mechanism.

2.4 WELFARE EFFECTS

2.4.1 Sender's Benefit from the Persuasion

In this section we first compare the sender's benefit from public persuasion with that from private persuasion. Then we demonstrate that the sender's welfare ranking is robust to the number of signal draws. Last, we show a stronger and more general result that the sender can achieve the upper bound of the set of her expected payoffs under public persuasion, regardless of the number of signal draws or the signals' correlated structure.

2.4.1.1 Comparing Sender's Expected Payoffs: Public versus Private Persuasion

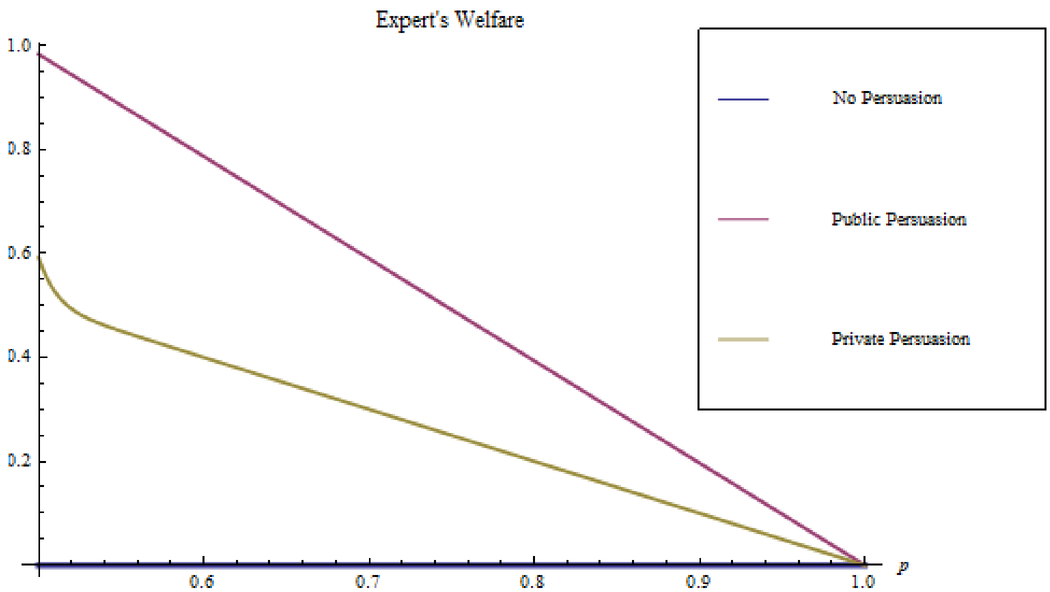
The sender's expected utility from either type of persuasion is represented by Equation 2.1. Maximizing her expected utility is equivalent to maximizing the summed probability of reaching her preferred decision in each state, as described by the sender's objective functions in Section 2.3.2.1 and 2.3.2.2. Given the sender's optimal choice of the persuasion mechanism $(\pi^*(\cdot|\alpha), \pi^*(\cdot|\beta))$ under each type of persuasion, we have the following comparison result:

Proposition 6. *The sender's optimal mechanisms under public persuasion and under private persuasion yield the sender different levels of expected utilities:*

$$EU^S_{PRI} < EU^S_{PUB}$$

The proof of this result is included in the Appendix. It shows that the sender's expected utility is higher under public persuasion and lower under private persuasion. Moreover, the latter term is strictly greater than the sender's utility when there is no persuasion and all receivers vote based on the common prior, i.e. $EU^S_{\text{PRI}} > u^S(A, t)$ since $\text{Prob}_{\text{PRI}}(v = B) > 0$. Comparison between the probability $\text{Prob}(v = B)$'s with no persuasion, public persuasion, and private persuasion as a function of the common prior p is shown in the left part of Figure 2.2.

In short, compared to Kamenica and Gentzkow (2011), the sender benefits from persuasion; but but the benefit is smaller under private persuasion. This result describes the sender's "cost" of communicating in a private environment: the receivers' strategic interactions in equilibrium reveal extra information about each others' private signals beyond what the sender's mechanism has transmitted. On the other hand, under public persuasion the same signal is observed by all receivers simultaneously. The receivers, at best, can make use of what the sender's mechanism has



Parameter values: $n = 60, m = 31$ (majority rule), $q'(\cdot) > 0$.

Figure 2.2: The sender's expected utility as a function of p

generated and transmitted. So public persuasion helps the sender achieve a higher level of expected utility.

2.4.1.2 Multiple Draws of Signals under Public Persuasion

In this section we address a potential problem associated with the above comparison between persuasion environments: under public persuasion the sender sets up a mechanism to generate one signal realization, whereas under private persuasion the sender chooses a mechanism to generate n independent signal realizations for n receivers. Now we demonstrate that the comparison result is robust to the number of signal draws.

Consider the institution which requires the sender to choose a mechanism $\{\pi(\cdot|t)\}_{t \in T}$ to draw n independent signal realizations; all receivers observe these realizations publicly. Despite the number of trials, all other elements of the game remain unchanged. Same as the public persuasion with one signal draw, receivers hold common posterior belief. Denote $\gamma(\ell, n)$ the posterior belief μ^b when there are ℓ signals $s = b$ out of all n signals. Assumption 2 implies that $\gamma(\ell - 1, n) < \gamma(\ell, n), \forall \ell \in \{1, ..., n\}$. A receiver i with threshold doubt q_i will vote for B when there are more than ℓ favorable signals ($s = b$) out of all n signals if and only if $\gamma(\ell - 1, n) < q_i < \gamma(\ell, n)$.

The sender chooses a mechanism such that at least m receivers will vote for B upon observing more than a threshold number ($\hat{\ell}$) of favorable signals, i.e. $\gamma(\ell, n) \leq q_m \leq \gamma(\ell+1, n)$. The sender's objective is to maximize the probability that at least $\hat{\ell}$ favorable signals are generated in each state:

$$\begin{aligned}
&\max_{\pi(b|\alpha),\pi(b|\beta)} p \cdot \sum_{j=\ell}^{n} (\pi(b|\alpha))^j (1-\pi(b|\alpha))^{n-j} \\
&+(1-p) \cdot \sum_{j=\ell}^{n} (\pi(b|\beta))^j (1-\pi(b|\beta))^{n-j} \\
&\text{subject to } \gamma(\ell, n) \leq q_m \leq \gamma(\ell+1, n)
\end{aligned}$$

Solving the sender's problem yields the following proposition:

Proposition 7. *Under public persuasion with multiple independent signal draws, the sender's optimal mechanism generates signals with probabilities:*

$$\pi^*_{MD}(b|\alpha) = \sqrt[n]{\pi^*_{PUB}(b|\alpha)}, \pi^*_{MD}(b|\beta) = 1$$

Moreover, the sender's expected utility is the same as that under public persuasion with a single signal draw.

$$EU^S_{MD} = EU^S_{PUB}$$

The proof of this proposition are included in the Appendix. The first part can also be expressed as $\pi^*_{\text{MD}}(b|\alpha) = \sqrt[n]{\frac{(1-q_m)(1-p)}{q_m p}} = \sqrt[n]{\pi^*_{\text{PUB}}(b|\alpha)}, \pi^*_{\text{MD}}(b|\beta) = \sqrt[n]{\pi^*_{\text{PUB}}(b|\beta)}$. Since $\pi^*_{\text{PUB}}(b|\alpha) < 1$, we have $\pi^*_{\text{MD}}(b|\alpha) > \pi^*_{\text{PUB}}(b|\alpha)$, which indicates that the optimal mechanism generates "wrong" signals with higher probability in state α when multiple signals instead of a single one are drawn. In other words, the signals become much less precise in the state where the players' interests are mis-aligned. Moreover, as $n \to \infty$, $\pi^*_{\text{MD}}(b|\alpha) \to 1$, which means the informativeness of the mechanism drops drastically as the number of trials grows. When the number of public draws increases towards the number of independent draws under private persuasion, the signals do not become as precise as those under private persuasion. Instead, they becomes even vaguer than the one under public persuasion with a single draw.

The second part of the proposition shows that the sender achieves the same level of expected utility as she does under public persuasion with a single trial. The reason is that as long as the persuasion channel remains public, receivers have no private information which might be revealed by the distribution of their votes in equilibrium. Drawing multiple signals increases the probability of the receivers observing unfavorable signals ($s = a$). Therefore, once the sender knows that multiple trials would be examined, she will optimally choose a mechanism that generates unfavorable signals much less frequently.

2.4.1.3 Public Persuasion and The Upper Bound of Sender's Expected Payoffs

In previous sections we have shown that the sender's expected utility is higher under public persuasion. In this section, we present a stronger result: public persuasion helps the sender to achieve the "highest possible payoff" for any given common prior p. Specifically, we shall drop Assumption 1 and show that the sender can always attain the concave closure of the set of her possible payoffs under public persuasion irrespective of the number of signal draws or the signals' correlated structure.

The following figures illustrate the sender's sets of possible payoffs as a function of the common prior p. The blue line in Figure 2.3 represents the sender's expected payoff without persuasion. When at least m receivers' threshold doubts are sufficiently low, i.e. $p < 1 - q_m$, the collective decision is B even without the persuasion. So the sender gets $\text{Prob}(v = B) = 1$. However, when $p > 1 - q_m$, the sender gets $\text{Prob}(v = B) = 0$ as the probability of getting her preferred alternative voted. Thus the sender's expected payoff is discontinuous at $p = 1 - q_m$.

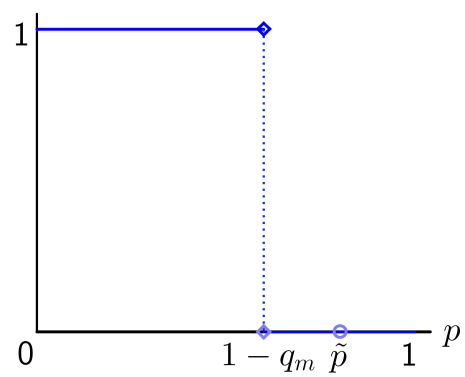


Figure 2.3: Sender's expected payoffs without persuasion

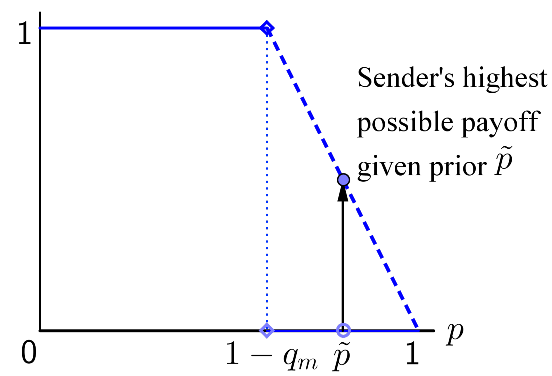


Figure 2.4: Sender's expected payoffs under public persuasion: the convex hull

The dashed blue line in Figure 2.4 illustrates the sender's payoff increment under public persuasion. To formally establish the result, we first define the *concave closure* of the graph of EU^S as:

Definition 6. *Denote* Λ *as the* concave closure *of* EU^S*:*

$$\Lambda(p) \doteq \sup\{u : (p, u) \in co(EU^S)\}$$

where $co(EU^S)$ *is the* convex hull *of the graph of* EU^S*.*

Then we have the following result:

Proposition 8. *For any voting rule* m *and any* $\tilde{p} \in [0,1]$*, under public persuasion the sender achieves an expected payoff on the concave closure of the set of all possible payoffs, i.e.*

$$EU^S_{PUB}(\tilde{p}) = \Lambda(\tilde{p})$$

Moreover, this result remains unchanged even if the sender is required to draw multiple signal realizations, or to draw correlated signals.

The proof of this result is included in the Appendix. Compared with the sender's payoff without persuasion in Figure 2.3, it is easy to see that with public persuasion, $\forall \tilde{p} < 1 - q_m$, the sender's payoffs remain the same; while $\forall \tilde{p} > 1 - q_m$, the sender's expected payoffs lie on the dashed blue line in Figure 2.4. The probability of getting her preferred voting outcome is $\text{Prob}(v = B) = \frac{1}{q_m}(1 - \tilde{p})$. In other words, public persuasion helps the sender to achieve the concave closure of the convex hull of the sender's set of possible payoffs.

It is also worth noting that the uniqueness of the *supremum* guarantees the uniqueness of the sender's optimal payoff for each given $p \in [0,1]$ under public persuasion. This is also true for the case in which multiple signal draws are generated, and/or the signals are correlated. This is because signals under public persuasion always give rise to common belief about the unknown state of the world. It does not matter whether all receivers observe a single signal draw, or observe a series of signal draws, or observe a series of correlated signal draws. In this situation the sender can appropriately choose a state-dependent persuasion mechanism to manipulate the common posterior beliefs. Her optimal mechanism generates noisy signals so that the posterior belief upon a signal realization $s = b$ is just above the threshold doubt of receiver-m.

Notice that the concave closure of the convex hull of the payoff set is the highest possible the sender is able to achieve. As shown in previous sections, adding private element worsens the sender's welfare. A combination of public and private persuasion also makes the sender worse off.

2.4.2 Receivers' Decision Quality

In this section we compare the receivers' decision quality under public persuasion to that under private persuasion. Specifically, we compare the summed probabilities of two types of errors, i.e. $\text{Prob}(v = B|\alpha) + \text{Prob}(v = A|\beta)$ as the voting outcome does not match the true state. Note that in any *subgame perfect equilibrium*, the first term is the probability that at least m receivers vote for B in state α. The second term is the probability that at most $m-1$ receivers vote for B in state β so that the default alternative remains as the voting outcome.

Under public persuasion, the sum of two types of errors equals the probability that the sender's mechanism generates a favorable signal ($s = b$) in state α plus the probability that the mechanism fails to generate such a signal in state β. Under private persuasion, it equals the probability that the mechanism generates at least $m-k$ favorable signals for receivers $\{R_k, ...R_r\}$ who vote informatively in state α plus the probability that the mechanism generates no more than $m-k-1$ favorable signals for $\{R_k, ...R_r\}$ in state β.

Moreover, if the receivers are more prone to decision mistakes under public persuasion, a natural question is whether the decision quality improves when more independent signals are drawn and observed by all receivers. Particularly, when the number of such independent trials approaches n, the number of separate draws under private persuasion, will the receivers' collective decision be as good as the one under private persuasion? We have the following result:

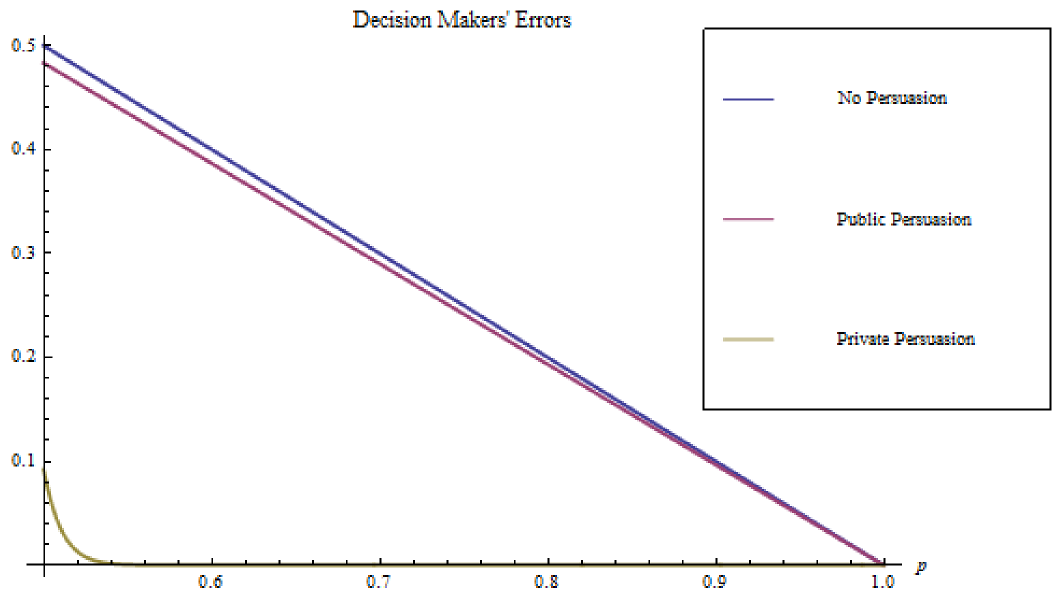
Proposition 9. *The probabilities of the receivers' decision errors is larger under public persuasion than that under private persuasion*

$$\psi_{PRI} < \psi_{PUB}$$

Moreover, the probability of receivers' decision errors under public persuasion with multiple signal draws is the same under public persuasion with a single signal draw.

$$\psi_{MD} = \psi_{PUB}$$

The proof of this result is included in the Appendix. The first part of the proposition shows that the probability of receivers' decision errors is higher under public persuasion and lower under private persuasion. The lower the probability, the higher the decision quality. Note that the sender's and the receivers' interests are perfectly-aligned in state β and completely mis-aligned in state α. From previous sections we show that public persuasion maximizes the sender's expected payoffs. In other words, it maximizes the probability of $s = b$ being generated in both states. Public persuasion maximizes receivers' type I error while minimizes their type II error. Nevertheless, in equilibrium the sender's optimal mechanism generates more precise signal realizations in the sender's favorable state β. Thus the receivers' gain from public persuasion's minimizing the type II error does not compensate the loss from its maximizing the type I error. Therefore, under public persuasion the receivers suffer from higher error probabilities overall.



Parameter values: $n = 60, m = 31$ (majority rule), $q'(\cdot) > 0$.

Figure 2.5: The receivers' decision mistakes as a function of p

Figure 2.5 shows the comparison between the error probabilities without persuasion, with public persuasion, and with private persuasion. The error probability under public persuasion, ψ_{PUB}, is sometimes higher than the error probability when there is no persuasion and the receivers vote based on their common prior. In the latter case, the error probability $\psi_N = 1 - p$. Specifically, if the m^{th} receiver's threshold doubt is lower than the common prior, i.e. $q_m \leq p$, we have $\psi_{\text{PUB}} \geq \psi_N$.

The second part of the proposition provides a negative answer to the question we raised above. Increasing the number of trials does not help the receivers to reduce the probability of decision errors. The reason is that as long as the persuasion channel remains public, receivers have no private information which might be revealed by their votes in equilibrium. Drawing multiple signals increases the probability of the receivers observing unfavorable signals ($s = a$). Therefore, once the sender knows that multiple trials would be examined, she will optimally choose a mechanism that generates unfavorable signals much less frequently.

In short, receivers make a better decision under private persuasion. This result captures the receivers' "benefit" of communicating in a private environment: when the receivers vote strategically, they update posterior beliefs based not only on their own private signals, but also on additional information revealed by others' equilibrium votes. On the other side, under public persuasion, publicly observed signal creates common belief among all. Difference in receivers' votes results only from their heterogeneous preferences, but do not reveal extra information regarding each others' private signals. Thus public persuasion from a biased sender makes the receivers worse off.

2.5 DISCUSSION AND EXTENSIONS

2.5.1 Correlated Private Signals

In this section we drop the assumption that the receivers' signal observations are independent under private persuasion. Instead, we consider a richer private persuasion environment which allows the sender to generate correlated realizations conditional on each state t. Consider the case in which the sender chooses a persuasion mechanism $\{\pi(\cdot|t)\}_{t\in T}$ to draw n correlated signal realizations. Each receiver privately observe one signal realization.

Let ν Denote the number of $s = b$ out of all n signals in each state. Denote the joint distribution of the n signal realizations in the following way:

in state α:

$P(\nu = n|\alpha) = \eta_n, P(\nu = n-1|\alpha) = \eta_{n-1}, ..., P(\nu = 0|\alpha) = \eta_0$

in state β:

$P(\nu = n|\beta) = \xi_n, P(\nu = n-1|\beta) = \xi_{n-1}, ..., P(\nu = 0|\beta) = \xi_0$

Marginalizing on the rest $n-1$ signal realizations gives the marginal probability of each one signal realization. Additionally, assume the receivers' signal observations are interchangeable. Thus the probability of all the event in which k receivers observe $s = b$ and $n-k$ observe $s = a$ can be written as $\binom{n}{k} \cdot \eta_k$ in state α and $\binom{n}{k} \cdot \xi_k$ in state β. For the joint probability we require:

$$\sum_{k=0}^{n} \binom{n}{k} \cdot \eta_k = 1, \sum_{k=0}^{n} \binom{n}{k} \cdot \xi_k = 1$$

Similar to Section 2.3.1, we first derive the receiver's posterior belief μ_i^β when k out of r signals are $s = b$ (while the rest $n-r$ receivers' votes are uninformative) as:

$$\gamma(k,r) = \frac{(1-p)\sum_{j=k}^{n-(r-k)} \binom{n}{j} \cdot \xi_j}{p\sum_{j=k}^{n-(r-k)} \binom{n}{j} \cdot \eta_j + (1-p)\sum_{j=k}^{n-(r-k)} \binom{n}{j} \cdot \xi_j}$$

All other analysis and results in Section 2.3.1 remain unchanged. We can write the sender's maximization problem as:

$$\max_{(\eta_0,...\eta_n),(\xi_0,...,\xi_n)} p\cdot(\sum_{j=m-k}^{r-k}\binom{r-k}{j}\eta_j)+(1-p)\cdot(\sum_{j=m-k}^{r-k}\binom{r-k}{j}\xi_j)$$

Incorporating correlation structure into the private persuasion allows us to link the two persuasion environments together. As the correlation between the signal realizations becomes larger, i.e. $\eta_n+\eta_0\to 1$ and $\xi_n+\xi_0\to 1$, the private persuasion environment approaches the public persuasion environment. Thus we have the following conjecture:

Conjectrure 1. *When the joint distributions* $\{(\eta_0,...\eta_n),(\xi_0,...,\xi_n)\}$ *satisfy*

$$\frac{\xi_k}{\xi_{k-1}}\leq\frac{\eta_k}{\eta_{k-1}},\forall k\in\{1,...,n\}$$

As $\eta_n+\eta_0\to 1$ *and* $\xi_n+\xi_0\to 1$*, we have* $EU^S_{PRI}\to EU^S_{PUB}$.

A sketch of the proof of this conjecture is included in the Appendix. The underlying intuition is that the receivers form common posterior belief once the persuasion environment become public. Upon observing his own signal realization, a receiver believes, with higher probability, that others also observe the same realization. This is a *confirming effect* that offsets the *pivotal voting effect* under private persuasion with independent signal observations. Upon observing a sender's favorable signal $s=b$, a receiver now becomes less skeptical about the true state being β. The *pivotal voting effect* forces the sender to generate more precise signal realizations and consequently decreases the sender's expected payoff; whereas this *confirming effect* drives the sender's expected payoffs to the opposite direction.

Moreover, as we show in previous sections, drawing multiple signal draws under public persuasion yields the sender the same expected payoff as drawing one signal draw, i.e. $EU^S_{\text{PUB}}=EU^S_{\text{MD}}$. Thus we conclude that the sender's expected payoff under private persuasion increases to that under public persuasion with the *same number of* signal draws when the signals becomes perfectly correlated.

2.5.2 Generalization of the Signal Realization Space

In this section we generalize the signal realization space to a continuous space, $S=[0,1]$. We characterize the receivers' voting behavior and then discuss the sender's optimal persuasion mechanisms in equilibrium.

2.5.2.1 Receivers' Voting Behavior

A strategy for the sender is to choose a family of conditional distributions from $\mathcal{F} = \{\pi(\cdot|t)\}_{t\in T}$ each element of which maps from the state space T to the simplex over S. A strategy for receiver i is a measurable mapping $\sigma_i : \mathcal{F} \times S \to \Delta\{A, B\}$, where $\sigma_i(s_i; \pi)$ denotes the probability that this receiver votes for β upon observing signal s_i from $[0, 1]$. Specifically, we are interested in a class of voting profiles where each receiver adopts a *cutoff strategy*:

Definition 7. *A receiver's voting strategy σ_i is called a* cutoff voting strategy *if $\exists \bar{s}_i \in [0, 1]$ such that*

$$\sigma_i(s) = \begin{cases} 1, & \textit{if } s \geq \bar{s}_i \\ 0, & \textit{if } s < \bar{s}_i \end{cases}$$

Next we show how the cutoff signal $\bar{s}_i$ is determined and what properties it satisfies. Given a voting strategy σ_i, the probability that receiver i votes for B in state α is $\int_0^1 \sigma_i(s)\mu(a|s)ds$ and the probability that i votes for B in state β is $\int_0^1 \sigma_i(s)\mu(b|s)ds$. The probabilities that receiver i is pivotal in state α and in state β are as follows, respectively:

$$P_{-i}(\text{piv}|\alpha) = \sum_{|M|=m-1, i\notin M} (\prod_{j\in M} (\int_0^1 \sigma_i(s)\mu(\alpha|s)ds) \prod_{j\notin M, j\neq i} (\int_0^1 (1-\sigma_i(s))\mu(\alpha|s)ds))$$

$$P_{-i}(\text{piv}|\beta) = \sum_{|M|=m-1, i\notin M} (\prod_{j\in M} (\int_0^1 \sigma_i(s)\mu(\beta|s)ds) \prod_{j\notin M, j\neq i} (\int_0^1 (1-\sigma_i(s))\mu(\beta|s)ds))$$

For receiver i, define

$$\Delta(\sigma_{-i}, s) = P_{-i}(\text{piv}|\alpha) \cdot \pi(s|\alpha) \cdot q_i \cdot p - P_{-i}(\text{piv}|\beta) \cdot \pi(s|\beta) \cdot (1-q_i) \cdot (1-p)$$

Similar to Duggan and Martinelli (Duggan and Martinelli 2001), we have the following lemmas:

Lemma 3. *Given the voting profile σ_{-i} of all other receivers, a strategy σ_i is a best response for receiver i if and only if*

$$\sigma_i(s) = \begin{cases} 1, & \textit{if } \Delta(\sigma_{-i}, s) < 0 \\ 0, & \textit{if } \Delta(\sigma_{-i}, s) > 0 \end{cases}$$

Moreover, it is equivalent to the following cutoff strategy

$$\hat{\sigma}_i(s) = \begin{cases} 1, & \textit{if } s \geq \bar{s}_i \\ 0, & \textit{if } s < \bar{s}_i \end{cases}$$

where $\bar{s}_i = \inf\{s \in [0, 1] | \Delta(\sigma_{-i}, s) \leq 0\}$

Lemma 4. *The cutoff signal* $\bar{s_i}$ *is monotone increasing in the receiver's threshold doubt* q_i*. Moreover, the cutoff signal* $\bar{s_i}$ *is monotone increasing in the ratio of the conditional probability density functions* $\frac{\pi(s|\alpha)}{\pi(s|\beta)}, \forall s \in [0,1]$*.*

The proofs of both lemmas are given in the Appendix. It follows directly that the cutoff signal values satisfy $0 \leq \bar{s_1} \leq ... \leq \bar{s_n} < 1$. All receivers vote for A when $s \in [0, \bar{s_1}]$ and vote for B when $s \in [\bar{s_n}, 1]$. For signal realizations $s \in [\bar{s_1}, \bar{s_n}]$, the number of receivers who vote for each alternative varies. Thus under public persuasion the sender chooses a persuasion mechanism which maximizes the probability of a signal realization s such that $\bar{s_m} \leq s < 1$. Under similar lines of reasoning as Proposition 2.3.2.1, we have the following corollary:

Corollary 4. *Under public persuasion with a continuous signal realization space* $S = [0,1]$*, suppose the receivers adopt cutoff voting strategy described in Lemma 3. Then for any voting rule* m *and any common prior* $p \in [0,1]$*, the sender can achieve the expected payoffs on the concave closure of the set of all possible payoffs, i.e.*

$$EU^S_{PUB}(p) = \Lambda(p)$$

where $\Lambda(p)$ is defined in Section 2.4.1.3. The proof of this result is included in the Appendix.

2.6 RELATED LITERATURE

Farrell and Gibbons (1989) are the first to compare public and private information transmission. They analyze a one-sender two-receiver cheap-talk game with binary state space. They show that whenever there exists an equilibrium in which the agent truthfully reveals the underlying state under private communication, there is an equilibrium under public communication in which the agent does the same. However, the reverse is not true[11]. Our result shows, in contrast, that the sender's signalling mechanism is always more informative under private persuasion and less informative under public persuasion.

There are three key differences between Farrell and Gibbons' (1989) setting and ours, which drive such differences in results. First, players' conflict of interests are different. Farrell and Gibbons (1989) adopt a cheap-talk model where not only the receivers', but also the sender's

[11] Goltsman and Pavlov (2011) extend Farrell and Gibbons' (1989) results with continuous state space and continuous action space. They also show that the sender is willing to reveal more information if she is allowed to combine the two messaging channels together.

utility is state-dependent. We analyze a persuasion game in which only the receivers prefer state-contingent decisions; the sender always intends to induce one decision regardless of the state. Second, we analyze different communication techniques. In Farrell and Gibbons (1989) the sender uses costless and non-verifiable messages; whereas in our model the sender has no direct control over what the receivers observe, though she can choose the signal-generating mechanism. Third, the receivers follow different decision rules. In Farrell and Gibbons (1989) each receiver takes an action separately, with no interaction of any form occurring between them. In our setting the final decision is made through q-rule voting. The sender has to incorporate the strategic interactions between receivers' beliefs. The rest of this literature review will discuss previous research on each of the three features, i.e. sender's state-independent preferences, persuasion mechanism, and strategic voting.

We first discuss the literature which involves a sender preferring one decision regardless of the state while receiver(s) desire state-contingent decisions. Grossman (1981) and Milgrom (1981) analyze the information disclosure via verifiable messages between a seller (sender) and buyers (receiver). Although babbling equilibria in which the receiver ignores any of the sender's messages always exists, in a sequential equilibrium a high-quality seller is able to distinguish herself by making a full disclosure. This is because such equilibria imposes restrictions on the receiver's off-equilibrium beliefs; the receiver forms rational expectation about the sender's true types. Chakraborty and Harbaugh (2011) investigate a sender's persuasion through cheap-talk messages. With one-dimensional state space and state-independent preferences, no non-babbling equilibrium exists in the cheap-talk model. Nevertheless, with a multidimensional state space a sender can convince a receiver by making credible comparative statements over the two states. However, when there are two receivers to make the decision, interactions between them might offset the benefit from the sender's persuasion.

The second strand of literature discusses persuasion mechanisms. This paper extends the one by Kamenica and Gentzkow[12] (2011) to a multi-receiver framework. Kamenica and Gentzkow (2011) analyzes a Bayesian persuasion game between one informed sender and one uninformed receiver. The sender chooses a state-dependent persuasion mechanism to provide signal realizations to the receiver. Both the mechanism and the generated signal are known to the receiver. They derive

[12] Two more papers discusses similar mechanisms. Rosar and Schulte (2010) look into the design of a device with which an imperfectly informed sender can generate public information about the underlying state. The set of the generated information is a superset of the set of the sender's private information. However, the designer's goal is to provide information as precise as possible, unlike in Kamenica and Gentzkow (2011) and ours the sender is biased towards one decision outcome. Rayo and Segal (2010) develop a sender's optimal disclosure rule with a multidimensional state space. In Rayo and Segal (2010) the receiver also has private information regarding the true state; yet in our model the receivers do not possess any relevant information.

necessary and sufficient conditions under which an optimal mechanism that yields the sender strictly positive benefit exists. The key difference between Kamenica and Gentzkow (2011) and ours arises from the strategic feature of the receivers' collective decision-making. In Kamenica and Gentzkow (2011) the sender faces a single receiver, who updates his belief upon observing one signal realization and makes a decision on his own. The crux of Kamenica and Gentzkow (2011) is whether the sender can commit to a persuasion mechanism so that the receiver will take the sender's preferred action upon receiving a more favorable signal. In our model multiple receivers vote for one final decision. A receiver makes implicit inference about the distribution of others' signal observations conditional on his own vote being pivotal. In turn the sender incorporates this strategic effect into her choice of mechanisms. The sender's problem remains as convincing enough receivers to vote for the sender-preferred alternative upon observing a favorable signal; but each receiver's incentive to do so changes[13].

Third, the distinction between the two types of persuasion is the strategic consideration associated with the receivers' voting behavior. Existing voting literature has extensively discussed how private information affects the quality of collective decisions. Austen-Smith and Banks (1996) (and Feddersen and Pesendorfer (1998)) point out, as a challenge to the Condorcet Jury Theorem[14], that it is not rational for each voter to vote only according his private signal. A rational voter has to take into account the information revealed by the event of one's vote being decisive[15]. Feddersen and Pesendorfer (1997) analyze a general voting model in which voters have heterogeneous preferences and receive noisy private signals from different information services. Feddersen and Pesendorfer (1997) fully characterizes the voters' equilibrium voting behavior and demonstrates that a q-rule voting fully aggregates information despite that the fraction of voters who vote informatively decreases to zero as the electorate grows to infinity. Gerardi and Yariv (2007) consider a committee

[13]The discussion of "persuasion mechanism" is also related to a broader scope of literature on persuasion rules. Glazer and Rubinstein (2004) study a persuasion game with one sender and one receiver, the latter of whom can verify the former's report for at most one of two aspects. The persuasion mechanism, with the objective to minimize the probability of the receiver's decision errors, specifies a set of cheap-talk messages for the sender to choose from, a device for the receiver to select the aspect to be checked, and a rule for the receiver to take the final action. Glazer and Rubinstein (2006) examine a similar setting but allow the receiver to randomize at the final stage of decision. Glazer and Rubinstein (2006) shows that there exists a persuasion rule with no randomization and all optimal rules satisfy ex-post optimality.

[14]Condorcet Jury Theorem states that if each voter's noisy private signal is with precision greater than 1/2 and if each votes according to this private signal, the probability of selecting the correct alternative approaches 1 as the electorate turns to infinity.

[15]In defense of the Condorcet Jury Theorem, McLennan (1998) show that for any common interest voting game, if with the sincere voting assumption collective decisions can successfully select the right alternative, there exists an equilibrium voting profile with everyone responding strategically. More importantly, outcomes from all such Nash equilibria are at least as good as the ones from sincere but non-optimal voting profiles.

voting with deliberation. Sequential equilibria exists with committee members truthfully revealing own private information and rationally adjust one's vote according to other members' information. In this paper the information service from which private noisy signals are generated is from the sender. The novelty is that the sender optimally chooses from a family of information service and optimally adjust the strength of the signals in each state.

We close this section by comparing this paper with Caillaud and Tirole (2007), who also examine a one-sender multiple-receiver model. In their paper, a sender, who lacks information regarding each receiver's type, approaches receivers in a group sequentially to get a project approved. The receiver being approached can investigate with a fixed cost and get informed. The sender's task is to design a mechanism which is incentive compatible for the selected receiver(s) to costly acquire the information and to approve the project if his type is high. The "persuasion mechanism" examined in ours is different from Caillaud and Tirole (2007). In the latter the mechanism selects key receiver(s) to conduct a costly investigation. The sender does not possess any relevant information regarding the underlying state. This paper looks into the "investigation process" itself, the details of which Caillaud and Tirole (2007) ignore. We show that as long as the sender can design the details of an investigation, the difference in persuasion environments matters even though the receivers have no investigation cost and are willing to examine the sender's report.

2.7 CONCLUSION

In this paper we analyze a Bayesian persuasion model with one sender persuading multiple receivers. We compare two institutions: under public persuasion, both the sender's choice of mechanism and the generated signals are observed by all receivers publicly; under private persuasion, the former remains commonly known while the latter is drawn independently and separately for each receiver. We show, in the state where hers and the receivers' interests are mis-aligned, that the sender's optimal mechanism generates more informative signals under private persuasion than under public persuasion. However, in the state where the players' interests are perfectly-aligned, the signal that represents the true state is not perfectly informative. This is because strategic interactions among receivers reveal more information than the mechanism itself has transmitted; and the sender incorporates this effect into her choice of the optimal mechanism.

It naturally follows that the sender achieves higher expected utility under public persuasion than under private persuasion while the receivers make a wrong decision with higher probability under public persuasion than under private persuasion. This result captures the sender's cost of transmitting information in a private environment and the receivers' benefit from getting informed in a private environment. Apparently a private environment provides the sender with incentives to tell different lies to different receivers; nevertheless it no longer benefits the sender in the same manner when she loses direct control over receivers' signal observations. It is also worth noting that as long as the persuasion channel remains public, increasing the number of independent draws does not help the receivers reduce decision errors. In fact, the sender's optimal mechanism generates very imprecise signals in the state where players' interests are mis-aligned. This mechanism yields the sender the same expected payoff as she can achieve under public persuasion with a single draw.

There are three effects underlying the difference between the two institutions: the *information precision effect*, which means the mechanism generates more precise signals in the state where interests are misaligned so as to convince more receivers to vote informatively; the *probability increment effect*, which refers to the fact that convincing more receivers to vote informatively offsets the decrement in probability caused by independent signal draws; and the *stratetic voting effect*, which suggests that convincing more receivers to vote informatively enables the receivers to learn more about each others' signal observations than what has actually been transmitted by the mechanism itself. Under public persuasion only the *information precision effect* is relevant, no matter there is a single draw or multiple independent draws. In contrast, under private persuasion the sender faces all three effects, the balance between which forces the mechanism to generate more precise signals in the state where players' interests are misaligned.

2.8 APPENDIX: PROOFS AND CALCULATIONS

2.8.1 The Implication from the Monotone Likelihood Ratio Property

Observation 5. *For any signal-generating mechanism* $\{\pi(\cdot|t)\}_{t\in T}$ *that satisfies the* Monotone Likelihood Ratio Property*, with a binary state space* $T = \{\alpha, \beta\}$*, the CDF* $\Pi(\cdot|\beta)$ first order stochastic dominates $\Pi(\cdot|\alpha)$.

Proof The MLRP implies that for every $s > s'$ and $t > t'$,

$$\frac{\pi(s|t)}{\pi(s|t')} > \frac{\pi(s'|t)}{\pi(s'|t')}$$

given any $s^* \in S$,

- if $s^* < s$, we have

$$\begin{aligned} \pi(s|\beta) \cdot \pi(s^*|\alpha) &> \pi(s^*|\beta) \cdot \pi(s|\alpha) \\ \text{so that } \int_{s \in S, s^* < s} \pi(s|\beta) \cdot \pi(s^*|\alpha) ds &> \int_{s \in S, s^* < s} \pi(s^*|\beta) \cdot \pi(s|\alpha) ds \\ \text{implies } \pi(s^*|\alpha) \cdot (1 - \Pi(s^*|\beta)) &> \pi(s^*|\beta) \cdot (1 - \Pi(s^*|\alpha)) \end{aligned}$$

- if $s^* \geq s$, similarly we have

$$\pi(s^*|\alpha) \cdot \Pi(s^*|\beta) < \pi(s^*|\beta) \cdot \Pi(s^*|\alpha)$$

Therefore we have $\forall s^* \in S$,

$$\begin{aligned} \frac{\Pi(s^*|\beta)}{\Pi(s^*|\alpha)} < \frac{\pi(s^*|\beta)}{\pi(s^*|\alpha)} &< \frac{1 - \Pi(s^*|\beta)}{1 - \Pi(s^*|\alpha)} \\ \text{implies } \Pi(s^*|\beta) &< \Pi(s^*|\alpha) \end{aligned}$$

The last inequality indicates that $\Pi(\cdot|\beta)$ FOSDs $\Pi(s|\alpha)$. We replace $\int$ with $\sum$ in this exercise when S is finite.

2.8.2 Proofs for Lemmas, Observations, and Propositions

2.8.2.1 Proof of Lemma 1 Fix a sender's persuasion mechanism $\{\pi(\cdot|t)\}_{t \in T}$. Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$. Assumption 2 is now equivalent to $y > x$. Receiver i's posterior belief $\mu(\beta|\cdot)$ upon observing each signal realization is

$$\begin{aligned} \mu(\beta|a) &= \frac{(1-y)(1-p)}{(1-x)p + (1-y)(1-p)} \\ \mu(\beta|b) &= \frac{y(1-p)}{xp + y(1-p)} \end{aligned}$$

First notice that $\mu(\beta|a) < \mu(\beta|b)$ under Assumption 2. As shown in Section 2.2.3, a receiver i votes for B if and only if $\mu(\beta|\cdot) \geq q_i$. Therefore take $\bar{q} = \mu(\beta|b) = \frac{y(1-p)}{xp+y(1-p)}$. It is obvious that $\bar{q} \in [0, 1]$. For all receivers with $q_i \leq \bar{q}$, we have $q_i \leq \mu(\beta|b)$, which indicates that these receivers will vote for B on signal $s = b$. For the receivers with $q_j > \bar{q}$, we have $q_j > \mu(\beta|b)$, which means they votes for A on signal $s = b$.

Note that this lemma does not exclude the possibility that some receivers vote *uninformatively* for B after observing either $s = a$ or b signal realization. For receivers whose q_j's are sufficiently small, it is possible that $\mu(\beta|a) > q_j$. However, the following claim shows that the number of receivers who always vote for B does not dominate the voting outcome:

Claim: Under Assumption 1 and 2, the number of receivers who vote for B regardless of the signal observation is less than m.

To see this, we only have to show $\mu(\beta|a) < q_m$.

$$\begin{aligned} &\Leftrightarrow \quad \frac{(1-y)(1-p)}{(1-x)p+(1-y)(1-p)} < q_1 \\ &\Leftrightarrow \quad x < y \text{ and } q_j \geq 1-p, \forall j \in \{m, ..., n\} \end{aligned}$$

which is always true under Assumption 1 and 2.

2.8.2.2 Proof of Lemma 2

Fix a sender's persuasion mechanism $\{\pi(\cdot|t)\}_{t\in T}$. Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$. Denote $\gamma(k,r)$ the receiver's posterior belief μ_i^β when k out of r signals are $s = b$:

$$\gamma(k,r) = \frac{(1-p)y^k(1-y)^{r-k}}{px^k(1-x)^{r-k} + (1-p)y^k(1-y)^{r-k}}$$

Assumption 2 implies that $\forall 1 \leq k \leq r \leq n$, the following hold:

$$\begin{aligned} \gamma(k-1,r) &< \gamma(k,r) \\ \gamma(k,r+1) &< \gamma(k,r) \\ \gamma(k,r) &< \gamma(k+1,r+1) \\ \gamma(k-1,r) &< \gamma(k,r+1) \end{aligned}$$

For any r independent draws, suppose receiver i knows that there are $k-1$ signals $s = b$. Then his posterior belief is $\mu_i^\beta = \gamma(k,r) \in (0,1)$ if himself observes $s = b$ and $\gamma(k-1,r) \in (0,1)$ if himself observes $s = a$. There exist receivers with high threshold doubts $q_j's$ such that $\gamma(k-1,r) < \gamma(k,r) < q_j$, which means this receiver votes for A regardless of his signal observation. On the other hand, unlike in lemma 1, $\gamma(k,r) \geq q_i$ does not necessarily imply $\gamma(k-1,r) < q_i$. There exist receivers with low threshold doubts $q_i's$ such that $q_i \leq \gamma(k-1,r) < \gamma(k,r)$, which means this receiver votes for B irrespective of his own observation. So there are a portion of the receivers $\{R_1, ..., R_k\}$ who vote for B uninformatively and a portion of receivers $\{R_r, ..., R_n\}$ who vote for A uninformatively. And notice that $k \geq 1$, the equality holds depending on the value of x, y.

Next we show that $k < m < r$. Suppose $k \geq m$. The voting outcome will be B in both states, which is wrong with probability p. Thus voting for B irrespective of the signal observation becomes a weakly dominated strategy for some receivers in $\{R_1, ..., R_k\}$. These receivers will be better off if they switch to an informative voting strategy. Similar argument holds for the case $r \leq m$. Thus if all receivers adopt undominated strategies, in any subgame the voting outcome will not be determined by the receivers who vote uninformatively for either alternative.

Now we consider the rational voting behavior in each subgame when every receiver acts strategically. First, only receivers $\{R_{k+1}, ..., R_r\}$ vote informatively, i.e. the distribution of their votes reveals the distribution of their private signals. For these $r-k$ receivers, observing $s=a$ and being pivotal indicate that there are $m-k-1$ signals that are $s=b$ out of all $r-k$ signals; and observing $s=b$ and being pivotal indicate that there are $m-k$ signals that are $s=b$ out of all $r-k$ signals. To have these receivers to vote informatively we need:

$$\gamma(m-k-1, r-k) \leq q_j \leq \gamma(m-k, r-k), j=k+1, ..., r$$

For receivers $\{R_1, ..., R_k\}$ who vote for B irrespective of the signal observation, observing $s=a$ and being pivotal indicate that there are $m-k$ signals that are $s=b$ out of all $r-k+1$ signals (including his own); and observing $s=b$ and being pivotal indicate that there are $m-k+1$ signals that are $s=b$ out of all $r-k+1$ signals (including his own). To have these receivers to vote uninformatively for B we need:

$$q_j \leq \gamma(m-k, r-k+1), q_j \leq \gamma(m-k+1, r-k+1), j=1, ..., k$$

For receivers $\{R_{r+1}, ..., R_n\}$ who vote for A irrespective of the signal observation, observing $s=a$ and being pivotal indicate that there are $m-k-1$ signals that are $s=b$ out of all $r-k+1$ signals (including his own); and observing $s=b$ and being pivotal indicate that there are $m-k$ signals that are $s=b$ out of all $r-k+1$ signals (including his own). To have these receivers to vote uninformatively for A we need:

$$q_j \geq \gamma(m-k-1, r-k+1), q_j \geq \gamma(m-k, r-k+1), j=r+1, ..., n$$

Summing up, we have:

$$\gamma(m-k-1, r-k) \leq q_k \leq \gamma(m-k, r-k+1) \leq q_r \leq \gamma(m-k, r-k)$$

From above, it can be verified that for all $(x, y) \in [0,1] \times [0,1]$, if q_k is fixed, then there is a unique cutoff value q_r that satisfies $\gamma(m-k, r-k+1) \leq q_r \leq \gamma(m-k, r-k)$; if q_r is fixed, then there is a unique cutoff value q_k that satisfies $\gamma(m-k-1, r-k) \leq q_k \leq \gamma(m-k, r-k+1)$. But note that the two cutoff values may change simultaneously. Consider a three-receiver example with $q_1 < q_2 < q_3$. There exists values of (x, y) such that the conditions $\gamma(2,3) \geq q_3, \gamma(1,3) < q_1$ (which means it is incentive compatible for all three receivers to vote informatively) and the conditions $q_1 \leq \gamma(1,2) \leq q_3, \gamma(0,1) < q_2 \leq \gamma(1,1)$ (which means R1 votes uninformatively for B, R2 votes informatively, and R3 votes uninformatively for A) are both satisfied. Thus for these values of (x, y)'s, there are multiple sets of best-response functions at the voting stage.

The above being mentioned, such multiplicity does not affect the conclusion of Lemma 2, nor the subsequent equilibrium characterization and welfare effects. This is because Lemma 2 has incorporated all the possible pairs of values $\{q_k, q_r\}$ that characterizes the receivers' best responses at the voting stage, and the sender's maximization problem under private persuasion in Equation 2.3 is well-defined for each pair of such $\{q_k, q_r\}$.

2.8.2.3 Proof of Proposition 4

Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$ denote the sender's choice of the conditional probabilities. Under public persuasion, all receivers observe the same signal realization s. As shown in lemma 1, no receiver votes for B upon observing signal realization $s = a$. Upon observing $s = b$, if at least m receivers vote for B, the final decision is B. This requires the receivers' posterior belief upon observing $s = b$ to satisfy:

$$\mu(\beta|b) = \frac{y(1-p)}{xp + y(1-p)} \geq q_m$$

since $q_m \geq ... \geq q_1$, this inequality implies that receivers $\{R_1, ..., R_m\}$ will vote for B upon observing $s = b$. So the sender's problem is:

$$\max_{x,y \in [0,1]^2} g(x,y) = p \cdot x + (1-p) \cdot y$$
$$\text{subject to } \frac{y(1-p)}{xp + y(1-p)} \geq q_m$$

where the objective function is the sum of the probabilities that signal $s = b$ is generated in each state.

First notice that the $g(x, y)$ is continuous in (x, y). The constraint set,
$\{(x, y) \in [0,1]^2 | \frac{y(1-p)}{xp+y(1-p)} \geq q_m\}$ is closed and bounded in $\mathbf{R}^2$; thus it is compact. By *Weierstrass's extreme value theorem*, the maximum of the objective function is attainable. We form the Lagrangian function

$$L(x, y, \lambda) = g(x, y) - \lambda(-\frac{y(1-p)}{xp + y(1-p)} + q_m)$$

the Kuhn-Tucker condition yields the following equations and inequalities:

$$
\begin{aligned}
&p+\frac{\lambda p(1-p)y}{(px+(1-p)y)^2}=0\\
&(1-p)+\lambda\frac{p(1-p)x}{(px+(1-p)y)^2}=0\\
&\lambda(-\frac{y(1-p)}{xp+y(1-p)}+q_m)=0\\
&\lambda\geq 0\\
&\frac{y(1-p)}{xp+y(1-p)}\geq q_m
\end{aligned}
$$

solving the equations and inequalities yields $x^*=\frac{(1-q_m)\cdot(1-p)}{q_m\cdot p}, y^*=1$. In fact, the constraint is binding, i.e. $\mu(\beta|b)=q_m$.

Now we show that there are exactly m receivers who vote for B upon observing $s=b$. Suppose there are only $m-1$ receivers who do so. In this case the voting outcome is always $v=A$ in each state, which yields the sender only $u^S(A,t)$. But the sender's optimal mechanism yields $u^S(A,t)+(u^S(B,t)-u^S(A,t))\cdot g(x^*,y^*)>u^S(A,t)$. Now suppose instead there are $m+1$ receivers who vote for B upon observing $s=b$, which changes the constraint to $\mu(\beta|b)\geq q_{m+1}$. Solving the sender's problem we get $x'=\frac{(1-q_{m+1})\cdot(1-p)}{q_{m+1}\cdot p}, y'=1$. Since $q_{m+1}\geq q_m$, $g(x',y')\leq g(x^*,y^*)$. Therefore, in equilibrium there are exact m receivers who vote for B upon observing $s=b$, i.e. $\bar{q}=\mu(\beta|b)=q_m$.

2.8.2.4 Proof of Observation 4

Let $x=\pi(b|\alpha)$ and $y=\pi(b|\beta)$ denote the sender's choice of the conditional probabilities. Under private persuasion, each receiver observes an independent draw of the signal realization. When a receiver votes sincerely, he update posterior belief based only on his own observation. As shown in lemma 1, no receiver votes for B upon observing signal realization $s=a$. Upon observing signal $s=b$, on the other hand, if the t^{th} receiver votes for B, i.e. $\mu(\beta|b)\geq q_t$, so do all receivers in $\{R_1,...,R_t\}$. The sender's problem can be formalized as maximizing the probability of $t\geq m$ receivers' observing a favorable signal realization in both states:

$$
\begin{aligned}
&\max_{x,y} g(x,y)=p\cdot\sum_{j=m}^{t}\binom{t}{j}\cdot x^j(1-x)^{t-j}+(1-p)\cdot\sum_{j=m}^{t}\binom{t}{j}\cdot y^j(1-y)^{t-j}\\
&\text{subject to }\frac{y(1-p)}{xp+y(1-p)}\geq q_t\\
&0\leq x\leq 1, 0\leq y\leq 1
\end{aligned}
$$

First notice that the $g(x,y)$ is continuous in (x,y). The constraint set, $\{(x,y) \in [0,1]^2 | \frac{y(1-p)}{xp+y(1-p)} \geq q_t\}$ is closed and bounded in $\mathbf{R}^2$; thus it is compact. By *Weierstrass's extreme value theorem*, the maximum of the objective function is attainable. We form the Lagrangian function:

$$L(x,y;\lambda_1,\lambda_x,\lambda_y,\nu_x,\nu_y) = g(x,y) - \lambda_1(q_t \cdot p \cdot x + (q_t-1)(1-p)y - 0) - \lambda_x \cdot (x-1) - \lambda_y \cdot (y-1) + \nu_x \cdot x + \nu_y \cdot y$$

with $p \in [0,1]; m,t \in \mathbf{N}, m < t; q_t = q(t) \in [1-p,1], q'(\cdot) > 0$. The Kuhn-Tucker condition for the above constrained maximization problem yields the following equations and inequalities.

$$p \cdot \sum_{j=m}^{t} \binom{t}{j} \cdot x^{j-1}(1-x)^{t-j-1} \cdot (j-tx) - \lambda_1 \cdot q_t \cdot p - \lambda_x + \nu_x = 0$$
$$(1-p) \cdot \sum_{j=m}^{t} \binom{t}{j} \cdot y^{j-1}(1-y)^{t-j-1} \cdot (j-ty) - \lambda_1 \cdot (q_t - 1) \cdot (1-p) - \lambda_y + \nu_y = 0$$
$$\lambda_1 \cdot (q_t \cdot p \cdot x + (q_t-1)(1-p)y) = 0$$
$$\lambda_x \cdot (x-1) = 0$$
$$\lambda_y \cdot (y-1) = 0$$
$$\nu_x \cdot x = 0$$
$$\nu_y \cdot y = 0$$
$$\lambda_1, \lambda_x, \lambda_y, \nu_x, \nu_y \geq 0$$

solving the equations and inequalities yields $x_{\text{SIN}} = \frac{(1-q_t)\cdot(1-p)}{q_t \cdot p}, y_{\text{SIN}} = 1$. In fact, the constraint is binding, i.e. $\mu(\beta|b) = q_t$.

Notice x_{SIN} is a function of q_t, so does the maximum $g(x_{\text{SIN}}, y_{\text{SIN}})$. We then have the following claim:

Claim: There exists a receiver t with threshold doubt q_t such that $g(x_{\text{SIN}}, y_{\text{SIN}})$ is maximized.

Proof of the claim: First by Envelope Theorem,

$$\frac{d}{dt} g(x_{\text{SIN}}(t), y_{\text{SIN}}; t) = \frac{\partial}{\partial t} g(x_{\text{SIN}}(t), y_{\text{SIN}}; t)$$

where q_t is monotone increasing in t. Then we simplify $g(x_{\text{SIN}}, y_{\text{SIN}})$ as:

$$\begin{aligned} g(x_{\text{SIN}}, y_{\text{SIN}}) &= p \cdot \sum_{j=m}^{t} \binom{t}{j} \cdot (x_{\text{SIN}})^j (1 - x_{\text{SIN}})^{t-j} + (1-p) \\ &= p \cdot (1 - \sum_{j=0}^{m-1} \binom{t}{j} \cdot (x_{\text{SIN}})^j (1 - x_{\text{SIN}})^{t-j}) + (1-p) \\ &= p \cdot (1 - \sum_{j=0}^{m-1} \frac{\Gamma(t+1)}{\Gamma(j+1)\Gamma(t-j+1)} (x_{\text{SIN}})^j (1 - x_{\text{SIN}})^{t-j}) + (1-p) \end{aligned}$$

where $\Gamma(t+1) = t!$ from the property of factorial functions. We need to show that $\frac{\partial g(x_{\text{SIN}}(t), y_{\text{SIN}}; t)}{\partial t} = 0$ for some t with $m \leq t \leq n$.

$$\frac{\partial g(x_{\text{SIN}}(t), y_{\text{SIN}}; t)}{\partial t} = -p \sum_{j=0}^{m-1} (\frac{\partial h(t,j)}{\partial t} (x_{\text{SIN}})^j (1 - x_{\text{SIN}})^{t-j}$$

$$+ h(t,j) \frac{d(x_{\text{SIN}})^k}{dt} (1 - x_{\text{SIN}})^{t-j} + h(t,j)(x_{\text{SIN}})^j \frac{d(1 - x_{\text{SIN}})^{t-j}}{dt})$$

where $h(t,j) = \frac{\Gamma(t+1)}{\Gamma(j+1)\Gamma(t-j+1)} > 0$. For $j = 0$, $h(t,j) = \frac{1}{\Gamma(1)}$ a constant. For $j \geq 1$, $h(t,j) = \frac{1}{j \cdot B(t-j+1,j)} = \frac{1}{j \cdot \int_0^1 \nu^{t-j}(1-\nu)^{j-1} d\nu}$ from the properties of Beta functions. Thus we have:

$$\begin{aligned} \frac{\partial B(t-j+1,j)}{\partial t} &= \int_0^1 \ln \nu \cdot \nu^{t-j} (1-\nu)^{j-1} d\nu < 0 \\ \Rightarrow \frac{\partial h(t,j)}{\partial t} &= -\frac{\partial B(t-j+1,j)}{\partial t} \cdot \frac{1}{j \cdot (B(t-j+1,j))^2} > 0 \end{aligned}$$

Thus the first term in the summation $\frac{\partial h(t,j)}{\partial t} \cdot (x_{\text{SIN}})^j (1 - x_{\text{SIN}})^{t-j} > 0$ where as the last two terms equal $h(t,j) \cdot (\frac{(1-p)q'(t)(t x_{\text{SIN}} - j)}{p q_t^2 (1 - x_{\text{SIN}})} + x_{\text{SIN}} \cdot \ln(1 - x_{\text{SIN}})) < 0$. Hence $\exists \hat{t}$ such that $\frac{d}{dt} g(x_{\text{SIN}}(\hat{t}), y_{\text{SIN}}; \hat{t}) = 0$. It follows that $q_m \leq q_{\hat{t}} \leq q_n$. Thus the sender's optimal mechanism with sincere receivers under private persuasion involves convincing $\hat{t}$ receivers, where $m \leq \hat{t} \leq n$, to vote informatively.

2.8.2.5 Proof of Proposition 5

Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$ denote the sender's choice of the conditional probabilities. Receivers' strategic voting behavior under private persuasion is characterized by lemma 2. The sender's problem is to maximize the probability of at least $m-k$ receivers' observing $s = b$ of all $r-k$ receivers who vote informatively:

$$\max_{x,y} g(x,y) = p \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot x^j (1-x)^{r-k-j} + (1-p) \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot y^j (1-y)^{r-k-j}$$

subject to $\gamma(m-k-1, r-k) \leq q_k \leq \gamma(m-k, r-k+1) \leq q_r \leq \gamma(m-k, r-k)$

$0 \leq x \leq 1, 0 \leq y \leq 1$

subject to receivers' posterior beliefs described in lemma 2. First notice that the $g(x, y)$ is continuous in (x, y). The constraint set, $\{(x, y) \in [0, 1]^2 | \gamma(m-k-1, r-k) \leq q_k \leq \gamma(m-k, r-k+1) \leq q_r \leq \gamma(m-k, r-k)\}$ is closed and bounded in $\mathbf{R}^2$; thus it is compact. By *Weierstrass's extreme value theorem*, the maximum of the objective function is attainable. We form the Lagrangian function:

$$L(x, y; \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_x, \lambda_y, \nu_x, \nu_y) = g(x, y) - \lambda_1 \cdot (\frac{(1-p)y^{m-k-1}(1-y)^{r-m+1}}{px^{m-k-1}(1-x)^{r-m+1} + (1-p)y^{m-k-1}(1-y)^{r-m+1}} - q_k)$$
$$-\lambda_2 \cdot (-\frac{(1-p)y^{m-k}(1-y)^{r-m+1}}{px^{m-k}(1-x)^{r-m+1} + (1-p)y^{m-k}(1-y)^{r-m+1}} + q_k)$$
$$-\lambda_3 \cdot (\frac{(1-p)y^{m-k}(1-y)^{r-m+1}}{px^{m-k}(1-x)^{r-m+1} + (1-p)y^{m-k}(1-y)^{r-m+1}} - q_r)$$
$$-\lambda_4 \cdot (-\frac{(1-p)y^{m-k}(1-y)^{r-m}}{px^{m-k}(1-x)^{r-m} + (1-p)y^{m-k}(1-y)^{r-m}} + q_r) - \lambda_x \cdot (x-1) - \lambda_y \cdot (y-1) + \nu_x \cdot x + \nu_y \cdot y$$

with $p \in [0, 1]; k, m, r \in \mathbf{N}, k < m < r; 1 - p \leq q_k < q_r \leq 1], q'(\cdot) > 0$. The Kuhn-Tucker condition for the above constrained maximization problem yields the following equations and inequalities.

$$p \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot x^{j-1}(1-x)^{r-k-j-1}(j-(r-k)x)$$
$$+\lambda_1 \cdot p \cdot q_k \cdot x^{m-k-2}(1-x)^{r-m}((m-k-1)-(r-k)x)$$
$$-\lambda_2 \cdot p \cdot q_k \cdot x^{m-k-1}(1-x)^{r-m}((m-k)-(r-k+1)x)$$
$$+\lambda_3 \cdot p \cdot q_r \cdot x^{m-k-1}(1-x)^{r-m}((m-k)-(r-k+1)x)$$
$$-\lambda_4 \cdot p \cdot q_r \cdot x^{m-k-1}(1-x)^{r-m-1}((m-k)-(r-k)x) - \lambda_x + \nu_x = 0$$

$$(1-p)\cdot\sum_{j=m-k}^{r-k}\binom{r-k}{j}\cdot y^{j-1}(1-y)^{r-k-j-1}(j-(r-k)y)$$

$$-\lambda_1\cdot(1-p)\cdot(1-q_k)\cdot y^{m-k-2}(1-y)^{r-m}((m-k-1)-(r-k)y)$$

$$+\lambda_2\cdot(1-p)\cdot(1-q_k)\cdot y^{m-k-1}(1-y)^{r-m}((m-k)-(r-k+1)y)$$

$$-\lambda_3\cdot(1-p)\cdot(1-q_r)\cdot y^{m-k-1}(1-y)^{r-m}((m-k)-(r-k+1)y)$$

$$+\lambda_4\cdot(1-p)\cdot(1-q_r)\cdot y^{m-k-1}(1-y)^{r-m-1}((m-k)-(r-k)y)-\lambda_y+\nu_y=0$$

$$\lambda_1(-p\cdot q_k\cdot x^{m-k-1}(1-x)^{r-m+1}+(1-p)\cdot(1-q_k)\cdot y^{m-k-1}(1-y)^{r-m+1}-0)=0$$

$$\lambda_2(p\cdot q_k\cdot x^{m-k}(1-x)^{r-m+1}-(1-p)\cdot(1-q_k)\cdot y^{m-k}(1-y)^{r-m+1}-0)=0$$

$$\lambda_3(-p\cdot q_r\cdot x^{m-k}(1-x)^{r-m+1}+(1-p)\cdot(1-q_r)\cdot y^{m-k}(1-y)^{r-m+1}-0)=0$$

$$\lambda_4(p\cdot q_r\cdot x^{m-k}(1-x)^{r-m}-(1-p)\cdot(1-q_r)\cdot y^{m-k}(1-y)^{r-m}-0)=0$$

$$\lambda_x\cdot(x-1)=0$$

$$\lambda_y\cdot(y-1)=0$$

$$\nu_x\cdot x=0$$

$$\nu_y\cdot y=0$$

$$\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_x,\lambda_y,\nu_x,\nu_y\geq 0$$

Although the solving the above conditions yields no close-form expression of x^*, y^*, it can be verified that $0 < y^* < 1$ and $0 < x^* < x_{\text{SIN}}$ where $x_{\text{SIN}} = \frac{(1-q_t)\cdot(1-p)}{q_t\cdot p}$ in proposition 4. Thus the result follows.

2.8.2.6 Proof of Proposition 6

Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$ denote the sender's choice of the conditional probabilities. Since $EU^S = u^S(A,t) + (u^S(B,t) - u^S(A,t)) \cdot \text{Prob}(v = B), \forall t \in T$, to compare the sender's expected utilities from the optimal mechanisms in both types of persuasion, we only need to compare the maximum values of $\text{Prob}(v = B)$ under each type of persuasion. Under public persuasion, sender's optimal mechanism satisfies $x^*_{\text{PUB}} = \frac{(1-q_m)\cdot(1-p)}{q_m\cdot p}, y^*_{\text{PUB}} = 1$. Then we have:

$$\text{Prob}_{\text{PUB}}(v=B) = p\cdot\frac{(1-q_m)\cdot(1-p)}{q_m\cdot p} + (1-p)\cdot 1$$

Under private persuasion, the probability of getting the final voting outcome as $v = B$ can be written as:

$$p\cdot\sum_{j=m-k}^{r-k}\binom{r-k}{j}\cdot(x^*_{\text{PRI}})^j(1-x^*_{\text{PRI}})^{r-k-j}+(1-p)\cdot\sum_{j=m-k}^{r-k}\binom{r-k}{j}\cdot(y^*_{\text{PRI}})^j(1-y^*_{\text{PRI}})^{r-k-j}$$

Since the following always holds:

$$\sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (y^*_{\text{PRI}})^j (1-y^*_{\text{PRI}})^{r-k-j} < 1$$

it remains to show that

$$\sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j} \leq \frac{(1-q_m)(1-p)}{q_m \cdot p}$$

which can be reduced to

$$(1 - \frac{(1-q_m)(1-p)}{q_m \cdot p}) \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j}$$
$$\leq \frac{(1-q_m)(1-p)}{q_m \cdot p} \cdot \sum_{j=0}^{m-k-1} \binom{r-k}{j} \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j}$$

And since $\frac{q_m+p-1}{q_m \cdot p} - \frac{(1-q_m)(1-p)}{q_m \cdot p}$ is a large negative number; the difference between the rest of the left-hand-side and the right-hand-side is small, we have the above inequality holds. Hence we have $\text{Prob}_{PRI}(v = B) < \text{Prob}_{PUB}(v = B)$ and the comparison between the sender's expected utilities follows.

2.8.2.7 Proof of Proposition 7

Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$ denote the sender's choice of the conditional probabilities. All receivers observe the same n public signal realizations; thus they all have the same posterior belief. Denote $\gamma(\ell, n)$ the posterior belief μ^β when there are ℓ signals $s = b$ out of all n signals. A receiver i with threshold doubt q_i will vote for B when there are more than ℓ favorable signals ($s = b$) out of all n signal realizations if and only if $\gamma(\ell - 1, n) < q_i < \gamma(\ell, n)$. The sender chooses the mechanism such that exact m receivers are convinced to vote for B when there are more than ℓ favorable signals, i.e. $\gamma(\ell, n) \leq q_m \leq \gamma(\ell + 1, n)$. Had the sender attempted to convince more than m receivers to vote informatively, e.g. $\gamma(\ell, n) \leq q_{m+j} \leq \gamma(\ell + 1, n), j = 1, 2, ..., x$ would decrease and the sender's objective function, $\text{Prob}(v = B)$, would decline, too. The sender's problem is:

$$\max_{x,y} g(x, y) = p \cdot \sum_{j=\ell}^{n} x^j (1-x)^{n-j} + (1-p) \cdot \sum_{j=\ell}^{n} y^j (1-y)^{n-j}$$

subject to $\gamma(\ell, n) \leq q_m \leq \gamma(\ell + 1, n)$

$0 \leq x \leq 1, 0 \leq y \leq 1$

First notice that the $g(x,y)$ is continuous in (x,y). The constraint set, $\{(x,y) \in [0,1]^2 | \gamma(\ell,n) \leq q_m \leq \gamma(\ell+1,n)\}$ is closed and bounded in $\mathbf{R}^2$; thus it is compact. By *Weierstrass's extreme value theorem*, the maximum of the objective function is attainable. We form the Lagrangian function:

$$
\begin{aligned}
&L(x,y;\lambda_1,\lambda_2,\lambda_x,\lambda_y,\nu_x,\nu_y) = g(x,y) \\
&-\lambda_1 \cdot (q_m \cdot p \cdot x^{\ell}(1-x)^{n-\ell} - (1-q_m)(1-p)y^{\ell}(1-y)^{n-\ell} - 0) \\
&-\lambda_2 \cdot (-q_m \cdot p \cdot x^{\ell-1}(1-x)^{n-\ell+1} + (1-q_m)(1-p)y^{\ell-1}(1-y)^{n-\ell+1} - 0) \\
&-\lambda_x \cdot (x-1) - \lambda_y \cdot (y-1) + \nu_x \cdot x + \nu_y \cdot y
\end{aligned}
$$

with $p \in [0,1]; l,n \in \mathbf{N}, 1-p \leq q_m \leq 1, q'(\cdot) > 0$. The Kuhn-Tucker condition for the above constrained maximization problem yields the following equations and inequalities.

$$
\begin{aligned}
&\frac{\partial g(x,y)}{\partial x} - \lambda_1 \cdot q_m \cdot p x^{\ell-1}(1-x)^{n-\ell-1}(\ell - nx) \\
&+\lambda_2 \cdot q_m \cdot p \cdot x^{\ell-2}(1-x)^{n-\ell}(\ell-1-nx) - \lambda_x + \nu_x = 0 \\
&\frac{\partial g(x,y)}{\partial y} + \lambda_1 \cdot (1-q_m) \cdot (1-p) y^{\ell-1}(1-y)^{n-\ell-1}(\ell - ny) \\
&-\lambda_2 \cdot (1-q_m) \cdot (1-p) \cdot y^{\ell-2}(1-y)^{n-\ell}(\ell-1-nx) - \lambda_y + \nu_y = 0 \\
&\lambda_1 \cdot (q_m \cdot p \cdot x^{\ell}(1-x)^{n-\ell} - (1-q_m)(1-p)y^{\ell}(1-y)^{n-\ell} - 0) = 0 \\
&\lambda_2 \cdot (-q_m \cdot p \cdot x^{\ell-1}(1-x)^{n-\ell+1} + (1-q_m)(1-p)y^{\ell-1}(1-y)^{n-\ell+1} - 0) = 0 \\
&\lambda_x \cdot (x-1) = 0 \\
&\lambda_y \cdot (y-1) = 0 \\
&\nu_x \cdot x = 0 \\
&\nu_y \cdot y = 0 \\
&\lambda_1, \lambda_2, \lambda_x, \lambda_y, \nu_x, \nu_y \geq 0
\end{aligned}
$$

solving the equations and inequalities yields $x^*_{MD} = \sqrt[n]{\frac{(1-q_m)(1-p)}{q_m p}}, y^*_{MD} = 1$ and $\hat{\ell} = n$. To see the latter, notice that $\forall \ell < n$, both $\gamma(\ell-1,n)$ and $\gamma(\ell,n) \to 0$ as $y \to 1$. Only when $\hat{\ell} = n$ we have $\gamma(\hat{\ell},n) = (1-p)/(p \cdot x^n + 1 - p) \geq q_m$ while $\gamma(\hat{\ell}-1,n) < q_m$.

Plug $x^*_{MD} = \sqrt[n]{\frac{(1-q_m)(1-p)}{q_m p}}, y^*_{MD} = 1$ into the sender's objective function, we have:

$$
\begin{aligned}
&\text{Prob}_{\text{MD}}(v = B) = p \cdot (x_{\text{MD}})^n + (1-p) \cdot (y_{\text{MD}})^n \\
&= p \cdot \frac{(1-q_m) \cdot (1-p)}{q_m \cdot p} + (1-p) = \text{Prob}_{\text{PUB}}(v = B)
\end{aligned}
$$

2.8.2.8 Proof of Proposition 8

Let $x = \pi(b|\alpha)$ and $y = \pi(b|\beta)$ denote the sender's choice of the conditional probabilities. We first show that for all $p \in [0, 1]$, public persuasion helps the sender to achieve $\Lambda(p)$ as illustrated in Figure 2.6. For any given (x, y), let $\underline{p} = 1/(\frac{1-x}{1-y} \cdot \frac{q_m}{1-q_m} + 1)$. We have the following three cases:

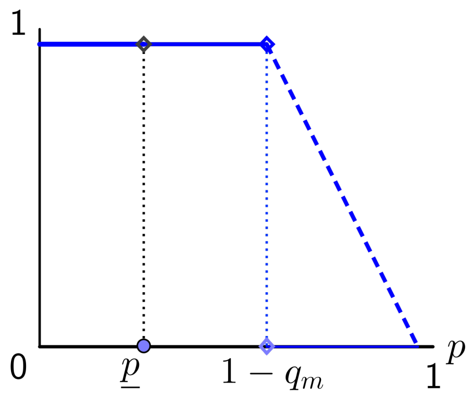


Figure 2.6: Attainability of the $\Lambda(p)$

Case 1: $p \in [0, \underline{p})$. Since $p < \underline{p} < 1 - q_m$, we have $q_m \leq \mu(b|\alpha) < \mu(b|\beta)$. This means at least m receivers vote for B irrespective of the signal realization. Thus the sender gets voting outcome B with probability 1.

Case 2: $p \in [\underline{p}, 1 - q_m)$. We have $\mu(b|\alpha) < q_m, \mu(b|\beta) > q_m$, which means more than m receivers vote for B on a signal realization b and less than m receivers vote for B on signal realization a. This yields the sender expected payoffs $(1 - p)/q_m$, which is less than $\Lambda(p), \forall p \in [\underline{p}, 1 - q_m)$. However, the sender can increase her payoffs by choosing $\pi(b|\alpha) = \pi(b|\beta)$. In this way the signal realizations become completely uninformative and the receivers make decision on the common prior. So the sender get the payoff as high as her expected payoff without persuasion, when the receivers cast votes based on their common prior.

Case 3: $p \in [1 - q_m, 1]$. Without persuasion, the sender's expected payoff is always 0. Under public persuasion, however, the sender chooses $\{\pi(\cdot|t)\}_{t \in T}$ as shown in Proposition 4. The sender gets

$$\text{Prob}_{\text{PUB}}(v = B) = (1 - p) \cdot \frac{1}{q_m}$$

which is the value of $\Lambda(p), \forall p \in [1 - q_m, 1]$.

From Proposition 7 we know that $\text{Prob}_{\text{MD}}(v = B) = (1 - p) \cdot \frac{1}{q_m}$. Now it remains to show that the same results holds if the sender is required to generate n correlated, not independent, public signal realizations. Denote the conditional joint distributions of the n signal realizations by η_j, ξ_j, where η_j represents the probability that the sender's persuasion mechanism generates j signal $s = b$ in state α; ξ_j represents the probability that the mechanism generates j signal $s = b$ in state β. Similar to the proof of Proposition 7, there exists cut off number of b signals ℓ such that

$$\gamma(\ell - 1, n) = \frac{(1-p)\sum_{j=\ell-1}^{n}\binom{n}{j}\xi_j}{p\sum_{j=\ell-1}^{n}\binom{n}{j}\eta_j + (1-p)\sum_{j=\ell-1}^{n}\binom{n}{j}\xi_j} < q_m$$
$$\gamma(\ell, n) = \frac{(1-p)\sum_{j=\ell}^{n}\binom{n}{j}\xi_j}{p\sum_{j=\ell}^{n}\binom{n}{j}\eta_j + (1-p)\sum_{j=\ell}^{n}\binom{n}{j}\xi_j} \geq q_m$$

and the sender's objective function is

$$p\sum_{j=\ell}^{n}\binom{n}{j}\eta_j + (1-p)\sum_{j=\ell}^{n}\binom{n}{j}\xi_j$$

Solving the maximization and we get $\text{Prob}_{\text{CRR}}(v = B) = (1-p)\cdot\frac{1}{q_m}$, the same expected payoff as $\text{Prob}_{\text{PUB}}(v = B)$.

2.8.2.9 Proof of Proposition 9

First we show that under public persuasion the probability that the receivers make decision errors is not necessarily lower than the one without any persuasion. Assumption 1 implies that without persuasion $\psi_N = 1 - p$ since all receivers vote for the default alternative A. Under public persuasion, the probability of decision errors equals the probability that the mechanism generates a signal $s = b$ in state α plus the probability that the mechanism fails to generate a signal $s = b$ in state β. Then $x^*_{\text{PUB}} = \frac{(1-q_m)\cdot(1-p)}{q_m\cdot p}, y^*_{\text{PUB}} = 1$ implies:

$$\psi_{\text{PUB}} = p\cdot x^*_{\text{PUB}} + (1-p)\cdot(1 - y^*_{\text{PUB}}) = p\cdot\frac{(1-q_m)\cdot(1-p)}{q_m\cdot p}$$

Hence when $q_m \leq p$ we have $\psi_{\text{PUB}} \geq \psi_N$, which indicates that the receivers are more likely to make mistakes under public persuasion than without any persuasion if the m^{th} receiver's threshold doubt is less than p.

Next, we show that $\psi_{\text{PRI}} < \psi_{\text{PUB}}$. Notice that under private persuasion with strategic-voting receivers, the probability of decision errors equals the probability that the mechanism generates at least $m - k$ favorable signals for $\{R_k, ...R_r\}$ who vote informatively in state α plus the probability that the mechanism generates up to $m - k - 1$ favorable signals in state β. From above we have $x^*_{\text{PRI}} < \frac{(1-q_{\hat{t}})\cdot(1-p)}{q_{\hat{t}}\cdot p}, y^*_{\text{PRI}} < 1$, where $\hat{t}$ is determined in the proof for proposition 4 . Then the probability of the receivers' decision

$$\psi_{\text{PRI}} = p \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j} + (1-p) \sum_{j=0}^{m-k-1} \binom{r-k}{j} \cdot (y^*_{\text{PRI}})^j (1-y^*_{\text{PRI}})^{r-k-j}$$

In fact, we have shown earlier that the first component is much smaller than $(1-q_m)(1-p)/q_m \cdot p$. Expand the binomial distribution, we have:

$$\begin{aligned} & p \cdot \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j} + (1-p)(1 - \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (y^*_{\text{PRI}})^j (1-y^*_{\text{PRI}})^{r-k-j}) \\ = \; & (1-p) + \sum_{j=m-k}^{r-k} \binom{r-k}{j} \cdot (p \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j} - (1-p)(y^*_{\text{PRI}})^j (1-y^*_{\text{PRI}})^{r-k-j}) \end{aligned}$$

Since j sums from $m-k$ to $r-k$, the receiver's optimal voting strategies described in Lemma 2 ensures that $p \cdot (x^*_{\text{PRI}})^j (1-x^*_{\text{PRI}})^{r-k-j} - (1-p)(y^*_{\text{PRI}})^j (1-y^*_{\text{PRI}})^{r-k-j} < 0$. So the second component of the above expression is negative. Therefore we have the whole expression less than $(1-p) \cdot (1-q_m)/q_m \cdot p$.

Last we show that $\psi_{\text{MD}} = \psi_{\text{PUB}}$. Plugging $x^*_{\text{MD}}, y^*_{\text{MD}}$ in proposition 7, we get the probability of the receivers' making decision mistakes is:

$$\begin{aligned} \psi_{MD} &= p \cdot (x_{MD})^n + (1-p) \cdot \sum_{j=0}^{\ell-1} (y_{MD})^j (1-y_{MD})^{n-j} \\ &= p \cdot \frac{(1-q_m) \cdot (1-p)}{q_m \cdot p} = \psi_{PUB} \end{aligned}$$

2.8.2.10 Sketch of Proof of Conjecture 1

Suppose under private persuasion with correlated signals, the correlation between observations becomes stronger, i.e. $\eta_n + \eta_0 \to 1$ and $\xi_n + \xi_0 \to 1$. A receiver observing $s = b$ ($s = a$) believes that other receivers also observes a b (a) signal realization. Thus the receiver's posterior belief upon observing $s = b$ ($s = a$) approaches:

$$\begin{aligned} \mu(\beta|a) &= \frac{\xi_0(1-p)}{\eta_0 p + \xi_0(1-p)} \\ \mu(\beta|b) &= \frac{\xi_n(1-p)}{\eta_n p + \xi_n(1-p)} \end{aligned}$$

As $\eta_n + \eta_0 \to 1$ and $\xi_n + \xi_0 \to 1$, the sender's problem reduces to choosing η_n and ξ_n to maximize

$$p \cdot \eta_n + (1-p) \cdot \xi_n$$

which gives the same solution as in Proposition 2.3.2.1. Thus the sender's optimal persuasion mechanism in this case yields her the same expected payoff as EU^S_{PUB}.

2.8.2.11 Proof of Lemma 3

We follow Duggan and Martinelli's (Duggan and Martinelli 2001) approach in this proof. Suppose the voting strategy $\sigma_i(s)$ described in Lemma 3 is not a best response for receiver i. Then there is another σ_i' that yields receiver i expected payoff as least as high as the expected payoff from σ_i. Specifically, consider two sets

$$V = \{s \in [0,1] | \Delta(\sigma_{-i}, s) < 0 \text{ and } \sigma_i'(s) < 1\}$$
$$W = \{s \in [0,1] | \Delta(\sigma_{-i}, s) > 0 \text{ and } \sigma_i'(s) > 0\}$$

The difference of expected payoffs between strategy σ_i' and σ_i is

$$\int_V (\sigma_i' - 1)(p(-q_i)P_{-i}(\text{piv}|\alpha)\pi(s|\alpha) - (1-p)(-(1-q_i))P_{-i}(\text{piv}|\beta)\pi(s|\beta))ds$$
$$+ \int_W \sigma_i'(p(-q_i)P_{-i}(\text{piv}|\alpha)\pi(s|\alpha) - (1-p)(-(1-q_i))P_{-i}(\text{piv}|\beta)\pi(s|\beta))ds$$

When $s \in V, \Delta(\sigma_{-i}, s) < 0$ and $\sigma_i' - 1 < 0$, so the first term of the integral is negative. When $s \in W, \Delta(\sigma_{-i}, s) > 0$ and $\sigma_i' > 0$, so the second term is also negative. Thus we get the contradiction, which indicates that the above specified voting strategy σ_i is a best response given σ_{-i}.

$\Delta(\sigma_{-i}, s)$ is continuous in s. Moreover, Assumption 2 implies that it is strictly decreasing in s. Thus it follows that there is a unique cutoff signal $\bar{s}_i$ for receiver i which solves $\Delta(\sigma_{-i}, s) = 0$ and $\inf\{s \in [0,1] | \Delta(\sigma_{-i}, s) \leq 0\}$ is a well defined. Therefore, σ_i being the best response implies that it is equivalent to the cutoff strategy $\hat{\sigma}_i$.

2.8.2.12 Proof of Lemma 4

Since $\Delta(\sigma_{-i}, s)$ is continuous in q_i, so does $\bar{s}_i$. Take any two threshold doubts $0 \leq q_j \leq q_i \leq 1$. We have

$$\Delta(\sigma_{-i}, s) = P_{-i}(\text{piv}|\alpha) \cdot \pi(s|\alpha) \cdot q_i \cdot p - P_{-i}(\text{piv}|\beta) \cdot \pi(s|\beta) \cdot (1-q_i) \cdot (1-p)$$
$$\Delta(\sigma_{-j}, s) = P_{-j}(\text{piv}|\alpha) \cdot \pi(s|\alpha) \cdot q_j \cdot p - P_{-j}(\text{piv}|\beta) \cdot \pi(s|\beta) \cdot (1-q_j) \cdot (1-p)$$

and the following cutoff strategies for i and j, respectively:

$$\sigma_i(s) = \begin{cases} 1, & \text{if } s \geq \bar{s}_i \\ 0, & \text{if } s < \bar{s}_i \end{cases}$$

where $\bar{s}_i = \inf\{s \in [0,1] | \Delta(\sigma_{-i}, s) \leq 0\}$.

$$\sigma_j(s) = \begin{cases} 1, & \text{if } s \geq \bar{s}_j \\ 0, & \text{if } s < \bar{s}_j \end{cases}$$

where $\bar{s_j} = \inf\{s \in [0,1] | \Delta(\sigma_{-j}, s) \leq 0\}$.

Suppose $\bar{s_j} = \bar{s_i}$. Since $\Delta(\sigma_{-j}, s)$ is continuous in s and q, there exists $\epsilon > 0$ such that at $s = \bar{s_i} - \epsilon, \Delta(\sigma_{-i}, s) > 0$, so receiver i will vote for B with probability 0. However, $s = \bar{s_i} - \epsilon = \bar{s_j} - \epsilon$ and $q_j < q_i$ implies $\Delta(\sigma_{-j}, s) \leq 0$, which indicates that receiver j is not willing to vote for A. The cutoff strategy $\sigma_j(s)$ is not a best response. A contradiction.

Now suppose $\bar{s_j} > \bar{s_i}$. Take $s \in [\bar{s_i}, \bar{s_j})$. Since $s \geq \bar{s_i}$, we have $\Delta(\sigma_{-i}, s) \leq 0$, so receiver i is willing to vote for B with probability 1. Moreover, $s < \bar{s_j}$ implies $\Delta(\sigma_{-j}, s) > 0$, which means receiver j will vote for A. Nonetheless, $q_j < q_i$ implies $\Delta(\sigma_{-j}, s) < \Delta(\sigma_{-i}, s) \leq 0$, which indicates receiver j is not willing to vote for A. A contradiction. Thus the first part of the lemma follows.

For the second part of the lemma, notice that for each receiver i, $\bar{s_i} = \inf\{s \in [0,1] | \Delta(\sigma_{-i}, s) \leq 0\}$. Expand the inequality $\Delta(\sigma_{-i}, s) \leq 0$, we have:

$$\begin{aligned}
& \sum_{|M|=m-1, i \notin M} (\prod_{j \in M} (\int_0^1 \sigma_i(s) \frac{\pi(s|\alpha)p(\alpha)}{\sum_{t' \in T} \pi(s|t')p(t')} ds) \prod_{j \notin M, j \neq i} (\int_0^1 (1-\sigma_i(s)) \frac{\pi(s|\alpha)p(\alpha)}{\sum_{t' \in T} \pi(s|t')p(t')} ds)) \pi(s|\alpha) \cdot q_i \cdot p \\
\leq & \sum_{|M|=m-1, i \notin M} (\prod_{j \in M} (\int_0^1 \sigma_i(s) \frac{\pi(s|\beta)p(\beta)}{\sum_{t' \in T} \pi(s|t')p(t')} ds) \prod_{j \notin M, j \neq i} (\int_0^1 (1-\sigma_i(s)) \frac{\pi(s|\beta)p(\beta)}{\sum_{t' \in T} \pi(s|t')p(t')} ds)) \pi(s|\beta) \cdot (1-q_i) \cdot (1-p)
\end{aligned}$$

Denote the ratio of the probability densities as $r = \frac{\pi(s|\alpha)}{\pi(s|\beta)}$. As r increases, the LHS of the inequality increases while the RHS decreases or remain unchanged. In both cases, for some $s \in [\bar{s_i}, 1]$, $\Delta(\sigma_{-i}, s) > 0$. Continuity implies that such $s's$ either satisfy $s \in [\bar{s_i}, \bar{s_i} + \eta]$ or satisfy $s \in [1 - \vartheta, 1]$. Suppose for $s \in [1 - \vartheta, 1]$ we have $\Delta(\sigma_{-i}, s) > 0$. But this contradicts the definition of a *cutoff strategy*. Thus for $s \in [\bar{s_i}, \bar{s_i} + \eta]$ the value of $\Delta(\sigma_{-i}, s)$ are positive. Hence the value of $\inf\{s \in [0,1] | \Delta(\sigma_{-i}, s) \leq 0\}$ increases.

2.8.2.13 Proof of Corollary 4

Under public persuasion, the sender chooses a set of conditional distributions $(\pi(s|\alpha), \pi(s|\beta))$ over $S = [0,1]$ to maximize the probability that $s \geq \bar{s_m}$:

$$\max_{(\pi(s|\alpha), \pi(s|\beta))} p \cdot P(s \geq \bar{s_m}|\alpha) + (1-p) \cdot P(s \geq \bar{s_m}|\beta)$$

The receivers' posterior belief after observing a signal realization $s \geq \bar{s_m}$ is:

$$\begin{aligned}
\mu(\beta | s \geq \bar{s_m}) &= \frac{(1-p) \cdot P(s \geq \bar{s_m}|\beta)}{p \cdot P(s \geq \bar{s_m}|\alpha) + (1-p) \cdot P(s \geq \bar{s_m}|\beta)} \\
&= \frac{(1-p) \cdot \int_{\bar{s_m}}^1 \pi(s|\beta) ds}{p \cdot \int_{\bar{s_m}}^1 \pi(s|\alpha) ds + (1-p) \cdot \int_{\bar{s_m}}^1 \pi(s|\beta) ds}
\end{aligned}$$

Having at least m receivers voting for B upon observing $s \geq \bar{s_m}$ requires $\mu(\beta|s \geq \bar{s_m}) \geq q_m$. Thus the sender's expected utility is maximized when the conditional distributions $(\pi(s|\alpha), \pi(s|\beta))$ satisfy:

$$(1-p) \cdot (1-q_m) \cdot \int_{\bar{s_m}}^{1} \pi(s|\beta)ds \geq p \cdot q_m \cdot \int_{\bar{s_m}}^{1} \pi(s|\alpha)ds$$

and $EU^S = p \cdot \frac{(1-p)(1-q_m)}{p \cdot q_m} + (1-p) = \Lambda(p), \forall p \in [0,1]$.

3.0 AN EXPERIMENTAL INVESTIGATION ON BELIEF AND HIGHER-ORDER BELIEF IN THE CENTIPEDE GAMES

3.1 INTRODUCTION

This paper studies rationality, belief of rationality, and higher-order belief of rationality in the centipede game experiment. Actual play in centipede experiments seldom ends as backward induction predicts. Existing literature attributes the departure from backward induction (BI thereafter) prediction either to players' lack of rationality, or to players' inconsistent beliefs and higher-order beliefs of others' rationality. In this paper, we evaluate these arguments in a more direct fashion. We elicit the first mover's belief about the second mover's strategy as well as the second mover's initial and conditional beliefs about the first mover's strategy and 1st-order belief. The measured beliefs help us to infer the conditional probability systems (CPS thereafter) of both players. The inferred CPS's and players' actual strategy choices identify *why* they fail to reach the BI outcomes.

The first strand of the existing experimental literature focuses on players' lack of rationality. It presumes presence of behavioral types who fail to or do not maximize monetary payoffs[1]. For example, McKelvey and Palfrey (1992) assume that ex-ante a player chooses to not play along the BI path with probability p. But assuming irrationality before a game starts is restrictive; people could be right but think others are wrong. In this paper, the inferred CPS and players' strategies allow us to directly examine players' rationality. We define rationality as a player's strategy best responding to the measured belief. We find, in all three treatments, the frequency of either player's being rational is significantly lower than 100 percent. But in the Constant-Sum treatment, which excludes the efficiency property as well as any possibility of mutual benefits, the frequency of the first-mover being rational is significantly higher than those in the other two treatments.

The other strand of literature attributes the experimental anomalies to lack of common knowl-

[1] See McKelvey and Palfrey (1992), Fey et.al.(1996), Zauner (1999), Kawagoe and Takizawa (2012).

edge of rationality. Two field centipede experiments (Palacios-Huerta and Volij (2009) and Levitt et.al (2009)) are in this fashion. Both use professional chess players as experimental subjects; the authors assume there is always rationality and common knowledge of rationality among chess players. The authors' approach is based on Aumann's (1995) claim[2] "if common knowledge of rationality holds then the backward induction outcome results." Nevertheless, the notion of "common knowledge" is not empirically verifiable; one can never ensure the existence of "common knowledge" among chess-players or the non-existence of it among ordinary laboratory subjects. This suggests that the knowledge-based approach may have limited explanatory power for the anomalies in the centipede experiments. Thus in this paper we follow an alternative approach, the belief-based epistemic game theory[3] to address the notion of common *belief* of rationality. The measured beliefs, high-order beliefs, and players' actual strategy choices help us to identify whether *rationality and common initial belief of rationality* and/or *rationality and common strong belief of rationality* hold.

We find, in fact, that common *initial* belief of rationality does not always exist in the laboratory. In all three treatments, the frequency of players' believing opponents' rationality is significantly less than 100 percent. Nevertheless, in the Constant-Sum treatment this frequency is significantly higher than that in the other two treatments; whereas the frequency in the Baseline Centipede treatment does not differ significantly form that in the No-Mutual-Benefit treatment, a treatment that excludes the mutually beneficial outcome but not the efficiency property from the Baseline game. Moreover, in all treatments the average frequency of the second mover's initially believing the first mover's rationality and 2nd-order rationality is significantly less than 100 percent. This frequency in the Constant-Sum treatment is significantly higher than those in the other two treatments. Also it gradually increases towards 100 percent as subjects gain experience in later rounds of the experiment; whereas in the other two treatments there is no such increasing pattern as more rounds are played.

Furthermore, we find that common *strong* belief of rationality is seldom observed in the laboratory, especially for the second-mover. In all three treatments, the the average frequency of the second mover's strongly believing the first-mover's rationality and 2nd-order rationality is significantly less than 100 percent. And this frequency in the Constant-Sum treatment does not significantly differ from those in the other two treatments. Notice that the second-movers are in-

[2]For more knowledge-based theoretical discussion on the "backward induction paradox," see Bicchieri (1988, 1989), Pettit and Sugden (1989), Reny (1988, 1992), Bonanno (1991), Aumann (1995, 1996, 1998), Binmore (1996, 1997).

[3]See Aumann and Brandenburger (1995), Battigalli (1997), Battigalli and Siniscalchi (1999, 2002), Ben-Porath (1997), Brandenburger (2007).

formed that the first-mover has chosen a non-BI strategy for the first stage before being asked to state their conditional beliefs. Thus our result indicates that once the second-movers observe the first-movers' deviating from the BI path, the former can hardly believe that the latter's rationality AND higher-order belief of rationality.

Last but not least, let us close this section by emphasizing the difference between this belief-based approach and the *level-k* model. The level-k analysis assumes the presence of behavior types *before* the game starts: there always exists a level-zero who is the least sophisticated; each player believes her opponent to be less sophisticated than herself and respond to those types optimally. Nevertheless, our approach does not impose any presumptions on players' beliefs and behavior: we elicit the true patterns of them. We do not assume *ex-ante* that players best respond to others' types; nor do we restrict players' beliefs about their opponents' degree of sophistication. Strategies and reported beliefs from our experiment can be used to examine the level-k model, but not vice versa.

The remainder of the paper is organized as follows. Section 3.2 formally defines players' beliefs, rationality, and beliefs of rationality in the centipede game. Section 3.3 presents the experimental design in detail, with Section 3.3.1 introducing experimental treatments and testing hypothesis and Section 3.3.2 introducing the procedure and belief elicitation method in the laboratory. Section 3.4 presents the experimental findings on players' strategies, players' beliefs about opponents' strategies, rationality, and higher-order beliefs of rationality. Section 3.5 reviews related theoretical literature on backward induction and epistemic game theory and previous experimental studies on the centipede games. Section 3.6 concludes.

3.2 DEFINING BELIEF, RATIONALITY, AND BELIEF OF RATIONALITY

We follow Brandenburger's (Brandenburger 2007) notation of players' belief types and epistemic states throughout this section. Denote the two-player (Ann and Bob) finite centipede game $\langle S^a, S^b, \Pi^a, \Pi^b \rangle$ where S^i and Π^i represent player i's set of pure strategies and set of payoffs, respectively.

Definition 8. *We call the structure* $\langle S^a, S^b; T^a, T^b; \lambda^a(\cdot), \lambda^b(\cdot) \rangle$ *a type structure for the players of a two-person finite game where* T^a *and* T^b *are compact metrizable space, and each* $\lambda^i : T^i \to \Delta(S^{-i} \times T^{-i}), i = a, b$ *is continuous. An element* $t^i \in T^i$ *is called a* ***type*** *for player* $i, (i = a, b)$.

An elements $(s^a, s^b, t^a, t^b) \in S \times T$ (where $S = S^a \times S^b$ and $T = T^a \times T^b$) is called a ***state.***

We first define *rationality* using the type-state language:

Definition 9. *A strategy-type pair of player $i, (i = a, b)$, (s^i, t^i) is* ***rational*** *if s^i maximizes player i's expected payoff under the measure $\lambda^i(t^i)$'s marginal on S^{-i}.*

Next, we define a player's believing an event as:

Definition 10. *Player i's type t^i believes an event $E \subseteq S^{-i} \times T^{-i}$ if $\lambda^i(t^i)(E) = 1, i = a, b$. Denote*

$$B^i(E) = \{t^i \in T^i : t^i \text{ believes } E\}, i = a, b$$

the set of player i's types that believe the event E.

For each player i, denote R_1^i the set of all rational strategy-type pairs (s^i, t^i). Thus R_1^{-i} stands for the set of all rational strategy-type pairs of opponent $-i$, i.e.

$$R_1^{-i} = \{(s^{-i}, t^{-i}) \in S^{-i} \times T^{-i} : (s^{-i}, t^{-i}) \text{ is rational.}\}, i = a, b$$

Now we can define a player's believing in his or her opponent's rationality as player i believes an event $E = R_1^{-i}$:

Definition 11. *Player i's type t^i* ***believes his or her opponent's rationality*** *$R_1^{-i} \subseteq S^{-i} \times T^{-i}$ if $\lambda^i(t^i)(R_1^{-i}) = 1$. Denote*

$$B^i(R_1^{-i}) = \{t^i \in T^i : t^i \text{ believes } R_1^{-i}\}, i = a, b$$

the set of player i's types that believe opponent $-i$'s rationality.

Then for all $m \in \mathbf{N}$ and $m > 1$, we can define R_m^i inductively by

$$R_m^i = R_{m-1}^i \cap (S^i \times B^i(R_{m-1}^{-i})), i = a, b$$

And write $R_m = R_m^a \times R_m^b$. Then players' higher order beliefs of rationality is defined in the following way:

Definition 12. *If a state $(s^a, s^b, t^a, t^b) \in R_{m+1}$, we say that there is* ***rationality and mth-order belief of rationality*** *(RmBR) at this state.*
If a state $(s^a, s^b, t^a, t^b) \in \cap_{m=1}^{\infty} R_m$, we say that there is ***rationality and common belief of rationality*** *(RCBR) at this state.*

For a perfect-information sequential move game such as the centipede game, in case the game situation involves the players not playing the backward-induction path (BI path thereafter), we also need to describe players' beliefs of probability-0 events. We use the tool of *conditional probability systems* (CPS thereafter) introduced by Renyi's (Rényi 1955). It consists of a family of conditional events and one probability measure for each of these events. For the centipede game under analysis, we define player i ***initially believes event E*** if i's CPS assigns probability 1 to event E at the *root* of the perfect-information game tree. We denote the set of player i's types that initially believe event E as $\mathrm{IB}^i(E), i = a, b$. We also define player i ***strongly believes event E*** if for any information set H that is reached, i.e. $E \cap (H \times T^{-i}) \neq \emptyset$, i's CPS assigns probability 1 to event E. We denote the set of a player's types who strongly believe event E as $\mathrm{SB}^i(E), i = a, b$.

3.3 EXPERIMENTAL DESIGN

We experimentally investigate players' rationality, beliefs and higher-order beliefs about opponents' rationality. Section 3.3.1 describes the treatments and testing hypothesis. Section 3.3.2 details the laboratory environments, belief elicitation, and other experimental procedures.

3.3.1 Treatments and Hypothesis

Our experiment consists of three treatments, each of which is a three-legged centipede game. The first treatment, "Baseline Centipede Game" is shown in Figure 3.1. Both the Nash equilibrium outcome and Subgame Perfect equilibrium outcome involve player A choosing OUT at the first stage and the two players ending up with a $20 - 10$ split of payoffs.

Here we emphasize three feature of this baseline game: first, the same as the "backward induction paradox" discussed in the theoretical literature, the sum of the players' payoffs grows at each stage. Had the players not played the BI path, the outcome would yield the players a larger *sum* of payoffs. We call this an *efficient* outcome. Second, had player A played IN at both of his/her decision stages and had player B played IN at his/her decision stage, the game would end up with a mutually beneficial $25 - 45$ payoff split. This is because 25 is greater than 20, the payoff that player A gets if he/she plays OUT at the first stage; and 45 is greater than 40, the payoff that player B gets if he/she plays OUT at the second stage. Third, allowing for probabilistic belief, it

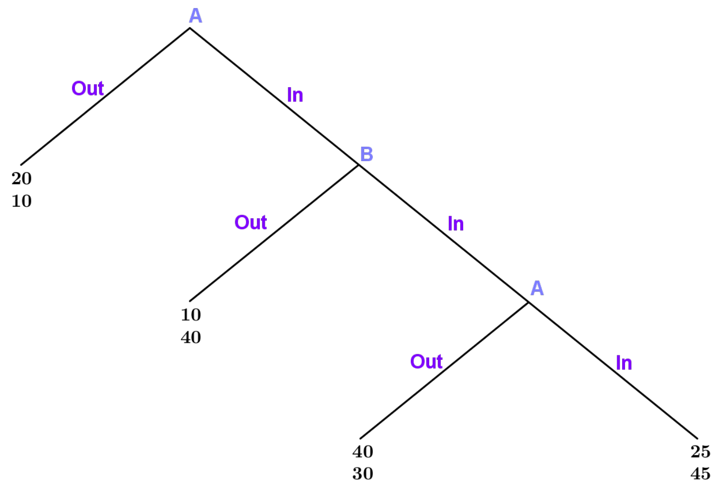


Figure 3.1: Baseline Centipede Game

is easy to calculate that if player A expects player B to play IN with a probability greater than $\frac{1}{3}$, his/her best response is to play IN for the first stage and OUT for the third stage. As for player B, if he/she expects player A to play IN at the third stage with a probability greater than $\frac{2}{3}$, his/her best response is to play IN for the second stage.

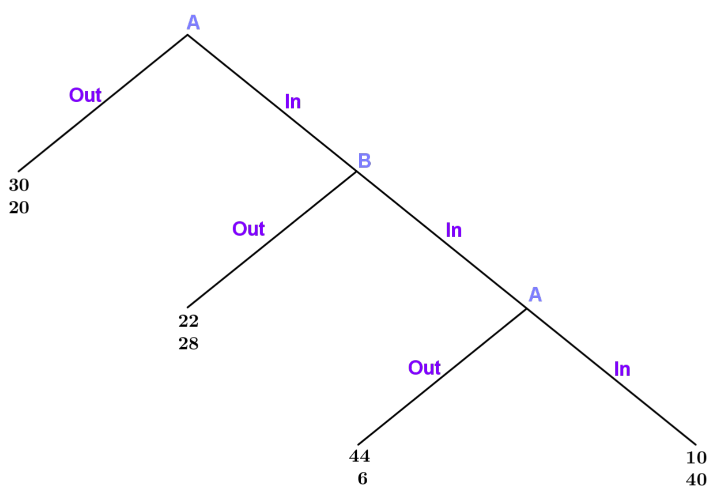


Figure 3.2: Constant-Sum Centipede Game

Our second treatment, the "Constant-Sum Centipede Game," is shown in Figure 3.2. The sum of the players' payoffs at all stages is a constant. This version of the centipede game eliminates the *efficient concern* presented in the Baseline Centipede treatment; and similar experimental treatments *without* examining players' beliefs has been conducted by Fey et.al.(1996), Levitt et.al.(2009). We choose the constant-sum payoff to be 50 because this is the actual average sum of the payoffs subjects earned in the laboratory in the Baseline Centipede treatment. And we choose the split of the players' payoffs at each stage such that the cutoff probabilistic belief for each player is the same as that in the Baseline Centipede treatment. Namely, if player A expects player B to play IN with $p \geq \frac{1}{3}$, his/her best response is to play IN-OUT; if player B expects player A to play IN at the third stage with $q \geq \frac{2}{3}$, his/her best response is to play IN for the second stage.

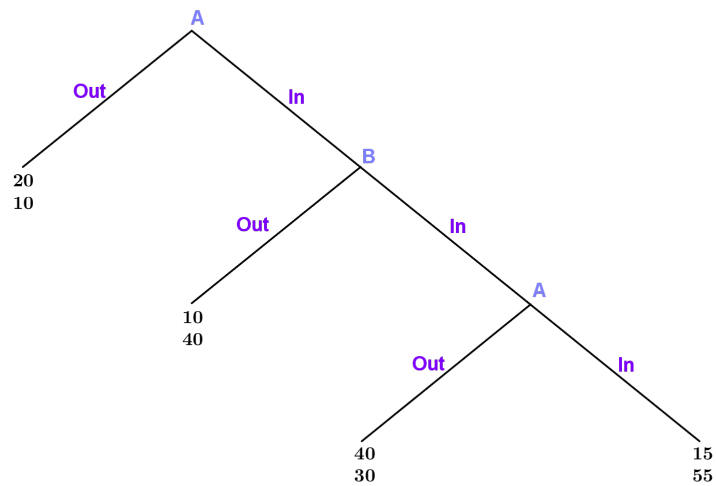


Figure 3.3: No-Mutual-Benefit Centipede Game

Notice that the Constant-Sum Centipede excludes both the *efficiency* property and *mutual-beneficial payoff* property from the Baseline Centipede. To further investigate the key driving force underlying the observed differences between the first two treatments, we conduct a "No-Mutual-Benefit Centipede" treatment as shown in Figure 3.3. The sum of the players' payoffs at each stage remains the same as in the Baseline Centipede; the only change is the $15 - 55$ split of payoffs had both players played the IN-IN-IN path. In this case player A's cutoff probabilistic belief for playing IN-OUT remains as $\frac{1}{3}$, whereas player B's cutoff probabilistic belief for playing IN changes to $\frac{2}{5}$.

Table 3.1 below summarizes the treatments and number of sessions, subjects, and matches of games for each treatment.

Next, we list the testing hypotheses as comparisons between the treatments, and as comparisons with theoretical predictions. Our first set of hypotheses are on the players' strategy choices. In all three treatments, the Nash equilibrium outcome involves the game ending as player A plays OUT at the first stage; and the Subgame Perfect equilibrium prescribes both player's choosing OUT at

Table 3.1: Experimental Treatments

Treatments	# of Sessions	# of Subjects	Total # of Games
Baseline Centipede	5	60	450
Constant-Sum	5	60	450
No-Mutual-Benefit	3	36	270

each's decision stage(s). Therefore we have the following hypotheses:

Hypothesis 1. *In all three treatments, the frequency of player A's choosing IN at the first stage does not differ significantly from* 0*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 2. *In all three treatments, the frequency of player B's choosing IN at the second stage does not differ significantly from* 0*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

The second set of hypotheses describes players' rationality. As defined in Section 3.2, a player is rational if his or her strategy choice maximizes the expected payoffs *given* his or her belief. A fully rational player best responds to both the initial belief and conditional belief with probability 1. In Appendix 3.7.1 we demonstrate the following hypotheses by proving five observations.

Hypothesis 3. *If player A is* ***rational****, then in all three treatments, the frequency of A's strategy best responding to A's belief does not significantly differ from* 1*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 4. *If player B is* ***rational****, then in all three treatments, the frequency of B's strategy best responding to B's belief does not significantly differ from* 1*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

We then move to players' belief about the opponents' rationality and common belief of rationality. As defined in Section 3.2, a player believes one's opponent being rational if he/she assigns probability 1 to all the states (s^{-i}, t^{-i}) in which opponent $-i$'s strategy best responds to the belief in that state. For player A, this probability is the one he/she states before the game starts. For player B, the probability he/she assigns to A's strategy-belief pair at the *root* of the game tree is the *initial* belief, while the probability he/she assigns once called upon to move at the second stage (if observed) is the *conditional* belief in the definition of "strong belief" in Section 3.2. Thus we have the following hypotheses. In Appendix 3.7.1 we prove them by demonstrating five observations.

Hypothesis 5. *If* ***rationality and common strong belief of rationality*** *holds, then in all three treatments, the frequency of A's believing B's choosing IN at the second stage does not significantly differ from* 0*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 5 comes from the fact that *rationality common strong belief of rationality (RCSBR)* implies that player A's believing in B's rationality, believing in B's (initially and conditionally) believing A's rationality, and so on. Thus as shown in Section 3.7.1, there is no state that involves player A's believing player B's choosing IN satisfying RCSBR.

Hypothesis 6. *If* ***common belief of rationality*** *holds, then in all three treatments, the frequency of B's believing A's rationality does not significantly differ from* 1*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 7. *If* ***rationality and common initial belief of rationality*** *holds, then in all three treatments, the frequency of B's* ***initially*** *believing in A's rationality* and 2*nd-Order rationality does not differ significantly from* 1*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 8. *If* ***rationality and common strong belief of rationality*** *holds, then in all three treatments, the frequency of B's* ***conditionally*** *believing in A's rationality* and 2*nd-Order rationality does not differ significantly from* 1*. Specifically, this frequency in the Baseline Centipede treatment does not differ significantly from that in the Constant-Sum treatment.*

Hypothesis 6 comes from the fact that *common belief of rationality* implies player B's (both initially and conditionally) assigning probability 1 to the event of A's rationality. Hypothesis 7 comes from the fact that *rationality and common* ***initial*** *belief of rationality* implies player B's assigning probability 1 to A's rationality AND A's believing B's rationality at the *root* of the game tree. Hypothesis 8 is from the fact that **rationality and common *strong* belief of rationality** implies player B's still assigning probability 1 to A's rationality and 2nd-order rationality even after observing A has chosen IN for the first stage.

3.3.2 Design and Procedure

All sessions were conducted at the Pittsburgh Experimental Economics Lab (PEEL) in Spring 2013. A total of 156 subjects are recruited from the undergraduate population of the University of Pittsburgh who have no prior experience in our experiment. The experiment adopts between-subject design, with 5 sessions for the Baseline Centipede treatment, 5 sessions for the Constant-Sum treatment, and 3 sessions for the No-Mutual-Benefit treatment. The experiment is programmed and conducted with z-Tree (Fischbacher (2007)).

Upon arrival at the lab, we seat the subjects at separate computer terminals. After we have enough subjects to start the session[4], we hand out instructions and then read the instruction aloud. A quiz which tests the subjects' understanding of the instruction follows. We pass the quiz's answer key after the subjects finish it, explaining in private to whomever have questions.

In each session, 12 subjects participate in 15 rounds of one variation of the centipede game. Half of the subjects are randomly assigned the role of Member A and the other half the role of Member B. The role remain fixed throughout the experiment. In each round, one Member A is paired with one Member B to form a group of two. The two members in a group would then play the centipede game in that treatment. Subjects are randomly rematched with another member of the opposite role after each round.

For the aim of collecting enough data, we first use strategy method to elicit the subjects' strategy choice[5]. We ask the subjects to specify their choice at *each* decision stage had it been reached. Then the subjects' choice(s) are carried out automatically by the programme and one would not have a chance to revise it if one's decision stage is reached.

After the subjects finish the choice task, they enter a "forecast task" phase which is to elicit their beliefs about opponent's choices. Member A is asked to choose from one of the two statements which he/she thinks more likely[6]: "Member B has chosen IN" or "Member B has chosen OUT." Member A's predictions are incentivized by a linear rule: 5 points if correct, 0 if incorrect. Member B is informed that his/her partner A has made a selection of choices for stage 1 and 3, AND have

[4]Each session has 12 subjects. We over-recruit as many as 16 subjects each time. By arrival time, from the 13th subject on, we pay them a $5.00 show-up fee and ask them to leave.

[5]Another advantage of the strategy method is to exclude subjects' incentives to signal, hedge, or bluff their opponent. Had we not adopted this method, in the baseline treatment we would have observed an even higher frequency of player A's choosing IN for the first stage. Player A might find it optimal to "bluff opponent" if player B is tempted by the *efficient* and *mutually beneficial* payoff split in the Baseline Centipede treatment AND B would not strongly believe A's rationality after observing A's choosing IN for the first stage.

[6]Since in all treatments player A's cutoff probabilistic belief is $\frac{1}{3}$, which is smaller than 50 percent, the point prediction Member A is making here is without loss of generality.

chosen a statement about Member B's choice. Then Member B's are asked to enter six numbers as the percent chance into a table, each cell of which represents a choice-forecast pair that Member A has chosen. For example, as shown in the table below, the upper-left cell represents the event that Member A has chosen OUT for the 1st stage and "Statement I."

Table 3.2: Member B's Estimation Task

Statement I	■		
Statement O			
	1st Stage *Out*, 3rd *In* or *Out*	1st Stage *In*, 3rd Stage *Out*	1st Stage *In*, 3rd Stage *In*

If B's decision stage is reached (which means his/her partner Member A has chosen IN), he/she will be asked to make a second forecast about the percent chance for each possible outcome of A's choices. Member B's predictions are incentivized by the quadratic rule:

$$5 - 2.5 \times [(1 - \beta_{ij})^2 + \sum_{kl \neq ij} \beta_{kl}^2]$$

where β_{kl} stands for Member B's stated percent chance in row k column l of the table, and i, j represents that row i column j is the outcome from Member A's choices[7].

At the end of the experiment, one round is randomly selected to count for payment. A subject's earning in each round is the sum of the points he/she earn from the choice task and the forecast task(s). The exchange rate between points and US dollars is 2.5 : 1. A subject receivers his/her earning in that selected round plus the $5.00 show-up fee.

3.4 EXPERIMENTAL FINDINGS

3.4.1 Players' Strategy Choices

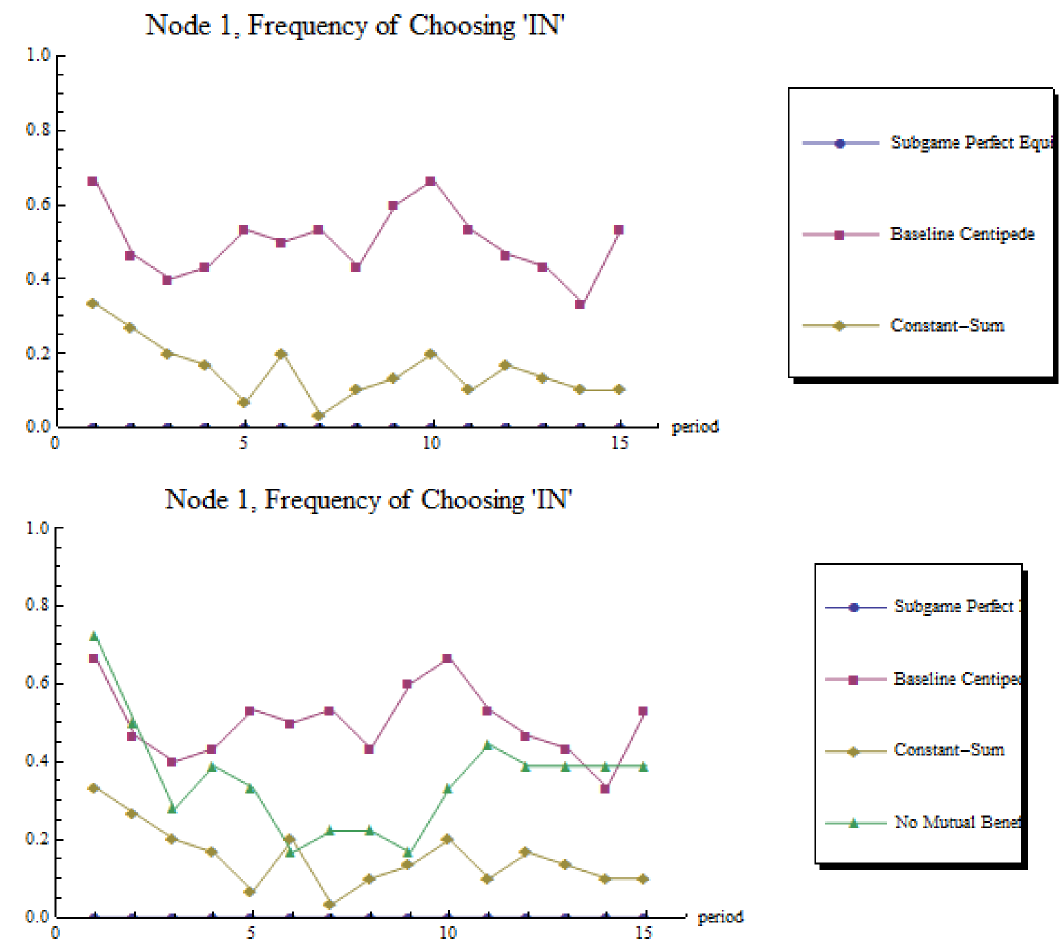
Our first set of results compares the frequency of players' strategy choices with that predicted by the Subgame Perfect equilibrium. We first state the result addressing Hypothesis 1, then move

[7]Palfrey and Wang (Palfrey and Wang 2009) and Wang (Wang 2011) have discussed eliciting subjects' beliefs using *proper* scoring rules. This is the major reason we adopt a quadratic scoring rule. We are also aware of the risk-neutrality assumption behind the quadratic rule and the possibility to use an alternative belief elicitation method proposed by Karni (Karni 2009). But concerning the complexity of explaining Karni's method to the subjects, we adopt the quadratic rule which is simpler in explanation.

to the result addressing Hypothesis 2.

Result 1. *(1) In all three treatments, the average frequency of A's choosing IN at the first stage is significantly higher than* 0. *(2) The average frequency of A's choosing IN at the first stage in the Constant-Sum treatment is significantly lower than that in the Baseline Centipede treatment.*

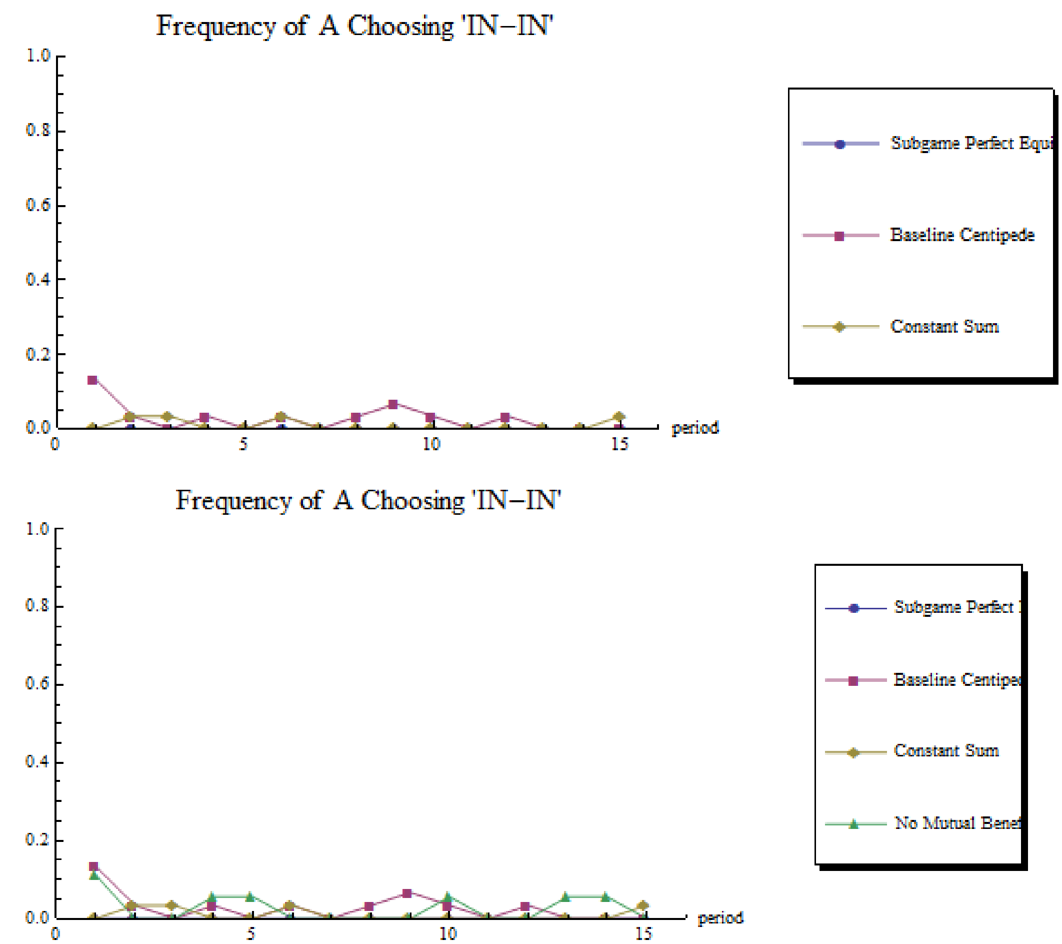
Result 1 addresses Hypothesis 1. Figure **??** depicts the treatment-average frequency of player A's choosing IN at the first stage across all periods. This frequency in the Constant-Sum treatment is significantly lower than that in the Baseline treatment; but both of them are significantly higher than 0, the Subgame Perfect equilibrium prediction.



Note: Figure on top compares the average frequency of A's choosing IN predicted by the Subgame Perfect equilibrium (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.4: Average Frequency of A's Strategy Choice, Across Periods

It is natural to ask what these player A's would play given that they had deviated from the equilibrium path. Namely, what is the frequency of choosing strategy IN-OUT versus the frequency of choosing IN-IN? Figure ?? depicts the treatment-average frequency of player A's choosing strategy IN-IN. It is interesting to note that the frequency in all three treatment is not significantly different from 0; and there is no significant difference across treatments. Notably, this is true even for the Baseline Centipede treatment. In other words, despite the *efficiency* property and *mutual benefit* property of the Baseline Centipede, actual plays seldom end up with the "mutually beneficial" $25 - 45$ payoff split. Conditional on the third node being reached, almost all player A's optimally choose OUT for the third stage.

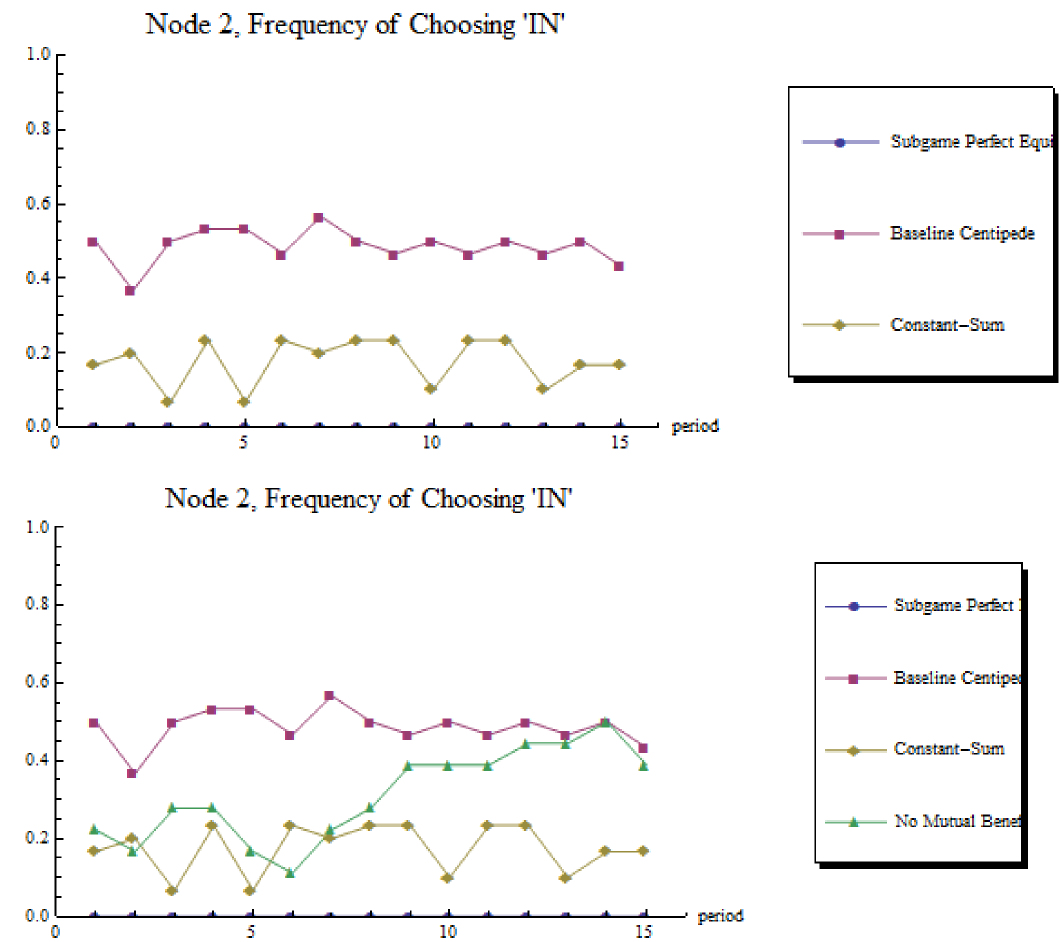


Note: Figure on top compares the average frequency of A's choosing IN at the first stage and IN at the third stage predicted by the Subgame Perfect equilibrium (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.5: Average Frequency of A's Choosing IN-IN at Both Decision Stages, Across Periods

Result 2. *(1) In all three treatments, the average frequency of B's choosing IN at the second stage* *is significantly higher than* 0*. (2) The average frequency of B's choosing IN at the second stage in the Constant-Sum treatment is significantly lower than that in the Baseline Centipede treatment.*

Result 2 addresses Hypothesis 2. Figure ?? depicts the treatment-average frequency of player B's choosing IN at the second stage across all periods. This frequency in the Constant-Sum treatment is significantly lower than that in the Baseline treatment; but both of them are significantly higher than 0, the Subgame Perfect equilibrium prediction.



Note: Figure on top compares the average frequency of B's choosing IN predicted by the Subgame Perfect equilibrium (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.6: Average Frequency of B's Strategy Choice, Across Periods

3.4.2 Rationality

In this section we present comparison results on players' rationality across treatments. We first examine the frequency of A's best responding to his/her stated belief. Notice that there are two data points from A's strategy-belief choices that can be identified as "rational." Either player A chooses strategy IN-OUT and believes that B has chosen IN, or chooses OUT for the first stage and believes that B has chosen OUT. We sum up the frequencies from the two cases as we calculate the overall frequency of A's being rational.

Result 3. *(1) In all three treatments, the average frequency of player A's being rational is significantly lower than* 1*. (2) The average frequency of player A's being rational in the Constant-Sum treatment is significantly higher than that in the Baseline Centipede treatment.*

Result 3 addresses Hypothesis 3. Figure ?? depicts the treatment-average frequency of player A's being rational across all periods. This frequency in the Constant-Sum treatment is significantly higher than that in the Baseline treatment; but both of them are significantly lower than 1 as required by the notion of *rationality*.

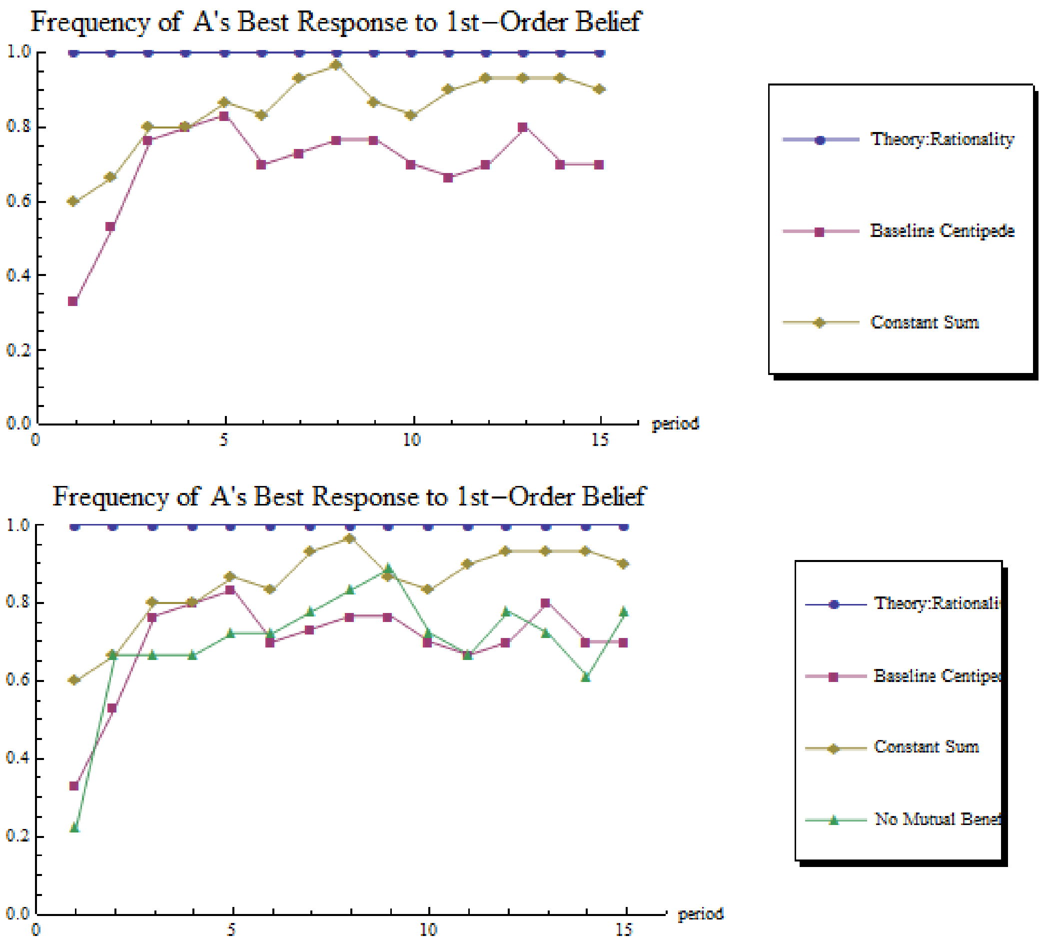
We then investigate the frequency of B's best responding to his/her stated belief. From B's stated belief, if the probability he/she assigns to A's choosing strategy IN-IN is greater than his/her cutoff probabilistic belief, it is rational for B to choose IN for the second stage; otherwise, it is rational to choose OUT for the second stage. We sum up the frequencies from the two cases as we calculate the overall frequency of B's being rational.

Result 4. *(1) In all three treatments, the average frequency of player B's being rational is significantly lower than* 1*. (2) The average frequency of player B's being rational in the Constant-Sum treatment is not significantly different from that in the Baseline Centipede treatment.*

Result 4 addresses Hypothesis 4. Figure ?? depicts the treatment-average frequency of player B's being rational across all periods. This frequency in the Constant-Sum treatment is not significantly higher than that in the Baseline treatment; and both of them are significantly lower than 1 as required by the notion of *rationality*.

3.4.3 Belief of Rationality and Higher-Order Belief of Rationality

In this section we present comparison results on players' belief of rationality and higher-order belief of rationality across treatments. We first examine the frequency of A's believing B's choosing



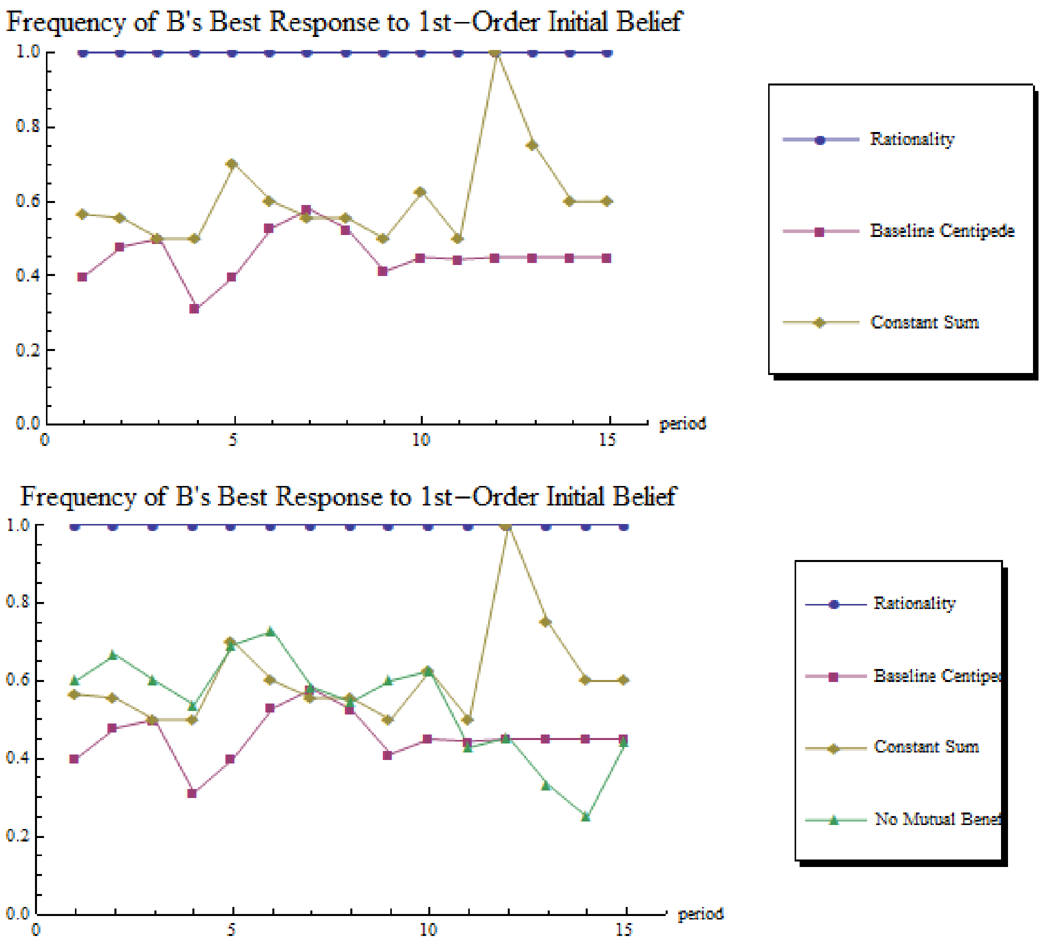
Note: Figure on top compares the average frequency of A's best responding to his/her stated belief if A is rational (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.7: Average Frequency of A's Best Responding to Own Belief, Across Periods

IN for the second stage.

Result 5. *(1) In all three treatments, the average frequency of player A's believing B's choosing IN is significantly higher than* 0*. (2) The average frequency of player A's believing B's choosing IN in the Constant-Sum treatment is significantly lower than that in the Baseline Centipede treatment.*

Result 5 addresses Hypothesis 5. As shown in Section 3.3.1, if *rationality and common strong belief of rationality* holds, in all states player A should not believe that B would ever chosen IN for the second stage. Figure **??** depicts the treatment-average frequency of player A's believing B's

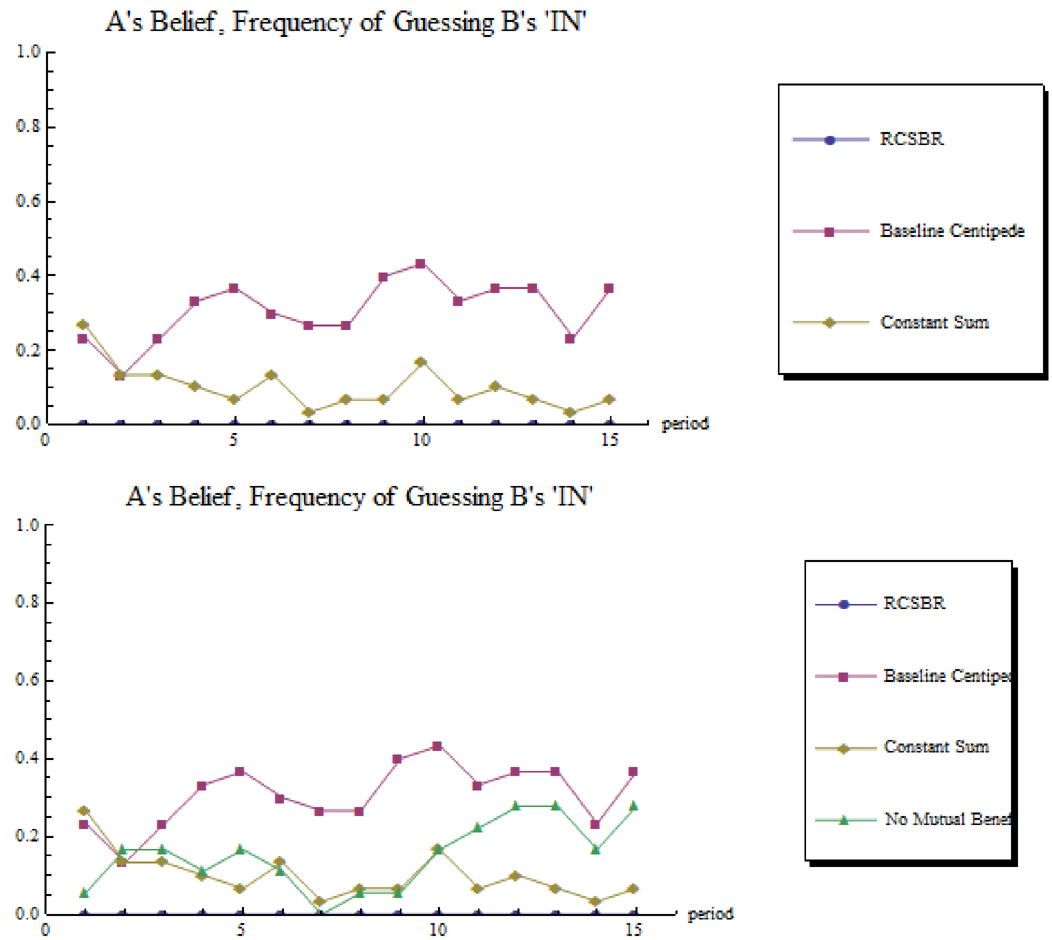


Note: Figure on top compares the average frequency of B's best responding to his/her own belief if B is rational (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.8: Average Frequency of B's Best Responding to Own Belief, Across Periods

choosing IN across all periods. This frequency in the Constant-Sum treatment is significantly lower than that in the Baseline treatment; but both of them are significantly higher than 0 as required by the notion of RCSBR. The comparison of player A's belief accuracy across treatments is included in Appendix 3.7.2.

We then examine player B's believing A's rationality. If player B's stated belief assigns a sum of probability 1 to the two cases in which player A is rational (either A chooses strategy IN-OUT and believes B has chosen IN, or A chooses OUT for the first stage and believes B has chosen OUT), we say that player B believes A's rationality.

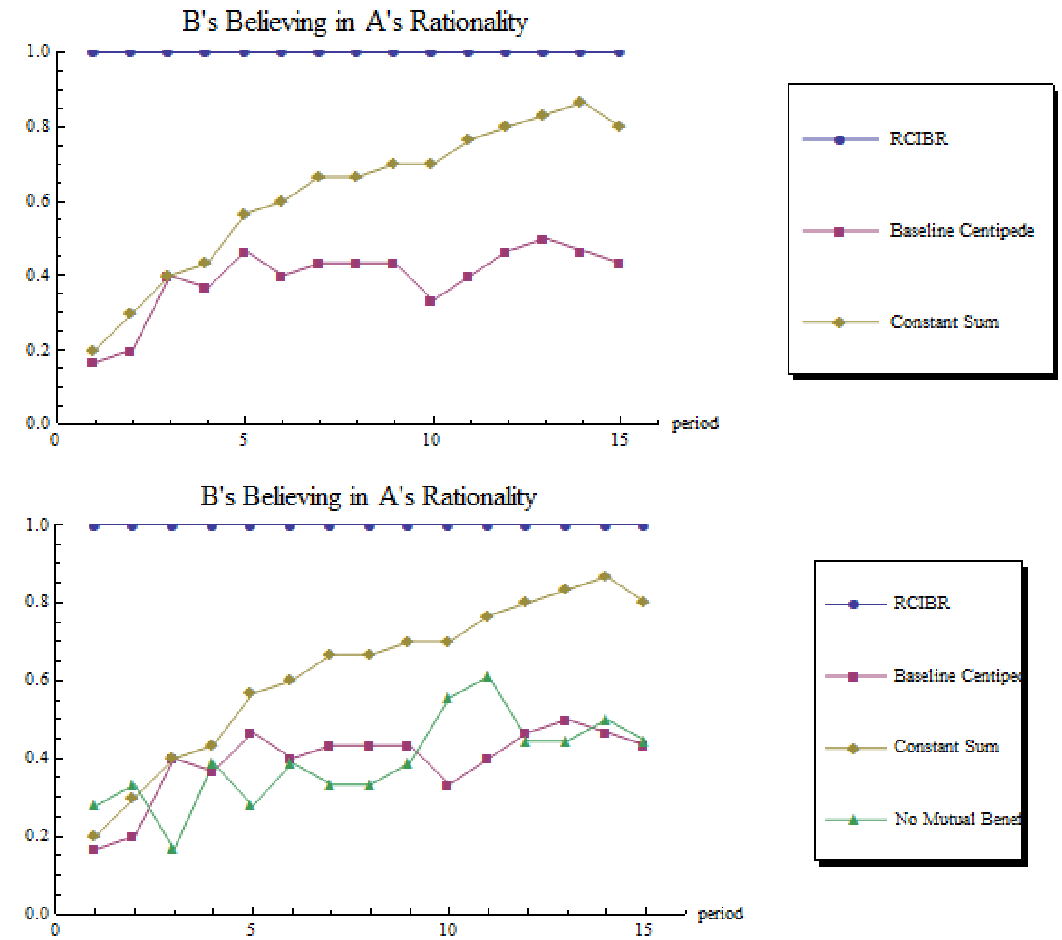


Note: Figure on top compares the average frequency of A's believing B's rationality and 2nd-Order rationality if RCSBR holds (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.9: Average Frequency of A's Believing B's Choosing IN, Across Periods

Result 6. *(1) In all three treatments, the average frequency of player B's believing A's rationality is significantly lower than* 1*. (2) The average frequency of player B's believing A's rationality in the Constant-Sum treatment is significantly higher than that in the Baseline Centipede treatment.*

Result 6 addresses Hypothesis 6. Figure **??** depicts the treatment-average frequency of player B's believing A's rationality across all periods. This frequency in the Constant-Sum treatment is significantly higher than that in the Baseline treatment; but both of them are significantly lower than 1 as required by the notion of *common belief of rationality.* It is also worth noting that in the Constant-Sum treatment this frequency increases towards 1 gradually as more rounds are played.

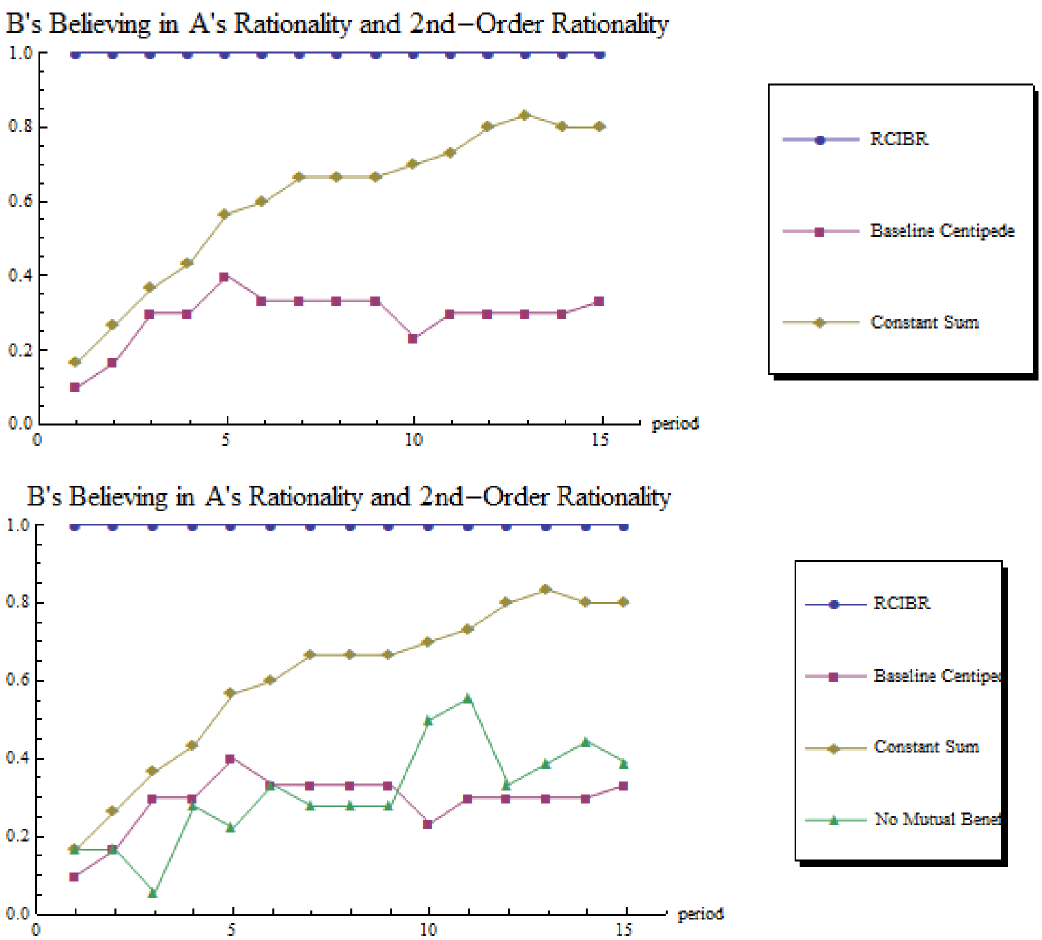


Note: Figure on top compares the average frequency of B's believing in A's rationality if *common belief in rationality* holds (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.10: Average Frequency of B's Believing A's Rationality, Across Periods

Next we examine player B's believing A's rationality AND believing A's believing B's rationality (2nd-Order rationality). If player B's initial belief assigns probability 1 to the event that player A chooses OUT for the first stage and believes B has chosen OUT, we say that player B initially believes A's rationality and 2nd-Order rationality.

Result 7. *(1) In all three treatments, the average frequency of player B's initially believing A's rationality and 2nd-order rationality is significantly lower than* 1*. (2) The average frequency of player B's initially believing A's rationality and 2nd-order rationality in the Constant-Sum treatment is significantly higher than that in the Baseline Centipede treatment.*



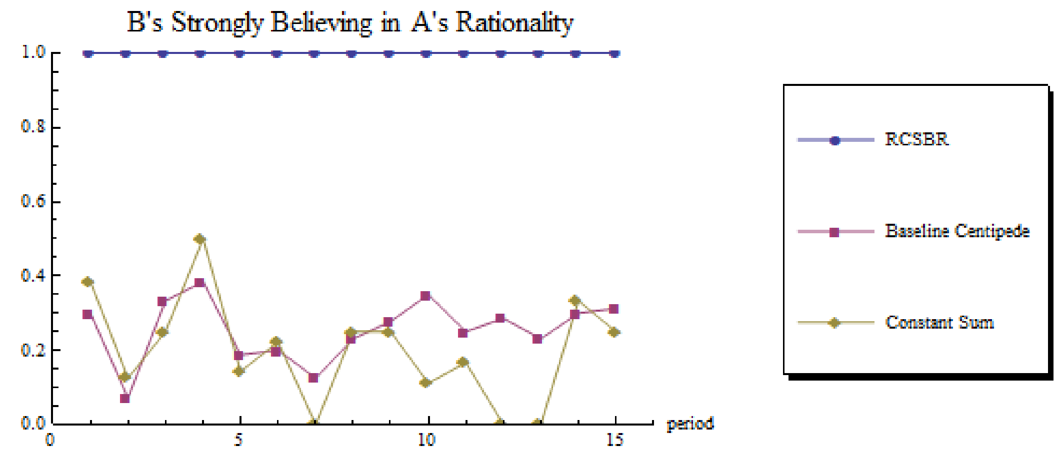
Note: Figure on top compares the average frequency of B's believing A's rationality and 2nd-Order rationality if RCIBR holds (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

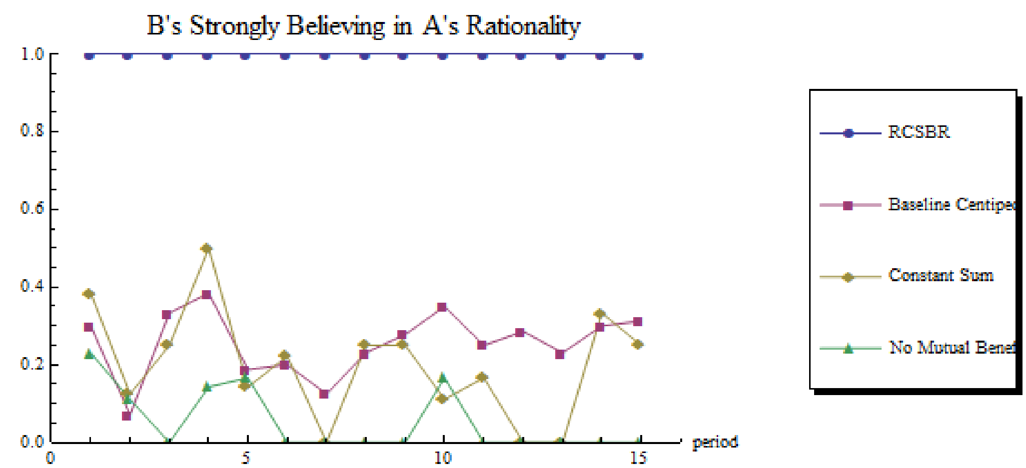
Figure 3.11: Average Frequency of B's Believing A's Rationality and 2nd-Order Rationality, Across Periods

Result 7 addresses Hypothesis 7. Figure ?? depicts the treatment-average frequency of player B's believing A's rationality and 2nd-order rationality across all periods. This frequency in the Constant-Sum treatment is significantly higher than that in the Baseline treatment; but both of them are significantly lower than 1 as required by the notion of *rationality and common initial belief of rationality*. It is also worth noting that in the Constant-Sum treatment this frequency increases towards 1 gradually as more rounds are played.

Last we look into player B's strongly believing A's rationality AND 2nd-Order rationality conditional on A has chosen IN for the first stage. If player B's conditional belief assigns probability 1 to the event that player A chooses strategy IN-OUT and believes B has chosen IN, we say that player B strongly believes A's rationality and 2nd-Order rationality.

Result 8. *(1) In all three treatments, the average frequency of player B's strongly believing A's rationality and 2nd-order rationality is significantly lower than* 1*. (2) The average frequency of player B's strongly believing A's rationality and 2nd-order rationality in the Constant-Sum treatment is not significantly different from that in the Baseline Centipede treatment.*





Note: Figure on top compares the average frequency of B's believing in A's rationality and 2nd-Order rationality if RCSBR holds (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.12: Average Frequency of B's Strongly Believing A's Rationality and 2nd-Order Rationality, Across Periods

Result 8 addresses Hypothesis 8. Figure ?? depicts the across-period treatment-average frequency of player B's strongly believing A's rationality and 2nd-order rationality conditional on B is informed that A has chosen IN for the first stage. This frequency in the Constant-Sum treatment is not significantly different from that in the Baseline treatment; and both of them are significantly lower than 1 as required by the notion of *rationality and common strong belief of rationality.* In other words, once player B observes player A's deviating from the equilibrium path, B hardly believes A's being rational AND A's believing B's rationality.

3.5 RELATED LITERATURE

McKelvey and Palfrey's (1992) seminal centipede game experiment shows individuals' behavior inconsistent with standard game theory prediction. Neither do they find convergence to subgame perfect equilibrium prediction as subjects gain experience in later rounds of the experiment. The authors attribute such inconsistent behavior to uncertainties over players' payoff functions; specifically, the subjects might believe a certain fraction of the population is altruist. They establish a structural econometric model to incorporate players' selfish/altruistic types, error probability in actions, and error probability in beliefs. If most of the players are altruistic, the altruistic type always chooses PASS except on the last node while the selfish type might mimic the altruist for the first several moves as in standard reputation models. As pointed out, the equilibrium prediction of this incomplete information game is sensitive to the beliefs about the proportion of the altruistic type. In our design we try to avoid this complication by allowing sorting.

Subsequent experimental studies on centipede games tend to view this failure of backward induction as individuals' irrationality. Fey et.al.(1996) examine a constant-sum centipede game which excludes the possibility of Pareto improvement by not backward inducting. Among the non-equilibrium models, they find that the Quantal Response Equilibrium, in which players err when playing their best responses, fit the data best. Zauner (1999) estimates the variance of uncertainties about players' preferences and payoff types and makes comparison between the altruism models and the quantal response models. Kawagoe and Takizawa (2012) offer an alternative explanation for the deviations in centipede games adopting level-k analysis. They claim that the level-k model provide good predictions for the major features in the centipede game experiment without the complication to incorporate incomplete information on "types." Nagel and Tang (1998) investigate centipede games in a variation of the strategy method: the game is played in the reduced normal form, which is considered as "strategically equivalent" to the extensive form counterpart, but precisely to identify "learning." They examine behavior across periods according to learning direction theory. They show significant differences in patterns of choices between the cases when a player has to split the cake before her opponent and when she moves after her opponent.

Consequently, other research tries to restore the subgame perfect equilibrium outcome by providing the subjects with aids in their decision-making processes. Bornstein et. al.(2004) show that groups tend to terminate the game earlier than individual players, once free communication is allowed within each group. Maniadis (2010) examines a set of centipede games with different stakes and finds that providing aggregate information causes strong convergence to the subgame perfect equilibrium outcome. However, after uncertainties are incorporated into the payoff structure, the effect of information provision shifts in the opposite direction. Rapoport et. al.(2003) study a three-person centipede game. They show that when the number of players increases and the stakes are sufficiently high, results converge to theoretical predictions more quickly. But when the game is played with low stakes, both convergence to equilibrium and learning across iterations of the stage game are weakened. Palacios-Huerta and Volij (2009) cast their doubt on average people's full rationality by recruiting expert chess players to play a field centipede. Strong convergence to subgame perfect prediction is observed.

3.6 CONCLUSION AND DISCUSSION

This paper explores people's beliefs behind non-backward induction behavior in laboratory centipede games. We elicit the first mover's belief about the second mover's strategy as well as the second mover's initial and conditional beliefs about the first mover's strategy and 1st-order belief. The measured beliefs help me infer the conditional probability systems of both players. The inferred CPS's and players' actual strategy choices identify why they fail to reach the BI outcomes. First, we examine whether the player's strategies are best response to the stated beliefs. In both the Baseline Centipede treatment and the Constant-Sum treatment, the frequency of players' best responding to own beliefs is significantly lower than 1. Specifically, the frequency in the Constant-Sum treatment is higher than that in the Baseline treatment; and the frequency in the No-Mutual-Benefit treatment is not significantly different from that in the Baseline treatment. Second, we investigate players' belief of opponents' rationality and higher-order belief of rationality. In all treatments, both the frequency of players' believing in others' rationality and the frequency of higher-order belief of rationality are significantly smaller than 1. Nevertheless, the frequency in the Constant-Sum treatment dominates that in the Baseline and the No-Mutual-Benefit treatment.

Third, when it comes to the second mover's conditional beliefs once the first-mover has chosen a non-BI strategy, the frequency of the second movers' strongly believing the first movers' rationality is very low; and there is no significant different across treatments.

3.7 APPENDIX

3.7.1 Proofs and Calculations

This section demonstrating the hypotheses specified in the main text by proving five observations. The first observation is about B's belief of A's rationality. The rest four observations identify the states that satisfy RCIBR and RCSBE. In summary, when the strategy choices and inferred CPS's constitute a state that satisfies *rationality and common strong belief of rationality* (RCSBR henceforth), players *do not* fail to reach the backward induction (BI henceforth) outcome in this state. But the reverse is not true. It is possible that Role A's strategy choice leads to the BI outcome, but Role B's strategy and belief are not consistent with RCSBR. Moreover, there exists a state in which Role B's strategy and belief are consistent with the BI outcome but Role A's are not. There also exists a state in which neither player's strategy and belief are consistent with the BI outcome, but a weaker notion of common belief in rationality, *rationality and common initial belief of rationality*, still holds.

For the east of demonstration, we alter the notations of the players' moves slightly, as shown in Figure 3.13. Since we shall prove the following observations for all three treatments, we use (x_j, y_j) to represent the players' payoffs associated with each terminal node. And u^A represents Statement OUT, t^A represents Statement IN in the instruction.

Observation 6. *From the measured initial belief of player B, if $\beta_{11} + \beta_{22} = 1$, player B initially believes player A's rationality.*

From the measured conditional belief of player B, if $\gamma_{22} = 1$, player B strongly believes player A's rationality and 2nd-Order rationality.

Observation 7. If the following data point is observed, the players' strategies and beliefs constitute a state that satisfies RCSBR:

- Role A chooses *Out* and statement u^A
- Role B chooses *Out* and the measured beliefs take the form:

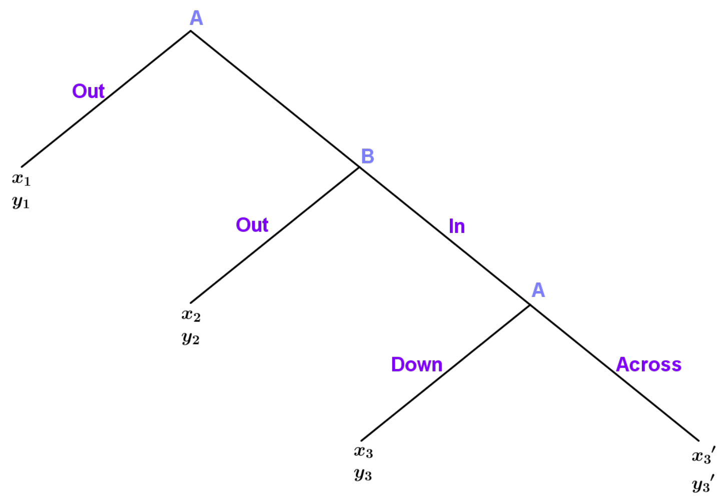


Figure 3.13: The Centipede Game

Table 3.3: Measured Initial Belief of Player B

	Out	Down	Across
u^A	β_{11}	β_{12}	β_{13}
t^A	β_{21}	β_{22}	β_{23}

Table 3.4: Measured Conditional Belief of Player B

	Down	Across
u^A	γ_{12}	γ_{13}
t^A	γ_{22}	γ_{23}

Observation 8. If the following data point is observed, Role B's strategy and belief are not consistent with RCSBR. Nevertheless, the BI outcome still obtains.

- Role A chooses *Out* and statement u^A

- Role B chooses *In* and the measured beliefs take the form:

Table 3.5: RCSBR Beliefs

u^A	1[0]	0[0]	0[0]
t^A	0[0]	0[1]	0[0]
	Out	Down	Across

Note: The first number in each cell represents Role B's belief in task (2). The second number in [] represents Role B's revised belief in task (3).

Table 3.6: Backward Induction without RCSBR Beliefs

u^A	1[0]	0[0]	0[0]
t^A	0[0]	0[0]	0[1]
	Out	Down	Across

Note: The first number in each cell represents Role B's belief in task (2). The second number in [] represents Role B's revised belief in task (3).

Remark: *1st-order strong belief of rationality* of both players because Role B's measured belief indicates that he does not *strongly* believe Role A's rationality. However, since Role A chooses *Out* at the first node, the BI outcome still obtains. Although RCSBR does not hold in this state, a weaker notion, *rationality and common initial belief of rationality* (RCIBR), still holds. RCIBR only requires the belief consistency given the root of the game tree.

Observation 9. If the following data point is observed, Role B's strategy and belief are consistent with the BI outcome. But the BI outcome does not obtain.

- Role A chooses *Down* and statement t^A
- Role B chooses *Out* and the measured beliefs take the form:

Remark: In this state there is no *1st-order* ***strong*** *belief of rationality*, nor *1st-order* ***initial*** *belief of rationality* because Role A's measured belief indicates that she does not strongly, nor initially believe Role B's rationality. Since Role A chooses *Down* and Role B chooses *Out*, the BI outcome

Table 3.7: Player B's Strong Belief of Rationality in Non-BI Outcome

	Out	Down	Across
u^A	1[0]	0[0]	0[0]
t^A	0[0]	0[1]	0[0]

Note: The first number in each cell represents Role B's belief in task (2). The second number in [] represents Role B's revised belief in task (3).

does not obtain. The game ends at the second node by Role B's playing *Out*.

Observation 10. If the following data point is observed, neither player's strategy and belief is consistent with the BI outcome. The BI outcome does not obtain. Nevertheless, the strategies and beliefs constitute a state that satisfies *rationality and common* ***initial*** *belief of rationality.*

- Role A chooses *Down* and statement t^A
- Role B chooses *In* and the measured beliefs take the form:

Table 3.8: No RCSBR and Non-Backward-Induction Outcome

	Out	Down	Across
u^A	1[0]	0[0]	0[0]
t^A	0[0]	0[0]	0[1]

Note: The first number in each cell represents Role B's belief in task (2). The second number in [] represents Role B's revised belief in task (3).

Remark: In this state there is no *1st-order strong belief of rationality* of both players because (1) Role B's measured belief indicates that he does not *strongly* believe Role A's rationality, and (2) Role A's measured beliefs indicates that she believes Role B's rationality in response to his belief, but she does not believe that Role B believes her rationality. Since Role A chooses *Down* and Role B chooses *In*, the BI outcome does not obtain. The game ends at the last node by Role A's choosing *Down*.

Compare observation 8 and 10. Role B's inferred CPS is the same, which assigns probability 0 to Role A's rationality if Role B's decision node is reached. Therefore, whenever Role B does not believe Role A's rationality conditional on a zero-probability event, Role A can attain a higher payoff by playing *Down* instead of *Out*. Both states satisfy *rationality and common* ***initial*** *belief of rationality*, but not *rationality and common* ***strong*** *belief of rationality*.

3.7.1.1 Proof for Observation 7 and 9

The inferred CPS's of both players are as follows:

Table 3.9: Proof for Observation 7 and 9

$\lambda^a(t^a)$

T^b		Out	In
	t^b	0	1

(S^b)

$\lambda^a(u^a)$

T^b		Out	In
	t^b	1[0]	0[1]

(S^b)

$\lambda^b(t^b)$

T^a		Out	Down	Across
	u^a	1[0]	0[0]	0[0]
	t^a	0[0]	0[1]	0[0]

(S^a)

We are going to show:

1. The state $(\text{Out}, u^a, \text{Out}, t^b)$ satisfies both RCIBR and RCSBR
2. The state $(\text{Down}, t^a, \text{Out}, t^b)$ satisfies neither RCIBR nor RCSBR

First notice that the strategy-type pair (Out, u^a) and (Down, t^a) are rational for player Ann. The strategy-type pair (Out, t^b) is rational for player Bob. For the **initial** belief we have:

$$\text{IB}^a(R_1^b) = \{u^a\}, \text{IB}^b(R_1^a) = \{t^b\},$$

then we have:

$$R_2^a = R_1^a \cap (S^a \times \text{IB}^a(R_1^b)) = \{(\text{Out}, u^a)\}$$
$$R_2^b = R_1^b \cap (S^b \times \text{IB}^b(R_1^a)) = \{(\text{Out}, t^b)\}$$

Inductively, we have $R^a_m\{(\text{Out}, u^a)\}$ and $R^b_m = (\text{Out}, t^b)$, $\forall m \in \mathbf{N}$. Therefore we have:

$$\cap_{m=1}^{\infty} R_m = (\text{Out}, u^a, \text{Out}, t^b)$$
$$\text{and } (\text{Down}, t^a, \text{Out}, t^b) \notin \cap_{m=1}^{\infty} R_m$$

As for **strong** beliefs, at the second node of the game, Bob's information set $H = \{\text{Ann would play "Down" or "Across"}\}$. Thus

$$H \times T^a = \{(\text{Down}, t^a), (\text{Down}, u^a), (\text{Across}, t^a), (\text{Across}, u^a)\}$$

Bob's type t^b is the only type who assigns probability 1 to any event E s.t. $E \cap (H \times T^a) \neq \emptyset$. So we have $\text{SB}^b(R^a_1) = \{t^b\}$.

At the first node of the game, $H = \emptyset$ for Ann. So Ann's strong beliefs at this node are degenerate. At the third node of the game, Ann's information set $H = \{\text{Bob played "In"}\}$. Both Ann's type assigns probability 1 to any event E s.t. $E \cap (H \times T^a) \neq \emptyset$. So we have $\text{SB}^a(R^b_1) = \{t^a, u^a\}$.

Inductively we have:

$$\begin{aligned} R^a_2 &= R^a_1 \cap (S^a \times \text{SB}^a(R^b_1)) = R^a_1 \\ R^b_2 &= R^b_1 \cap (S^b \times \text{SB}^b(R^a_1)) = \{(\text{Out}, t^b)\} \end{aligned}$$

Iterate one more level, we have:

$$\begin{aligned} \text{SB}^b(R^a_2) &= \text{SB}^b(R^a_1) = \{t^b\} \\ \text{SB}^a(R^b_2) &= \{t^a \in T^a : \forall H \text{ s.t. } R^b_2 \cap (H \times T^b) \neq \emptyset, \lambda^a(t^a)(R^b_2) = 1\} \\ &= \{u^a\} \end{aligned}$$

and

$$R^a_3 = R^a_2 \cap (S^a \times \text{SB}^a(R^b_2)) = \{(\text{Out}, u^a)\}$$
$$R^b_3 = R^b_2 \cap (S^b \times \text{SB}^b(R^a_2)) = \{(\text{Out}, t^b)\}$$

Then we have $\text{SB}^b(R^a_3) = \{t^b\}$ and $\text{SB}^a(R^b_3) = \{u^a\}$. For any $m \geq 3$, we have $R^a_m = \{(\text{Out}, u^a)\}$ and $R^b_m = \{(\text{Out}, t^b)\}$. Thus:

$$\cap_{m=1}^{\infty} R_m = \{(\text{Out}, u^a, \text{Out}, t^b)\}$$

Therefore, the only state that satisfies *rationality and common strong belief of rationality* is $(\text{Out}, u^a, \text{Out}, t^b)$.

The results are summarized in the following table:

Table 3.10: Summary for Proof for Observation 7 and 9

State	RCIBR	RCSBR
$(\text{Down}, t^a, \text{Out}, t^b)$	$\times$	$\times$
$(\text{Out}, u^a, \text{Out}, t^b)$	$\surd$	$\surd$

Table 3.11: Proof for Observation 8 and 10

$\lambda^a(t^a)$

T^b	Out	In
u^b	0	0
t^b	0	1

S^b

$\lambda^a(u^a)$

T^b	Out	In
u^b	1[0]	0[0]
t^b	0[0]	0[1]

S^b

$\lambda^b(t^b)$

T^a	Out	Down	Across
u^a	1[0]	0[0]	0[0]
t^a	0[0]	0[0]	0[1]

S^a

3.7.1.2 Proof for Observation 8 and 10

The inferred CPS's of both players are:

We are going to show:

- Both states $(\text{Down}, t^a, \text{In}, t^b)$ and $(\text{Out}, u^a, \text{In}, t^b)$ satisfy RCIBR but not RCSBR.

We first identify strategy-type pairs that are rational. For player Ann, it is easy to show that strategy $s^a = \text{Down}$ maximizes type t^a's expected payoff and strategy $s^a = \text{Out}$ maximizes type u^a's expected payoff. For player Bob, strategy $s^b = \text{In}$ maximizes type t^b's expected payoff. Thus we have:

$$R_1^a = \{(\text{Down}, t^a), (\text{Out}, u^a)\}$$
$$R_1^b = \{(\text{In}, t^b)\}$$

Initial beliefs:

Both Ann's type t^a and u^aassign probability 1 to (In, t^b), so we have $\text{IB}^a(R_1^b) = \{t^a, u^a\}$. Bob's type t^b assigns probability 1 to $(\text{Out}, u^a) \in R_1^a$, so we have $\text{IB}^b(R_1^a) = \{t^b\}$.

Then we have:

$$\begin{aligned} R_2^a &= R_1^a \cap (S^a \times \text{IB}^a(R_1^b)) \\ &= \{(\text{Down}, t^a), (\text{Out}, u^a)\} \cap (\{\text{Out}, \text{Down}, \text{Across}\} \times \{t^a, u^a\}) \\ &= R_1^a \end{aligned}$$

Similarly, $R_2^b = R_1^b$, and $R_3^a = R_2^a \cap (S^a \times \text{IB}^a(R_2^b)) = R_2^a = R_1^a$. Mathematical induction gives:

$$R_m^a = R_{m-1}^a = R_1^a \Rightarrow R_{m+1}^a = R_m^a = R_1^a$$

Similar result for Bob. Therefore we have

$$\begin{aligned} R_m &= R_m^a \times R_m^b = R_{m-1} \\ &\Rightarrow \cap_{m=1}^{\infty} R_m = R_1^a \times R_1^b \\ &= \{(\text{Down}, t^a, \text{In}, t^b), (\text{Out}, u^a, \text{In}, t^b)\} \end{aligned}$$

Thus both states satisfy *rationality and common initial belief of rationality.*

Strong beliefs:

At the second node of the game, Bob's information set $H = \{\text{Ann would play "Down" or "Across"}\}$. Thus

$$H \times T^a = \{(\text{Down}, t^a), (\text{Down}, u^a), (\text{Across}, t^a), (\text{Across}, u^a)\}$$

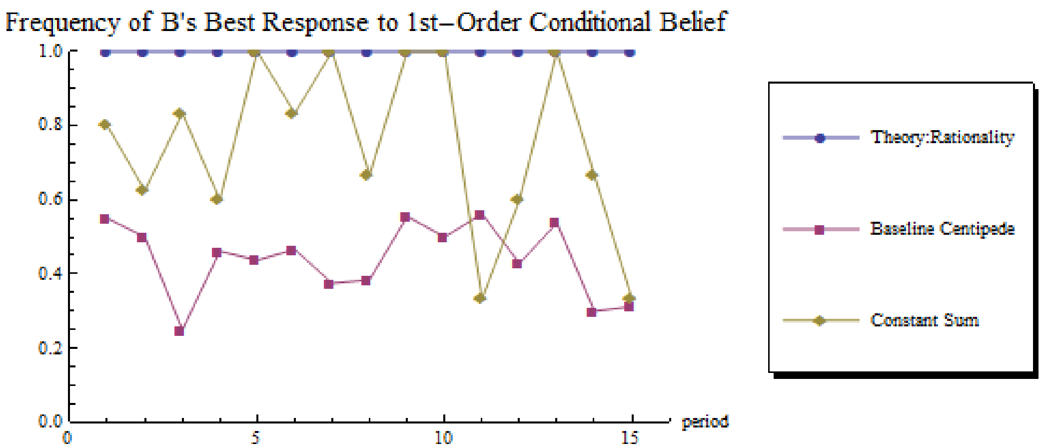
Bob's type t^b assigns probability 0 to $(\text{Down}, t^a) \in R_1^a$, but assigns probability 1 to $(\text{Across}, t^a) \notin R_1^a$. So we have $\text{SB}^b(R_1^a) = \emptyset$. Thus $\cap_{m=1}^{\infty} R_m = \emptyset$. No state belongs to $\emptyset$. Hence neither state satisfies RCSBR.

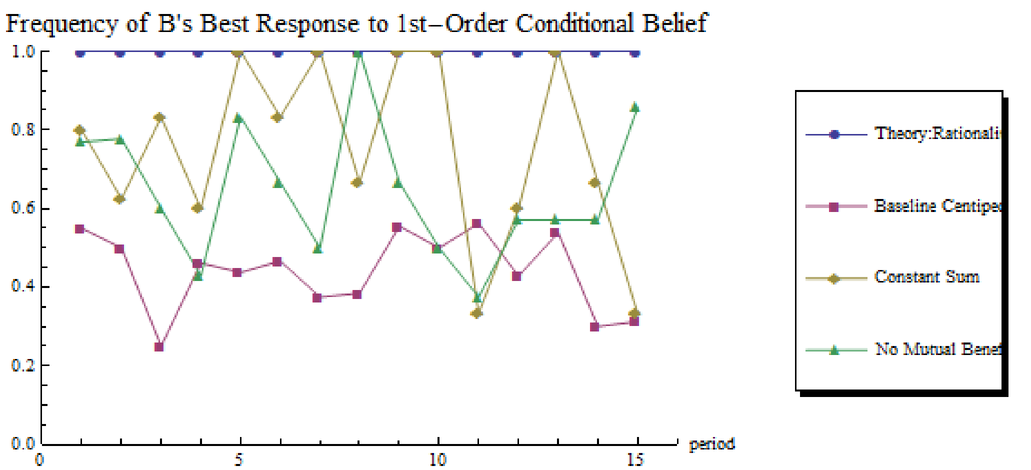
The results are summarized in the following table:

Table 3.12: Summary of Proof for Observation 8 and 10

State	RCIBR	RCSBR
$(\text{Down}, t^a, \text{In}, t^b)$	$\surd$	$\times$
$(\text{Out}, u^a, \text{In}, t^b)$	$\surd$	$\times$

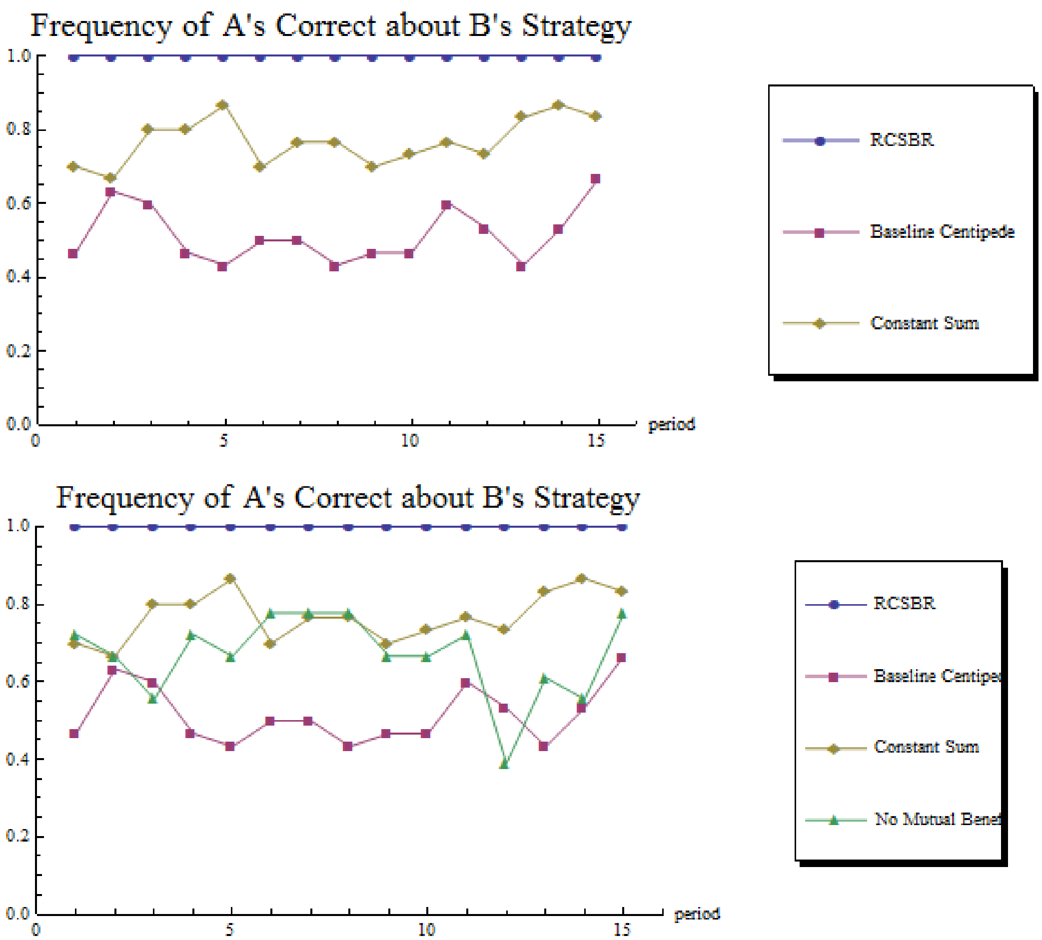
3.7.2 Other Figures and Tables





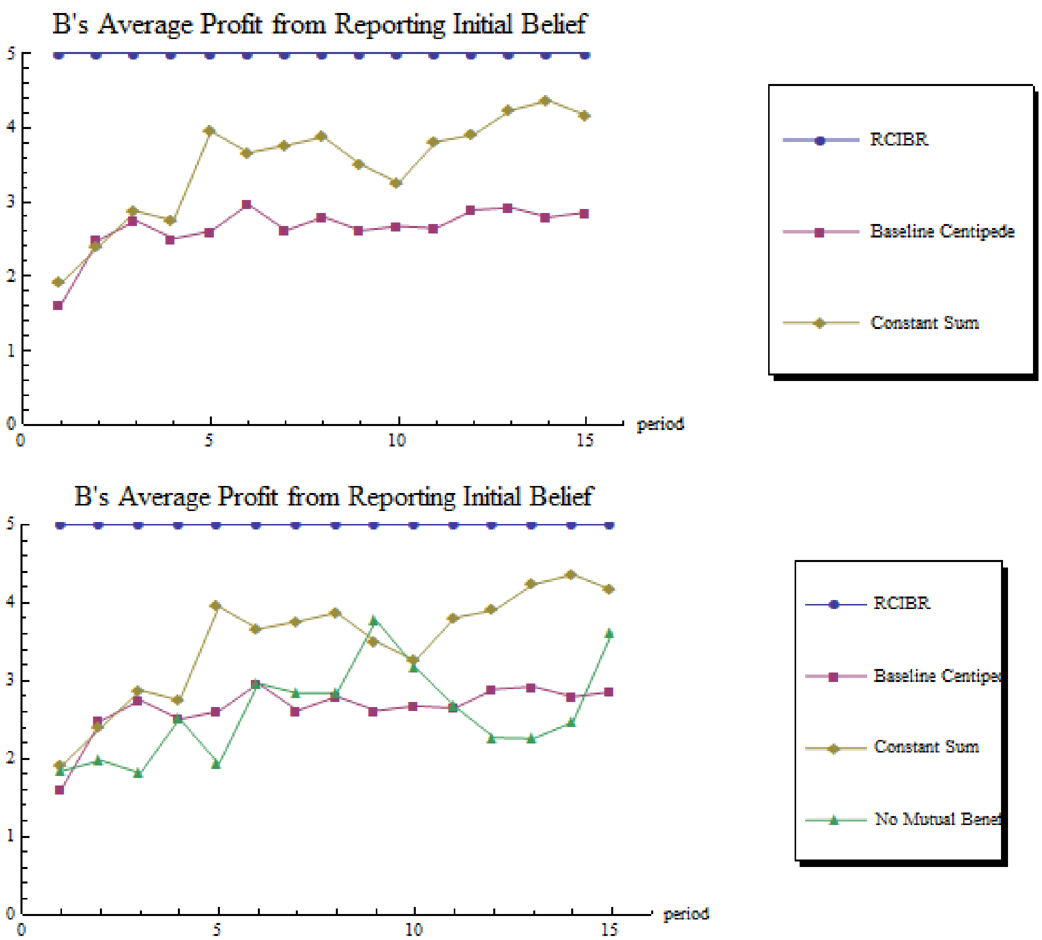
Note: Figure on top compares the average frequency of B's best responding to his/her stated conditional belief if B is conditionally consistent (blue curve), the frequency from the Baseline Centipede treatment (purple curve), and the frequency from the Constant-Sum treatment (yellow curve). Figure at bottom adds the average frequency from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.14: Average Frequency of B's Best Responding to Own Conditional Belief, Across Periods



Note: Figure on top compares the average accuracy of A's belief if RCBR holds (blue curve), the actual accuracy from the Baseline Centipede treatment (purple curve), and the actual accuracy from the Constant-Sum treatment (yellow curve). Figure at bottom adds the actual accuracy from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.15: Accuracy of A's Belief, Across Periods



Note: Figure on top compares the average accuracy of B's initial belief if RCIBR holds(blue curve), the actual accuracy from the Baseline Centipede treatment (purple curve), and the actual accuracy from the Constant-Sum treatment (yellow curve). Figure at bottom adds the actual accuracy from the No-Mutual-Benefit treatment (green curve), to the comparison.

Figure 3.16: Accuracy of B's Belief, Across Periods

3.7.3 Laboratory Instructions

INSTRUCTIONS

Welcome! Thank you for participating in this experiment. This experiment studies decision-making between two individuals. In the following one hour or less, you will participate in 15 rounds of decision making. Please read the instructions carefully; the cash payment you earn at the end of the experiment may depend on how well you understand the instructions and make your decisions accordingly.

Your Role and Decision Group

Half of the participants will be randomly assigned the role of Member A and half will be assigned the role of Member B. Your role will remain fixed throughout the experiment. In each round, one Member A will be paired with one Member B to form a group of two. The two members in a group make decisions that will affect their earnings in the round. Participants will be randomly rematched with another member of the opposite role after each round.

Your Choice Task(s) in Each Round

In each round, each group will face the three-stage decision task shown in Figure ??. The nodes represent choice stages, the letters above the nodes represent the member who is going to make a choice, and the numbers represent the points one will earn, with A's points on top and B's points at bottom.

- In the 1st stage A must decide between two options: *Out* or *In*. If A chooses *Out*, the task ends with A receiving 20 and B 10 points. If A chooses *In*, the task proceeds to the 2nd stage.
- In the 2nd stage B must decide between two options: *Out* or *In*. If B chooses *Out*, the task ends with A receiving 10 and B 40 points. If B chooses *In*, the task proceeds to the 3rd stage.
- In the 3rd stage A must choose again between two options: *Out* or *In*. If A chooses *Out*, A will receive 40 and B 30 points. If A chooses *In*, A will receive 25 and B 45 points.

Member A's Choice Task

You will be asked to specify your choices for **both** stage 1 and 3 through a computer interface. For each stage, you can choose one and only one option. Note that you will be making your choices **at the same time** your partner B is making his or her choice. So you don't know what B chooses. The choices you make here will be carried out automatically by the computer later on. You will not have an opportunity to revise them.

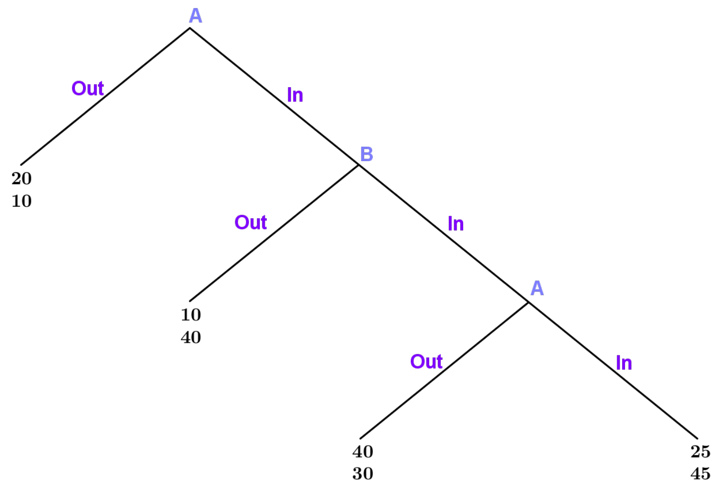


Figure 3.17: The Decision Task

Member B's Choice Task

You will be asked to specify your choice for stage 2 through a computer interface. You can choose one and only one option. Note that you will be making your choice **at the same time** your partner A is making his or her choices. So you don't know what A chooses. The choice you make here will be carried out automatically by the computer later on. You will not have an opportunity to revise it.

Forecast Tasks in Each Round

Besides having the opportunity to earn points in the choice task, you will also be given the opportunity to earn extra points by making forecast(s).

Member A's Forecast Task

Your partner, Member B, has made a choice for stage 2. Please select the statement that you believe is more likely:

- **Statement I**: Member B has chosen *In*.
- **Statement O**: Member B has chosen *Out*.

You will earn 5 points if your forecast is correct (i.e. if Member B chooses *In* and you select Statement I, or B chooses *Out* and you select Statement O). You will earn nothing otherwise.

Member B's Forecast Task(s)

Your partner, Member A, has made choices for both stage 1 and 3; also, he or she is selecting between **Statement I** and **Statement O**, each of which is a statement about the choice you just made for stage 2. Which choices do you think your partner A has made for his or her stages, **and** which statement do you think your partner A is selecting?

Notice that A's selections can be expressed in the table below. The column represents A's selection of statement, the row represents A's choices for 1st and 3rd stages. So each cell represents an outcome of A's choices **and** statement. For example, the upper-left cell represents the outcome that A has chosen *Out* for 1st stage, *In* or *Out* for 3rd stage, **and** *Statement I*.

Statement I	■		
Statement O			
	1st Stage *Out*, 3rd *In* or *Out*	1st Stage *In*, 3rd Stage *Out*	1st Stage *In*, 3rd Stage *In*

<u>Your first forecast task</u>

Your first task is to forecast the percent chance that each of the six outcomes happens. A percent chance is a number between 0 and 100, where 100 means that you are certain that such outcome is the correct one, and 0 means that you are certain that such outcome is *not* the correct one. Enter the percent chance of each outcome into the corresponding cell. If you leave any cell as blank it will be viewed as 0. Make sure the six numbers sum up to 100.

You will earn 5 points if your forecast exactly coincide with your partner A's statement **and** choices. If your forecast does not exactly coincide with your partner A's choice and statement, you will receive 5 points minus 2.5 times a penalty amount. The penalty amount is the sum of squared distances between each of the six numbers you entered and the correct answer, i.e. the outcome from A's selection.

Example: Suppose you believe that with 80 percent chance A has chosen to play *In* for 1st and *Out* for 3rd stage, **and** has selected *Statement I*; with 15 percent chance A has chosen to play *In* for 1st and *Out* for 3rd stage, **and** has selected *Statement O*; with 5 percent chance A has chosen to play *In* for 1st and *In* for 3rd stage, **and** has selected *Statement O*, you should enter the numbers as below:

Statement I	0	80	0
Statement O	0	15	5 ■
	1st Stage *Out*, 3rd *In* or *Out*	1st Stage *In*, 3rd Stage *Out*	1st Stage *In*, 3rd Stage *In*

Now suppose your partner A has chosen *In* for 1st and *In* for 3rd stage, and has selected *statement O*. The penalty amount is $(100/100 - 5/100)^2 + (0 - 80/100)^2 + (0 - 15/100)^2 = 1.54$. So you earn $5 - 2.5 * 1.54 = 1.15$ from this forecast.

<u>Your second forecast task</u>

After the computer carries out your partner's and your choices, you will be informed if your partner

A has chosen *In* for stage 1. Now you have a chance to make a second forecast. A four-cell table will be presented to you. (The first column of the table in your first forecast task is removed because A has chosen *In* for stage 1.) Please make a percent chance forecast again. Your penalty amount and earning point are calculated in the same way as in your first forecast task.

Final Comments

At the end of this experiment one round will be randomly selected to count for payment. Your earning in each round is the sum of the points you earn from the choice task and the forecast task(s). The exchange rate between points and US dollars is 2.5 : 1. Your cash payment will be your earning in US dollars plus the $5 show-up fee.

Your decisions and your payment will be kept confidential. You have to make decisions entirely on your own. Please do not talk to others. If you have any question at any time, raise your hand and the experimenter will come and assist you individually. Please turn off your cell phone and other electronic devices.

If you have any question, please raise your hand now. Otherwise we will proceed to the quiz.

BIBLIOGRAPHY

Ainsworth, S. and I. Sened, "The role of lobbyists: entrepreneurs with two audiences," *American Journal of Political Science*, 1993, pp. 834–866.

Alonso, R., W. Dessein, and N. Matouschek, "When does coordination require centralization?," *The American economic review*, 2008, *98* (1), 145–179.

Aumann, R. and A. Brandenburger, "Epistemic conditions for Nash equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 1161–1180.

Aumann, R.J., "Backward induction and common knowledge of rationality," *Games and Economic Behavior*, 1995, *8* (1), 6–19.

____ , "Reply to binmore," *Games and Economic Behavior*, 1996, *17* (1), 138–146.

____ , "On the centipede game," *Games and Economic Behavior*, 2004, *23* (1), 97–105.

Austen-Smith, D. and J.S. Banks, "Information aggregation, rationality, and the Condorcet jury theorem," *American Political Science Review*, 1996, pp. 34–45.

____ **and T.J. Feddersen**, "Information aggregation and communication in committees," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009, *364* (1518), 763–769.

Austen-Smith, David and T. Feddersen, "Deliberation and Voting Rules," *The Center for Mathematical Studies in Economics and Management Science at Northwestern University Discussion Paper No. 1359*, 2002.

____ **and** ____ , "The Inferiority of Deiliberaition under Unanimity," *The Center for Mathematical Studies in Economics and Management Science at Northwestern University Discussion Paper No. 1359*, 2002.

Battigalli, P., "On rationalizability in extensive games," *Journal of Economic Theory*, 1997, *74* (1), 40–61.

____ **and M. Siniscalchi**, "Hierarchies of conditional beliefs and interactive epistemology in dynamic games," *Journal of Economic Theory*, 1999, *88* (1), 188–230.

____ **and** ____ , "Strong belief and forward induction reasoning," *Journal of Economic Theory*, 2002, *106* (2), 356–391.

Ben-Porath, E., "Rationality, Nash equilibrium and backwards induction in perfect-information games," *The Review of Economic Studies*, 1997, *64* (1), 23–46.

Bicchieri, C., "Strategic behavior and counterfactuals," *Synthese*, 1988, *76* (1), 135–169.

____, "Self-refuting theories of strategic interaction: A paradox of common knowledge," *Erkenntnis*, 1989, *30* (1), 69–85.

Binmore, K., "Rationality and backward induction," *Journal of Economic Methodology*, 1997, *4* (1), 23–41.

Binmore, Ken, "A note on backward induction," *Games and Economic Behavior*, 1996, *17* (1), 135–137.

Board, Oliver and Tiberiu Dragu, "Expert Advice with Multiple Decision Makers," *University of Pittsburgh, Department of Economics Working Paper*, 2008.

Bonanno, G., "The logic of rational play in games of perfect information," *Economics and Philosophy*, 1991, *7* (1), 37–65.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer, "Individual and group decisions in the centipede game: Are groups more "rational" players?," *Journal of Experimental Social Psychology*, 2004, *40* (5), 599–605.

Brandenburger, A., "The power of paradox: some recent developments in interactive epistemology," *International Journal of Game Theory*, 2007, *35* (4), 465–492.

Caillaud, B. and J. Tirole, "Consensus building: How to persuade a group," *The American Economic Review*, 2007, *97* (5), 1877–1900.

Chakraborty, A. and R. Harbaugh, "Persuasion by cheap talk," *The American Economic Review*, 2010, *100* (5), 2361–2382.

Coughlan, P.J., "In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting," *American Political Science Review*, 2000, pp. 375–393.

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

DellaVigna, S. and M. Gentzkow, "Persuasion: Empirical Evidence," *National Bureau of Economic Research*, 2009.

Duggan, J. and C. Martinelli, "A Bayesian model of voting in juries," *Games and Economic Behavior*, 2001, *37* (2), 259–294.

Farrell, J. and R. Gibbons, "Cheap talk with two audiences," *The American Economic Review*, 1989, *79* (5), 1214–1223.

Feddersen, T. and W. Pesendorfer, "Voting behavior and information aggregation in elections with private information," *Econometrica: Journal of the Econometric Society*, 1997, pp. 1029–1058.

___ and ___ , "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting," *American Political Science Review*, 1998, pp. 23–35.

___ and ___ , "Elections, information aggregation, and strategic voting," *Proceedings of the National Academy of Sciences*, 1999, *96* (19), 10572.

Fey, M., R.D. McKelvey, and T.R. Palfrey, "An experimental study of constant-sum centipede games," *International Journal of Game Theory*, 1996, *25* (3), 269–287.

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental economics*, 2007, *10* (2), 171–178.

Galeotti, Christian Ghiglino Andrea and Francesco Squintani, "Strategic Information Transmission in Networks," *University of Essex, Department of Economics Working Paper*, 2009.

Gerardi, D. and L. Yariv, "Deliberative voting," *Journal of Economic Theory*, 2007, *134* (1), 317–338.

Gigler, F., "Self-enforcing voluntary disclosures," *Journal of Accounting Research*, 1994, *32* (2), 224–240.

Glazer, J. and A. Rubinstein, "On optimal rules of persuasion," *Econometrica*, 2004, *72* (6), 1715–1736.

___ and ___ , "A study in the pragmatics of persuasion: a game theoretical approach," *Theoretical Economics*, 2006, *1* (4), 395–410.

Goltsman, M. and G. Pavlov, "How to talk to multiple audiences," *Games and Economic Behavior*, 2011, *72* (1), 100–122.

Green, J.R. and J.J. Laffont, "Partially verifiable information and mechanism design," *The Review of Economic Studies*, 1986, pp. 447–456.

___ **and N.L. Stokey**, "A two-person game of information transmission," *Journal of Economic Theory*, 2007, *135* (1), 90–104.

Grossman, S.J., "The informational role of warranties and private disclosure about product quality," *Journal of Law and economics*, 1981, *24* (3), 461–483.

Hagenbach, J. and F. Koessler, "Strategic communication networks," *Review of Economic Studies*, 2010, *77* (3), 1072–1099.

Johns, L., "A servant of two masters: communication and the selection of international bureaucrats," *International Organization*, 2007, *61* (2), 245–275.

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

Kamenica, Emir and Matthew Gentzkow, "Competition in Persuasion," *University of Chicago Working Paper*, 2011.

Karni, Edi, "A mechanism for eliciting probabilities," *Econometrica*, 2009, *77* (2), 603–606.

Kawagoe, Toshiji and Hirokazu Takizawa, "Level-k analysis of experimental centipede games," *Journal Of Economic Behavior & Organization*, 2012.

Kawamura, Kohei, "A Model of Public Consultation: Why is Binary Communication So Common?," *University of Edinburgh, School of Economics Working Paper*, 2010.

Koessler, F., "Lobbying with two audiences: Public vs private certification," *Mathematical Social Sciences*, 2008, *55* (3), 305–314.

Koessler, Frederic and David Martimort, "Multidimensional Communication Mechanisms: Cooperative and Conflicting Designs," *Paris School of Economics Working Paper*, 2008.

Ladha, K.K., "The Condorcet jury theorem, free speech, and correlated votes," *American Journal of Political Science*, 1992, pp. 617–634.

Levitt, Steven D, A List John, and Sally Sadoff, "Checkmate: Exploring Backward Induction among Chess Players," *American Economic Review*, 2011, *101* (2), 975–990.

Maniadis, Zacharias, "Aggregate Information and the Centipede Game: a Theoretical and Experimental Study," *UCLA Department of Economics Working Paper*, 2010.

McKelvey, R.D. and T.R. Palfrey, "An experimental study of the centipede game," *Econometrica: Journal of the Econometric Society*, 1992, pp. 803–836.

McLennan, A., "Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents," *American Political Science Review*, 1998, pp. 413–418.

Meirowitz, A., "In defense of exclusionary deliberation: communication and voting with private beliefs and values," *Journal of Theoretical Politics*, 2007, *19* (3), 301–327.

Milgrom, P. and J. Roberts, "Relying on the information of interested parties," *The RAND Journal of Economics*, 1986, pp. 18–32.

Milgrom, P.R., "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 1981, pp. 380–391.

Morris, S. and H.S. Shin, "Social value of public information," *The American Economic Review*, 2002, *92* (5), 1521–1534.

___ **and** ___ , "Optimal communication," *Journal of the European Economic Association*, 2007, *5* (2-3), 594–602.

Nagel, R. and F.F. Tang, "Experimental results on the centipede game in normal form: an investigation on learning," *Journal of Mathematical Psychology*, 1998, *42* (2), 356–384.

Newman, P. and R. Sansing, "Disclosure policies with multiple users," *Journal of Accounting Research*, 1993, *31* (1), 92–112.

Palacios-Huerta, I. and O. Volij, "Field centipedes," *The American Economic Review*, 2009, *99* (4), 1619–1635.

Palfrey, Thomas R and Stephanie W Wang, "On eliciting beliefs in strategic games," *Journal of Economic Behavior & Organization*, 2009, *71* (2), 98–109.

Pettit, P. and R. Sugden, "The backward induction paradox," *The Journal of Philosophy*, 1989, pp. 169–182.

Rapoport, A., W.E. Stein, J.E. Parco, and T.E. Nicholas, "Equilibrium play and adaptive learning in a three-person centipede game," *Games and Economic Behavior*, 2003, *43* (2), 239–265.

Rayo, Luis and Ilya Segal, "Optimal Information Disclosure," *Stanford University, Department of Economics Working Paper*, 2010.

Reny, P.J., "Common knowledge and games with perfect information," in "PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association" JSTOR 1988, pp. 363–369.

——, "Backward induction, normal form perfection and explicable equilibria," *Econometrica: Journal of the Econometric Society*, 1992, pp. 627–649.

Rényi, A., "On a new axiomatic theory of probability," *Acta Mathematica Hungarica*, 1955, *6* (3), 285–335.

Rosar, Frank and Elisabeth Schulte, "Imperfect Private Information and the Design of Information-Generating Mechanisms," *University of Bonn Working Paper*, 2010.

Spence, M., "Job market signaling," *The quarterly journal of Economics*, 1973, *87* (3), 355–374.

Wang, S.W., "Incentive effects: The case of belief elicitation from individuals in groups," *Economics Letters*, 2011, *111* (1), 30–33.

Zauner, K.G., "A payoff uncertainty explanation of results in experimental centipede games," *Games and Economic Behavior*, 1999, *26* (1), 157–185.