THE UNIVERSITY OF CHICAGO

ESSAYS ON STRATEGIC COMMUNICATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
TAKUMA HABU

CHICAGO, ILLINOIS
JUNE 2023

Contents

List of Tables

List of Figures

Abstract

Effective communication requires trust between the parties, and the consequences of miscommunication can sometimes be dire—"The Boy Who Cried Wolf" was ultimately eaten by the wolf, not because the villagers could not hear him, but because they did not believe he was telling the truth. Situations with mistrust between communicating parties are commonplace: e.g., voters are often suspicious of politicians' statements because they worry that politicians are insincere; the court generally questions the credibility of witnesses; buyers might worry that a seller is not telling the truth in advertisements.

In this thesis, I use game theory to study the role of trust in communication between a *sender*, who is the source of information about the state of the world, and an *audience of receivers* that make decisions based on the information provided. Formally, I capture trust, or lack thereof, between the sender and the audience as whether or not the sender can be trusted to follow a communication strategy. Thus, the sender's communication is *Bayesian persuasion* when there is trust (Kamenica and Gentzkow, 2011) and is *cheap talk* (Crawford and Sobel, 1982) when there is no trust. Throughout, I focus on the case when the sender's preference is state independent so that the audience knows the sender's motives for providing information; e.g., voters know that politicians campaign to obtain votes, the court knows that plaintiffs and defendants provide evidence in support of their case, and the buyers know that sellers advertise to sell their products.The first

chapter studies a novel way in which communication can be effective when there is a lack of trust between the sender and an audience of receivers. Specifically, I explore how the presence of diverse opinions in the audience can make the sender's cheap talk credible by making statements have stakes. Building on this intuition, I study how the sender can optimally communicate *semi-publicly*; i.e., by partitioning the audience into strategically formed groups and communicating publicly within each group but privately across groups. Using a canonical game of persuasion, I show that it is optimal for the sender to separate her audience into two groups based on whether the sender needs to persuade the receiver in the first place. The sender can further benefit by partitioning the group that consists of those that need persuading into differently diverse (sub)groups by trading off the desire to tailor communication to individuals and the desire to gain credibility by ensuring appropriate diversity of opinions in the groups. A practical implication of these results is that, while there is no need to ensure diversity of opinions in political rallies that are held for the supporters, a politician can be more persuasive to swing voters by campaigning across multiple events in which the audience consists of groups of swing voters that care about different sets of issues.

Effective communication can be problematic because effectiveness is about whether the sender benefits, and not whether the audience benefits. Thus, when there is a conflict of interest between the sender and the audience, the sender might not be willing to provide sufficient information that would benefit the audience. The second chapter studies how the audience can induce the sender to provide more information than the sender is otherwise willing. Specifically, I consider the case in which there is doubt about the sender's trustworthiness and the audience is able to investigate and learn about the sender's trustworthiness before making a decision. While learning about the sender allows the receiver to avoid making decisions based on unreliable information, it can also affect the sender's incentive to provide information in the first place. I introduce doubts about the sender's trustworthiness in a canonical game of persuasion and show that the receiver's optimal investigations necessarily involve the receiver avoiding learning about the sender. This, in turn, means that the receiver must be able to commit to ignorance to directly implement optimal investigations. Alternatively, I show that the receiver can indirectly implement investigations by delegating investigations to a third party who is partially adversarial toward the sender and also partially aligned with the receiver. Interpreting the receiver's investigations as cross-examination of witnesses in courts or audit of financial information, these results shed light on the importance of the investigator's incentives in enabling courts or buyers to obtain more information from strategic sources of information.

While the previous chapters focus on communication games whereby players can communicate directly using costless messages, the last chapter studies a repeated game in which players do not have access to such explicit forms of communication. Specifically, the final chapter concerns characterisation of equilibrium payoffs in a two-player, undiscounted, infinitely repeated game in which only one player is informed about the state of the world, and players observe only each other's actions in each stage of the repeated game (Aumann, Maschler and Stearns, 1968). Such repeated games are related to communication games because the absence of discounting means that the informed player's actions in any initial finite stages of the game can be interpreted as costless messages sent by the sender in cheap-talk games. The main result of the chapter is a new characterisation of the informed player's equilibrium payoffs when the informed player's preferences are state independent. Unlike in the general case in which infinite stages of communication are sometimes needed to obtain equilibrium payoffs (Hart, 1985), with state-independent informed player preferences, only a finite number of stages of communication are needed to characterise the informed player's equilibrium payoffs. Thus, the result simplifies the identification of the set of equilibrium payoffs for a sender with known incentives to provide information when explicit communication is not possible.

Acknowledgements

I would like to express my deepest gratitude to my advisor Emir Kamenica and the rest of my committee, Elliot Lipnowski, Doron Ravid and Lars Stole, for guiding me throughout my PhD journey. I am also grateful to the members of the University of Chicago economics community for providing a stimulating and supportive environment. I thank my friends Agustín Gutiérrez, Eyo Herstad, Ricardo Quineche, Andy Choi and Christoph Schlom for putting up with me over the many, lengthy and invaluable discussions. I am especially grateful to my partner Nancy P. Lin for the unconditional support and encouragement that kept me going as I worked toward my PhD. Finally, I am forever in debt to my family who has helped me in every way possible throughout my life.

Chapter 1

Audience Design

Consider a politician who is trying to persuade a voter by claiming that her and the voter's interests are aligned.[1] Would the voter find such an argument persuasive? A valid concern for the voter is that the politician could be making the same claim to others whose interests are not aligned with that of the voter's. Such a concern might mean that the voter would not find the politician's argument persuasive. But what if the politician was making the same claim in the presence of another voter with diametrically opposed interests? The politician's claim that her interest is aligned with that of the original voters is then also a statement that her and the other voter's interests are not aligned. Thus, the same statement—that "our interests are aligned"—is more credible in the presence of the other voter because the claim now has stakes for the politician.

How does this intuition extend to the case with many voters that have diverse opinions? While the same intuition suggests that the politician would benefit from speaking publicly in front of all the voters, she may in fact be better off by making

[1]This chapter is based on a paper titled "The Power of Semi-Public Communication" that I wrote for a class in my third year. I am especially indebted to Nancy Stokey and Thomas Winberry for leading the class and providing invaluable advice and suggestions. In addition to the members of my committee, I am also grateful to my classmates, Ben Brooks, Alex Frankel, Mike Gibbs, Canice Prendergast, Phil Reny, and Kai-Hao Yang for their helpful comments.

tailored arguments to differently diverse groups of voters by, for example, inviting different sets of voters to multiple events. In other words, the politician may wish to design her audience by partitioning the audience into strategically formed groups and communicating publicly within each group but privately across groups. I refer to this mode of communication as *semi-public communication*.[2] The goal of the chapter is to formalise this mode of communication and to characterise the sender's optimal semi-public communication in a canonical persuasion game.

Specifically, I study a persuasion game in which a sender wishes to persuade multiple receivers to take an action using cheap talk (Crawford and Sobel, 1982). Players have a common prior belief about a payoff-relevant state and the sender wants to maximise the number of receivers taking one of two actions. In contrast to existing literature, the sender communicates with the receivers by first choosing a partition of receivers, and then choosing a state-continent message to send to each group of the partition. Semi-public communication allows the sender to trade off her desire to tailor communication specifically to individuals and her desire to gain credibility by ensuring diversity of opinions in the group.

I first show it is always optimal for the sender to separate the audience into at least two groups: one group consisting of those that do not need persuading (i.e., receivers that take the sender-preferred action without any information) and the other consisting of those that do need persuading (i.e., receivers that do not take the sender-preferred action without any information). Since the sender can always remain silent to any group of receivers, the sender's problem of finding optimal communication is nontrivial only for the latter group of receivers consisting of those that do need persuading. In doing

[2]Both public and private communication are special cases of semi-public communication in which the partition of the audience is either the entire audience or the union of singleton sets of individuals in the audience. In what follows, I use the term semi-public communication to mean communication that differs from public or private communication.

so, I also establish that the sender's problem is equivalent to *Coalition Structure Generation (CSG) problems* that are known to be computationally difficult to solve.

To make progress, I specialise the receivers' preferences in two ways. First, I consider receivers who are "single-minded," i.e., they take the sender-preferred action if and only if their belief that the state is their preferred one is above a threshold. I show that, when receivers are less sceptical (i.e., receivers' thresholds are low), the sender has more scope to benefit from semi-public communication; however, as receivers become more sceptical, the scope to benefit diminishes and, in the limit, public communication is optimal. I provide a sharp characterisation of a class of optimal partitions when the receivers who care about the same issues are all equally sceptical, and when there are two and three possible states of the world. I also demonstrate that when the number of possible states is strictly greater than three, there exists an algorithm to solve for optimal partitions efficiently if receivers are sufficiently sceptical.

Second, I consider the case in which the receivers' preferences are "spatial," i.e., the state space forms a spectrum and the receivers take the sender-preferred action if and only if the state is expected to be sufficiently close to their preferred end of the spectrum. I show that there exists an optimal partition consisting of pairs and/or singletons of receivers in which pairs consist of receivers who prefer the opposite ends of the spectrum with moderate preferences. Moreover, I show that the pairs exhibit a type of negative assortativity in which, among receivers who are paired, the most extreme of one type is paired with the least extreme of the other type.

These results imply, for example, that there is no need to ensure diversity of opinions in political rallies that are held for the supporters. In other words, there is no need to "preach to the choir." However, the politician can be more persuasive to swing voters by campaigning across multiple events in which the audience consists of groups of voters that care about different sets of issues to differing degrees. The results are also applicable to other contexts. For example, a seller attempting to persuade buyers to purchase a product by sending individual or group emails, or a manager attempting to induce effort from her workers by holding a single meeting or several meetings.

Related literature The rich literature on cheap talk began with Crawford and Sobel (1982) who consider the case with a single informed sender and a single receiver.[3] Farrell and Gibbons (1989) analyse a cheap-talk model with two receivers and shows, *inter alia*, that the sender can prefer to communicate publicly (instead of privately) due to an effect they call *mutual discipline* whereby the presence of one receiver disciplines the communication with the other receiver and vice versa, giving credibility to the sender's communication (as in the example in the introductory paragraph). One can thus think of this chapter as extending the idea of mutual discipline to a setting with more than two receivers, which allows for much richer modes of communication—semi-public communication—and, as I demonstrate in the chapter, much richer sources of the mutual disciplining effect.[4]

In Crawford and Sobel (1982)'s and related cheap-talk games, the sender's preference depends on the state so that the credibility of the sender's communication can arise from the endogenous costliness of messages as in signalling games (Spence, 1973). In this chapter (and also in subsequent chapters), I focus instead on the case in which the sender's preference is state independent to remove the possibility of signalling to generate credibility for the sender.[5] This allows me to focus on semi-public communication as the sole way in which the sender can gain credibility her

[3]See surveys by Sobel (2013); Özdogan (2016); Kamenica (2019); Bergemann and Morris (2019); Forges (2020).

[4]Goltsman and Pavlov (2011) study the two-receiver version of Crawford and Sobel (1982)'s uniform-quadratic model. Battaglini (2002) studies cheap-talk models with multiple senders.

[5]When the sender's preference is state independent, the receivers have complete information about the sender's preference. Hence, some authors describe such a sender as having an extreme bias (Chakraborty and Harbaugh, 2010) or transparent motives (Lipnowski and Ravid, 2020).

communication. A number of authors have shown alternative ways in which the sender with state-independent preferences in cheap talk models can gain credibility. Chakraborty and Harbaugh (2010) show that, when the state is multidimensional, a sender who faces a single receiver can gain credibility by trading off different dimensions of the state. Lipnowski and Ravid (2020) observe that a sender facing a single receiver gains credibility by degrading self-serving information. Schnakenberg (2015; 2017) studies how a sender facing multiple receivers can credibly communicate information to multiple receivers in collective choice settings (e.g., voting) by public cheap talk. Salcedo (2019) considers a similar problem in which a sender, who faces many receivers, cares about persuading only a subset of the receivers.

CSG problems have been studied extensively in the computer science literature (see, for example, a survey by Rahwan et al., 2015). In economics, Sandholm (1999) shows that the determination of winners in combinatorial auctions is a CSG problem. Although, general CSG problems have been shown to be computationally hard to solve, the literature has identified classes of CSG problems that are tractable. Of particular relevance is a class of CSG problems called *coalition skill games* (Ohta et al., 2006; Bachracht et al., 2010) in which players have "skills" and the value of a coalition is given by the union of skills among the players in the coalition. The latter authors give a condition under which coalition skill games can be solved in polynomial time that I rely on in the chapter.

The distinguishing feature of the model is that the sender here is able to communicate in groups. While the literature has compared public versus private communication as well as a combination of private and public communication (e.g., Goltsman and Pavlov, 2011; Arieli and Babichenko, 2019; Mathevet, Perego and Taneva, 2020), the idea that the sender communicates in strategically formed groups is new to the literature.[6]

[6]I briefly comment on the sender's ability to communicate via multiple partitions of the receivers in

The remainder of the chapter is structured as follows. Section 1.1 gives a simple example that explains the intuition for the main results. In section 1.2, I set up the formal model and describe the relevant equilibrium concept and state the sender's problem. Section 1.3 contains the formal statements of the results and section 1.4 provides a discussion. Section 1.5 gives a conclusion.

1.1 A simple example

In this section, I first provide an example that demonstrates how the sender's communication can be credible when there is a diversity of opinions in the audience. I then give an example in which the sender strictly benefits from communicating semi-publicly to the receivers by forming differently diverse group of receivers.

Example 1.1 (Private versus private communication)**.** Suppose that the sender is a politician and that she faces three receivers (i.e., voters), $N = \{1, 2, 3\}$. There are three possible states (i.e., issues), $\Theta = \{\theta_1, \theta_2, \theta_3\}$, that the receivers care about. The sender's type is a state $\theta \in \Theta$ and there is a uniform prior belief $\mu_0 \in \Delta\Theta$ about the sender's type.[7] Each receiver $i \in N$ is a "single-issue voter" and votes for the sender if and only if his belief that the sender's type is θ_i is greater than a threshold $\gamma \in (\frac{1}{2}, 1]$. The threshold γ can be thought of as receivers' scepticism toward the sender or their strength of preferences toward their respective issue. Initially, no receiver would vote for the sender and the sender's objective is to maximise the number of votes by communicating with the receivers via cheap talk.

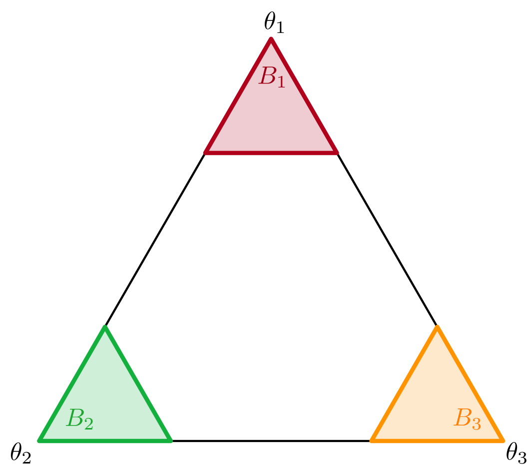
To understand what the sender can achieve via private, public and semi-public communication, let us take a belief-based approach so that the sender's cheap-talk

the discussion. Such an extension would include a combination of private and public communication as a special case.

[7]Given a set X, ΔX denotes the set of probability distributions over the set X.

communication strategy is expressed as a distribution of posterior beliefs that can be induced. Towards this goal, note that any belief about the sender's type can be expressed as a point in a belief simplex as shown in Figure 1.1, where each vertex labelled $\theta \in \Theta$ corresponds to the receiver having a certain belief that the sender's type is θ, denoted $\delta_\theta \in \Delta\Theta$. Each shaded region labelled $B_i := \{\mu \in \Delta\Theta : \mu(\theta_i) \geq \gamma\}$ for $i \in N$ in the figure is the receiver i's *voting region*; i.e., the set of beliefs under which receiver i would vote for the sender. The assumptions of uniform prior belief and $\gamma > \frac{1}{2}$ mean that the prior belief μ_0 must not be contained in any of the voting regions. Moreover, that $\gamma > \frac{1}{2}$ also means that the voting regions do not intersect; i.e., $B_i \cap B_j = \varnothing$ for all distinct $i, j \in N$.

Figure 1.1: Example: Private versus public communication.



It is well-known that the sender can induce any distribution of posterior beliefs that is a mean-preserving spread of the prior belief μ_0 using some communication strategy (Aumann and Maschler, 1968; Kamenica and Gentzkow, 2011). Geometrically, this means that the sender can induce any set of posterior beliefs whose convex hull contains the prior belief μ_0. In addition, the fact that the sender's communication is cheap talk means that, for communication to be credible, the sender must obtain the same number of voters under all posterior beliefs that could be induced. We are now ready to consider what the sender can achieve via private, public and semi-public communication.

With private communication, the sender cannot persuade any receiver to vote for her in equilibrium. If there were a private message sent in equilibrium that can persuade a receiver to vote, then all types of the sender would send the same message independently of the true type. Thus, such a message must be uninformative about the sender's type, and given the prior belief, the receiver would not be persuaded to vote for the sender—contradicting the initial assertion that the message was persuasive. Geometrically, private communication is not beneficial because any mean-preserving spread of μ_0 cannot be entirely contained in B_i for any $i \in N$.

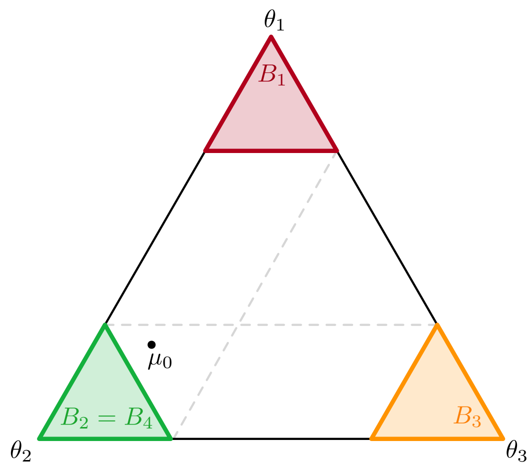
Suppose instead that the sender communicates publicly and that she simply tells the receivers her true type. Then, exactly one receiver votes for the sender independently of whether the sender tells the truth or lies. Hence, the sender does not have the incentive to lie about her type. In other words, an equilibrium exists in which all receivers believe the type that the sender claims to be and exactly one receiver always votes for the sender. Equivalently, we can see that truthful communication can ensure one vote for the sender in equilibrium because the convex hull of posterior beliefs $\{\delta_{\theta_1}, \delta_{\theta_2}, \delta_{\theta_3}\}$ contains μ_0, and exactly one receiver votes for the sender at each possible posterior belief.

In Example 1.1, public communication gives the sender credibility because the presence of other receivers disciplines the sender's communication with each receiver. For example, a type-θ_1 sender's incentive to lie about her type to receiver 2 (or 3) to gain a vote is offset by her incentive to be truthful to receiver 1. In a cheap-talk game with two receivers, Farrell and Gibbons (1989) refers to this phenomenon as *mutually discipline*. The example demonstrates how mutually discipline extends to the case with more than two receivers—what is important is the presence of diversity of opinions in the audience such that the sender's incentive to lie to a receiver with a particular opinion is offset by her incentive to be truthful to other receivers with different opinions.

In Example 1.1, the sender would never strictly prefer semi-public communication.[8] In the next example, I add a receiver and change the assumption regarding the prior belief to demonstrate the case in which the sender strictly benefits from semi-public communication.

Example 1.2 (Semi-public communication)**.** Suppose now that the sender faces an additional receiver, receiver 4, who has an identical preference to receiver 2; i.e., $B_4 = B_2$. Suppose further that the prior belief μ_0 lies in the convex hulls of $\{B_1, B_2\}$ and $\{B_3, B_4\}$ as shown in the figure below.

Figure 1.2: Example: Semi-public communication.



Observe that the sender can only obtain one vote via public communication in Example 1.1. If it were possible to obtain two voters via public communication, then the set of all posterior beliefs that are induced must all lie in the voting regions of receivers 2 and 4; however, the convex hull of such sets cannot contain μ_0. Let us now argue that the sender can obtain two votes via semi-public communication by partitioning N as

[8]To see why, note that any semi-public communication must partition the receivers into a pair and a singleton group. Since private communication is never effective, the sender can only hope to persuade receivers in the pair to vote for her. However, the sender can never persuade both receivers of the pair to vote for her—if this were possible, it must be that the receivers' posterior beliefs (whose convex hull must contain μ_0) must all lie in the intersection of the pair's voting regions. But that would contradict the assumption that the voting regions do not intersect.

$\{\{1,2\},\{3,4\}\}$. By assumption, convex hulls of $B_1 \cup B_2$ and $B_3 \cup B_4$ both contain μ_0. Consequently, there exist group-specific communication strategies that can ensure that one of the receivers in each of the pairs votes for the sender. It follows that the sender can always ensure that exactly two receivers vote for her.

Example 1.2 highlights the trade-off that the sender faces when deciding on how to partition the receivers. On the one hand, a larger group—which must be (weakly) more diverse—raises the possibility for the sender to gain credibility. On the other hand, a larger group ties the sender's hand by forcing her to send the same message—and thus induce the same set of posterior beliefs—to everyone in the group. Communicating in smaller but sufficiently diverse groups allows the sender to benefit from tailoring messages to each group. Thus, any sender's optimal communication trades off these two forces optimally.

The examples also suggest that the sender benefits from semi-public communication only when she can gain credibility via differently diverse group of receivers. In Example 1.2, the prior belief is such that the sender can benefit from communicating to groups in three different ways: by grouping those who care about issue θ_2 with those who care about θ_1 or θ_3, or by ensuring that the group contains those that care about each of the three issues. In contrast, in Example 1.1, semi-public communication cannot improve upon public communication because the sender can only benefit from communicating to groups that contain those that care about each of the three issues. The belief-based approach enables geometric characterisation of these cases in terms of convex hulls of appropriate unions (and intersections) of voting regions.

In the next section, I provide a formal set-up, the relevant equilibrium concept, and the sender's problem to solve for optimal communication.

1.2 A cheap-talk persuasion game with multiple receivers

In this chapter, I focus on a cheap-talk game in which a sender with state-independent preference wishes to persuade as many receivers as possible to take one of two possible actions. Formally, there is a single *Sender* (S) and a finite set $N := \{1, 2, \ldots, n\}$ of *Receivers*. The state, which I simply refer to as the *Sender's type*, is an element $\theta \in \Theta$ from a finite set Θ. Each receiver $i \in N$ can take one of two actions $A := \{0, 1\}$. Receivers have heterogeneous preferences and each Receiver i's payoff depends only on his own action $a_i \in A$ and the Sender's type θ. Let $u_i : A \times \Theta \to \mathbb{R}$ denote Receiver i's payoff. The Sender simply wants to maximise the number of Receivers taking action $a = 1$. Thus, I define the Sender's preference as

$$u_S(a_1, \ldots, a_n) := \sum_{i \in N} a_i.$$

Let $\mu_0 \in \Delta\Theta$ be the common prior belief about the Sender's type. Given a set X, let $\Pi(X)$, denote the set of all partitions of X.

The timing of the game is as follows. First, the Sender chooses a partition $\mathcal{P} \in \Pi(N)$ of receivers.[9] Then, the Sender's type θ is drawn according to μ_0. Having observed her realised type, the Sender chooses a message profile $\{m_P\}_{P \in \mathcal{P}} \in M^{\mathcal{P}}$, where M is the set of possible messages that is sufficiently rich.[10] Each Receiver i who belongs in group $P_i \in \mathcal{P}$ observes (P_i, m_{P_i}) and then takes an action $a_i \in A$.

Public communication corresponds to the case where $\mathcal{P} = \{N\}$ and *private* communication corresponds to the case where $\mathcal{P} = \{\{i\}\}_{i \in N}$. I refer to communication that are neither public nor private as *semi-public* communication.

[9] I assume that the Sender chooses a partition before she learns her type to simplify the definition of equilibrium. As I discuss later, assuming instead that the Sender chooses a partition after she learns her type does not change the results.

[10] For example, it will suffice that $|M| \geq |\Theta|$ (see Lipnowski and Ravid, 2020).

1.2.1 Equilibrium

Given a partition $\mathcal{P} \in \Pi(N)$, let $\sigma : \Theta \to \Delta(M^{\mathcal{P}})$ denote the Sender's messaging strategy; $\alpha_i : M \to \Delta A$ denote Receiver $i \in N$'s action strategy with $\alpha = (\alpha_i)_{i \in N}$; $\mu_P : M \to \Delta\Theta$ denote the belief map for Receivers in group $P \in \mathcal{P}$ with $\mu = (\mu_P)_{P \in \mathcal{P}}$. I define a $\mathcal{P}$-*equilibrium* as a perfect Bayesian equilibrium (PBE) of the game in which the Sender is restricted to sending the same message to receivers that are in the same group.[11] Thus, a tuple (σ, α, μ) is a $\mathcal{P}$*-equilibrium* if it satisfies the following conditions.

(i) For each $P \in \mathcal{P}$, the belief map μ_P, is derived by updating μ_0 via Bayes rule whenever possible; i.e, for all$m_P \in M$,

$$\mu_P\left(\cdot|m_P\right) \sum_{\theta \in \Theta} \sum_{m_{-P} \in M_{-P}} \sigma\left(m_P; m_{-P}|\theta\right) \mu_0\left(\theta\right) = \sum_{m_{-P} \in M_{-P}} \sigma\left(m_P; m_{-P}|\cdot\right) \mu_0\left(\cdot\right), \quad (1.1)$$

where $M_{-P} \coloneqq M^{\mathcal{P} \setminus P}$.

(ii) Each receiver $i \in N$'s action strategy, α_i, is optimal given μ; i.e., for all $i \in N$ and all $m_{P_i} \in M$,

$$\text{supp}\left(\alpha_i\left(\cdot|m_{P_i}\right)\right) \subseteq \arg\max_{a_i \in \{0,1\}} \sum_{\theta \in \Theta} u_i\left(a_i, \theta\right) \mu_{P_i}\left(\theta|m_{P_i}\right), \quad (1.2)$$

where $P_i \in \mathcal{P}$ is the group that i belongs in.

(iii) Sender's messaging strategy, σ, is incentive compatible given α; i.e., for all $\theta \in \Theta$,

$$\text{supp}\left(\sigma\left(\cdot|\theta\right)\right) \in \arg\max_{m \in M^{\mathcal{P}}} \sum_{P \in \mathcal{P}} \sum_{i \in P} \alpha_i\left(1|m_P\right). \quad (1.3)$$

The last condition captures the fact that the Receivers do not inherently trust the Sender

[11] One can interpret the Sender's ability to communicate in groups as a limited form of commitment to communication strategies. In particular, given a partition, the sender is able to commit to communication strategies that sends the same messages to receivers who are in the same group.

and implies that the Receiver only finds the sender's cheap-talk communication strategy σ to be credible if it is in the Sender's best interest to follow σ.

The goal of the chapter is to characterise partitions that maximise the Sender's $\mathcal{P}$-equilibrium payoff among all partitions. i.e., to characterise $\mathcal{P}^*$ and $(\sigma^*, \alpha^*, \mu^*)$ that solves

$$\max_{\mathcal{P}\in\Pi(N),\ (\sigma,\alpha,\mu)} \sum_{\theta\in\Theta} \sum_{m\in M^{\mathcal{P}}} \sum_{P\in\mathcal{P}} \sum_{i\in P} \alpha_i \left(1|m_P\right) \sigma\left(m|\theta\right) \mu_0\left(\theta\right) \tag{1.4}$$
$$\text{s.t.} \qquad (\sigma, \alpha, \mu) \text{ is a } \mathcal{P}\text{-equilibrium.}$$

1.2.2 Simplifying the Sender's problem

In general, the Sender can send messages that are correlated across groups. However, Lemma 1.1 below shows that it is without loss to assume that the messages are independent across groups. Therefore, to compute $W(\mathcal{P})$, one can adopt a "divide-and-conquer" approach and solve for the optimal messaging strategy with respect to each group $P \in \mathcal{P}$ separately.[12]

Lemma 1.1 (Divide and conquer)**.** *Let (σ, α, μ) be a $\mathcal{P}$-equilibrium. Then, the Sender can achieve the same $\mathcal{P}$-equilibrium payoff with a conditionally independent messaging strategy $(\sigma_P : \Theta \to \Delta M)_{P\in\mathcal{P}}$ that is also a $\mathcal{P}$-equilibrium.*

Proof. Fix $\mathcal{P} \in \Pi(N)$ and let (σ, α, μ) be a $\mathcal{P}$-equilibrium. It suffices to show that the sender can induce the same posterior beliefs via a conditionally independent messaging strategy. Define $\sigma'_P : \Theta \to \Delta M$ and $\sigma' : \Theta \to \Delta(M^{\mathcal{P}})$ as

$$\sigma'_P\left(m_P|\theta\right) := \sum_{\tilde{m}_{-P}\in M_{-P}} \sigma\left(m_P; \tilde{m}_{-P}|\theta\right),\ \sigma'\left(\left(m_P\right)_{P\in\mathcal{P}}|\theta\right) := \prod_{P\in\mathcal{P}} \sigma'_P\left(m_P|\theta\right).$$

[12] Arieli and Babichenko (2019) obtains an analogous result (Theorem 4) in the case of Bayesian persuasion; i.e., without requiring incentive compatibility condition for the Sender, (1.3), in the definition of $\mathcal{P}$-equilibrium.

Let $\mu'_P : M \to \Delta\Theta$ be the posterior belief after observing (only) m_P sent according to σ':

$$\mu_P\left(\cdot|m_P\right)\sum_{\tilde{\theta}\in\Theta}\sum_{\tilde{m}_{-P}\in M_{-P}}\sigma'\left(m_P;\tilde{m}_{-P}|\tilde{\theta}\right)\mu_0\left(\tilde{\theta}\right)=\sum_{\tilde{m}_{-P}\in M_{-P}}\sigma'\left(m_P;\tilde{m}_{-P}|\cdot\right)\mu_0\left(\cdot\right).$$

The above is equivalent to the first condition in the definition of $\mathcal{P}$-equilibrium, (1.1), because

$$\sum_{m_{-P}\in M_{-P}}\sigma'\left(m_P;m_{-P}|\theta\right)=\sigma'_P\left(m_P|\theta\right)=\sum_{\tilde{m}_{-P}\in M_{-P}}\sigma\left(m_P;\tilde{m}_{-P}|\theta\right).$$

Thus, $\mu_P = \mu'_P$ as desired. ■

Let us redefine the Sender's problem for each group $P \subseteq N$ using the belief-based approach (Kamenica, 2019; Forges, 2020). Therefore, I replace the Sender's messaging strategy with respect to group P, σ_P, with the distribution of posterior beliefs that is induced by σ_P that I denote as τ_P. It is well-known that any distribution of posterior beliefs that is Bayes plausible can be induced by some messaging strategy; i.e., $\tau_P \in \mathcal{T} := \{\tau \in \Delta\Delta\Theta : \int \mu \mathrm{d}\tau(\mu) = \mu_0\}$. For each $i \in N$, let $V_i : \Delta\Theta \rightrightarrows [0,1]$ denote the best-response correspondence for Receiver i given belief $\mu \in \Delta\Theta$ so that $V_i(\mu)$ denotes the set of probabilities that Receiver i will optimally take action $a_i = 1$ given belief μ. Then, the Sender's problem with respect to any group P is given by

$$w\left(P\right):=\max_{\tau_P\in\mathcal{T}} w_P \quad \text{s.t.} \quad w_P\in\bigcap_{\tilde{\mu}_P\in\text{supp}(\tau_P)}\sum_{i\in P}V_i\left(\tilde{\mu}_P\right),$$

where the constraint reflects the Sender's incentive compatibility constraint that the Sender's payoffs from inducing any posterior beliefs in the support of τ_P must be the same.

Observe that the Sender's problem with respect to P is equivalent to the problem in which she communicates with a single "representative" Receiver against whom the

Sender's value correspondence is given by $V_P := \sum_{i\in P} V_i$.[13] Hence, the single-receiver results from Lipnowski and Ravid (2020) can be applied. To that end, define $v_P : \Delta\Theta \to \mathbb{Z}_+$ for any $P \subseteq N$ as[14]

$$v_P(\mu_P) := \max\left\{ w_P \in \sum_{i\in P} V_i(\mu_P) \right\};$$

and say that a payoff w_P *is P-securable* if there exists $\tau_P \in \mathcal{T}$ such that

$$v_P(\mu_P) \geq w_P \ \forall \mu_P \in \operatorname{supp}(\tau_P).$$

Let $B_i \subseteq \Delta\Theta$ denote the set of beliefs under which Receiver i takes action $a_i = 1$; i.e.,

$$B_i := \{\mu \in \Delta\Theta : 1 \in V_i(\mu)\};$$

and define $B_P := \bigcap_{i\in P} B_i$ for any $P \subseteq N$. The following geometric definition of P-securability, which specialises to the geometric conditions mentioned in the example in section 1.1, is immediate from Lipnowski and Ravid (2020).[15] Given a set $X \subseteq \Delta\Theta$, let co(X) denote the convex hull of the set X.

Lemma 1.2. *$w_P \in \mathbb{Z}_+$ is P-securable if and only if*

$$\mu_0 \in \operatorname{co}\left(\bigcup_{S\subseteq P:|S|=w_P} B_S \right), \tag{1.5}$$

Moreover, μ_0 is at most a convex combination of $|\Theta|$ elements from $(B_S)_{S\subseteq P:|S|=w_P}$.

Proof. See Appendix 1.A.1. ■

[13]Since V_i is a correspondence, $\sum$ denotes the sumset; i.e., $\sum_{i\in P} V_i(\mu_P) \equiv \{\sum_{i\in P} v_i : v_i \in V_i(\mu_P)\ \forall i \in P\}$.

[14]That $v_P \in \mathbb{Z}_+$ follows from the fact that $v_P = \sum_{i\in P} v_{\{i\}}$ and each $v_{\{i\}}$ is an (upper semicontinuous) indicator function.

[15]Schnakenberg (2015) provides a condition for the case with $w_P = 1$ and $P = N$.

By Theorem 1 in Lipnowski and Ravid (2020), the Sender can obtain a payoff $w_P \geq v_P(\mu_0)$ from group $P \subseteq N$ in equilibrium if and only if w_P is P-securable. Consequently, $w(P)$ is the maximum payoff that is P-securable. Thus, any Sender-preferred partition $\mathcal{P}^*$ solves the following problem:

$$\max_{\mathcal{P} \in \Pi(N)} \sum_{P \in \mathcal{P}} w\left(P\right). \tag{1.6}$$

1.3 Optimal communications

Formulated as (1.6), the problem of finding Sender-optimal communication is equivalent to maximising the social surplus in a cooperative game in which players (i.e., Receivers) form coalitions (i.e., groups), and the social surplus associated with coalition $P \subseteq N$ is given by $w(P)$. Such problems are called *coalition structural generation* (CSG) problems, and they have been studied extensively in the computer science literature.[16] The solution is immediate if $w(\cdot)$ is supper-additive or sub-additive.[17] In the former case, optimal partition corresponds to public communication; in the latter case, optimal partition corresponds to private communication. However, as the examples in section 1.1 demonstrate, $w(\cdot)$ is not, in general, super-additive or sub-additive. It is well-known that when $w(\cdot)$ is neither super- nor sub-additive, CSG problems are computationally hard to solve and its solutions difficult to characterise.[18]

In what follows, I first establish a result that enables the Sender to focus attention to the problem of partitioning only receivers who need to be persuaded; i.e., $i \in N$ such that $\mu_0 \notin B_i$. To provide further characterisations of optimal partitions, I then specialise the receivers' preferences in two ways. First, I consider receivers who are "single-minded";

[16]See, for example, Rahwan et al. (2015) for a survey.

[17]$w(\cdot)$ is superadditive (resp. subadditive) if, for any disjoint $P, P' \subseteq N$, $w(P \cup P') \geq w(P) + w(P')$ (resp. $w(P \cup P') \leq w(P) + w(P')$). Superadditivity (subadditivity) is implied by supermodularity (submodularity).

[18]For example, Sandholm et al. (1999) showed that CSG problems are NP-complete with oracle access to function w.

i.e., they take the sender-preferred action if and only if their belief that the state is their preferred one is above a threshold. Second, I consider the case in which the state space forms a spectrum and receivers have "spatial" preferences so that they take the sender-preferred action if and only if the state is expected to be sufficiently close to their preferred end of the spectrum.

1.3.1 Don't preach to the choir

Define N_0 as the set of Receivers who are willing to take action $a = 1$ under the prior belief,

$$N_0 := \{i \in N : 1 \in V_i(\mu_0)\},$$

and $N_1 := N \backslash N_0$. Thus, N_0 is the group that consists of all Receivers who do not require persuading and N_1 is the group consisting of all those who need persuading. For each $z \in \{0, 1\}$, let w_z^* denote the maximum payoff that the Sender can obtain by optimal communicating with set N_z of receivers; i.e.,

$$w_z^* := \max_{\mathcal{P}_z \in \Pi(N_z)} \sum_{P \in \mathcal{P}_z} w(P), \tag{1.7}$$

and let $\mathcal{P}_z^*$ denote a partition that attains w_z^*. The following establishes that it is without loss to split the problem of optimally communicating with N receivers into two independent problems of communicating with N_0 and N_1 receivers. Moreover, since remaining silent is always optimal for the Sender to any groups of Receivers in N_0, any partition of N_0 is optimal.

Proposition 1.1. *$w^* = w_0^* + w_1^* = |N_0| + w_1^*$ and the Sender can attain w^* with any $\mathcal{P}^* = \mathcal{P}_0 \cup \mathcal{P}_1^*$, where $\mathcal{P}_1^*$ solves the Sender's problem (1.7) with respect to N_1 and $\mathcal{P}_0 \in \Pi(N_0)$.*

Proof. See Appendix 1.A.2. ■

The proof involves first showing that the Sender can only persuade at most one more Receiver by adding a Receiver to any group. Hence, if an optimal partition involves a group that contains Receivers from both N_0 and N_1, then removing one Receiver in N_0 from the group cannot lower the Sender's payoff—because the Sender can guarantee a payoff of one from the Receiver that she removes by remaining silent. Iterating this process gives an optimal partition in which no group contains Receivers from both N_0 and N_1. Moreover, by staying silent, the Sender can obtain a payoff of $w_0^* = |N_0|$ from any partition of Receivers in N_0. Together, these imply that there always exists an optimal partition in which the Sender first separates Receivers into two groups, N_0 and N_1, and then considers how to optimally partition the Receivers in N_1.

As already mentioned, characterising partitions of N_1 that attains w_1^* is difficult because $w(\cdot)$ need not be super- or sub-additive. Thus, in the following sections, I specialise the Receiver's preferences to obtain further properties of optimal partitions (of N_1).

1.3.2 Single-minded Receivers

Suppose now that Receivers are "single-minded" so that each Receiver $i \in N$ takes action $a_i = 1$ if and only if he believes with sufficiently high probability that the Sender's type is his preferred one. Thus, any single-minded Receiver can be described by a pair of parameters $(t_i, \gamma_i) \in \Theta \times [0,1]$, where t_i denotes Receiver i's preferred type of the Sender and γ_i denotes Receiver i's threshold belief.[19] As in the example in section 1.1, the threshold γ_i can be thought of as Receiver i's scepticism toward the sender or his strength of preference toward his preferred Sender type. Given two distinct Receivers

[19] An example of Receiver preference that yields such best response is the following:

$$u_i\left(a_i, \theta\right) := a_i\left(\mathbb{1}_{\{\theta=\theta_i\}} - \gamma_i\right),$$

where the Receiver's payoff from choosing $a_i = 0$ is normalised to be zero.

$i, j \in N$ who prefer the same Sender type (i.e., $t_i = t_j$), I say that Receiver i is *more sceptical* than Receiver j if $\gamma_i \geq \gamma_j$.

With single-minded Receivers, the set of posterior beliefs under which Receiver i is a half-space in $\Delta\Theta$:

$$B_i = \{\mu \in \Delta\Theta : \mu(t_i) \geq \gamma_i\}.$$

The following lemma characterises when the Sender can persuade all Receivers in a group to take action $a = 1$.

Lemma 1.3. *Suppose Receivers are singled-minded. For any $P \subseteq N$, $w(P) = |P|$ if and only if $\mu_0 \in B_i$ for all $i \in P$.*

Proof. If $\mu_0 \in B_i$ for all $i \in P$, then the Sender can remain silent and all Receivers would take action so that $w(P) = |P|$. Conversely, suppose that $w(P) = |P|$ which means that $|P|$ is P-securable and, by Lemma 1.2, it must be that $\mu_0 \in \text{co}(\bigcap_{i \in P} B_i) = \bigcap_{i \in P} B_i$, where the equality follows from the convexity of B_i. ■

The above lemma, in particular, implies that the Sender cannot persuade any Receivers in a group consisting of those in N_1 and who all prefer the same Sender type. Therefore, for private communication to be optimal, it must be that the Sender cannot persuade any Receivers in any group of Receivers in N_1 to take action $a = 1$. In other words, there cannot be any scope to persuade Receivers who need persuading. The following proposition gives a condition for private communication to be optimal. Note that whenever private communication is optimal, the Sender is indifferent across all possible partitions of the Receivers.

Proposition 1.2. *Suppose Receivers are single-minded. Private communication is optimal if and only if $\mu_0 \notin \text{co}(\bigcup_{i \in N_1} B_i)$.*

Proof. By the argument above, private communication yields a payoff of $|N_0|$. Recall that the Sender cannot gain by grouping Receivers in N_0 with Receivers in N_1. Hence, it suffices to show that $W(\mathcal{P}) = 0$ for any $\mathcal{P} \in \Pi(N_1)$ if and only if $\mu_0 \notin \mathrm{co}(\bigcup_{i \in N} B_i)$. Suppose the latter condition holds. Then, by Lemma 1.2, a payoff of one is not N_1-securable. Recall that any payoff that is P'-securable is also P-securable for any $P \supseteq P'$—equivalently, any payoff that is not P'-securable is not P-securable for any $P \supseteq P'$. Thus, a payoff of one is not securable with any subset of N_1. Hence, $W(\mathcal{P}) = 0$ for any $\mathcal{P} \in \Pi(N_1)$. To prove the converse, towards a contraction, suppose that $w(P) > 0$ for some $P \subseteq N_1$, then a payoff of one is P-securable and also N_1-securable. Lemma 1.2 gives the desired condition. ■

Whenever there is scope to persuade Receivers in N_1, i.e., $\mu_0 \in \mathrm{co}(\bigcup_{i \in N_1} B_i)$, public or semi-public communication is preferred over private communication. Combining Proposition 1.1 and Lemma 1.3 gives that the Sender's payoff from any partition $\mathcal{P}_1 \in \Pi(N_1)$ is bounded above by $|N| - |\mathcal{P}_1|$. This bound is greatest if public communication is preferred over semi-public communication with respect to Receivers in N_1.[20] For this to be the case, it must be that there is no scope for the Sender to benefit from forming differently diverse groups of Receivers. The following proposition show that Receivers being extremely sceptical is a sufficient condition for the Sender to prefer public communication. Let n_θ denote the number of Receivers who prefer Sender type θ; i.e., $n_\theta := |\{i \in N : t_i = \theta\}|$.

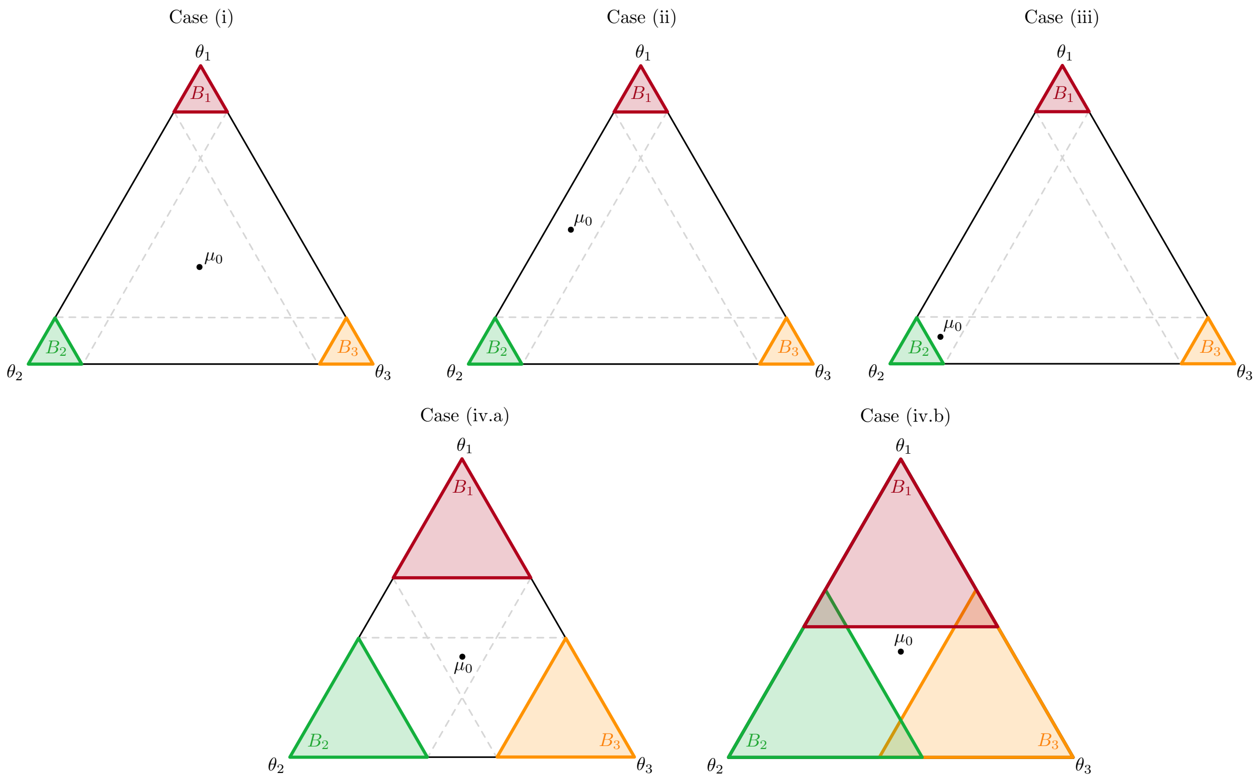
Proposition 1.3. *Suppose Receivers are single-minded and μ_0 is interior. Then, as Receivers become extremely sceptical (i.e., $\gamma_i \to 1$ for all $i \in N$), optimal partitions consist of groups that contain Receivers who prefer every possible Sender type, and $w^* = \min_\theta n_\theta = w(\{N\})$.*

[20] The fact that the Sender strictly prefers public communication with respect to Receivers in N_1 need not mean that public communication with respect to all Receivers (i.e., N) is optimal for the Sender. This is because public communication that is optimal with respect to Receivers in N_1 may result in Receivers in N_0 not always taking action.

Proof. Note that $B_i \to \delta_{t_i}$ as $\gamma_i \to 1$. Thus, for sufficiently large thresholds, $N_0 = \varnothing$ given that μ_0 is interior. That μ_0 is interior implies that $\mu_0(\theta) > 0$ for all $\theta \in \Theta$ such that, for sufficiently large thresholds, μ_0 can only be expressed as a convex combination of B_is whose union contain $\{\delta_\theta\}_{\theta \in \Theta}$. Since $B_i \cap B_j = \varnothing$ for any district $i, j \in N$ with $t_i \neq t_j$ for sufficiently large γ_i and γ_j, $B_P \neq \varnothing$ if and only if $P \subseteq N$ contains Receivers who prefer the same Sender type. Together with Lemma 1.2, it follows that a group can secure a positive payoff only if it contains Receivers who prefer every possible Sender type, and there can be at most $\min_\theta n_\theta$ many such groups in any partition. Let $\mathcal{P}^*$ be an optimal partition such that $|\mathcal{P}^*| > 1$. Previous argument implies that $\{t_i : i \in P\} = \Theta$ for every $P \in \mathcal{P}^*$ such that $w(P) > 0$. Moreover, combining any $P, P' \in \mathcal{P}^*$ such that $w(P), w(P') > 0$ would $P \cup P'$-secure $w(P) + w(P')$ because there are $w(P) + w(P')$-many Receivers who prefer every possible types in $P \cup P'$. Hence, $w^* = w(\{N\})$ by induction. ■

Consider the case in which there are three possible states, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, and the Receivers are all equally sceptical; i.e., $\gamma_i = \gamma$ for all $i \in N$. Figure 1.3 below presents the possible ways in which differently diverse groups of Receivers can secure a strictly positive payoff under these assumptions. Case (i) corresponds to the case of extreme scepticism (as in Proposition 1.3), where γ is sufficiently high so that any group that secures a positive payoff must contain Receivers who prefer all possible Sender types. In particular, no pairs of Receivers can secure a positive payoff. In case (ii), a pair consisting of Receivers who prefer Sender types θ_1 and θ_2 is the only type of pairs that can secure a payoff. Case (iii) corresponds to Example 1.2, where pairs of Receivers who prefer Sender types θ_1 and θ_2, as well as pairs consisting those who prefer Senders types θ_2 and θ_3 can secure a positive payoff. Case (iv.a) corresponds to the case in which any pairs of Receivers who prefer distinct Sender types can secure a positive payoff. Note that in all cases, a trio of Receivers who prefer all possible Sender types can secure a payoff of one. However, in case (iv.b), such a trio can secure a payoff of two.

Figure 1.3: Types of groups that can secure positive payoffs with three states.



Restricting attention to Receivers in N_1, the assumption on prior belief μ_0 determines the set of differently diverse groups that can secure a positive payoff. However, note that cases (iv.a) and (iv.b) can only arise if $\gamma \leq \frac{2}{3}$ and $\gamma \leq \frac{1}{2}$, respectively. These conditions on thresholds mean that the Sender has greater scope to benefit from communication when Receivers are less sceptical. It turns out that when the number of states is three, the Sender does not benefit from forming groups that contain more than three Receivers.[21] It follows that in case (i) through (iv.a), any partition that maximises the number of pairs of Receivers that can secure a payoff of one is optimal. In case (iv.b), an optimal partition first maximises the number of trios of Receivers (that secures a payoff of two), before maximising the number of pairs that can secure a payoff of one. I characterise the Sender optimal partition in this manner while allowing the thresholds to differ across Receivers

[21] In Appendix 1.A.4, I give an example with $|\Theta| = 4$ and $|N_1| = 5$ (and $|N_0| = 0$) in which the Sender strictly benefits from public communication.

who prefer different Sender types. In doing so, identify the type of any group of Receivers by the set of Sender types preferred by the Receivers in the group. Let $n_{1,\theta} := |\{i \in N_1 : t_i = \theta\}|$ for each $\theta \in \Theta$.

Proposition 1.4. *Suppose Receivers are single-minded, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and Receivers in N_1 who prefer the same Sender type have a common threshold; i.e., for each $\theta \in \Theta$, $\gamma_i = \gamma_\theta$ for all $i \in N_1$ such that $t_i = \theta$.*

(i) Suppose no types of pairs can secure a payoff of one; i.e., $\mu_0 \notin \bigcup_{\theta,\theta' \in \Theta} \mathrm{co}(B_\theta, B_{\theta'})$. Then, $w_1^* = \min_{\theta \in \Theta}\ n_{1,\theta}$ attainable by forming trios of Receivers who prefer all possible Sender types.

(ii) Suppose only one type of pairs can secure a payoff of one; i.e., there exists a unique pair of $\theta, \theta' \in \Theta$ such that $\mu_0 \in \mathrm{co}(B_\theta, B_{\theta'})$. Then, $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'}\}$ attainable by forming pairs consisting of Receivers who prefer Sender types θ and θ'.

(iii) Suppose there are two possible types of pairs that can secure a payoff of one; i.e., there exists unique $\theta \in \Theta$ such that $\mu_0 \in \mathrm{co}(B_\theta, B_{\theta'})$ and $\mu_0 \in \mathrm{co}(B_\theta, B_{\theta''})$ for distinct $\theta', \theta'' \in \Theta \backslash \{\theta\}$. Then, $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'} + n_{1,\theta''}\}$ attainable by forming as many pairs consisting of Receivers who prefer either Sender types θ and θ' or Sender types θ and θ''.

(iv) Suppose any types of pairs can secure a payoff of one; i.e., $\mu_0 \in \mathrm{co}(B_\theta, B_{\theta'})$ for all distinct $\theta, \theta' \in \Theta$. Let $n_{1,\theta} \geq n_{1,\theta'} \geq n_{1,\theta''}$. Suppose further that a trio of Receivers who prefer all possible Sender types...

  (a) cannot secure a payoff of two; i.e., $\sum_{\theta \in \Theta} \gamma_\theta \leq 2$ and $\gamma_\theta + \gamma_{\theta'} > 1$ for all distinct $\theta, \theta' \in \Theta$. Then, $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'} + n_{1,\theta''}\}$ attainable by forming as many pairs consisting of Receivers who prefer either Sender types θ and θ', or Sender types θ and θ''.

(b) *can secure a payoff of two; i.e.,* $\gamma_\theta + \gamma_{\theta'} \leq 1$ *for all distinct* $\theta, \theta' \in \Theta$. *Then,* $w_1^* = 2n_{1,\theta''} + \min\{n_{1,\theta} - n_{1,\theta''}, n_{1,\theta'} - n_{1,\theta''}\}$ *attainable by forming* $n_{1,\theta''}$ *trios of Receivers who prefer all possible Sender types, and* $\min\{n_{1,\theta} - n_{1,\theta''}, n_{1,\theta'} - n_{1,\theta''}\}$ *pairs of Receivers consisting of either Sender types* θ *and* θ'' *or Sender types* θ' *and* θ''.

The proposition highlights the ways in which the Sender benefits from semi-public communication. In cases (i) and (ii), semi-public communication is unnecessary—the Sender can attain the optimal payoff by publicly communicating within N_1. In cases (iii) and (iv.a), semi-public communication is beneficial to the Sender because it allows her to adopt two different communication strategies across two types of pairs. Finally, in case (iv.b), semi-public communication is beneficial because it allows the Sender to adopt three different communication strategies over two types of pairs and a trio.

The proposition above also implies that when the state is binary (i.e., $|\Theta| = 2$), $w_1^* = \min_{\theta \in \Theta} n_{1,\theta}$. The coarsest partition of N_1 that achieves this optimal payoff corresponds to public communication (within N_1) and the finest partitions of N_1 that achieve this optimal payoff has as many pairs of Receivers as possible of those who prefer "opposite" states.

Recall that the Sender's problem of finding an optimal partition is computationally difficult when $w(\cdot)$ is neither super- nor sub-additive. Thus, the hypothesis of Proposition 1.4 represents one set of conditions that allows for explicit characterisations of optimal partition by restricting the set of types of groups that can secure positive payoffs. If one is instead content with finding optimal partition computationally, it is possible to relax the restrictions on the state space while focusing on Receivers who are sufficiently sceptical.

Proposition 1.5. *Suppose Receivers are singled-minded, Receivers who prefer the same Sender type have a common threshold, $\gamma_\theta + \gamma_{\theta'} > 1$ for all distinct $\theta, \theta' \in \Theta$, and Θ is finite. Then, there exists an optimal partition in which each group contains at most one Receiver who prefers each possible Sender type. Moreover, the sender's problem is solvable in polynomial time.*

The lemma above implies that it is without loss to restrict attention to groups of Receivers of size no more than $|\Theta|$ that contains at most one of each type of Receiver and can secure at most a payoff of one. The Sender's problem is then a coalition skill game (Ohta et al., 2006) and can be shown to satisfy a condition for the problem to be solvable in polynomial time (Bachracht et al., 2010). More generally, if each Receiver's threshold can differ, then the types of groups must be identified not only by each Receiver's preferred sender type but also with the Receiver's threshold. With arbitrary thresholds, the set of possible types of groups becomes large making the optimal partitions difficult to characterise.

1.3.3 Receivers with spatial preferences

Suppose now that the Sender's types are ordered and so forms a spectrum, and that the Receiver has a spatial preference over the spectrum. Specifically, suppose that the Sender's type, θ, lies on the interval $\Theta = [\ell, r] \subseteq \mathbb{R}$; ℓ representing the "left" end of the spectrum and r representing the "right" end of the spectrum. Assume, for simplicity, that μ_0 is atomless.[22] There can be two types of Receivers: type-ℓ Receivers who take action $a = 1$ if they believe θ to be sufficiently close to ℓ and type-r Receivers who take action $a = 1$ if they believe that θ is sufficiently close to r. Let $t_i \in \{\ell, r\}$ denote the type of Receiver i and $\gamma_i \in \Theta$ denote the strength of i's preference. Given Receivers i and j of the same type (i.e., $t_i = t_j$), say that Receiver i is *more moderate* (equivalently, *less extreme*)

[22]The results would not change even if the prior belief contained atoms.

than Receiver j if $|t_i - \gamma_i| \geq |t_j - \gamma_j|$.[23] Define

$$\beta_i(E) = \begin{cases} \{1\} & \text{if } |t_i - \gamma_i| > |t_i - E| \\ [0,1] & \text{if } |t_i - \gamma_i| = |t_i - E| \\ \{0\} & \text{if } |t_i - \gamma_i| < |t_i - E| \end{cases},$$

$\Upsilon_P := \sum_{i \in P} \beta_i$, and $v_P := \max \Upsilon_P$ so that $\beta_i(\mathbb{E}_\mu[\theta]) \equiv V_i(\mu)$, $\Upsilon_P(\mathbb{E}_\mu[\theta]) \equiv V_P(\mu)$, and $v_P(\mathbb{E}_\mu[\theta]) \equiv v_P(\mu)$. Observe that, when $N_0 = \varnothing$, v_P is quasiconvex. Therefore, by Claim 5 in Lipnowski and Ravid (2020) for any $P \subseteq N_1$, a payoff of w_P is P-securable if and only if there exists $k \in \Theta$ such that

$$v_P\left(\overline{\theta}(k)\right) \geq w_P \text{ and } v_P\left(\underline{\theta}(k)\right) \geq w_P.$$

where $\overline{\theta}(k) := \mathbb{E}_{\mu_0}[\theta|\theta \geq k]$ and $\underline{\theta}(k) := \mathbb{E}_{\mu_0}[\theta|\theta \leq k]$. That is, w_P is securable if and only if w_P is P-securable with a *k-cutoff* τ_P, *denoted* τ_P^k, in which the Sender tells Receivers in P whether θ is above or below the cutoff k.

As was the case with single-minded Receivers, the Sender can secure a positive payoff from a group only if the group is diverse, which, in this case, means that the group must consist of both ℓ- and r-type Receivers. Moreover, the following lemma shows that if a pair consisting of type-ℓ and type-r Receivers secures a payoff of one, then the Sender can secure a payoff of one from a pair consisting of more moderate Receivers.

---

[23]For example, we may let the payoff for each Receiver $i \in N$ of type t_i and threshold γ_i from taking action a_i when the Sender's type is θ to be

$$u_i(a_i, \theta) = \begin{cases} a_i(\gamma_i - \theta) & \text{if } t_i = \ell \\ a_i(\theta - \gamma_i) & \text{if } t_i = r \end{cases},$$

where Receiver i's payoff from $a = 0$ is normalised to be zero.

Lemma 1.4. *Suppose $w(\{\ell, r\}) = 1$ where $\ell, r \in N_1$ and $t_\ell = \ell$ and $t_r = r$. Then, $w(\{\ell', r'\}) = 1$ for any $\gamma_{\ell'} \geq \gamma_\ell$ and $\gamma_{r'} \leq \gamma_r$ (where $t_{\ell'} = \ell$ and $t_{r'} = r$).*

Proof. Since $w(\{\ell, r\}) = 1$, there exists $k \in \Theta$ such that $\gamma_\ell \geq \underline{\theta}(k)$ and $\gamma_r \leq \overline{\theta}(k)$. Thus, any type-$\ell$ Receiver ℓ' with $\gamma_{\ell'} \geq \gamma_\ell$ will also take action $a = 1$ when $\underline{\theta}(k)$ is realised and any type-r Receiver r' with $\gamma_{r'} \leq \gamma_r$ will take action $a = 1$ when $\overline{\theta}(k)$ is realised. Hence, the same τ_P^k ensures that $v_{\{\ell', r'\}}(\overline{\theta}(k)), v_{\{\ell', r'\}}(\underline{\theta}(k)) \geq 1$; i.e., $w(\{\ell', r'\}) = 1$. ■

Since B_i, the set of beliefs under which Receiver i takes action $a = 1$, remains convex, Proposition 1.2 remains true. The condition for when there is no scope for the Sender to benefit from mutual discipline can be written as follows using Lemma 1.4.

Corollary 1.1. *Private communication is optimal if and only if $w(\{\overline{\ell}, \underline{r}\}) = w(\{\underline{\ell}, \overline{r}\}) = 0$, where $\overline{t} \in N_1$ (resp. $\underline{t} \in N_1$) denote the most extreme (resp. least extreme) type-t Receiver in N.*

In this environment, semi-public communication allows the Sender to adopt different cutoffs $(k_P)_{P \in \mathcal{P}}$ for each group; in contrast, under public communication, the Sender is restricted to using the same cut-off for all Receivers.

The following establishes an existence of optimal partitions in which moderate Receivers are paired in a negatively assortative manner.

Proposition 1.6. *There exists an optimal partition with the following properties.*

(i) Every Receiver is either paired with another Receiver or unpaired;

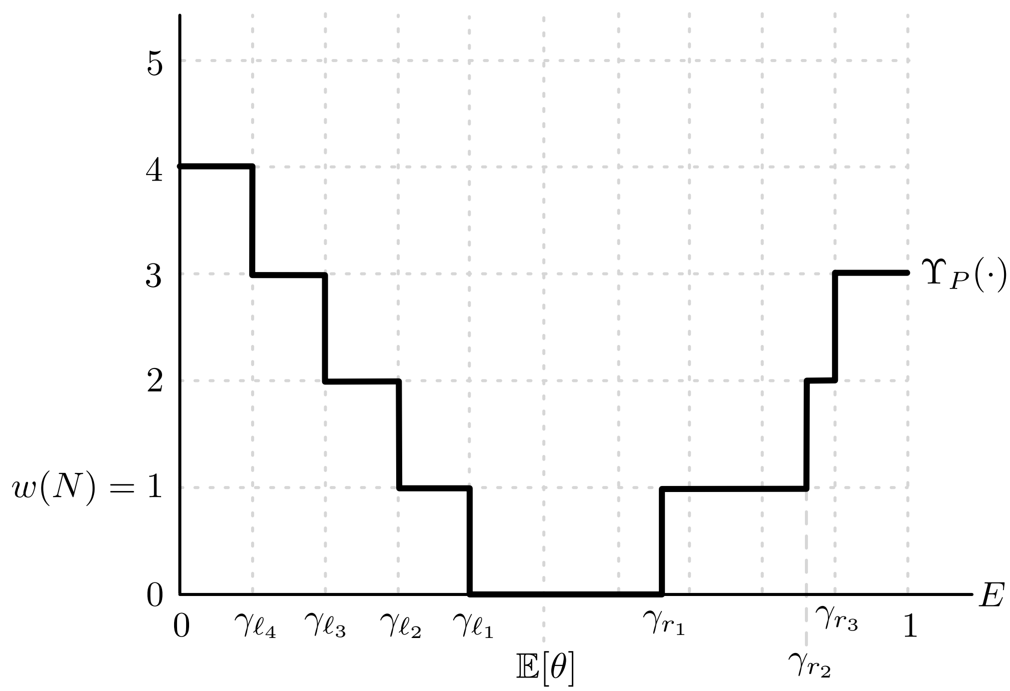
(ii) Unpaired Receivers are more extreme than any paired Receivers;

(iii) Receivers are paired negatively assortatively; i.e., among Receivers in pairs, the most extreme type-t Receiver is paired with the least extreme type-$t' \neq t$ Receiver, and so on.

The proof proceeds by showing that any optimal partition can be further partitioned to satisfy (i). I then show that property (iii) can be satisfied among Receivers who are paired using Lemma 1.4. The same lemma then can then be used to show any unpaired Receivers who are less extreme than some paired Receivers can be swapped around to satisfy (ii). The following example demonstrates a constructive process that leads to an optimal partition.

Example 1.3. Suppose $\Theta = [0, 1]$ and μ_0 is a uniform distribution over Θ. Let $N = N_1 = \{\ell_1, \ell_2, \ell_3, \ell_4, r_1, r_2, r_3\}$. Suppose the thresholds are such that the Sender's value function, v_N, is given by the figure below. Observe that $w(N) = 1$ can be secured by a cut-off τ_P^k with the cutoff, for example, at $k = \mathbb{E}[\theta]$.

Figure 1.4: Example: Receivers with spatial preferences.



Let $P_0 = N$ and $\mathcal{P} = \varnothing$.

▷ Step 1: The first substep is to remove the most extreme Receiver in P_0 who cannot be paired to secure a positive payoff. Observe that ℓ_4 cannot be paired with any r-type to secure a positive payoff since the maximum k such that $\underline{\theta}(k) \geq \gamma_{\ell_4}$ is $k = \gamma_{\ell_3}$, which implies $\overline{\theta}(\gamma_{\ell_3}) < \gamma_{r_i}$ for all $i \in \{1,2,3\}$. In contrast, r_4 can be paired with ℓ_1 to secure a payoff of one (with cutoff at γ_{h_3}). $\tilde{P}_1 = N\backslash\{\ell_4\}$ and $\mathcal{P} = \{\{\ell_4\}\}$. The next substep is to pair the most extreme type-ℓ Receiver with the least extreme type-r to secure a payoff of one from $\tilde{P}_1$; i.e., $\{\ell_3, r_1\}$. Then, set $P_1 = \tilde{P}_1\backslash\{\ell_3, r_1\}$ and $\mathcal{P} = \{\{\ell_4\}, \{\ell_3, r_1\}\}$.

▷ Step 2: We apply the first substep to P_1. Observe that ℓ_2 cannot be paired with remaining type-r Receivers in P_1 to secure a payoff of one. Thus, $\tilde{P}_2 = P_1\backslash\{\ell_2\}$ and $\mathcal{P} = \{\{\ell_4\}, \{\ell_3, r_1\}, \{\ell_2\}\}$. From $\tilde{P}_2$, the most extreme type-ℓ Receiver is ℓ_1 who can be paired with either r_2 or r_3 to secure a payoff of one. Since r_2 is the least extreme between the two, $P_2 = \tilde{P}_2\backslash\{\ell_1, r_2\}$ and $\mathcal{P} = \{\{\ell_4\}, \{\ell_3, r_1\}, \{\ell_2\}, \{\ell_1, r_2\}\}$.

▷ Step 3: No type-ℓ Receiver is left in P_2 and the algorithm terminates while setting

$$\mathcal{P} = \{\{\ell_4\}, \{\ell_3, r_1\}, \{\ell_2\}, \{\ell_1, r_2\}, \{r_3\}\}.$$

The result of the algorithm yields a payoff of 2 for the Sender, which is optimal in this case. However, observe that ℓ_2 is unpaired despite being less extreme than ℓ_3 who is paired. But by lemma 1.4, ℓ_3 and ℓ_2 can be switched without affecting the payoff. Thus, the following is an optimal partition that satisfies all three properties of the proposition.

$$\mathcal{P}^* = \{\{\ell_4\}, \{\ell_3\}, \{\ell_2, r_1\}, \{\ell_1, r_2\}, \{r_3\}\}.$$

1.4 Discussions

The desire for the Sender to communicate semi-publicly arises from the fact that the Receivers do not inherently trust the Sender to communicate "truthfully". This lack of trust, formally captured as the Sender's inability to commit a communication strategy, means that the sender's communication is not credible unless it satisfies the incentive compatibility constraint, (1.3), in the definition of equilibrium.[24] Semi-public communication benefits the Sender by giving more ways to satisfy the incentive compatibility constraint. If, instead, the Receivers trusted the Sender to communicate truthfully as in Bayesian persuasion (Kamenica and Gentzkow, 2011), the Sender's communication is credible even if it is not incentive compatible for the Sender. Consequently, there is no gain from communicating in groups and so private communication is always optimal for the Sender if the Receivers do not trust the Sender.

An implicit assumption in the model is that the Sender chooses the partition of Receivers prior to observing her type. This assumption ensures that the Sender's choice of partition does not convey any information about the state. One may also consider the case in which the Sender chooses the partition *after* observing her type, in which case it is possible for the Sender to signal her type via the choice of the partition. Allowing for such signalling does not affect the results. More concretely, in such a signalling version of the game, there always exists a pooling equilibrium in which the sender's payoff and the optimal partition (on the equilibrium path) corresponds to the equilibrium payoff and partition in the original game.[25]

Recall that communication is costless in the model. In particular, not only messages are assumed not to be payoff-relevant, there are also no (marginal) costs associated with forming groups. While this is realistic in some situations (e.g., communication via emails),

[24] The Sender here first can claim to follow the strategy σ, but after observing the realised θ as well as a message m drawn from $\sigma(\theta)$, she can lie to the Receiver by misreporting m. In contrast, if the Sender can commit to a communication strategy σ, she always communicates the m drawn from $\sigma(\theta)$ truthfully to the Receiver.

[25] To see this, suppose toward a contradiction that there is a type $\theta \in \Theta$ that selects a different partition from all other types in equilibrium. Then, it must be that that type-θ Sender gets at least as high payoff as others. If type-θ Sender get a strictly higher payoff, then other types would deviate and choose the same partition as type-θ. Hence, if there was a separating/hybrid equilibrium in the signalling version, there must also exists a pooling equilibrium in which type-θ Sender selects the same partition as all other types. Moreover, any deviation from the on-path partition can be punished by an off-path belief assumes any communication by the Sender to be uninformative.

in other situation such as the sender communicating with receivers by holding meetings, it may be more plausible to include costs that depend on the number of groups in each partition. Costs of forming groups would be an additional force that pushes the sender to prefer public communication. Given that there are often many partitions that gives rise to the same equilibrium payoff for the Sender, costs of forming groups represents a way to justify the sender selecting the least fine partition among optimal partitions.[26]

While I have focused on the case in which the sender selects a single partition of the receivers, it is also possible that the sender selects multiple partitions of the receivers and a messaging strategy for each partition. Such a generalisation allows the sender, for example, to send both public and private messages to the receivers (Goltsman and Pavlov, 2011; Arieli and Babichenko, 2019; Mathevet, Perego and Taneva, 2020). While a characterisation of optimal multiple partitions of the audience is left for future research, in appendix 1.A.7, I provide an example in which the ability for the Sender to adopt two nontrivial partitions can strictly benefit the Sender.

1.5 Conclusion

In this chapter, I explore how a sender who lacks receivers' trust can benefit from a mode of communication that I call *semi-public* communication in which the sender partitions the receivers into groups, and communicates publicly within each group but privately across groups. The benefit from communicating in groups arises from the fact that cheap-talk communication can be credible in front of an audience with diverse opinions because the sender's incentive to lie to some members of the audience can be offset by her incentive to be truthful to the other members of the audience. Semi-public communication enables the sender to communicate more effectively than private or

[26]On the other hand, if the Sender has doubts about the Receivers' levels of scepticism, she may prefer to adopt the finest partition among optimal partitions.

public communication because it allows the sender to form differently diverse groups that trades off her desire to tailor communication to individuals with her desire to obtain credibility by ensuring appropriate diversity in her audience.

In a canonical game of persuasion with multiple receivers, I show that it is optimal for the sender to separate her audience into two groups based on whether the sender needs to persuade the receiver in the first place. The sender can further benefit by partitioning the group consisting of those that need persuading and I provide various characterisations of optimal partitions under different assumptions on the receivers' preferences. A practical implication of my results is that, while there is no need to ensure diversity of opinions in political rallies that are held for the supporters, a politician can be more persuasive to swing voters by campaigning across multiple events in which the audience consists of groups of swing voters that care about different sets of issues.

1.A Appendix

1.A.1 Proof of Lemma 1.2

Lemma 1.2. *$w_P \in \mathbb{Z}$ is P-securable if and only if*

$$\mu_0 \in \text{co}\left(\bigcup_{S \subseteq P : |S| = w_P} B_S \right),$$

Moreover, μ_0 is at most a convex combination of $|\Theta|$ elements from $(B_S)_{S \subseteq P : |S| = w_P}$.

Proof. Fix $P \subseteq N$. Suppose (1.5) holds. Then, μ_0 is a convex combination of $(\mu_S)_S \subseteq \Delta\Theta$, where for each μ_S, all $i \in S$ takes action $a_i = 1$; i.e., $v_P(\mu_S) = |S| = w_P$ for each S. Let $(\tau_S)_S \subseteq [0, 1]$ be the corresponding coefficients so that $\mu_0 = \sum_S \tau_S \mu_S$. Then, $\tau = (\tau_S)_{S \in \mathcal{S}} \in \mathcal{T}$. Thus, w_P is P-securable. Conversely, suppose that w_P is P-securable and let $\tau \in \mathcal{T}$ be such that it secures w_P. Fix any $\mu \in \text{supp}(\tau)$. Since $v_P(\mu) = \sum_{i \in P} v_{\{i\}}(\mu) = \sum_{i \in P} \mathbb{1}_{\{\mu \in B_i\}} \geq w_P$, there exists $S \subseteq P$ such that $|S| = w_P$ and $\mu \in B_S$. Since $\tau \in \mathcal{T}$, it follows that μ_0 is a convex combination of elements in $\text{supp}(\tau)$ and so (1.5) follows.

Finally, Recall $B_i \subseteq \Delta\Theta \subseteq \mathbb{R}_+^{|\Theta|}$. Define $\text{proj} : \mathbb{R}^{|\Theta|} \to \mathbb{R}^{|\Theta|-1}$ as a projection of $(\mu_\theta)_{\theta \in \Theta}$ to the first $|\Theta| - 1$ coordinates. Since $\sum_{\theta \in \Theta} \mu_\theta = 1$ for any $\mu = (\mu_\theta)_{\theta \in \Theta} \in \Delta\Theta$, $\text{proj}(\mu)$ uniquely identifies an element in $\Delta\Theta$. Thus, (1.5) is equivalent to

$$\text{proj}(\mu_0) \in \text{co}\left(\bigcup_{S \subseteq P : |S| = w_P} \bigcap_{i \in S} \text{proj}(B_i) \right).$$

Since $\text{proj}(B_i) \subseteq \mathbb{R}^{|\Theta|-1}$, by Carathéodory's theorem, above implies that there exists $\{\text{proj}(\mu_r) \in \bigcap_{i \in S_r} \text{proj}(B_i)\}_{r=1}^{|\Theta|}$, where $\{S_r\}_{r=1}^{|\Theta|} \subseteq \{S \subseteq P : |S| = w_P\}$, and

$\{\tau_r\}_{r=1}^{|\Theta|} \subseteq [0,1]$ with $\sum_{r=1}^{|\Theta|} \tau_r = 1$ such that

$$\text{proj}\left(\mu_0\right) = \sum_{r=1}^{|\Theta|} \tau_r \text{proj}\left(\mu_r\right),$$

which also implies that μ_0 is a convex combination of $\{\mu_r \in \bigcap_{i \in S_r} B_i\}_{r=1}^{|\Theta|}$. ∎

1.A.2 Proof of Proposition 1.1

To prove the proposition, I first bound the benefit that the Sender can gain from adding a Receiver to the group.

Lemma 1.5. *For any $i \in N$ and $P \subseteq N$, $w(P \cup \{i\}) - w(P) \leq 1$.*

Proof. Observe first that the result is trivial if $i \in P$ and so fix $P \subset N$ and $i \in N \backslash P$. Note that any $\tau \in \mathcal{T}$ that $P \cup \{i\}$-secures a value s necessarily P-secures $s - 1$ since $v_{P \cup \{i\}}(\mu) = v_{\{i\}}(\mu) + \sum_{j \in P} v_{\{j\}}(\mu)$ and $v_{\{i\}} \in [0,1]$. Hence, P secures $w(P \cup \{i\}) - 1$ and the result follows from the definition of $w(P)$ as the maximum P-securable payoff. ∎

Proposition 1.1. *$w^* = w_0^* + w_1^* = |N_0| + w_1^*$ and the Sender can attain w^* with any $\mathcal{P}^* = \mathcal{P}_0 \cup \mathcal{P}_1^*$, where $\mathcal{P}_1^*$ solves the Sender's problem (1.7) with respect to N_1 and $\mathcal{P}_0 \in \Pi(N_0)$.*

Proof of Proposition 1.1. Let $\mathcal{P}_1^*$ be the maximiser that achieves w_1^* and $\mathcal{P}^*$ be the maximiser that achieves w^*. As already noted, the Sender's maximum payoff from Receivers in N_0 is $|N_0|$, and the payoff is achievable via an uninformative messaging strategy. For any $\mathcal{P}_1 \in \Pi(N_1)$, $N_0 \cup \mathcal{P}_1$ is a partition of N. Thus, by definition of w^*, we must have

$$w^* \geq |N_0| + w_1^*.$$

Take any $P \in \mathcal{P}^*$ such that $P \cap N_0 \neq \varnothing$. By Lemma 1.5,

$$w(P) - w(P \backslash \{i\}) \leq 1 \; \forall i \in P \cap N_0.$$

The Sender can obtain a payoff of one from Receivers in $P \cap N_0$ via uninformative messaging strategy; i.e., $w(\{i\}) = 1$ for all $i \in P \cap N_0$. Hence,

$$\begin{aligned} & w(P) \leq w(P \backslash \{i\}) + w(\{i\}) \;\; \forall i \in P \cap N_0 \\ \Leftrightarrow\; & w(P) \leq w(P \backslash \{P \cap N_0\}) + \sum_{i \in P \cap N_0} w(\{i\}) \\ & = w(P \backslash \{P \cap N_0\}) + |P \cap N_0|. \end{aligned}$$

Since this holds for all $P \in \mathcal{P}^*$,

$$\begin{aligned} w^* = \sum_{P \in \mathcal{P}^*} w(P) & \leq \sum_{P \in \mathcal{P}^*} \left[w(P \backslash \{P \cap N_0\}) + |P \cap N_0| \right] \\ & \leq |N_0| + w_1^*, \end{aligned}$$

where the last line follows from the fact that $\bigcup_{P \in \mathcal{P}^*} P \backslash \{P \cap N_0\}$ is a partition of N_1. ■

1.A.3 Proof of Proposition 1.4

Proposition 1.4. *Suppose Receivers are single-minded, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and Receivers in N_1 who prefer the same Sender type have a common threshold; i.e., for each $\theta \in \Theta$, $\gamma_i = \gamma_\theta$ for all $i \in N_1$ such that $t_i = \theta$.*

(i) Suppose no types of pairs can secure a payoff of one; i.e., $\mu_0 \notin \bigcup_{\theta, \theta' \in \Theta} \text{co}(B_\theta, B_{\theta'})$. Then, $w_1^* = \min_{\theta \in \Theta} n_{1,\theta}$ attainable by forming trios of Receivers who prefer all possible Sender types.

(ii) Suppose only one type of pairs can secure a payoff of one; i.e., there exists a unique pair of $\theta, \theta' \in \Theta$ *such that* $\mu_0 \in \text{co}(B_\theta, B_{\theta'})$. *Then,* $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'}\}$ *attainable by forming pairs consisting of Receivers who prefer Sender types* θ *and* θ'.

(iii) Suppose there are two possible types of pairs that can secure a payoff of one; i.e., there exists unique $\theta \in \Theta$ *such that* $\mu_0 \in \text{co}(B_\theta, B_{\theta'})$ *and* $\mu_0 \in \text{co}(B_\theta, B_{\theta''})$ *for distinct* $\theta', \theta'' \in \Theta \backslash \{\theta\}$. *Then,* $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'} + n_{1,\theta''}\}$ *attainable by forming as many pairs consisting of Receivers who prefer either Sender types* θ *and* θ'*, or Sender types* θ *and* θ''.

(iv) Suppose any types of pairs can secure a payoff of one; i.e., $\mu_0 \in \text{co}(B_\theta, B_{\theta'})$ *for all distinct* $\theta, \theta' \in \Theta$. *Let* $n_{1,\theta} \geq n_{1,\theta'} \geq n_{1,\theta''}$. *Suppose further that A trio of Receivers who prefer all possible Sender types...*

(a) cannot secure a payoff of two; i.e., $\sum_{\theta \in \Theta} \gamma_\theta \leq 2$ *and* $\gamma_\theta + \gamma_{\theta'} > 1$ *for all distinct* $\theta, \theta' \in \Theta$. *Then,* $w_1^* = \min\{n_{1,\theta}, n_{1,\theta'} + n_{1,\theta''}\}$ *attainable by forming as many pairs consisting of Receivers who prefer either Sender types* θ *and* θ'*, or Sender types* θ *and* θ''.

(b) can secure a payoff of two; i.e., $\gamma_\theta + \gamma_{\theta'} \leq 1$ *for all distinct* $\theta, \theta' \in \Theta$. *Then,* $w_1^* = 2n_{1,\theta''} + \min\{n_{1,\theta} - n_{1,\theta''}, n_{1,\theta'} - n_{1,\theta''}\}$ *attainable by forming* $n_{1,\theta''}$ *trios of Receivers who prefer all possible Sender types, and* $\min\{n_{1,\theta} - n_{1,\theta''}, n_{1,\theta'} - n_{1,\theta''}\}$ *pairs of Receivers consisting of either Sender types* θ *and* θ'' *or Sender types* θ' *and* θ''.

I first show that it is without loss to focus on partitions of N_1 such that each group contains no more than $|\Theta| = 3$ Receivers.

Lemma 1.6. *Suppose Receivers are single-minded and* $|\Theta| = 3$. *For any* $P \subseteq N_1$*, there exists a partition* $\mathcal{S} \in \Pi(P)$*, such that* $|S| \leq |\Theta|$ *for all* $S \in \mathcal{S}$ *and* $\sum_{S \in \mathcal{S}} w(S) \geq w(P)$.

Proof. Suppose Receivers are single-minded and $|\Theta| = 3$ and fix $P \subseteq N_1$ with $w_P \equiv w(P) > 0$. By Lemma 1.2, μ_0 is a convex combination of beliefs in B_Ss associated with exactly (i) two or (ii) three elements from $\{S \subseteq P : |S| = w_P\}$.

Case (i): $\mu_0 = \alpha\mu_{S_1} + (1-\alpha)\mu_{S_2}$ for some $\alpha \in (0,1)$ and $\mu_{S_r} \in B_{S_r}$ with $|S_r| = w_P$ and $S_r \subseteq P$ for each $r \in \{1,2\}$. By convexity of each B_i, B_{S_r} is also convex and so $S_1 \cap S_2 = \varnothing$ (recall $P \subseteq N_1$). It follows that $|P| \geq 2w_P$. Take any pair $\{s_1, s_2\}$ where for each $r \in \{1,2\}$, $s_r \in S_r$ and observe that $\mu_{S_r} \in B_r$. By 1.2, payoff of one is $\{s_1, s_2\}$-securable. Since w_P many such pairs can be created, partitioning of P into such pairs together yields a payoff of at least w_P to the Sender.

Case (ii): $\mu_0 = \alpha_1\mu_{S_1} + \alpha_2\mu_{S_2} + (1-\alpha_1-\alpha_2)\mu_{S_3}$ for some $\alpha_1, \alpha_2 > 0$, $\alpha_1 + \alpha_2 < 1$ and $\mu_{S_r} \in B_{S_r}$ with $|S_r| = w_P$ and $S_r \subseteq P$ for each $r \in \{1,2,3\}$. Convexity of B_i and the fact that $P \subseteq N_1$ means $S_1 \cap S_2 \cap S_3 = \varnothing$, which, in turn, means that no S_r can contain all types of Receiver; i.e., S_r contains only one type of Receiver (if $\gamma_\theta + \gamma_{\theta'} > 1$ for all distinct $\theta, \theta' \in \Theta$) or contains at most two distinct types of Receivers, say $\theta, \theta' \in \Theta$ (if $\gamma_\theta + \gamma_{\theta'} \leq 1$).

Consider the following figure, where columns represent Receivers in P and the rows represent each S_r. If $i \in S_r$, then the coordinate is marked with ● and if not ○. Without loss of generality, assume that Receivers $\{1, \dots, w_P\}$ belong in S_1 (note $|P| > w_P$).

	1	2	$\cdots$	w_P	$w_P + 1$	$\cdots$	$\|P\|$
S_1	●	●	$\cdots$	●	○	$\cdots$	○
S_2							
S_3							

For each $i \in S_1$, the possibilities are: i belongs in just S_1 or i belongs in either $S_1 \cap S_2$ or

$S_1 \cap S_3$. Reorder i's such that

	$\{i\}_{i=1}^{n_1}$	$\{i\}_{i=n_1+1}^{n_1+n_{12}}$	$\{i\}_{i=n_1+n_{12}+1}^{n_1+n_{12}+n_{13}=w_P}$	w_P+1	$\cdots$	$\lvert P\rvert$
S_1	●	●	●	○	$\cdots$	○
S_2	○	●	○			
S_3	○	○	●			

where n_r denotes the number of Receivers in S_r that belongs only in S_r and n_{rt} denote the number of Receivers in S_r that belongs in both S_r and S_t. Given this notation, there must be $w_P - n_{12}$ many Receivers in $\{w_P+1, \ldots, |P|\}$. Such Receivers can either belong in just S_2 (n_2 many of them) or in S_{23} (n_{23} many of them). Reorder $\{w_P+1, \ldots |P|\}$ so that

	$\{i\}_{i=1}^{n_1}$	$\{i\}_{i=n_1+1}^{n_1+n_{12}}$	$\{i\}_{i=n_1+n_{12}+1}^{n_1+n_{12}+n_{13}=w_P}$	$\{i\}_{i=w_P+1}^{w_P+n_2}$	$\{i\}_{i=w_P+n_2+1}^{w_P+n_2+n_{23}}$	$\cdots$	$\lvert P\rvert$
S_1	●	●	●	○	○	$\cdots$	○
S_2	○	●	○	●	●		
S_3	○	○	●	○	●		

Then, $i \in \{w_P+n_2+n_{23}+1, \ldots |P|\}$, can belong in just S_3; i.e.,

	$\{i\}_{i=1}^{n_1}$	$\{i\}_{i=n_1+1}^{n_1+n_{12}}$	$\{i\}_{i=n_1+n_{12}+1}^{n_1+n_{12}+n_{13}=w_P}$	$\{i\}_{i=w_P+1}^{w_P+n_2}$	$\{i\}_{i=w_P+n_2+1}^{w_P+n_2+n_{23}}$	$\{i\}_{i=w_P+n_2+n_{23}+1}^{\lvert P\rvert}$
S_1	●	●	●	○	○	○
S_2	○	●	○	●	●	○
S_3	○	○	●	○	●	●

Since w_P is P-securable, we must have

$$|S_1| = n_1 + n_{12} + n_{13} = w_P, \tag{1.8}$$

$$|S_2| = n_2 + n_{12} + n_{23} = w_P, \tag{1.9}$$

$$|S_3| = n_3 + n_{13} + n_{23} = w_P. \tag{1.10}$$

Define $S_{rt} := S_r \cap S_t$ (and so $n_{rt} = |S_{rt}|$). Consider first a trio consisting of one Receiver each from $a \in S_{12}$, $b \in S_{13}$ and $c \in S_{23}$. The trio $\{a, b, c\}$ secures a payoff of two since

$$\mu_0 = \alpha\mu_1 + \beta\mu_2 + (1 - \alpha - \beta)\,\mu_3$$

for some $\mu_1 \in B_a \cap B_b$, $\mu_2 \in B_a \cap B_c$ and $\mu_3 \in B_b \cap B_c$. There can be at most $\min\{n_{12}, n_{13}, n_{23}\}$ many such trios. Without loss of generality, suppose that $\min\{n_{12}, n_{13}, n_{23}\} = n_{12}$ (i.e., $n_{13}, n_{23} \geq n_{12}$) and so there are $n_{13} - n_{12}$ and $n_{23} - n_{12}$ many Receivers left in S_{13} and S_{23}, respectively. Observe that $i \in S_{13}$ can be paired with $j \in S_2 \cup S_{23}$ to secure a payoff of one since $\mu_1, \mu_3 \in B_i$ and $\mu_2 \in B_j$. Similarly, $i \in S_{23}$ can be paired with $j \in S_1 \cup S_{13}$ to secure a payoff of one. The number of pairs that can be formed are as follows.

- ▷ $S_{13} \times S_2$: $\min\{n_{13} - n_{12}, n_2\}$.
- ▷ $S_{23} \times S_1$: $\min\{n_{23} - n_{12}, n_1\}$.
- ▷ $S_{13} \times S_{23}$: $\min\{n_{13} - n_{12}, n_{23} - n_{12}\}$.

Suppose that the Sender creates $\min\{n_{13} - n_{12}, n_2\}$ pairs of $S_{13} \times S_2$ first, followed by pairs of $S_{23} \times S_1$.

- ▷ If $\min\{n_{13} - n_{12}, n_2\} = n_{13} - n_{12}$, then there may be Receivers left over in S_2's but no more $j \in S_{13}$ to pair them with. However, there are still $n_{23} - n_{12}$ many Receivers left in S_{23} who can be paired with $j \in S_1$ to secure a payoff of one; $\min\{n_{23} - n_{12}, n_1\}$ many such pairs can be formed. Consider now the Sender's ability to produce other pairs of $S_{23} \times S_1$ that secure a payoff of one.

- ▷ If $\min\{n_{23} - n_{12}, n_1\} = n_1$, then Sender can secure a total payoff of

$$w(\mathcal{S}) = 2n_{12} + (n_{13} - n_{12}) + n_1 = w_P,$$

where the last equality follows from (1.8).

- ▷ If $\min\{n_{23} - n_{12}, n_1\} = n_{23} - n_{12}$, then there may be Receivers in S_1, S_2 and S_3—a trio consisting of each such Receiver can secure a payoff of one. Hence, by (1.8)–(1.10),

$$\begin{aligned} w(\mathcal{S}) &= 2n_{12} + (n_{13} - n_{12}) + (n_{23} - n_{12}) + \min\{n_1, n_2, n_3\} \\ &= \begin{cases} \underbrace{n_{12} + n_{13} + n_1}_{=w_P} + \underbrace{n_{23} - n_{12}}_{\geq 0} & \text{if } \min\{n_1, n_2, n_3\} = n_1 \\ \underbrace{n_{12} + n_{23} + n_2}_{=w_P} + \underbrace{n_{13} - n_{12}}_{\geq 0} & \text{if } \min\{n_1, n_2, n_3\} = n_2 \\ \underbrace{n_{13} + n_{23} + n_3}_{=w_P} + \underbrace{n_{12} - n_{12}}_{=0} & \text{if } \min\{n_1, n_2, n_3\} = n_3 \end{cases} \\ &\geq w_P. \end{aligned}$$

▷ If $\min\{n_{13} - n_{12}, n_2\} = n_2$, then there may be Receivers left over in S_{13} but no more Receivers in S_2 to pair them with. However, they can be paired with Receivers left in S_{23} to secure a payoff of one; $\min\{n_{13} - n_{12} - n_2, n_{23} - n_{12}\}$ many such pairs can be formed.

- ▷ If $\min\{n_{13} - n_{12} - n_2, n_{23} - n_{12}\} = n_{23} - n_{12}$, then there is no Receivers that can be paired with Receivers left over in S_{13} to secure a payoff of one. Hence, by (1.10),

$$w(\mathcal{S}) = 2n_{12} + n_2 + n_{23} - n_{12} = n_{12} + n_2 + n_{23} = w_P.$$

▷ If $\min\{n_{13} - n_{12} - n_2, n_{23} - n_{12}\} = n_{13} - n_{12} - n_2$, then there may be Receivers left over in S_{23} who can be paired with Receivers in S_1 to secure a payoff of one; $\min\{n_{23} - n_{12} - (n_{13} - n_{12} - n_2), n_1\}$ many such pairs can be formed.

- If $\min\{n_{23} - n_{12} - (n_{13} - n_{12} - n_2), n_1\} = n_1$, then, by (1.8),

$$w(\mathcal{S}) \geq 2n_{12} + n_2 + (n_{13} - n_{12} - n_2) + n_1 = n_{12} + n_{13} + n_1 = w_P.$$

- If $\min\{n_{23} - n_{12} - (n_{13} - n_{12} - n_2), n_1\} = n_{23} - n_{12} - (n_{13} - n_{12} - n_2)$, then, by (1.9),

$$\begin{aligned} w(\mathcal{S}) &\geq 2n_{12} + n_2 + (n_{13} - n_{12} - n_2) + n_{23} - n_{12} - (n_{13} - n_{12} - n_2) \\ &= n_{12} + n_{23} + n_2 = w_P. \end{aligned}$$

Hence, we can construct a partition of P in which each group consists of no more than three Receivers that together can secure at least w_P. ■

The following Lemma establishes that it is without loss to focus on groups that contain at most one Receiver of any type.

Lemma 1.7. *Suppose Receivers are single-minded and $|\Theta| = 3$. For any $P \subseteq N_1$ such that $|P| \leq |\Theta|$ and $w(P) > 0$, there exits $S \subseteq P$ such that $w(S) = w(P)$ and S contains at most one Receiver of any type.*

Proof. It suffices to consider the case when (i) $w(P) = 1$ and $|P| = 2$; (ii) $w(P) = 1$ and

$|P| = 3$; and (iii) $w(P) = 2$ and $|P| = 3$.

(i) If $S = P$ contains two Receivers of the same type, then $P \subseteq N_1$ implies $w(P) = 0$, a contradiction.

(ii) Label $P = \{1, 2, 3\}$ and suppose $t_1 = t_2 \in \Theta$. Letting $S = \{1, 3\} = P \backslash \{2\}$, since $\text{co}(B_{t_1} \cup B_{t_2} \cup B_{t_3}) = \text{co}(B_{t_1} \cup B_{t_3})$, by Lemma 1.2, $w(S) = 1$.

(iii) Label $P = \{1, 2, 3\}$ and suppose $t_1 = t_2 \in \Theta$. Then, by Lemma 1.2, there exists $\mu_{ij} \in B_i \cap B_j$ and $\alpha, \beta \in (0, 1)$ with $\alpha + \beta \leq 1$ such that

$$\mu_0 = \alpha\mu_{12} + \beta\mu_{13} + (1 - \alpha - \beta)\,\mu_{23}.$$

But $\mu_{12}, \mu_{13}, \mu_{23} \in B_1, B_2$ so that by convexity of B_i, $\mu_0 \in B_1, B_2$, contradicting $P \subseteq N_1$. ■

The following lemma allows us to define the cases above as conditions on $\gamma = (\gamma_\theta)_{\theta \in \Theta}$.

Lemma 1.8. *Suppose Receivers are single-minded, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and Receivers in N_1 who prefer the same Sender type have a common threshold; i.e., for each $\theta \in \Theta$, $\gamma_i = \gamma_\theta$ for all $i \in N_1$ such that $t_i = \theta$.*

$$\bigcap_{\theta, \theta' \in \Theta : \theta \neq \theta'} \text{co}\left(B_\theta \cup B_{\theta'}\right) \neq \varnothing \Leftrightarrow \sum_{\theta \in \Theta} \gamma_\theta \leq 2 \tag{1.11}$$

and

$$B_\theta \cap B_{\theta'} \neq \varnothing \Leftrightarrow \gamma_\theta + \gamma_{\theta'} \leq 1. \tag{1.12}$$

Proof. Note that $\text{co}(B_\theta \cup B_{\theta'})$ can be characterised as a hyperplane:

$$\text{co}\left(B_\theta \cup B_{\theta'}\right) = \left\{\mu \in \Delta\Theta : \mu\left(\theta\right)\gamma_{\theta'} + \mu\left(\theta'\right)\gamma_\theta \geq \gamma_\theta\gamma_{\theta'}\right\}.$$

For the intersections of these to be nonempty, by symmetry, it suffices to show that the intersection of two hyperplanes representing $\text{co}(B_{\theta_1} \cup B_{\theta_2})$ and $\text{co}(B_{\theta_1} \cup B_{\theta_3})$ lies inside $\text{co}(B_{\theta_2} \cup B_{\theta_3})$. Thus, take any $\mu \in \Delta\Theta$ that lies in the intersection; i.e.,

$$\mu(\theta_1)\gamma_{\theta_2} + \mu(\theta_2)\gamma_{\theta_1} = \gamma_{\theta_1}\gamma_{\theta_2} \text{ and } \mu(\theta_1)\gamma_{\theta_3} + \mu(\theta_3)\gamma_{\theta_1} = \gamma_{\theta_1}\gamma_{\theta_3}.$$

Solving yields

$$\mu(\theta_2) = \frac{\gamma_{\theta_2}\left(\gamma_{\theta_1} - 1\right)}{\gamma_{\theta_1} - \gamma_{\theta_2} - \gamma_{\theta_3}} \text{ and } \mu(\theta_3) = \frac{\gamma_{\theta_3}\left(\gamma_{\theta_1} - 1\right)}{\gamma_{\theta_1} - \gamma_{\theta_2} - \gamma_{\theta_3}}.$$

Since $\gamma_{\theta_1} < 1$ and $\mu_{\theta_2} \geq 0$, it must be that $\gamma_{\theta_1} - \gamma_{\theta_2} - \gamma_{\theta_3} < 0$. For $\mu \in \text{co}(T_{\theta_2} \cup T_{\theta_3})$,

$$\gamma_{\theta_2}\gamma_{\theta_3} \geq \mu(\theta_2)\gamma_{\theta_3} + \mu(\theta_3)\gamma_{\theta_2} \Leftrightarrow \gamma_{\theta_1} + \gamma_{\theta_2} + \gamma_{\theta_3} \leq 2.$$

To prove (1.12), observe that if $\mu \in B_\theta \cap B_{\theta'}$, then $\mu(\theta) \geq \gamma_\theta$ and $\mu(\theta') \geq \gamma_\theta$ so that

$$1 \geq \mu(\theta) + \mu\left(\theta'\right) \geq \gamma_\theta + \gamma_{\theta'}.$$

Conversely, if $\gamma_\theta + \gamma_{\theta'} \leq 1$, then we may choose $\mu \in \Delta\Theta$ such that $\mu(\theta) = \gamma_\theta$ and $\mu(\theta') = \gamma_{\theta'}$ so that $\mu \in B_\theta \cap B_{\theta'}$. ■

Proof of Propositon 1.4. The proposition is immediate from lemmata above. ■

1.A.4 A counterexample with four states

The next example shows that when $|\Theta| = 4$, the maximum size of groups need not be less than or equal to $|\Theta|$ and groups might contain more than one Receiver of the same type.

Example 1.4. Suppose $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, $N = N_1 = \{1, 2, 3, 4, 5\}$ with

$$t_1 = t_5 = \theta_1,\ t_2 = \theta_2,\ t_3 = \theta_3,\ t_4 = \theta_4,$$
$$\gamma_{\theta_1} = \frac{3}{4},\ \gamma_{\theta_2} = \gamma_{\theta_3} = \gamma_{\theta_4} = \frac{1}{4},$$

and $\mu_0 = (\frac{1+\epsilon}{2}, \frac{1-\epsilon}{6}, \frac{1-\epsilon}{6}, \frac{1-\epsilon}{6})$ for some $\epsilon > 0$ sufficiently small. Then, $\mu_0(\theta) \le \gamma_\theta$ for all $\theta \in \Theta$. Moreover, given

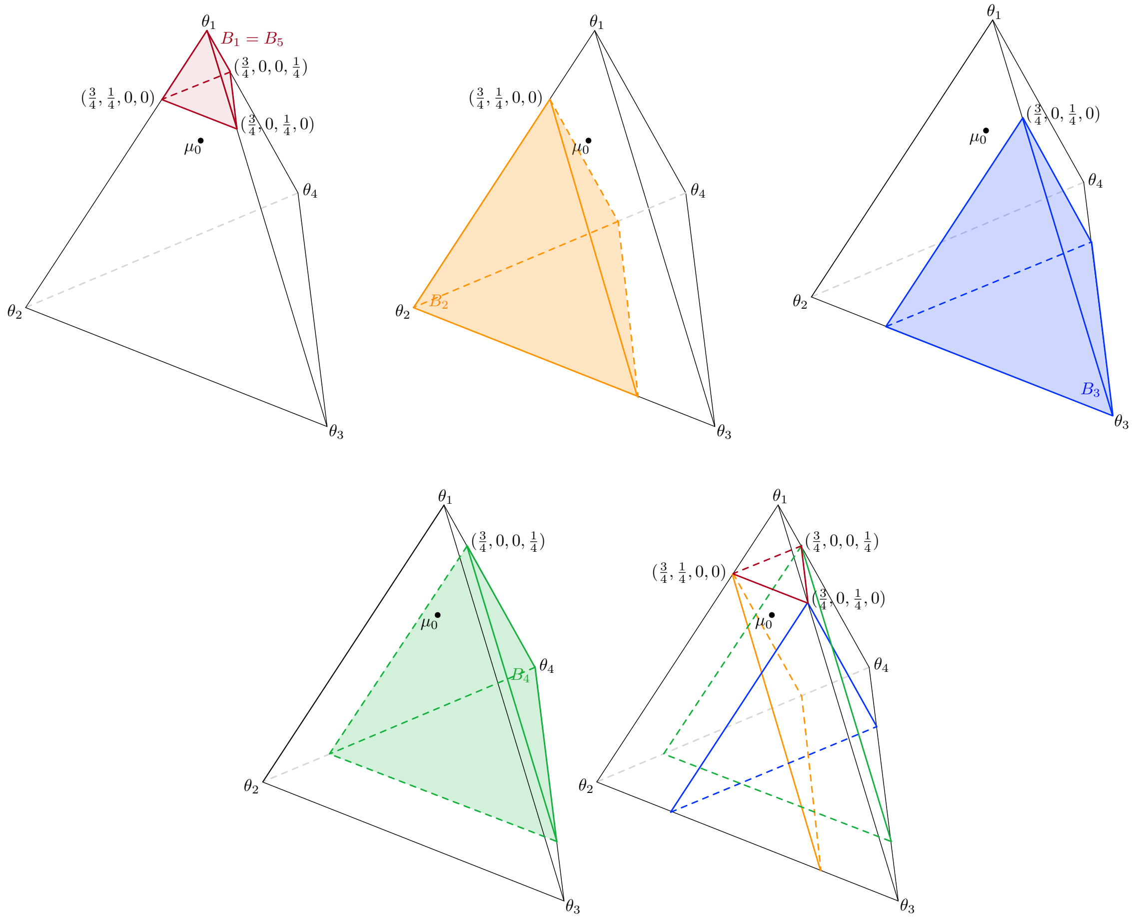
$$\mu_1 = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \in B_2 \cap B_3 \cap B_4, \quad \mu_3 = \left(\frac{3}{4}, 0, \frac{1}{4}, 0\right) \in B_1 \cap B_3 \cap B_5,$$
$$\mu_2 = \left(\frac{3}{4}, \frac{1}{4}, 0, 0\right) \in B_1 \cap B_2 \cap B_5, \quad \mu_4 = \left(\frac{3}{4}, 0, 0, \frac{1}{4}\right) \in B_1 \cap B_4 \cap B_5,$$

μ_0 can be expressed as

$$\mu_0 = \frac{1-2\epsilon}{2}\mu_1 + \frac{1-2\epsilon}{6}\mu_2 + \frac{1-2\epsilon}{6}\mu_3 + \frac{1-2\epsilon}{6}\mu_4.$$

Therefore, by Lemma 1.2, $w(N_1) = 3$. However, no (nontrivial) partition of N_1 can guarantee a payoff of 3 for the Sender. Since a partition of N_1 can secure 3 if and only if (i) a group consisting of four Receivers secures a payoff of three or (ii) a group consisting of three Receivers, say $P_1 \subseteq N_1$, secures a payoff of two, and another group consisting of two Receivers, say $P_2 = N_1 \backslash P_1$, secures a payoff of one.

Figure 1.5: Counterexample: Maximum size of groups need not be less than $|\Theta|$ when $|\Theta| = 4$.



(i) Note that $B_2 \cap B_3 \cap B_4$, $B_1 \cap B_2 \cap B_5 = \mu_2$, $B_1 \cap B_3 \cap B_5 = \mu_3$, $B_1 \cap B_4 \cap B_5 = \mu_4$ are the only (set of) beliefs in which three Receivers would take action $a = 1$. Removing Receiver 1 or 5 from N would leave only $B_2 \cap B_3 \cap B_4$ so that, given $\mu_0 \notin B_i$ for all $i \in N$, the remainder of Receivers cannot secure a payoff of one for the Sender. Consider removing Receiver 4 (the argument is symmetric for Receivers 2 and 3), which leaves $B_1 \cap B_2 \cap B_5 = \mu_2$ and $B_1 \cap B_3 \cap B_5 = \mu_3$. However, for any $\alpha \in [0, 1]$,

$$\alpha\mu_2 + (1 - \alpha)\,\mu_3 = \left(\frac{3}{4}, \frac{\alpha}{4}, \frac{1-\alpha}{4}, 0\right)$$

so that $\mu_0 \notin \text{co}(\{\mu_2, \mu_3\})$. Hence, by Lemma 1.2, $\{1,2,3,5\}$ cannot secure a payoff of three. It follows that group of four Receivers can secure a payoff of three for the Sender.

(ii) Since $P \subseteq N_1$, $P_2 \neq \{1,5\}$. Consider two cases: (a) P_2 consists of type-θ_1 Receiver and a type-θ_2 Receiver (symmetric argument for θ_3 and θ_4) or (b) P_2 consists of type-θ_2 and type-θ_3 Receivers (symmetric argument in the case P_2 consists of types (θ_2, θ_4) or (θ_3, θ_4)).

(a) In the first case, $P_1 = \{1,3,4\}$ and the set of beliefs under which two Receivers in P_1 take action $a = 1$ are: $B_3 \cap B_4$, $B_1 \cap B_3 = \mu_3$ and $B_1 \cap B_4 = \mu_4$. Toward a contradiction, suppose there exists $\mu_{34} \in B_3 \cap B_4$ such that

$$\begin{aligned}
\mu_0(\theta_1) &= (1-\alpha-\beta)\,\mu_{34}(\theta_1) + \alpha\mu_3(\theta_1) + \beta\mu_3(\theta_1) \\
&= (1-\alpha-\beta)\,\mu_{34}(\theta_1) + \frac{3}{4}(\alpha+\beta), \\
\mu_0(\theta_2) &= (1-\alpha-\beta)\,\mu_{34}(\theta_2) + \alpha\mu_3(\theta_2) + \beta\mu_4(\theta_2) \\
&= (1-\alpha-\beta)\,\mu_{34}(\theta_2). \qquad (1.13)
\end{aligned}$$

Adding the two together gives

$$\mu_0(\theta_1) + \mu_0(\theta_2) = \frac{3}{4}(\alpha+\beta) + (1-\alpha-\beta)\left(\mu_{34}(\theta_1) + \mu_{34}(\theta_2)\right).$$

Since $\mu_{34} \in B_3 \cap B_4$, $\mu_{34}(\theta_1) + \mu_{34}(\theta_2) \leq 1 - \gamma_{\theta_3} - \gamma_{\theta_4} = \frac{1}{2}$ so that

$$\mu_0(\theta_1) + \mu_0(\theta_2) \leq \frac{3}{4}(\alpha+\beta) + \frac{1}{2}(1-\alpha-\beta) = \frac{1}{2} + \frac{1}{4}(\alpha+\beta).$$

Solving (1.13) and using the fact that $\mu_{34}(\theta_2) \leq \mu_{34}(\theta_1) + \mu_{34}(\theta_2) \leq \frac{1}{2}$ gives

$$\alpha + \beta = 1 - \frac{\mu_0(\theta_2)}{\mu_{34}(\theta_2)} \in [\mu_0(\theta_2), 1 - 2\mu_0(\theta_2)]$$

so that

$$\mu_0(\theta_1) + \mu_0(\theta_2) \leq \frac{1}{2} + \frac{1}{4}(1 - 2\mu_0(\theta_2)) = \frac{3}{4} - \frac{1}{2}\mu_0(\theta_2)$$
$$\Leftrightarrow \mu_0(\theta_1) \leq \frac{3}{4} - \frac{3}{2}\mu_0(\theta_2).$$

Substituting the values for $\mu_0(\theta_1)$ and $\mu_0(\theta_2)$ gives

$$\frac{1+\epsilon}{2} \leq \frac{3}{4} - \frac{3}{2}\left(\frac{1-\epsilon}{6}\right) = \frac{1}{4}(2+\epsilon) \Leftrightarrow \epsilon \leq 0,$$

which contradicts the assumption that $\epsilon > 0$.

(b) In the second case, $P_1 = \{1, 4, 5\}$. However, since $B_1 = B_5$, that $P \subseteq N_1$ implies that $w(P_1) = 1$.

1.A.5 Proof of Proposition 1.5

Proposition 1.5. *Suppose Receivers are singled-minded, Receivers who prefer the same Sender type have a common threshold, $\gamma_\theta + \gamma_{\theta'} > 1$ for all distinct $\theta, \theta' \in \Theta$, and Θ is finite. Then, there exists an optimal partition in which each group contains at most one Receiver who prefers each possible Sender type. Moreover, the sender's problem is solvable in polynomial time.*

I first prove the following lemma.

Lemma 1.9. *Suppose $\gamma_\theta + \gamma_{\theta'} > 1$ for all distinct $\theta, \theta' \in \Theta$. Then, for any $P \subseteq N_1$ such that P contains at most k of each type of Receiver, $w(P) \leq k$.*

Proof. By way of contradiction, suppose that $w(P) > k$. By Lemma 1.2, there exist $\mathcal{S} \subseteq \{S \subseteq P : |S| = w(P)\}$, $(\tau_S)_{S\in\mathcal{S}} \in \Delta^{|\mathcal{S}|}$ and $(\mu_S)_{S\in\mathcal{S}} \in (\Delta\Theta)^{\mathcal{S}}$ such that $\mu_0 = \sum_{S\in\mathcal{S}} \tau_S \mu_S$, $|\mathcal{S}| = \Theta$, and each μ_S belongs in $w(P) > k$ many B_i's. Since $B_i \cap B_j = \varnothing$ if $\theta_i \neq \theta_j$, each S must contain the same type of Receivers and since $|S| = w(P) > k$, each S contains strictly more than k Receivers of the same type; a contradiction. ∎

Proof of Proposition 1.5. $\mathcal{S} \subseteq \{S \subseteq P : |S| = w(P)\}$, $(\tau_S)_{S\in\mathcal{S}} \in \Delta^{|\mathcal{S}|}$ and $(\mu_S)_{S\in\mathcal{S}} \in (\Delta\Theta)^{\mathcal{S}}$ such that $\mu_0 = \sum_{S\in\mathcal{S}} \tau_S \mu_S$, $|\mathcal{S}| = \Theta$, and each μ^S belongs in $w(P)$ many B_i's. In particular, for any distinct $S, S' \in \mathcal{S}$ such that $\tau_S, \tau_{S'} > 0$, S and S' contains $w(P)$ many of Receivers of different types. Construct a group P_1 consisting of one Receiver from each $S \in \mathcal{S}$ such that $\tau_S > 0$. By construction, P_1 secures a payoff of one. With the remaining Receivers, $P \backslash P_1$, we can construct a group P_2 consisting again of one Receiver from each $S \in \mathcal{S}$ such that $\tau_S > 0$. We can inductively construct $(P_k)_{k=1}^{w(P)}$ such that each P_k secures a payoff of one. Hence,

$$w\left(P\right) \leq \underbrace{\sum_{k=1}^{w(P)} w\left(P_k\right)}_{=w(P)} + w\left(P \backslash \bigcup_{k=1}^{w(P)} P_k\right).$$

However, since $(\{P_k\}_{k=1}^{w(P)}) \cup (P \backslash \bigcup_{k=1}^{w(P)} P_k)$ is a partition of P, the right-hand side must be less than $w(P)$ by optimality of P so that $w(P \backslash \bigcup_{k=1}^{w(P)} P_k) = 0$. Hence, no subset of $P \backslash \bigcup_{k=1}^{w(P)} P_k$ can secure a payoff of one and thus we may partition $P \backslash \bigcup_{k=1}^{w(P)} P_k$ such that each (sub)group contains (at most) one of each type of Receiver.

Finally, observe that the treewidth of the hypergraph associated with the Receiver is one as each set of Receivers has a unique type (i.e., skill) and so forms a tree. Thus, Bachracht et al. (2010)'s algorithm can solve for the optimal partition in polynomial time.

∎

1.A.6 Proof of Proposition 1.6

Proposition 1.6. *There exists an optimal partition with the following properties:*

(i) every Receiver is either paired with another Receiver or unpaired;

(ii) unpaired Receivers are more extreme than paired Receivers;

(iii) Receivers are paired negatively assortatively; i.e., among Receivers in pairs, the most extreme type-t Receiver is paired with the least extreme type-$t' \neq t$ Receiver, and so on.

Proof. Fix an optimal partition $\mathcal{P}^*$. I will first construct a partition of N_1 from $\mathcal{P}^*$ that consists of pairs and singleton Receivers. I will then show that Receivers in pairs can be exchanged to ensure negative assortativity. To that end, for each group $P \in \mathcal{P}^*$ such that $|P| > 2$, by definition, there exists a k-cutoff τ_P^k that secures a payoff of $w(P)$. For each $t \in \{\ell, r\}$, let $P^t \subseteq P$ be the type-t Receivers that take action $a = 1$ with τ_P^k (note$|P^t| = w(P)$). Since a payoff of one is securable with any pair $(i, j) \in P^\ell \times P^r$ with the same τ_P^k, $P^\ell \cup P^r$ can be decomposed into pairs that each secure a payoff of one that together secures a payoff of $w(P)$. Let $\mathcal{S}_P$ denote the partition of P consisting of unions of such pairs of Receivers with singleton sets of Receivers in $P \backslash \{P^\ell \cup P^r\}$. Then, $\mathcal{S} := \bigcup_{P \in \mathcal{P}^*} \mathcal{S}_P$ is a partition of N consisting of pairs of Receivers that each secure a payoff of one and singleton sets of Receivers such that $W(\mathcal{S}) = w^*$.

Let $\{(\ell_1, r_1), (\ell_2, r_2), \ldots, (\ell_{w^*}, r_{w^*})\}$ denote the pairs in $\mathcal{S}$ such that ℓ_i is more extreme than ℓ_{i+1} for each $i \in \{1, \ldots, w^* - 1\}$. Suppose that ℓ_1 (i.e., the most extreme type-ℓ Receiver among those paired) is not paired with the most moderate type-r Receiver among those paired (if not, then repeat the process for ℓ_2); i.e., there exists $i \in \{2, \ldots, w^*\}$ such that r_i is strictly more moderate that r_1. Let r_i' be the least moderate among all such i's. By Lemma 1.4, r_i' and r_1 can be switched without affecting payoffs. Now repeat the process for ℓ_2 and so on. The process clearly terminates. Let $\tilde{\mathcal{S}}$ denote pairs after this process terminates and observe that the pairs satisfy property (iii) among all Receivers who are paired in $\tilde{S}$.

Now suppose there exists an unpaired type-t Receiver, say t_i, who is more moderate than some type-t Receiver who is paired. By Lemma 1.4, such a Receiver can be swapped with the most extreme type-t who is paired and is also more moderate than Receiver t_i without affecting payoff. Repeat this process until there are no such t_is. Then, all the unpaired Receivers are more moderate than paired Receivers (of the same type). ■

1.A.7 Multiple partitions

The following example demonstrates that the ability to adopt communication strategies over multiple partitions can strictly benefit the Sender by allowing her to guarantee that two Receivers take action $a = 1$ when a single partition could only guarantee one Receiver to take action $a = 1$.

Example 1.5. Suppose $N = \{1, 2, 3\}$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mu_0 = (\frac{1}{5}, \frac{3}{5}, \frac{3}{5})$, $t_i = \theta_i$ for each $i \in \{1, 2, 3\}$, and $\gamma_{\theta_1} = \frac{2}{5}$, $\gamma_{\theta_2} = \frac{13}{20}$, and $\gamma_{\theta_3} = \frac{1}{4}$. Since $\gamma_{\theta_1} + \gamma_{\theta^2} > 1$, $w(N) \leq 1$. However, we will show that if the Sender can partition the Receivers in multiple ways, she can guarantee a payoff of two. Specifically, suppose $\mathcal{P}_1 = \{\{1, 2\}, \{3\}\}$ and $\mathcal{P}_2 = \{\{1\}, \{2, 3\}\}$ and let $m_1 \in \{\ell_1, r_1\}$ be the message that the Sender sends to the group $\{1, 2\}$ under $\mathcal{P}_1$ and $m_2 \in \{\ell_2, r_2\}$ be the message that the Sender sends to group $\{2, 3\}$ under $\mathcal{P}_3$. Thus, Receiver 1 observes message m_1, Receiver 3 observes message m_2, and Receiver 2 observes message (m_1, m_2). Let $\pi : \Theta \to \Delta(\{\ell_1, r_1\} \times \{\ell_2, r_2\})$ be as shown in the table below.

Table 1.1: Example: Messaging strategy over multiple partition.

$\pi(m_1, m_2 \mid \theta)$	θ_1	θ_2	θ_3
(ℓ_1, ℓ_2)	0	0	0
(ℓ_1, r_2)	0	$\frac{3}{5}$	$\frac{4}{5}$
(r_1, ℓ_2)	0	$\frac{2}{5}$	0
(r_1, r_2)	1	0	$\frac{1}{5}$

The posterior beliefs for Receiver 1 who observes $m_1 \in \{\ell_1, r_1\}$ are

$$\begin{aligned} 0 = \mu(\theta_1 \mid \ell_1) &= \frac{\pi(\ell_1, r_2 \mid \theta_1)}{\sum_{\tilde{\theta} \in \Theta} \pi(\ell_1, r_2 \mid \tilde{\theta})} \\ &< \gamma_{\theta_1} = \frac{2}{5} \\ &\leq \frac{\pi(r_1, \ell_2 \mid \theta_1) + \pi(r_1, r_2 \mid \theta_1)}{\sum_{\tilde{\theta} \in \Theta} \pi(r_1, \ell_2 \mid \tilde{\theta}) + \pi(r_1, r_2 \mid \tilde{\theta})} = \mu(\theta_1 \mid r_1) = \frac{5}{12} \end{aligned}$$

so that Receiver 1 takes action $a = 1$ if and only if $m_1 = r_1$. The posterior beliefs for Receiver 3 who observes $m_2 \in \{\ell_2, r_2\}$ are:

$$\begin{aligned} 0 = \mu(\theta_3 \mid \ell_2) &= \frac{\pi(r_1, \ell_2 \mid \theta_3)}{\sum_{\tilde{\theta} \in \Theta} \pi(r_1, \ell_2 \mid \tilde{\theta})} \\ &< \gamma_{\theta_3} = \frac{1}{4} \\ &\leq \frac{\pi(\ell_1, r_2 \mid \theta_3) + \pi(r_1, r_2 \mid \theta_3)}{\sum_{\tilde{\theta} \in \Theta} \pi(\ell_1, r_2 \mid \tilde{\theta}) + \pi(r_1, r_2 \mid \tilde{\theta})} = \mu(\theta_3 \mid r_2) = \frac{5}{19} \end{aligned}$$

so that Receiver 3 takes action $a = 1$ if and only if $m_2 = r_2$. Posterior beliefs for Receiver

2 who observes $(m_1, m_2) \in \{\ell_1, r_1\} \times \{\ell_2, r_2\}$ are

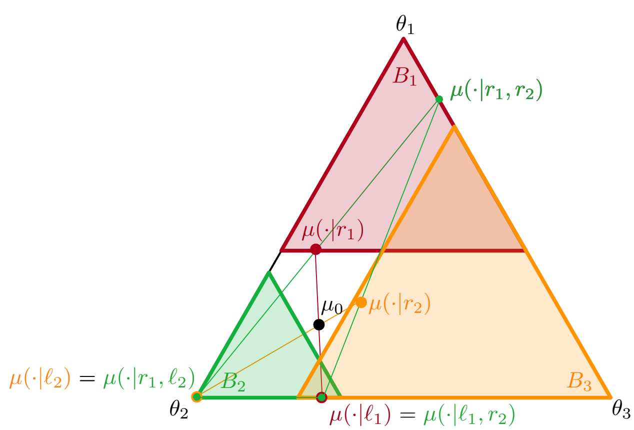
$$
\begin{aligned}
0 = \mu(\theta_2 | r_1, r_2) &= \frac{\pi(r_1, r_2|\theta_2)}{\sum_{\tilde{\theta}\in\Theta} \pi(r_1, r_2|\tilde{\theta})} \\
< \gamma_{\theta_2} &= \frac{13}{20} \\
< \mu(\theta_2 | \ell_1, r_2) &= \frac{\pi(\ell_1, r_2|\theta_2)}{\sum_{\tilde{\theta}\in\Theta} \pi(\ell_1, r_2|\tilde{\theta})} = \frac{9}{13} \\
< \mu(\theta_2 | r_1, \ell_2) &= \frac{\pi(r_1, \ell_2|\theta_2)}{\sum_{\tilde{\theta}\in\Theta} \pi(r_1, \ell_2|\tilde{\theta})} = 1
\end{aligned}
$$

so that Receiver 2 takes action $a = 1$ if and only if $(m_1, m_2) \in \{(r_1, \ell_2), \{\ell_1, r_2\}\}$. Therefore,

- ▷ if $(m_1, m_2) = (\ell_1, r_2)$, Receivers 2 and 3 take action $a = 1$;
- ▷ if $(m_1, m_2) = (\ell_1, r_2)$, Receivers 1 and 2 take action $a = 1$;
- ▷ if $(m_1, m_2) = (r_1, r_2)$, Receivers 1 and 3 take action $a = 1$.

Hence, the Sender has no incentive to deviate from π and she can guarantee a payoff of two from the group of two.

Figure 1.6: Example: Sender strictly benefits from multiple partitions.



Chapter 2

Agreeing to be fooled: Optimal ignorance about information sources

Suppose a receiver obtains a piece of information from a sender and is worried that the information might not be reliable.[1] Should the receiver learn whether the information is reliable before deciding on an action?[2]

The question arises in many economically important situations. For example, courts can learn about the reliability of witness evidence provided by the parties via cross-examination, buyers can audit the marketing material provided by sellers, and voters with doubts about the sincerity of a politician's statement can look up what the politician has said in the past. In these situations, learning about reliability is beneficial because it allows the receiver to avoid making decisions based on unreliable information. However, learning might also hurt the receiver by altering the sender's incentive to provide information in the first place. Such strategic considerations give rise

[1] This chapter is based on my job market paper with the same title. In addition to my committee members, I am also grateful to Ben Brooks, Modibo Camara, Alex Frankel, Marina Halac, Daniel Rappoport, Phil Reny, Joseph Root, Kai Hao Yang, and the participants of the micro-theory seminar at The University of Chicago for their helpful comments.

[2] Throughout, I refer to the receiver and the sender using male and female pronouns, respectively.

to the possibility that the receiver might wish to avoid learning to induce the sender to provide more useful information.

To see the intuition, suppose that the receiver can investigate the sender's reliability. To discipline the sender, the receiver can conduct a "harsh" investigation if the sender does not provide useful information, and conduct a "lenient" investigation if she provides useful information. Thus, by varying the harshness or the leniency of investigations based on the sender's information, the receiver could induce the sender to provide more useful information. Importantly, I show that harsh and lenient investigations require the receiver to strategically avoid learning that the sender is reliable or that she is unreliable. However, avoiding learning about reliability is never sequentially rational for the receiver. Thus, to implement harsh and lenient investigations directly, the receiver must be able to fight his inherent desire to learn by, for example, being able to commit to being ignorant. Alternatively, the receiver can delegate investigations to a third party with appropriate incentives who can (indirectly) implement harsh and lenient investigations on the receiver's behalf. To that end, I show that the receiver benefits from delegating investigations to the sender's adversary. I further show that, whenever the sender is sufficiently unreliable, the receiver can implement optimal commitment outcomes via delegation if the adversary third party is also partially aligned with the receiver.

In the court contest, these results suggest that if a cross-examiner is only concerned about the discovery of truth (e.g., the cross-examiner is the judge), cross-examination may not help the court obtain additional information. Perhaps surprisingly, for the judge to be an effective cross-examiner, the judge must paradoxically commit to sometimes not discovering the truth about the witness' reliability. In most jurisdictions, however, witnesses are cross-examined not by the judge but by an adversary to the party that calls on the witness to testify. Thus, courts delegate investigations to an adversarial third party which, as my results demonstrate, can induce more information to be provided to the courts. Moreover, in some jurisdictions, provisions exist to ensure that the cross-examiner would have incentives akin to an adversarial third party who is also partially received aligned. For example, in the US, prosecutors have a dual role as advocates seeking convictions and as "ministers of justice" (Fisher, 1988). My results highlight how such provisions can increase the efficacy of cross-examination by enabling the court to obtain more, and sometimes maximal, amount of information from the parties.

The results are also relevant outside of the court context. Investigations can also be thought of as audits of information such as financial statements or investment appraisals provided by a seller to investors. In this context, my results imply that an unfettered audit that fully reveals the seller's reliability might lead the seller to provide less information leaving investors no better off than if they did not audit at all. Instead, investors can induce the seller to provide more information by conducting audits that do not always reveal reliability either by conducting audits themselves (if they can commit to ignorance) or by choosing an auditor with adversarial incentives. Alternatively, realisations of investigations can be thought of as arguments about the sender, rather than arguments about the sender's argument. The former type of argument is an example of *ad hominem* (meaning "to the person") arguments. In this light, the results I obtain suggest that *ad hominem* "attacks" on politicians by an opposing political campaign or a media outlet can benefit the voters by inducing politicians to speak more truthfully.

Concretely, I study a binary-action, binary-state sender-receiver game of persuasion. The receiver is uncertain about the state and wants to match his action to the state. The sender wants the receiver to always take an action, and she can persuade the receiver by using any public signal about the state. With some probability, the sender is *unreliable* and can manipulate the "realisation" that the receiver sees. With complementarity probability, the sender is *reliable* and no manipulation occurs. By default, the receiver observes a realisation without knowing whether the sender is reliable. However, before choosing an action, the receiver can investigate the sender using any signal about the sender's reliability. The receiver can thus avoid learning about the sender's reliability in two ways: either by not always finding out that the sender is reliable when she is reliable, or by not always finding out that the sender is unreliable when she is unreliable. I refer to the sender's signal structure (about the state) as an *experiment* and the receiver's choice of signal structure (about the sender's reliability) as an *investigation*. The main result of this chapter is a characterisation of the receiver's optimal *investigation strategy* that specifies the investigation to be conducted conditional on the experiment chosen by the sender.

To derive the optimal investigation strategy, I begin by establishing a result that simplifies the analysis. Specifically, I show that it is without loss to restrict attention to a canonical set of experiments that are Blackwell ordered (Blackwell, 1953). Moreover, the sender's incentives are such that she chooses the least (Blackwell) informative experiment from this set that could persuade the receiver. In contrast, the receiver prefers the sender to choose the most informative experiment from the canonical set.

Now consider how the receiver could induce the sender to choose the most informative experiment in an equilibrium. As already mentioned, one way to achieve this is to reduce the sender's payoff from choosing less informative experiments by conducting harsh investigations. Harsh investigations take the form of the receiver strategically not finding out that the sender is reliable to reduce the sender's chance of successfully persuading the receiver. Alternatively, the receiver can increase the sender's payoff from choosing the most informative experiment. The receiver can do so by conducting a lenient investigation, which takes the form of the receiver not finding out that the sender is unreliable (to increase the chance that the receiver is persuaded). The receiver's ideal outcome is to find out whether the sender is reliable after the sender has chosen the most informative experiment. Thus, the receiver does not conduct a lenient investigation unless he must; i.e., unless the sender's payoff from choosing the most informative experiment and the receiver finding out her reliability is strictly lower than what she can obtain by deviating to another experiment. It is possible that the receiver cannot induce the sender to choose the most informative experiment using the two aforementioned uses of strategic ignorance. In such cases, the receiver must demand a less informative experiment on the equilibrium path from the sender.

The argument above implies that, while the receiver always uses ignorance off the equilibrium path as a punishment, the receiver only sometimes uses ignorance on the equilibrium path as a minimal "reward." Whether the receiver uses ignorance on path depends on the sender's prior incentive to provide information to the receiver. For example, when the prior belief that the sender is reliable is high, the sender knows that she can persuade the receiver with a relatively uninformative experiment. Thus, to induce the sender to choose the most informative experiment in this case, the receiver is more likely to also have to rely on strategic ignorance that increases the sender's payoff. If the sender's prior belief is sufficiently high, the receiver must also demand a less informative experiment. Put differently, on the equilibrium path, the receiver exploits the strategic sender's incentive to provide more information to a receiver with (more) doubts about the sender's reliability.

Because more information is always better for the receiver, given any experiment chosen by the sender, the receiver's sequentially rational investigation fully reveals the sender's reliability. However, always finding out reliability induces the sender to provide less information that exactly offsets the receiver's benefit from finding out. In other words, the receiver does not benefit from the ability to investigate reliability if he cannot commit to ignorance.

While the receiver's optimal investigation strategy is not sequentially rational for the receiver, it would be sequentially rational for a purely adversarial third party except possibly for the investigation chosen on the equilibrium path. I show that, as long as the prior belief that the sender is reliable is sufficiently high, the receiver benefits from delegating investigations to a purely adversarial third party. I further allow the third party's preference to be a linear combination of the (negative of the) sender's payoff and the receiver's payoff. I find that the receiver can do better if the adversarial third party also cares about the receiver's payoff in this way. Such balanced preferences of the third party help the receiver because they make it sequentially rational for the third party to conduct fully revealing investigations (that are never optimal for a purely adversarial third party) whenever the sender's choice of an experiment is sufficiently informative. In other words, delegating investigations to a third party with appropriately balanced preferences enables the receiver to credibly commit to finding out the sender's reliability on the equilibrium path, while punishing the sender for deviating to any would-be profitable experiments. Moreover, the receiver's payoff from delegating to such a third party coincides with his payoff under commitment if the sender is a priori sufficiently likely to be unreliable.

Related literature This chapter contributes to the literature on strategic communication.[3] In addition to a payoff-relevant state, I introduce uncertainty about the sender's (manipulative) behaviour in a probabilistic manner akin to models that relaxes the "commitment assumption" in Bayesian persuasion (Kamenica and Gentzkow, 2011) models (e.g., Frechette, Lizzeri and Perego, 2019; Min, 2021; and Lipnowski, Ravid and Shishkin, 2022).[4] The distinguishing feature of my model is that, in effect, there is a

[3]See surveys by Sobel (2013); Özdogan (2016); Kamenica (2019); Bergemann and Morris (2019); Forges (2020).

[4]The commitment assumption in Bayesian persuasion refers to the sender's ability to commit to a communication strategy. Equivalently, it is the assumption that the sender will truthfully communicate

second sender who can design information (only) about the uncertainty regarding the sender's behaviour. The type of information that the second sender provides, which is not about the payoff-relevant state, also sets this chapter apart from the existing literature on multiple senders (e.g., Gerardi and Yariv, 2008; Che and Kartik, 2009; Gentzkow and Kamenica, 2017*a*; Gentzkow and Kamenica, 2017*b*; and Dworczak and Pavan, 2022), on cheap-talk games in which the receiver can design information about the payoff-relevant state (Ivanov, 2010; Krähmer, 2021; and Ivanov and Sam, 2022), or on communication games in which the receiver can learn about the veracity of the sender's messages (e.g., Dziuda and Salas, 2018; Balbuzanov, 2019; Ederer and Min, 2022; Levkun, 2022; and Sadakane and Tam, 2022).[5]

That ignorance, or avoidance of information, can be beneficial for strategic reasons has been observed in other contexts.[6] Taylor and Yildirim (2011) study whether a decision-maker prefers an informed review of an agent's effort (i.e., the agent's type, as well as a signal about his effort, is observed) or a blind review (i.e., only a signal about the agent's effort is observed). They show that the decision-maker may prefer blind reviews as they can provide better strategic incentives to the agent. McAdams (2012) shows how bidders may prefer not learning their values prior to a second-price auction to deter others from bidding. Roesler and Szentes (2017) show that a buyer can secure more favourable terms from a monopolist by committing to a single, partially revealing signal about the value of the good. Similarly, in a disclosure game, Onuchic (2022) shows that the buyer can obtain more information about the value of the good from the seller when there is uncertainty about the sender's profitability from the sale of the good (i.e., knowing about the sender's profitability from the sale is not beneficial for the buyer).

realisation of a chosen statistical experiment to the receiver.

[5] Veracity refers to whether the sender uses messages in a way consistent with an exogenously given meaning of messages (e.g., whether the sender sends a message "good" or "bad" when the state is in fact bad).

[6] Golman, Hagmann and Loewenstein, 2017 provides a recent survey.

This chapter demonstrates similar results in a sender-receiver game with unrestricted, costless communication while also allowing the receiver to vary his signal (i.e., investigation) based on the sender's choice of an experiment.[7]

Following Schelling (1960), strategic delegation has been studied as a way to mitigate or eliminate commitment issues in many contexts, including industrial organisations (Vickers, 1985; Fershtman and Judd, 1987; Sklivas, 1987) and macroeconomics (Rogoff, 1985). I demonstrate that delegation can also be helpful in a strategic information game with uncertainty about the sender's reliability. I also bring new arguments based on strategic information considerations to literature that compares approaches to evidence across legal systems (Shin, 1998; Dewatripont and Tirole, 1999; Posner, 1999), complementing a recent contribution by Lichtig (2020). The model also brings new insight into how audits can incentivise companies to provide more information in equilibrium.[8] In contrast to existing models, in my model, audits are costless and transfers are not allowed (e.g., Townsend, 1979; Mookherjee and Png, 1989; Border and Sobel, 1987), and audits are about the sender's reliability type and not about the veracity of the sender's messages.

This chapter is most closely related to Lichtig (2020) who studies a verifiable disclosure game in which two senders move sequentially and the latter could detect undisclosed information by the first sender with some probability. He finds that the receiver sometimes prefers an adversarial second sender. In my model, I do not restrict the sender's message to be verifiable, and I also consider an adversarial second sender who also cares about the receiver's payoff. Importantly, in contrast to Lichtig (2020), I derive the receiver's optimal commitment payoff, characterise how commitment can

[7] As I discuss later, the receiver generally benefits from a partially revealing investigation even without the ability to vary investigations based on the sender's choice of an experiment. The latter restriction is imposed by both Taylor and Yildirim (2011) and Roesler and Szentes (2017) as a form of limited commitment on the part of the agent.

[8] See Ye (2021) for a survey of economic models that describe the role of audits.

help the receiver, and compare the receiver's payoff from delegation with the optimal commitment payoff.

The remainder of the chapter is structured as follows. Section 2.1 provides a simplified version of the model that highlights the intuition for my main results. In section 2.2, I set out the strategic communication game that I study and the definitions of equilibria in cases where the receiver can and cannot commit to investigation strategies. I also explain how I simplify the problem of finding equilibrium strategies and payoffs. In section 2.3, I characterise the receiver's equilibrium payoff under commitment, while also showing that commitment is necessary for the receiver to benefit from investigations. Section2.4 contains the results on how the receiver benefits from delegating investigations to a third party. In section 2.5, I discuss how my results shed light on the efficacy of cross-examination of witnesses, benefits of *ad hominem* arguments, and how audits can incentivise the auditee to provide more information in equilibrium. I also discuss the robustness of the results with respect to certain alternative assumptions and some extensions. Section 2.6 gives a conclusion.

2.1 A simple example

This section explains how the receiver can strategically use ignorance to induce the sender to reveal more information using a simplified version of the model. I describe the intuition in the context of a seller of a financial asset persuading an institutional investor to buy the asset using an investment appraisal. Investigations can therefore be thought of as the investor conducting an audit of the appraisal. This section intentionally borrows from the courtroom example in Kamenica and Gentzkow (2011).

Consider an institutional investor (receiver, he) deciding whether to *buy* or *not buy* an asset. There are two states of the world: the asset is either *good* or *bad*. The investor gets utility 1 for making the correct investment decision (*buy* when *good* and *not buy* when *bad*) and 0 for making the wrong decision (*buy* when *bad* and *not buy* when *good*). In contrast, the seller of the asset (sender, she) wants the investor to buy the asset regardless of the state. Assume she gets utility 1 from the investor buying and 0 from not buying. The investor and the seller share a prior belief 0.3 that the asset is *good* meaning that the investor would not buy the asset under the prior belief. To persuade the investor to buy, the seller provides an investment appraisal (e.g., as part of the prospectus) to the investor that sets out an appraisal method as well as the result of the appraisal analysis. Formally, an appraisal method is a signal structure, $\xi(\cdot|good)$ and $\xi(\cdot|bad)$, specifying a distribution over realisations of the appraisal conditional on the state. An appraisal is then a pair consisting of an appraisal method, $\xi = (\xi(\cdot|good), \xi(\cdot|bad))$ and the result. Suppose that there are only two possible results of the appraisal: g (meaning good) or b (meaning bad) and that the seller can only select between three appraisal methods: the "highly" informative (ξ^H), the "mildly" informative (ξ^M), and the "least" informative (ξ^L) methods. In particular, ξ^H fully reveals the quality of the asset. In contrast, while ξ^M and ξ^L reveal that the asset is good, they do not always reveal that the asset is bad. Specifically, ξ^M and ξ^L are unable to detect that the asset is bad with probabilities $\frac{1}{7}$ and $\frac{3}{7}$, respectively. The three appraisal methods are:

$$\xi^H(g|good) = 1, \quad \xi^M(g|good) = 1, \quad \xi^L(g|good) = 1,$$
$$\xi^H(g|bad) = 0, \quad \xi^M(g|bad) = \frac{1}{7}, \quad \xi^L(g|bad) = \frac{3}{7}.$$

Observe that the three methods differ only on their rate of false positives (i.e., the probability that the appraisal result is g when the asset is *bad*). Thus, the investor, who prefers appraisal methods that are more informative, strictly prefers ξ^H over ξ^M over ξ^L. In contrast, the seller, who wishes to maximises the probability that investor buys the asset, prefers to select the least informative method that can persuade the investor into buying. Having chosen a method $\xi \in \{\xi^H, \xi^M, \xi^L\}$, with probability 0.2, the seller is unreliable and can falsify the result of the analysis to her benefit. Because only the result g could persuade the investor to buy the asset, given the three possible appraisal methods, the seller always manipulates the appraisal result to be g when she can. On the other hand, with probability 0.8, the seller is reliable and cannot manipulate the result so that the true result of the appraisal method ξ is communicated to the investor. The investor is unable to tell whether the appraisal result he observes has been falsified without conducting an audit.

No audit versus full audit Let us first compare the case in which the investor never audits the appraisal and the case in which the investor always conduct a fully-revealing audit. Without an audit, after the seller has chosen $\xi \in \{\xi^H, \xi^M, \xi^L\}$, whenever the investor observes the result g, he believes that g was drawn according to ξ with probability 0.8 and that g was chosen independently of the state with probability 0.2. The seller's (ex ante) payoff from a method is the probability that the investor buys the asset, and her payoffs from methods (ξ^H, ξ^M, ξ^L) are $(0.44, 0.52, 0)$, respectively.[9] Hence, the seller selects ξ^M when the investor cannot investigate the seller.

Suppose now that the investor always finds out whether the seller is reliable using a fully revealing audit. Whenever the investor finds out that the seller is unreliable, the investor ignores the appraisal and does not buy the asset. Alternatively, when the investor finds out that the seller is reliable, then the investor knows that the appraisal result was obtained using the stated method and the investor can be persuaded to buy the asset. The

[9]Bayes rule gives that the investor's posterior beliefs after seeing g under methods (ξ^H, ξ^M, ξ^L) are $(\frac{15}{22}, \frac{15}{26}, \frac{15}{34})$, respectively. Thus, the investor would not buy if the seller selects ξ^L, and the seller's payoff from this method is 0. When $\xi \in \{\xi^H, \xi^M\}$, the seller's payoff equals the probability that the investor observes g: $0.2 + 0.8(0.3\xi(g|good) + 0.7\xi(g|bad)))$.

seller's payoffs from methods (ξ^H, ξ^M, ξ^L) are $(0.24, 0.32, 0.48)$, respectively.[10] Therefore, the seller now selects ξ^L so that, from the receiver's perspective, finding out the seller's reliability leads the seller to select a worse (for the investor) appraisal method.[11]

Partially ignorant audits as punishments Let us now consider how the investor can attain the ideal outcome—i.e., finding out whether the seller is reliable after she has chosen ξ^H—by avoiding learning about the seller's reliability. Specifically, suppose that the investor chooses an *audit strategy* that specifies the audit that will be conducted as a function of the seller's choice of an appraisal method. Note first that the seller's payoff from the investor's ideal outcome is 0.24, which is the product of the two prior beliefs. To attain this outcome, the investor must ensure that the seller's payoffs from selecting ξ^M and ξ^L are lower than 0.24. Observe that the investor can simply select not to audit the seller when ξ^L is selected in which case the seller's payoff from selecting ξ^L is $0 < 0.24$. To prevent the seller from selecting ξ^M, the investor can conduct an audit that reveals that the seller is unreliable with probability 1, but only reveal that the seller is reliable with some probability $p \in [0, 1)$.[12] That $p < 1$ implies that the investor does not always find out whether the seller is reliable, and this type of audit makes it less likely that the investor is persuaded to buy the asset. In particular, letting $p = 0.5$ minimises the seller's payoff from selecting ξ^M to $0.16 < 0.24$.[13] Hence, the investor can obtain the

[10]Bayes rule gives that the investor's posterior beliefs after seeing g for methods (ξ^H, ξ^M, ξ^L) are $(1, \frac{3}{5}, \frac{1}{2})$, respectively. Hence, the investor buys upon seeing g and finding out that the seller is reliable. The seller's payoff is therefore given by $0.8[0.3\xi(g|good) + 0.7\xi(g|bad)]$ for each $\xi \in (\xi^H, \xi^M, \xi^L)$.

[11]In this example, the investor is strictly worse off when finding out the seller's reliability. However, without the specific restrictions on the seller's appraisal methods (e.g., if the seller can also select a method, such that $\xi(g|good) = 1$ and $\xi(g|bad) = \frac{2}{7}$), the investor is indifferent between finding out the seller's reliability and not finding out.

[12]In this example, an audit is a binary-support signal structure about the sender's reliability "type". Hence, an audit can be characterised by a pair $(p, q) \in [0, 1]^2$ with $p \geq 1 - q$, where p (resp. q) is the probability that the audit reveals that the sender is reliable (resp. unreliable) when she is reliable (resp. unreliable). A fully revealing audit is the pair $(1, 1)$ and a fully non-revealing audit is the pair $(\frac{1}{2}, \frac{1}{2})$.

[13]When $p = 0.5$, the investor's posterior belief that the seller is reliable conditional on seeing g is either 1 (with probability $\frac{2}{5}$) or $\frac{2}{3}$ (with probability $\frac{3}{5}$). Given ξ^M, the investor's posterior belief that the asset is

ideal outcome by using ignorance—in the form of avoiding learning that the seller is reliable or avoiding learning altogether—as a way to punish the seller for selecting ξ^M and ξ^L.

Delegating audits Once the seller has selected a method, there is no strategic advantage to the investor from being ignorant. Hence, the sequentially rational audit for the investor always fully reveals the seller's reliability. It therefore follows that, for the investor to directly implement the audit strategy that induces the ideal outcome as described in the previous paragraph, the investor would have to commit to being (sometimes) ignorant. However, even without the ability to commit to an investigation strategy, the investor can still induce the seller to select the most informative method, ξ^H, by delegating the audit to a third party with different preferences.

For example, suppose that the third party is purely adversarial to the seller. Given any appraisal method, the sequentially rational audit for such an adversary is to minimise the probability that the investor buys; i.e., the adversary's sequentially rational audit is maximally punishing. When audits are sequentially rationally conducted by an adversary, the seller's payoffs from methods (ξ^H, ξ^M, ξ^L) are $(0.16, 0.16, 0)$, respectively.[14] Hence, it is possible for the investor to induce the seller to select ξ^H by delegating audits to a purely adversarial third party. However, delegating audits to a pure adversary does not induce the ideal outcome because the investor never fully learns the seller's reliability. More generally, it can be shown that the investor can attain the ideal outcome if the third party is not only adversarial to the seller but also cares

good after observing g and he believes that the seller is reliable with probabilities 1 and $\frac{2}{3}$ are $\frac{3}{4}$ and $\frac{1}{2}$, respectively. Assuming that the investor break ties by not buying (alternatively, the investor sets p to be slightly above 0.5), then the investor only buys the asset after observing g when his belief that the seller is reliable is 1. The seller's payoff is thus $\frac{2}{5}[0.3\xi(g|good) + 0.7\xi(g|bad)] = 0.16$.

[14]The maximally punishing audit following ξ^M and ξ^L are as in the audit strategy that induces the ideal outcome as described in the previous paragraph. Maximally punishing audit following ξ^H involves setting $p = \frac{2}{3}$ assuming that the investor break ties by not buying (alternatively, the investor sets p to be slightly above $\frac{2}{3}$).

about the investor's payoff,[15] and the prior belief that the seller is reliable is sufficiently low.

Partially ignorant audits as rewards Suppose now that the probability that the seller is unreliable is 0.1. The seller's payoff from the investor's ideal outcome is then 0.27. It turns out that the lowest payoff that the investor can induce by being strategically ignorant when the seller chooses ξ^M is $0.28 > 0.27$.[16] Thus, it is no longer possible to induce the seller to choose ξ^H while also finding out the seller's reliability. To induce the seller to choose ξ^H, the investor can strategically use ignorance differently by choosing investigations that do not fully reveal that the seller is unreliable; i.e., investigations that reveal that the seller is reliable with probability 1, but reveal that the seller is unreliable with probability $q \in [0, 1)$. Then, letting $q \leq \frac{2}{3}$ ensures the seller's payoff from choosing ξ^H is greater than 0.28.[17] Thus, in this case where the seller is less likely to be unreliable, to obtain his ideal outcome, the investor strategically uses ignorance as a way to reward the seller for selecting ξ^H as well as a way to punish the seller for selecting ξ^M and ξ^L.

However, when the probability that the seller is unreliable is even lower (e.g., 0.05), it becomes impossible for the investor to ensure that the seller's payoff from choosing ξ^H is greater than her payoff from choosing ξ^M. In this case, the investor's best option is to demand ξ^M and find out whether the seller is reliable (while not auditing ξ^L).

Beyond the simple example In what follows, I generalise the example by allowing the seller (i.e., the sender) to choose any appraisal method (i.e., a statistical experiment)

[15]In particular, I consider a third party whose preference is a linear combination of the seller's and the investor's payoffs.

[16]The seller's payoff when $p < \frac{7}{9}$ is $0.9p[0.3\xi(g|good) + 0.7\xi(g|bad)]$.

[17]With $q \in [0, 1)$, the investor's posterior belief that the seller is reliable conditional on seeing g is either $\frac{0.9}{0.9+0.1(1-q)}$ (with probability $0.9 + 0.1(1 - q)$) or 0 (with probability $0.1q$). Given ξ^H, the investor's posterior belief that the asset is good after observing g and he believes that the seller is reliable with probabilities $a(q) = \frac{0.9}{0.9+0.1(1-q)}$ and 0 are $\frac{0.3}{0.3+0.7(1-a(q))} > \frac{1}{2}$ and $0.3 < \frac{1}{2}$, respectively. Hence, the investor buys after observing g when his belief that the seller is reliable is 1. The seller's payoff is thus: $0.3[0.9 + 0.1(1 - q)]$.

about the state and allowing the receiver (i.e., the investor) to choose any audit (i.e., investigation). I show that assuming that the sender selects among all "one-sided" experiments (i.e., appraisals that only has a possibility of false positives as in the methods considered in the example) is without loss. I use this simplification to prove that the two uses of ignorance as punishment for the sender choosing "bad" experiments and possibly as a reward for choosing the "good" experiment carry over to the general case. I also study the extent to which the investor can implement commitment outcomes by delegating investigations to a third party whose preference is a linear combination of the sender's and the receiver's payoffs. In particular, I show that, when the receiver's optimal investigation strategy is fully revealing on the equilibrium path (e.g., when the probability that the seller is unreliable is 0.2 in the example above), the receiver can obtain the ideal outcome by delegating investigations to an appropriate third party.

2.2 A persuasion game with doubts about the sender's reliability

2.2.1 Set up

There are two players: a *Sender* (S) and a *Receiver* (R). The Receiver can take one of two actions, denoted $a \in A := \{0,1\}$, and the Receiver's payoff from each action depends on the binary states of the world, $\theta \in \Theta := \{0,1\}$. The preferences are such that the Receiver's optimal action is $a = 1$ (i.e., to *take action*) whenever he believes that the state is $\theta = 1$ with probability at least $\mu^* \in (0,1)$; otherwise the Receiver's optimal action is to choose $a = 0$ (i.e., to *not take action*). The Sender would like the Receiver to choose $a = 1$ no matter the state. Let $u_S : A \to \mathbb{R}$ and $u_R(a,\theta) : A \times \Theta \to \mathbb{R}$ denote theSender and the Receiver's payoffs, respectively. I normalise the Receiver's payoff from choosing $a = 0$ to be zero and his payoff from choosing $a = 1$ when $\theta = 0$ to be -1. Specifically, I assume the following payoffs for the players:

$$u_{\mathrm{R}}(a, \theta) := a\frac{\theta - \mu^*}{\mu^*},\ u_{\mathrm{S}}(a) := a.$$

Let $\mu_0 \in (0, 1)$ denote the common prior probability that $\theta = 1$. For brevity, I abuse notation and sometimes use μ_0 to denote the probability measure in $\Delta\Theta$ consistent with the prior belief μ_0.[18] To make the problem interesting, I assume that the Receiver does not take action under the prior belief; i.e.,

$$\mu_0 < \mu^*. \tag{2.1}$$

If this condition does not hold, the Sender has no incentive to provide any information and thus concerns about the reliability of the Sender become moot. Given the normalisations, both the Sender's and the Receiver's default payoffs under the prior beliefs are zero.

To persuade the Receiver to take action, the Sender publicly chooses a signal structure $\xi \in \Xi := (\Delta M)^{\Theta}$, where M is a finite set of messages with at least two elements. I refer to ξ as an *experiment*. There are two types of Senders $T := \{r, n\} \ni t$ and I let $\rho_0 \in (0, 1)$ denote the common prior probability that $t = r$. A *reliable Sender*, $t = r$, truthfully communicates realisations from the experiment ξ to the Receiver. In contrast, an *unreliable Sender*, $t = n$, can communicate any message $m \in M$. The Receiver observes the Sender's message m without observing the Sender's type.

Notice that a reliable Sender can commit to the announced experiment as in Bayesian persuasion models (Kamenica and Gentzkow, 2011).[19] In contrast, the unreliable Sender

[18]Given an arbitrary set X, I use ΔX to denote the set of probability measures on the set X.
[19]Forges (2020) describes Bayesian persuasion as the case in which the statistical experiment chosen by the sender is "fully reliable"

lacks such commitment power. Thus, the higher ρ_0 is, the more that the Sender is committed to the announced experiment. One can also interpret ρ_0 as capturing the imperfectness in the enforcement of truthful communication (Min, 2021), or the possibility that the Sender can indirectly alter the realisation of the experiment by influencing the experimenter that is carrying out of the experiment (Lipnowski, Ravid and Shishkin, 2022). More generally, ρ_0 can be thought of as capturing the conflicting incentives that an experimenter might have in truthfully communicating the results of the experiment to the Receiver.[20] In addition, ρ_0 can also be interpreted as the probability that the experimenter is competent; i.e., that the experimenter is capable of carrying out the experiment. Another interpretation of ρ_0 is that it represents the probability with which the Sender is simply corrupt and alters the result of the experiment.

Importantly, the Receiver can investigate the Sender's type by using any signal structure about the Sender's type. I take the belief-based approach (Kamenica, 2019; Forges, 2020) and express investigations as Bayes-plausible distributions over posterior beliefs about the Sender's type. Thus, an investigation is an element in $\mathcal{I} := \{\iota \in \Delta([0,1]) : \int \rho \mathrm{d}\iota(\rho) = \rho_0\} \ni \iota$. The Receiver's investigation strategy is a mapping from the Sender's choice of an experiment ξ to an investigation i, which I denote as $i : \Xi \to \mathcal{I}$.[21]

I consider two cases that reflect differing assumptions on the ability of the Receiver to commit to an investigation strategy. In the *commitment case*, I allow the Receiver to commit to any investigation strategy. In contrast, in the *no-commitment case*, I assume that

[20] For example, on the one hand, the experimenter may have reputational or moral concerns that guide them towards communicating truthfully. On the other hand, they may also have financial or relational concerns (either via explicit payment or implicit payment in the form of future interactions with the Sender) that guide them towards lying on behalf of the Sender. Under this interpretation, ρ_0 captures the Sender's and the Receiver's common uncertainty about the combined effect of these incentives on the experimenter's behaviour.

[21] As I discuss in section 2.5, allowing the investigation strategy to also depend on the realisation of the Sender's experiment does not alter the result.

the Receiver chooses an investigation after observing ξ. I consider the *delegation case*, in which a third party chooses an investigation after observing the Sender's experiment ξ in section 2.4.

The timing of the game is as follows. In the commitment case, the Receiver first publicly commits to an investigation strategy $i(\cdot)$. The Sender then publicly chooses an experiment $\xi \in \Xi$. Nature then independently draws the state and the Sender's type, $\theta \sim \mu_0$ and $t \sim \rho_0$, respectively.[22] If the Sender is reliable (i.e., $t = r$), then the Sender truthfully communicates m drawn from $\xi(\theta)$. If the Sender is unreliable (i.e., $t = n$), then the Sender chooses $m \in M$ to maximise her payoff without observing the realised θ.[23] Finally, the Receiver observes the realisation of the investigation $\rho \sim i(\xi)$ as well as the message from the Sender m, and optimally chooses an action $a \in A$. All players update beliefs using Bayes rule whenever possible.

In the no-commitment case, the Sender first publicly chooses an experiment and then Receiver chooses an investigation $\iota \in \mathcal{I}$. The rest of the play is the same except that the Receiver now observes the realisation of the investigation $\rho \sim \iota$.

2.2.2 Equilibrium

I define an equilibrium under commitment and no-commitment cases as a weak perfect Bayesian equilibrium (PBE) of an appropriately defined game. Let $\sigma : \Xi \times \mathcal{I} \to \Delta M$ denote the unreliable Sender's messaging strategy, $\alpha : \Xi \times \mathcal{I} \times [0,1] \times M \to \Delta A$ denote the Receiver's action rule, and $\mu : \Xi \times \mathcal{I} \times [0,1] \times M \to \Delta\Theta$ denote the Receiver's belief map. Given a profile $(\xi, \iota, \sigma, \alpha)$, let $V_j(\cdot)$ denote player $j \in \{S, R\}$'s associated ex ante

[22]The results do not change if the Sender observes her type prior to choosing an experiment (see section 2.5).

[23]The results do not change if the unreliable Sender can observe the realised state (see discussion in section 2.5).

payoff, i.e.,

$$
\begin{aligned}
&V_j\left(\xi, \iota, \sigma, \alpha\right) \\
&:= \int_0^1 \sum_{m\in M}\sum_{\theta\in\Theta}\sum_{a\in A} u_j\left(\cdot\right)\alpha\left(a|\xi,\iota,\rho,m\right)\left[\rho\xi\left(m|\theta\right)+\left(1-\rho\right)\sigma\left(m|\xi,\iota\right)\right]\mu_0\left(\theta\right)\mathrm{d}\iota\left(\rho\right),
\end{aligned}
$$

where $\xi(\cdot) = \sum_{\theta\in\Theta}\xi(\cdot|\theta)\mu_0(\theta)$.

An equilibrium under the no-commitment case is a PBE of the extensive-form game in which the Sender moves before the Receiver. A *no-commitment equilibrium* is thus a tuple $(\xi, i, \sigma, \alpha, \mu)$ that satisfies the following conditions.

(i) For each $(\xi', \iota') \in \Xi \times \mathcal{I}$, belief map, $\mu(\xi', \iota')$, is derived by updating μ_0 using the signal structure

$$\rho\xi' + (1-\rho)\,\sigma\left(\xi', \iota'\right) : \Theta \to \Delta M$$

via Bayes rule whenever possible.

(ii) Receiver's action rule, α, is optimal given μ; i.e., for all $(\xi', \iota', \rho, m) \in \Xi \times \mathcal{I} \times [0,1] \times M$,

$$\operatorname{supp}\left(\alpha\left(\cdot|\xi', \iota', \rho, m\right)\right) \subseteq \arg\max_{a\in A}\sum_{\theta\in\Theta} u_{\mathrm{R}}\left(a,\theta\right)\mu\left(\theta|\xi',\iota',\rho,m\right).$$

(iii) Unreliable Sender's messaging strategy, σ, is incentive compatible given α; i.e., for all $(\xi', \iota') \in \Xi \times \mathcal{I}$,

$$\operatorname{supp}\left(\sigma\left(\cdot|\xi', \iota'\right)\right) \subseteq \arg\max_{m\in M}\int\sum_{a\in A} u_{\mathrm{S}}\left(a\right)\alpha\left(a|\xi',\iota',\rho,m\right)\frac{1-\rho}{1-\rho_0}\mathrm{d}\iota'\left(\rho\right). \tag{IC}$$

(iv) Receiver's investigation strategy $i : \Xi \to \mathcal{I}$ is sequentially rational given σ and ρ; i.e., for all $\xi' \in \Xi$, $i(\xi')$ solves

$$\max_{\iota'\in\mathcal{I}} V_{\mathrm{R}}\left(\xi', \iota', \sigma, \alpha\right).$$

(v) Sender's experiment ξ is optimal given i, σ and ρ; i.e., ξ solves

$$\max_{\xi' \in \Xi} V_{\mathrm{S}}\left(\xi', i\left(\xi'\right), \sigma, \alpha\right).$$

In this case, it is without loss to focus on no-commitment equilibria that are preferred by the Sender by the usual argument (Kamenica and Gentzkow, 2011).

An equilibrium with commitment can be thought of as a no-commitment equilibrium without the requirement that the investigation strategy be always sequentially rational, i.e., without requirement (iv). Toward defining the no-commitment equilibrium formally, given an investigation strategy $i(\cdot)$, call a tuple $(\xi, \sigma, \alpha, \mu)$ an i-commitment equilibrium if it satisfies conditions (i), (ii), (iii) and (v), where the qualifier (ξ', ι') is replaced with $(\xi', i(\xi'))$. I define a *commitment equilibrium* as a Receiver-preferred i-commitment equilibrium; i.e., a tuple $(\xi, i, \sigma, \alpha, \mu)$ such that $(\xi, \sigma, \alpha, \mu)$ is an i-commitment equilibrium such that

$$V_{\mathrm{R}}\left(\xi, i\left(\xi\right), \sigma, \alpha\right) = \max_{i', \left(\xi', \sigma', \alpha', \mu'\right)} V_{\mathrm{R}}\left(\xi', i'\left(\xi'\right), \sigma', \alpha'\right)$$

$$\text{s.t. } \left(\xi', \sigma', \alpha', \mu'\right) \text{ is an } i'\text{-equilibrium.}$$

Allowing the Receiver to select an i-equilibrium given any $i(\cdot)$ (as implied by definition above) is without loss of generality and ensures that a solution to the problem above exists.

Observe that, given any no-commitment equilibrium $(\xi, i, \sigma, \alpha, \mu)$, a tuple $(\xi, \sigma, \alpha, \mu)$ is an i-equilibrium; i.e., the Receiver can commit to any no-commitment equilibrium. Hence, the Receiver's commitment-equilibrium payoff is an upper bound on the Receiver's no-commitment-equilibrium payoff.

2.2.3 Simplifying the problem

The following lemma greatly simplifies the analysis allowing me to reduce the Sender's choice of an experiment to a canonical set of experiments and to treat the unreliable Sender as being non-strategic.

Lemma 2.1. *Fix some $m_0, m_1 \in M$. Any commitment and no-commitment equilibrium payoffs are achievable via some $(\xi, \overline{\sigma}) \in \Xi^2$ pair such that* $\text{supp}(\xi) = \{m_0, m_1\}$*, $\xi(m_1|1) = 1$, $\xi(m_1|0) \leq 1 - \underline{\rho}$, and $\overline{\sigma}(m_1|\cdot) = 1$.*

The lemma establishes that it is without loss to focus on experiments in which the Sender makes action recommendations to the Receiver. I denote the recommendation to take action $a \in A$ as m_a so that the unreliable Sender always chooses the message m_1. The key step in proving the lemma is establishing a recommendation principle in games induced by any $(\xi, \iota) \in \Xi \times \mathcal{I}$. Given any $(\xi, \iota) \in \Xi \times \mathcal{I}$, call a tuple (σ, α, μ) an (ξ, ι)-equilibrium if it is a PBE of the game induced by (ξ, ι). The main step of the proof is to show that any tuple (σ, α, μ) that is a (ξ, ι)-equilibrium can be reduced to a payoff equivalent tuple $(\tilde{\sigma}, \tilde{\alpha}, \tilde{\mu})$ that is a $(\tilde{\xi}, \iota)$-equilibrium where $\tilde{\xi} \in \Xi$ makes action recommendations, $\tilde{\sigma} \in \Delta M$ always recommends the Receiver to take action, and ι remains optimal for the Receiver against $\tilde{\xi}$. Unfortunately, the fact that the Receiver is choosing an investigation means that neither existing results nor standard arguments apply to this setting. I therefore provide an original proof in the Appendix and provide a brief sketch here.

Given any $\pi \in \Xi$, let μ^π denote the posterior belief induced by signal structure π. First, observe that $\mu^\sigma = \mu_0$ because the unreliable Sender's strategy cannot depend on the realised state θ. It follows that the unreliable Sender would only send messages that would induce the Receiver to take action if the message is known to have been sent by the reliable Sender; i.e., $\text{supp}(\sigma) \subseteq M_1^\xi := \{m \in M : \mu^\xi(m) > \mu^*\}$. Because $\mu^\sigma(m) = \mu_0 < \mu^*$, for any $m \in M_1^{\xi}$, there exists a threshold belief about the Sender's type, $\overline{\rho}_m \in [0,1]$, at which the Receiver is indifferent between the two actions after observing m, and would only take action if $\rho > \overline{\rho}_m$. The unreliable Sender's payoff is (weakly) decreasing in $\overline{\rho}_m$. Moreover, (IC) implies that $\iota([\overline{\rho}_m, \overline{\rho}_{m'}]) = 0$ for all such $m, m' \in M_1^{\xi}$. Because pooling messages in M_1^{ξ} (in both ξ and σ) results in a threshold that is a weighted average of the cutoffs $\{\overline{\rho}_m\}_{m \in M_1^{\xi}}$, it follows that the unreliable Sender's payoff remains unchanged. This, in turn, implies that both the Sender's and the Receiver's ex ante payoffs are unaffected when pooling messages in M_1^{ξ}. That the unreliable Sender never sends messages in $\{m \in M : \mu^{\xi}(m) \leq \mu^*\}$ means that these messages can also be pooled without affecting payoffs. Let $\tilde{\xi}$ and $\tilde{\sigma}$ denote the strategies after pooling. The proof is completed by showing that the pooling of messages does not affect the choice of investigation. Specifically, I show that if there exists $\tilde{\iota} \in \mathcal{I}$ such that the Receiver's $(\tilde{\xi}, \tilde{\iota})$-equilibrium payoff is different from his $(\tilde{\xi}, \iota)$-equilibrium payoff, then one can construct a $(\xi, \tilde{\iota})$-equilibrium in which the Receiver's payoff is the same as in the $(\tilde{\xi}, \tilde{\iota})$-equilibrium.

The class of experiments in Lemma 2.1 can be summarised by a single parameter $\widehat{\rho} \in [\underline{\rho}, 1]$, where $\underline{\rho} := \frac{\mu^* - \mu_0}{\mu^*(1-\mu_0)}$ by defining $\xi_{\widehat{\rho}} : [\underline{\rho}, 1] \times \Theta \to \Delta(\{m_0, m_1\})$ as

$$\xi_{\widehat{\rho}}(m_1|1) = 1, \ \xi_{\widehat{\rho}}(m_1|0) = 1 - \frac{1}{\widehat{\rho}} \underline{\rho}.$$

Observe that any pair $(\xi_{\widehat{\rho}}, \overline{\sigma})$ induces posterior beliefs about θ, $\mu^{(1-\rho)\overline{\sigma} + \rho \xi_{\widehat{\rho}}}(\cdot)$, that are either 0 or some $\mu > \mu_0$, where μ depends on $\widehat{\rho}$. In fact, $\widehat{\rho} \in [\underline{\rho}, 1]$ is the threshold belief about the Sender's reliability above which the Receiver takes action after observing m_1; i.e.,

$$\mu^{(1-\rho)\overline{\sigma} + \rho \xi_{\widehat{\rho}}}(m_1) \geq \mu^* \Leftrightarrow \rho \geq \widehat{\rho}. \tag{2.2}$$

In what follows, I refer to the Sender's choice of an experiment by its associated threshold

$\widehat{\rho} \in [\underline{\rho}, 1]$ (instead of $\xi_{\widehat{\rho}}$). I also refer to commitment and no-commitment equilibria using $\widehat{\rho}$. With slight abuse of notation, I say that $(\widehat{\rho}, i)$ is a (resp. no-) commitment equilibrium if $(\xi_{\widehat{\rho}}, i)$ is a (resp. no-) commitment equilibrium, where $i : [\underline{\rho}, 1] \to \mathcal{I}$. I call a fully revealing investigation, $\rho_0\delta_1 + (1-\rho_0)\delta_0 \in \mathcal{I}$, a *full investigation* and say that the Receiver does not investigate if the Receiver chooses $\delta_{\rho_0} \in \mathcal{I}$. I refer to all other investigations as *partial investigations*.

Notice that ξ_1 is the Sender-optimal experiment when the Sender is known to be reliable (i.e., $\rho_0 = 1$).[24] I therefore refer to ξ_1 as the *Sender-optimal full-reliability experiment*. Since $\xi_{\underline{\rho}}$ is the fully informative experiment, $\widehat{\rho}$ uniquely identifies an experiment that is a weighted average of ξ_1 and $\xi_{\underline{\rho}}$, and a lower $\widehat{\rho} \in [\underline{\rho}, 1]$ corresponds to a more informative experiment within the class of experiments identified in Lemma 2.1.

I now derive the players' continuation payoffs as a function of the Sender's choice of threshold $\widehat{\rho} \in [\underline{\rho}, 1]$ and the Receiver's choice of an investigation $\iota \in \mathcal{I}$. Towards this goal, consider first the Receiver's payoff from $\widehat{\rho} \in [\underline{\rho}, 1]$ while fixing his belief about the Sender's reliability at some $\rho \in [0, 1]$. The Receiver's payoff is zero if $\rho \le \widehat{\rho}$ because the Sender cannot induce the Receiver to take action in this case. If $\rho > \underline{\rho}$, the Receiver takes action after observing m_1 and obtains a payoff $\frac{1-\mu^*}{\mu^*}$ if $\theta = 1$ and -1 if $\theta = 0$. Thus, the Receiver's payoff is given by

$$\begin{aligned}
&\left(\frac{1-\mu^*}{\mu^*}\right)\left[\rho\xi_{\widehat{\rho}}(m_1|1) + (1-\rho)\right]\mu_0 + (-1)\left[\rho\xi_{\widehat{\rho}}(m_1|0) + (1-\rho)\right](1-\mu_0) \\
&= \frac{\mu^* - \mu_0}{\mu^*}\left(\frac{\rho}{\widehat{\rho}} - 1\right).
\end{aligned}$$

Notice that the Receiver obtains a payoff of 0 under the Sender-optimal full-reliability experiment. As ρ increases, the probability of observing m_1 when $\theta = 1$ increases so that

[24] Since the model reduces to that of Bayesian persuasion when $\rho_0 = 1$, ξ_1 is the Sender-optimal experiment under Bayesian persuasion.

the Receiver's payoff increases linearly. In contrast, the Receiver's payoff is decreasing in $\widehat{\rho}$ (pointwise) as higher $\widehat{\rho}$ increases the probability of observing m_1 when $\theta = 0$; i.e., the Receiver prefers more informative experiments. The continuation payoffs for the Receiver as a function of $\widehat{\rho} \in [\underline{\rho}, 1]$ and $\iota \in \mathcal{I}$ can be written as

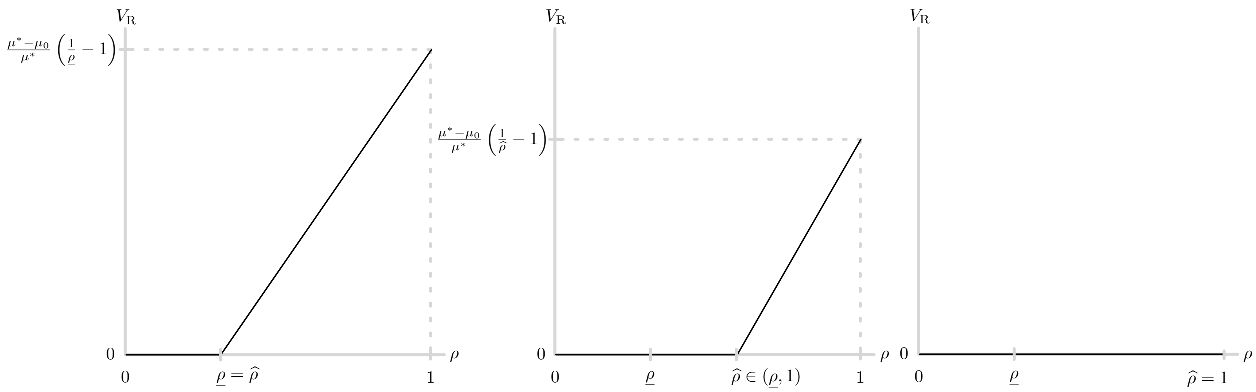
$$V_{\mathrm{R}}\left(\widehat{\rho}, \iota\right) = \int_0^1 \mathbb{1}_{\{\rho > \widehat{\rho}\}} \frac{\mu^* - \mu_0}{\mu^*} \left(\frac{\rho}{\widehat{\rho}} - 1\right) \mathrm{d}\iota\left(\rho\right). \tag{2.3}$$

The Receiver attains his maximal payoff when $\widehat{\rho} = \underline{\rho}$ and ι is fully revealing.

Once the Sender has chosen an experiment $\widehat{\rho}$, the Receiver's problem of choosing $\iota \in \mathcal{I}$ is to design information about ρ. Moreover, because ρ is a belief about the binary type of the Sender, the problem can be studied graphically (Aumann and Maschler, 1968; Kamenica and Gentzkow, 2011). With this in mind, it is helpful to consider the players' payoffs as a function of ρ while fixing $\iota = \delta_\rho$ for the three cases: $\widehat{\rho} = \underline{\rho}$, $\widehat{\rho} \in (\underline{\rho}, 1)$, and $\widehat{\rho} = 1$.[25] Figure 2.1 plots the Receiver's payoffs in these cases. Observe that the Receiver's payoff is continuous and convex in ρ.

[25] δ_ρ is a Dirac measure at $\rho \in [0,1]$ on $\Delta([0,1])$.

Figure 2.1: Receiver's continuation payoff, $V_{\mathrm{R}}(\widehat{\rho}, \delta_\rho)$.



The figure shows the Receiver's continuation payoff (as a function of ρ) when the Receiver's posterior belief that the Sender is reliable is degenerate at ρ and the Sender has chosen (from left to right): $\widehat{\rho} = \underline{\rho}$ (the fully informative experiment), $\widehat{\rho} \in (\underline{\rho}, 1)$, and $\widehat{\rho} = 1$ (the Sender-optimal full-reliability experiment).

Let us now consider the Sender's payoff from $\widehat{\rho} \in [\underline{\rho}, 1]$ while fixing the Receiver's belief about the Sender's type at some $\rho \in [0,1]$. Recall from (2.2) that the Sender cannot induce the Receiver to take action if $\rho < \widehat{\rho}$. If $\rho > \widehat{\rho}$, by (2.2), the message m_1 induces the Receiver to take action so that the Sender's payoff when $\rho > \widehat{\rho}$ is simply the probability that the Receiver observes m_1:

$$\rho\xi_{\widehat{\rho}}(m_1) + (1-\rho)\,\overline{\sigma}(m_1) = \rho\xi_{\widehat{\rho}}(m_1) + (1-\rho)\cdot 1 = 1 - \frac{\mu^* - \mu_0}{\mu^*}\frac{\rho}{\widehat{\rho}}.$$

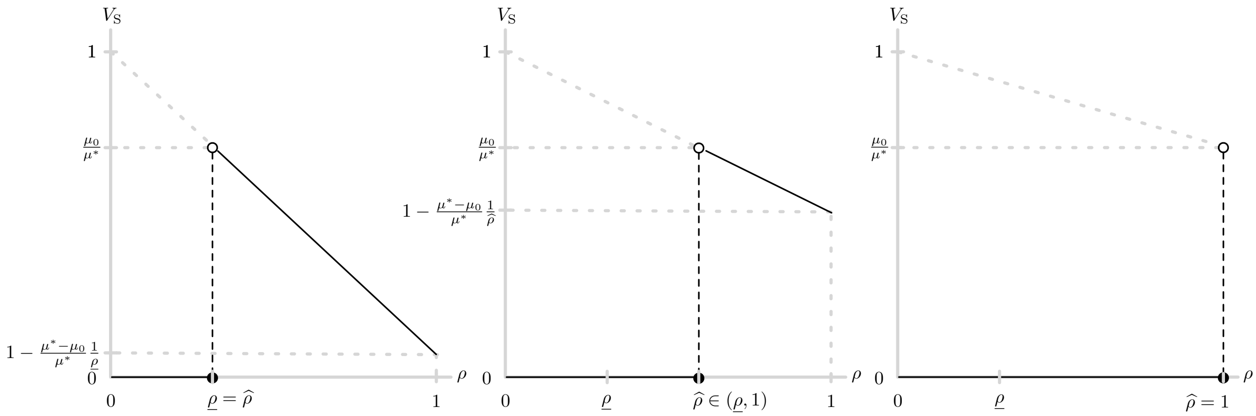
The Sender's payoff under the Sender-optimal full-reliability experiment is $\frac{\mu_0}{\mu^*}$. I refer to this payoff as the Sender-optimal full-reliability payoff. As ρ increases, the probability of observing m_1—and thus the Sender's payoff—falls linearly because a greater weight is put on $\xi_{\widehat{\rho}}(m_1)$ relative to $\overline{\sigma}(m_1) = 1 > \xi_{\widehat{\rho}}(m_1)$. As $\widehat{\rho}$ increases, the probability of observing m_1 increases so that the Sender's payoff when $\rho > \widehat{\rho}$ is increasing in $\widehat{\rho}$; in other words, the Sender prefers less informative experiments in this case. Assuming that the Sender break ties by not taking action,[26] the continuation payoffs for the Sender as a function of $\widehat{\rho} \in [\underline{\rho}, 1]$ and $\iota \in \mathcal{I}$ can therefore be written as

$$V_{\mathrm{S}}\left(\widehat{\rho}, \iota\right) = \int_0^1 \mathbb{1}_{\{\rho > \widehat{\rho}\}} \left(1 - \frac{\mu^* - \mu_0}{\mu^*} \frac{\rho}{\widehat{\rho}}\right) \mathrm{d}\iota\left(\rho\right). \tag{2.4}$$

Figure 2.2 plots the Sender's payoff fixing $\iota = \delta_\rho$ for the three cases: $\widehat{\rho} = \underline{\rho}$, $\widehat{\rho} \in (\underline{\rho}, 1)$, and $\widehat{\rho} = 1$. In general, the Sender's payoff is piecewise linear but is neither convex nor concave in ρ, and it has a discontinuity at $\rho = \widehat{\rho}$ when the Sender's payoff "jumps" from zero to the Sender-optimal full-reliability payoff $\frac{\mu_0}{\mu^*}$, and decreases linearly as ρ increases. Importantly, the Sender's payoff at $\rho = 1$ (i.e., when the Receiver is certain that the Sender is reliable) is increasing in the threshold $\widehat{\rho}$ because the Sender prefers less informativeness experiments (i.e., a higher $\widehat{\rho}$). Observe that, for any $\rho_0 \in (\widehat{\rho}, 1]$, the Sender's maximal payoff is achieved by any $\iota \in \mathcal{I}$ such that $\operatorname{supp}(\iota) \subseteq (\widehat{\rho}, 1]$ and in particular, it can be achieved by $\iota = \delta_{\rho_0}$; i.e., fixing $\widehat{\rho}$, the Sender's payoff is highest when the Receiver does not investigate the Sender's reliability.

[26] If the Sender break ties by taking action, then the strict inequality in the condition the indicator function in (2.4) is replaced with a weak inequality.

Figure 2.2: Sender's continuation payoff, $V_{\mathrm{S}}(\widehat{\rho}, \delta_\rho)$.



The figure shows the Sender's continuation payoff (as a function of ρ) when the Receiver's posterior belief that the Sender is reliable is degenerate at ρ and the Sender has chosen (from left to right): $\widehat{\rho} = \underline{\rho}$ (the fully informative experiment), $\widehat{\rho} \in (\underline{\rho}, 1)$, and $\widehat{\rho} = 1$ (the Sender-optimal full-reliability experiment).

2.3 Optimal ignorance about the sender's reliability

2.3.1 No-commitment case

Consider first the case in which the Receiver cannot commit to investigation strategies; i.e., the Receiver chooses an investigation after observing the Sender's choice of an experiment. In this case, there is no strategic consideration for the Receiver when choosing an investigation. Thus, the standard argument that more information is always better means that the Receiver's optimal investigation is fully revealing. Moreover, because the Receiver ignores the Sender's message whenever he learns that the Sender is unreliable, the Sender's optimal experiment in the no-commitment case corresponds to the case optimal experiment when she is known to be fully reliable; i.e., the Sender-optimal full-reliability experiment, $\widehat{\rho} = 1$, is also optimal for the Sender in the no-commitment case. The following result is then immediate.

Theorem 2.1. *In any no-commitment equilibrium, the Sender chooses her optimal experiment when she is known to be fully reliable,* $\widehat{\rho}^* = 1$*, and the Receiver always conducts the fully revealing investigation,* $i^* = \rho_0\delta_1 + (1-\rho_0)\delta_0$*. The Sender's no-commitment equilibrium payoff is* $\rho_0\frac{\mu_0}{\mu^*}$ *(i.e., the prior probability that the Sender is reliable times the Sender-optimal full-reliability payoff), and the Receiver's no-commitment equilibrium payoff is zero.*

To get some intuition, compare the result above with the case in which the Receiver cannot investigate the Sender's reliability (i.e., $i = \delta_{\rho_0}$). When the Receiver cannot learn, it is clear from the Sender's payoff (2.4) that the optimal experiment for the Sender is to set $\widehat{\rho} = \rho_0$, which yields the Sender a payoff of $\frac{\mu_0}{\mu^*}$ and the Receiver a payoff of zero. The Receiver's payoff is zero because the Sender provides "just enough" information (i.e., the smallest $\widehat{\rho}$) such that the message m_1 would induce the Receiver to take action and no more. In contrast, when the Receiver fully learns, while the Receiver benefits from learning about Sender's reliability, this benefit is exactly offset by the Sender providing less information (i.e., $\widehat{\rho} = 1 > \rho_0$) in the first place.[27] Combining with the observation that the Receiver's payoff is zero without any information from the Sender gives the corollary below. In particular, the corollary means that the Receiver is no better when he can (fully) learn about the Sender's reliability.

Corollary 2.1. *The Receiver's payoff is zero in each of the following cases: (i) the Sender provides no information; (ii) the Receiver cannot investigate; and (iii) the Receiver chooses investigation sequentially rationally (i.e., no-commitment case).*

Finally, notice that the Sender choosing the Sender-optimal full-reliability experiment, $\widehat{\rho} = 1$, and the Receiver choosing a full investigation is an equilibrium in the

[27]That the Receiver is no better off when fully learning relies on the fact that the Receiver's action is binary (see Proposition 5 in Kamenica and Gentzkow, 2011). Moreover, the same argument means that even if there was an upper bound on the informativeness of the Receiver's investigations, the Receiver's no-commitment equilibrium payoff would still be zero—the Receiver without the ability to commit to ignorance would continue to investigate to the full extent possible and the Sender can provide less information to exactly offset the Receiver's benefit from learning about reliability.

simultaneous-move version of the game.

2.3.2 Commitment case

Towards characterising the Receiver's optimal investigation strategy and the payoff under commitment, let us first consider how committing to ignorance can help the Receiver.[28]

For the Receiver to obtain a strictly positive payoff, the Receiver must strictly prefer to take action over not taking action; i.e., the posterior belief about the Sender's reliability, ρ, must be strictly higher than the Sender's experiment $\widehat{\rho}$. Thus, to benefit from a full investigation that induces posterior beliefs $\rho \in \{0,1\}$, the Receiver must induce the Sender to choose $\widehat{\rho} < 1$. To ensure that the Sender will not choose the Sender-optimal full-reliability experiment $\widehat{\rho} = 1$, the Receiver can commit to an investigation that lowers the Sender's payoff from choosing $\widehat{\rho} = 1$ as a way to punish the Sender from choosing $\widehat{\rho} = 1$. Because a full investigation leads the Sender to choose $\widehat{\rho} = 1$, any investigation that leads the Sender to choose some $\widehat{\rho} < 1$ must be a partial investigation and thus involve ignorance. I call an investigation strategy that maximally punishes the Sender for any choice of $\widehat{\rho} \in [\underline{\rho}, 1]$ as a *punishing investigation strategy* and denote it as $i^{\min}$.

Note that, when the prior belief about the Sender's reliability is below $\underline{\rho}$, even a fully informative experiment is not sufficient to induce the Receiver to take action; i.e.,

$$\mu^{(1-\rho)\overline{\sigma}+\rho\xi_{\widehat{\rho}}}(m_1) < \mu^* \; \forall \rho \in [0, \underline{\rho}). \tag{2.5}$$

Thus, when $\rho_0 \leq \underline{\rho}$, both the Sender's and the Receiver's payoffs are zero. Moreover, by

[28] The commitment problem that the Receiver faces is non-generic in the sense that it is not sufficient for the Receiver to be able to commit to punishing the Sender against finite values of $\widehat{\rho}$. For example, even if the Receiver could commit to punishing the Sender when the Sender chooses $\widehat{\rho} = 1$, the Sender could obtain approximately the same payoff by choosing $\widehat{\rho}$ that is arbitrarily smaller than 1.

definition of $\widehat{\rho}$, the Receiver would not take action if $\rho \leq \widehat{\rho}$. These observations mean that if $\rho_0 \leq \underline{\rho}$ or $\rho_0 \leq \widehat{\rho}$, the Receiver can simply not investigate the Sender (i.e., remaining ignorant) to ensure that the Sender's payoff is minimised at zero. If, instead, $\rho_0 \in (\underline{\rho}, \widehat{\rho})$, the Receiver's punishing investigation minimises the probability that message m_1 induces the Receiver to take action. Thus, a punishing investigation involves the Receiver avoiding learning that the Sender is reliable (or not learning at all), and induces posterior beliefs at $\widehat{\rho}$ and 1. In fact, the Receiver's punishing investigation strategy convexifies the Sender's continuation payoffs for every $\widehat{\rho} \in [\underline{\rho}, 1]$:

$$i^{\min}\left(\widehat{\rho}\right) := \begin{cases} \frac{\rho_0 - \widehat{\rho}}{1-\widehat{\rho}}\delta_1 + \frac{1-\rho_0}{1-\widehat{\rho}}\delta_{\widehat{\rho}} & \text{if } \rho_0 \geq \underline{\rho} \text{ and } \widehat{\rho} \in (\underline{\rho}, \rho_0) \\ \delta_{\rho_0} & \text{otherwise} \end{cases}. \tag{2.6}$$

While punishing Sender's "bad" choice of experiments using $i^{\min}$ minimises the Sender's incentive to deviate from the "good" choice, the Receiver may not be able to induce the Sender to choose any $\widehat{\rho} \in [\underline{\rho}, 1]$ using $i^{\min}$. This is because the Sender can guarantee a certain payoff by choosing a *maxmin experiment*, denoted $\widehat{\rho}^{\text{maxmin}}$, that maximises her payoff given that the Receiver uses $i^{\min}$. Therefore, in any commitment equilibrium, the Receiver must ensure that the Sender obtains her *maxmin payoff* defined by

$$V_{\text{S}}^{\text{maxmin}} := V_{\text{S}}\left(\widehat{\rho}^{\text{maxmin}}, i^{\min}\left(\widehat{\rho}^{\text{maxmin}}\right)\right).$$

I characterise the Sender's maxmin experiment in the lemma below.

Lemma 2.2. *The Sender's maxmin experiment is given by*

$$\widehat{\rho}^{\text{maxmin}} = \max\left\{\underline{\rho}, \left(1 + \sqrt{\frac{\mu_0}{\mu^* - \mu_0}\frac{1-\rho_0}{\rho_0}}\right)^{-1}\right\} \in [\underline{\rho}, \rho_0).$$

One can further show that $\widehat{\rho}^{\text{maxmin}} = \underline{\rho}$ holds if $\rho_0 \in [\underline{\rho}, \rho_{0,0}]$ for a unique $\rho_{0,0} \in (\underline{\rho}, 1)$, and that V_S^{maxmin} is strictly increasing in $\rho_0 \geq \underline{\rho}$ (and equals zero otherwise).

It is worth noting at this point that the Sender's maxmin payoff is the same against a Receiver who can (additionally) condition his investigation strategy on the realisation of the Sender's experiment. To see this, first observe that, given the simplification in Lemma 2.1, the Receiver knows with probability one that the Sender is reliable after observing the message m_0. Therefore, the only investigation that can affect the Receiver's belief is the one after the Receiver observes the message m_1. Hence, so long as the simplification holds, the assumption that the investigation strategies do not depend on realisations is without loss. Second, notice that the Receiver can always choose not to vary investigations based on realisations even if he could. Together, these observations mean that the Sender must be weakly worse off against a Receiver that minimises the Sender's payoff if she chose more complex experiments than the canonical experiments described in Lemma 2.1.

Notice that any commitment equilibrium can be achieved by an investigation strategy that punishes the Sender for deviating to any off-the-equilibrium-path experiment. Define

$$i_{\widehat{\rho},\iota}^{\min}\left(\widehat{\rho}'\right) := \begin{cases} \iota & \text{if } \widehat{\rho}' = \widehat{\rho} \\ i^{\min}\left(\widehat{\rho}\right) & \text{if } \widehat{\rho}' \neq \widehat{\rho} \end{cases},$$

so that $i_{\widehat{\rho},\iota}^{\min}$ punishes the Sender for choosing any $\widehat{\rho}' \neq \widehat{\rho}$ and ι is the investigation that the Receiver conducts if the Sender chooses $\widehat{\rho}$. For any $z \in (\rho_0, 1]$, I also define

$$\iota_z := \frac{\rho_0}{z}\delta_z + \frac{\rho_0}{z - \rho_0}\delta_0 \in \mathcal{I},$$

where $\iota_1 \equiv \rho_0\delta_1 + (1-\rho_0)\delta_0$ represents a full investigation and I define $\iota_{\rho_0} := \delta_{\rho_0}$ to represent no investigation. I note that distributions of posterior beliefs of the form ι_z are induced by signal structures that fully reveals that the Sender is reliable when the

Sender is reliable but does not fully reveal that the Sender is unreliable when the Sender is unreliable. Using this notation, the Receiver's maximal payoff is $\overline{V}_{\mathrm{R}} := V_{\mathrm{R}}(\underline{\rho}, \iota_1)$. Theorem 2.2 characterises the commitment equilibrium.

Theorem 2.2. $(\widehat{\rho}^*, i^{\min}_{\widehat{\rho}^*, \iota_{z^*}}(\cdot))$ *is a commitment equilibrium, where*

$$\widehat{\rho}^* = \max\left\{\underline{\rho}, \frac{\mu^* - \mu_0}{\mu^*}\frac{\rho_0}{1 - V_{\mathrm{S}}^{\mathrm{maxmin}}}\right\},$$
$$z^* = \max\left\{\min\left\{\frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1-\mu_0)\,\rho_0}, 1\right\}, \rho_0\right\}.$$

The players' commitment equilibrium payoffs are

$$V_{\mathrm{R}}\left(\widehat{\rho}^*, \iota_{z^*}\right) = \min\left\{\overline{V}_{\mathrm{R}}, \frac{\mu_0}{\mu^*}\left[(1-\mu_0)\,\rho_0 + V_{\mathrm{S}}^{\mathrm{maxmin}}\right] - V_{\mathrm{S}}^{\mathrm{maxmin}}, \frac{\mu_0}{\mu^*} - V_{\mathrm{S}}^{\mathrm{maxmin}}\right\}$$
$$V_{\mathrm{S}}\left(\widehat{\rho}^*, \iota_{z^*}\right) = \max\left\{V_{\mathrm{S}}\left(\underline{\rho}, \iota_1\right), V_{\mathrm{S}}^{\mathrm{maxmin}}\right\}.$$

In particular, both payoffs are strictly positive for any $\rho_0 \in (0, 1)$.

The optimal investigation strategy shown in the theorem involves punishing the Sender for deviating to any off-equilibrium-path experiment. The punishment, given by (2.6), involves using ignorance about the Sender being reliable to reduce the Sender's payoff. On the equilibrium path, the Receiver conducts an investigation given by ι_{z^*}. When the Sender's maxmin payoff is low, which is the case when the Sender has a strong prior incentive to provide information (e.g., ρ_0 is low), the Receiver's on-the-equilibrium-path investigation is fully revealing (i.e., $z^* = 1$). However, when the Sender's prior incentive to provide information is low (e.g., ρ_0 is high), the Receiver uses ignorance about the Sender being unreliable to ensure that the Sender's payoff on the equilibrium path is at least her maxmin payoff. Here, ignorance is used to increase the Sender's payoff. Thus, the Receiver's optimal investigation strategy under commitment uses ignorance on- and off-equilibrium paths to different effects.

To understand the expressions above and to see how each of the different cases arise, let us fix some $\mu^* > \mu_0$ and vary the prior belief ρ_0 about the Sender's reliability.[29]

Recall that when $\rho_0 \leq \underline{\rho}$, the players' payoffs are zero when the Receiver cannot investigate the Sender's reliability. In particular, this means that the Sender's maxmin payoff is zero so that the Receiver can obtain his maximal payoff of $\overline{V}_{\mathrm{R}}$.

Suppose instead that $\rho_0 > \underline{\rho}$. Fixing $\widehat{\rho}^{\text{maxmin}}$, both the Receiver and the Sender benefit from a more informative investigation than $i^{\min}(\widehat{\rho}^{\text{maxmin}})$. Hence, the Receiver's can increase his payoff while ensuring that the Sender choose $\widehat{\rho}^{\text{maxmin}}$ by fully investigating the Sender upon seeing $\widehat{\rho}^{\text{maxmin}}$. It therefore follows that, if $\widehat{\rho}^{\text{maxmin}} = \underline{\rho}$ (i.e., $\rho_0 \in [\underline{\rho}, \rho_{0,0}]$), the Receiver can obtain his maximal payoff of $\overline{V}_{\mathrm{R}}$ by choosing $i^{\min}_{\underline{\rho},\iota_1}$; i.e., by fully investigating if $\widehat{\rho} = \underline{\rho}$ and otherwise punishing the Sender.

Now suppose that $\widehat{\rho}^{\text{maxmin}} > \underline{\rho}$ (i.e., $\rho_0 > \rho_{0,0}$), and the Receiver chooses the same investigation strategy, $i^{\min}_{\underline{\rho},\iota_1}$. Because a full investigation strictly increases the Sender's payoff relative to $i^{\min}$, the Receiver can choose $i^{\min}_{\widehat{\rho}',\iota_1}$ for some $\widehat{\rho}' \in [\underline{\rho}, \widehat{\rho}^{\text{maxmin}})$ while ensuring that the Sender obtains $V_{\mathrm{S}}^{\text{maxmin}}$. Let $\widehat{\rho}^+$ denote the smallest $\widehat{\rho}$ (i.e., the most informative experiment) such that the Sender's payoff is at least $V_{\mathrm{S}}^{\text{maxmin}}$ when the Receiver conducts a full investigation; i.e.,

$$\widehat{\rho}^+ := \min\left\{\widehat{\rho}' \in \left[\underline{\rho}, 1\right] : \left(\widehat{\rho}', \iota_1\right) \geq V_{\mathrm{S}}^{\text{maxmin}}\right\}.$$

If $\widehat{\rho}^+ \leq \underline{\rho}$, the Receiver obtains his maximal payoff of $\overline{V}_{\mathrm{R}}$. One can show that there exists a unique $\rho_{0,1} \in [\rho_{0,0}, 1)$ such that $\widehat{\rho}^+ \leq \underline{\rho}$ for all $\rho_0 \in (\rho_{0,0}, \rho_{0,1}]$. If $\widehat{\rho}^+ > \underline{\rho}$ (i.e., $\rho_0 > \rho_{0,1}$), however, the Receiver faces a trade-off between continuing to conduct a full investigation versus conducting a partial investigation that can induce the Sender to choose a more

[29]Similar comparative statics results hold if I instead fix ρ_0 and vary $\mu^* - \mu_0$.

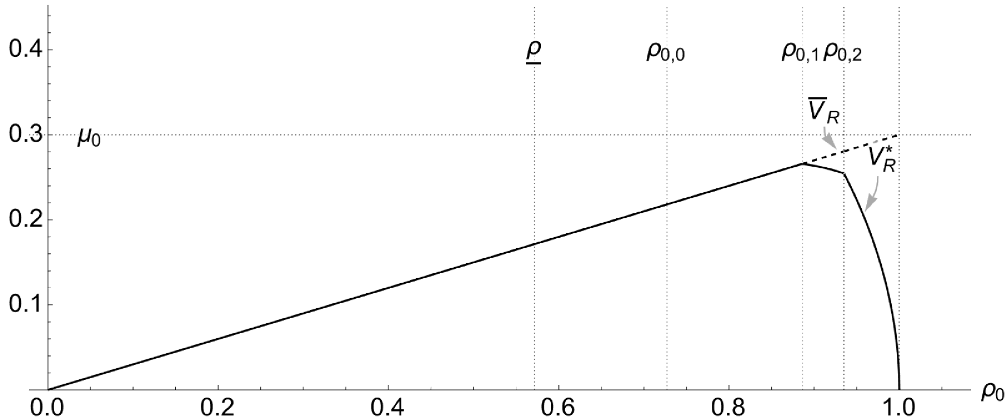
informative experiment, say $\widehat{\rho}' \in [\underline{\rho}, \widehat{\rho}^+)$. It turns out that the Receiver always prefers to do the latter. Moreover, when $\widehat{\rho}^+$ is sufficiently larger than $\underline{\rho}$—i.e., ρ_0 is greater than a unique $\rho_{0,2} \in (\rho_{0,1}, 1)$—the Receiver would not investigate the Sender on the equilibrium path to induce the Sender to choose a fully informative experiment.

Note that the Receiver's commitment equilibrium payoff is necessarily non-monotonic in ρ_0 because Receiver's equilibrium payoffs when $\rho_0 = 0$ or $\rho_0 = 1$ are both zero. The intuition for the result is the following. Recall that when the Receiver cannot investigate (i.e., ι is fixed at δ_{ρ_0}), the Sender chooses $\widehat{\rho} = \rho_0$ so that when there is more doubt about the Sender's reliability (i.e., a lower ρ_0), the Sender chooses a more informative experiment (than the Sender-optimal full-reliability experiment). In other words, the Sender has a stronger prior incentive to choose a more informative experiment when ρ_0 is low. Thus, when ρ_0 is low (i.e., $\rho_0 \leq \rho_{0,1}$), the Receiver need not give up the benefit of being able to distinguish the Sender's type; i.e., the Receiver can obtain his maximal payoff. However, when ρ_0 is high (i.e., $\rho_0 > \rho_{0,1}$), the Sender has a weaker prior incentive to choose the fully informative experiment, and the Receiver must give up the benefit of being able to distinguish the Sender's type to induce the Sender to choose the fully informative experiment. When ρ_0 is sufficiently high (i.e., $\rho_0 > \rho_{0,2}$), the Receiver is willing to give up all the benefits from investigating.

Fixing $\mu_0 = 0.3$ and $\mu^* = 0.5$, Figure 2.3 below shows how varying ρ_0 affects the Receiver's and Sender's commitment equilibrium payoffs in panels (a) and (b), respectively. Figure 2.3(a) shows that the Receiver can obtain the maximal payoff, $\overline{V}_{\mathrm{R}}$, for any $\rho_0 \leq \rho_{0,1}$. This is because, for any $\rho_0 \leq \rho_{0,1}$, the Sender's payoff associated with the Receiver's maximal payoff is lower than her maxmin payoff (see Figure 2.3(b) which shows that the Sender's commitment equilibrium payoff V_{S}^* lies above $V_{\mathrm{S}}^{\mathrm{maxmin}}$ when $\rho_0 \leq \rho_{0,1}$). However, when $\rho_0 > \rho_{0,1}$, the Receiver must give up some payoffs in the form of partial investigation, ι_{z^*}, in order to continue to induce the Sender to choose the fully informative experiment (see Figure 2.3(a) which shows that V_R^* lies below $\overline{V}_R$ for $\rho_0 > \rho_{0,1}$). Moreover, when $\rho_0 > \rho_{0,2}$, even conducting no investigation (which gives the Sender her maximal payoff given $\widehat{\rho}$) is insufficient to ensure that the Sender obtains her maxmin payoff from choosing the fully informative experiment. Thus, in this case, the Receiver can only induce the Sender to choose a partially informative experiment, $\widehat{\rho} > \underline{\rho}$ (see Figure 2.4). In the limit, as $\rho_0 \to 1$, the players' commitment payoffs converge to their payoffs in the full reliability case, which are zero for the Receiver and $\frac{\mu_0}{\mu^*}$ for the Sender.

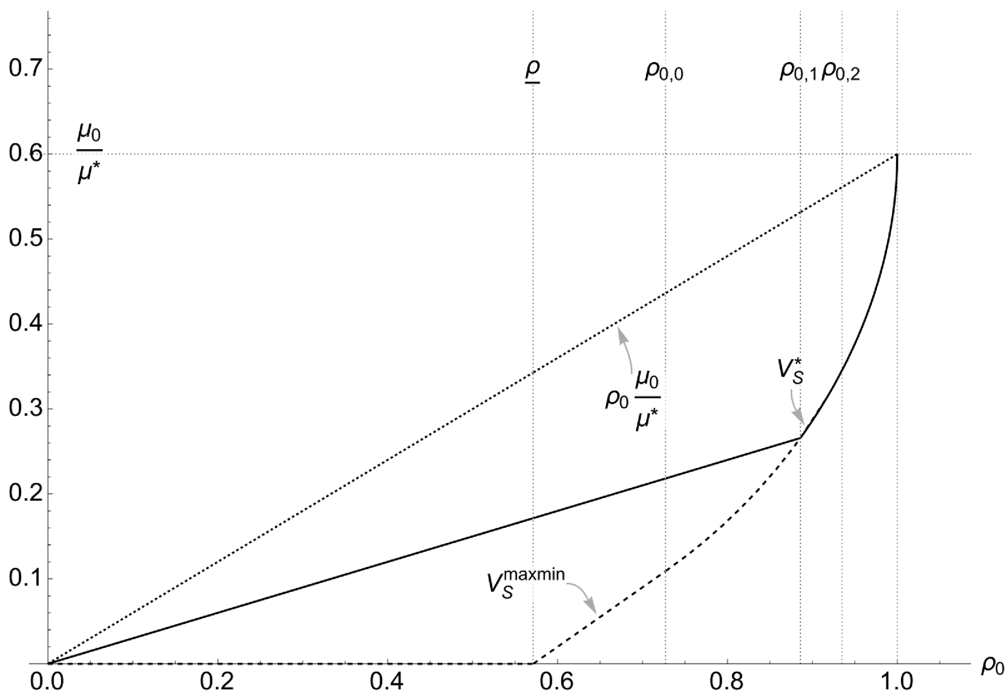
Figure 2.3: Commitment equilibrium payoffs.

(a) Receiver's commitment equilibrium payoff, $V_R^* := V_R(\widehat{\rho}^*, \iota_{z^*})$.



Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, panel (a) shows (as a function of ρ_0) the Receiver's: commitment equilibrium payoff, V_R^*; maximal payoff, $\overline{V}_R$. The Receiver's no-commitment equilibrium and Sender-optimal full-reliability payoffs are zero, and μ_0 is his payoff when he can directly learn the state.

(b) Sender's commitment equilibrium payoff, $V_S^* := V_S(\widehat{\rho}^*, \iota_{z^*})$.

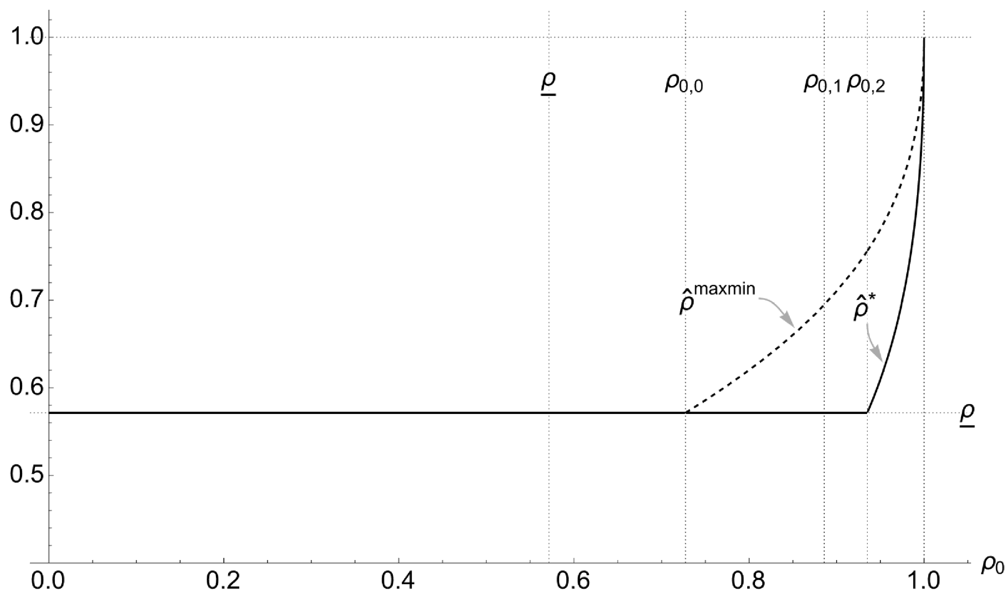


Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, panel (b) shows (as a function of ρ_0) the Sender's: commitment equilibrium payoff, V_S^*; maxmin payoff, V_S^{maxmin}; no-commitment equilibrium payoff, $\frac{\mu_0}{\mu^*}\rho_0$. The Sender-optimal full-reliability payoff is $\frac{\mu_0}{\mu^*}$.

Figure 2.3(a) also shows the Receiver's payoff is concave and achieves a maximum when $\rho_0 = \rho_{0,2}$, which is the point at which the Receiver can no longer induce the Sender to choose the fully informative experiment. Notice how this compares with the Receiver's payoff (of 0.3) when the Receiver can simply learn the state by himself.

Finally, while the Receiver benefits from the ability to commit to investigation strategies, the Sender's payoff is strictly lower (for any interior prior ρ_0) when the Receiver can commit (see Figure 2.3(b) and observe that U_S^* lies below the Sender's no-commitment equilibrium payoff of $\rho_0 \frac{\mu_0}{\mu^*}$).

Figure 2.4: Experiment induced in commitment equilibrium, $\widehat{\rho}^*$.



Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, the figure shows (as a function of ρ_0) the experiment induced in the commitment equilibrium, $\widehat{\rho}^*$, and the Sender's maxmin experiment, $\widehat{\rho}^{\text{maxmin}}$. $\underline{\rho}$ is the fully informative experiment and $\widehat{\rho} = 1$ is the Sender-optimal full-reliability experiment.

Figure 2.4 shows the experiment chosen by the Sender in the commitment equilibrium. For any $\rho_0 \in [0, \rho_{0,2}]$, the Sender's prior incentive to provide information is sufficiently high so that the Receiver can induce the Sender to choose the fully informative experiment $\underline{\rho}$. For $\rho_0 \in (\rho_{0,2}, 1]$, the Sender's prior incentive to provide information is low and the combinations of on- and off-equilibrium-path ignorance are insufficient to ensure that the Sender obtains her maxmin payoff from choosing the fully informative experiment. Thus, in this case, the Receiver must demand a less informative experiment. In the limit as $\rho_0 \to 1$, the Sender's prior incentive to provide information vanishes and the Receiver cannot induce the Sender to choose an experiment that is more informative than the Sender-optimal full-reliability experiment.

2.4 Implementing ignorance via delegation

Because being ignorant is never sequentially rational for the Receiver, the Receiver may not be able to achieve the commitment equilibrium payoff. In this section, I consider the extent to which the Receiver can implement commitment outcomes by learning about the Sender's reliability from a third party whose preferences differ from that of the Receiver. In particular, I focus on a third party whose preference is a linear combination of the Sender's and the Receiver's preferences.

2.4.1 Set up

I now introduce a third player, whom I refer to as the *Third Party* (T, *it*), into the original game. I refer to the modified game as the *delegation game.* In this game, the Third Party (instead of the Receiver) has the discretion to choose investigations about the Sender's reliability. Thus, in the delegation game, the Receiver simply chooses his action after observing the Sender's message and the realisation of the investigation. The timing is as in the no-commitment case of the original game, except that it is now the Third Party who chooses $\iota \in \mathcal{I}$ as a function of the Sender's experiment in a sequentially rational manner according to its preferences. To be clear, the Third Party cannot commit to investigation strategies. The Third Party's preference is a linear combination of the Sender's and the

Receiver's payoffs with weights $\lambda_i \in \mathbb{R}^2$ on player $i \in \{S, R\}$'s payoff. I say that the Third Party is *purely Receiver-aligned* if $\lambda_S = 0$ and $\lambda_R > 0$, and *purely adversarial* if $\lambda_S < 0$ and $\lambda_R = 0$.[30]

I define an equilibrium of the delegation game analogously to the no-commitment equilibrium of the original game. Importantly, the simplification in Lemma 2.1 continues to apply with respect to equilibrium payoffs in the delegation game.[31] In this section, I focus on the case where the Third Party is adversarial; i.e., $\lambda_S < 0$. I therefore normalise the weight on the Sender's preference as $\lambda_S = -1$ and let $\lambda \in \mathbb{R}_+$ denote the weight on the Receiver's preference. I refer to a Third Party whose weights are $(\lambda_S, \lambda_R) = (-1, \lambda)$ as a *λ-balanced Third Party*. I describe a 0-balanced Third Party as being *purely adversarial* and a ∞-balanced Third Party as being *purely Receiver aligned*. I say that tuple $(\xi, i, \sigma, \alpha, \mu)$ is a *λ-delegation equilibrium* if it satisfies the same conditions as a no-commitment equilibrium with the exception that, the investigation strategy i has to be sequentially rational given σ and a for the Third Party; i.e., condition (iv) becomes: for all $\xi' \in \Xi$ $i(\xi')$ solves

$$\max_{\iota' \in \mathcal{I}} -V_S(\cdot) + \lambda V_R(\cdot).$$

For convenience, I also define a λ-balanced Third Party's continuation payoff as $V_T^{\lambda}(\widehat{\rho}, \iota) := -V_S(\widehat{\rho}, \iota) + \lambda V_R(\widehat{\rho}, \iota)$.

2.4.2 Purely adversarial versus purely Receiver-aligned Third Party

If the Third Party is purely Receiver aligned, then the players' (∞-)equilibrium payoffs correspond to their no-commitment equilibrium payoffs in the original game. In

[30] See discussion in section 2.5 for the case when the Third Party is *purely Sender-aligned* (i.e., $\lambda_S > 0$ and $\lambda_R = 0$).

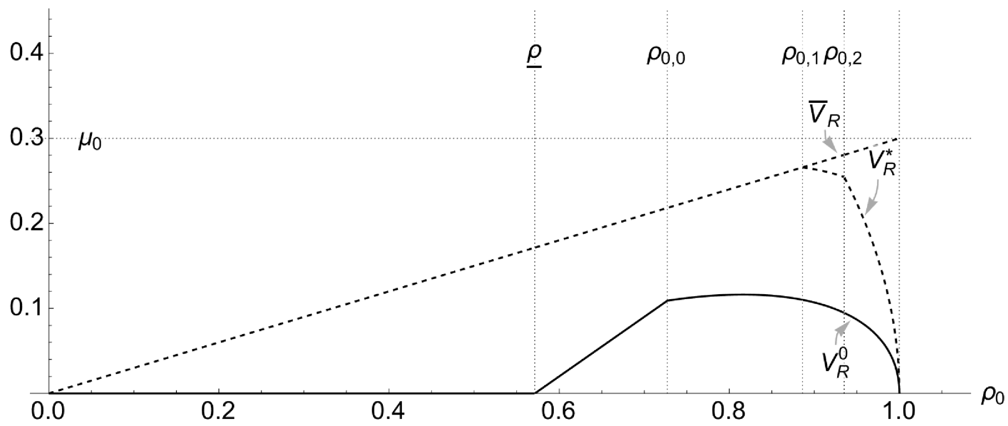
[31] Because any Receiver-aligned Third Party behaves exactly like the Receiver, I prove Lemma 2.1 assuming that the Third Party chooses an investigation strategy.

particular, this means that the Receiver's payoff is zero when the Third Party is purely Receiver aligned (Theorem 2.1). On the other hand, if the Third Party is purely adversarial, the Sender's (0-)equilibrium payoff is her maxmin payoff, V_S^{maxmin}. Moreover, because the support of $i^{\min}(\widehat{\rho}^{\text{maxmin}})$ is $\{\widehat{\rho}^{\text{maxmin}}, 1\}$ when $\rho_0 \in (\underline{\rho}, 1)$, the Receiver obtains a strictly positive payoff when $\rho_0 \in (\underline{\rho}, 1)$. On the other hand, if $\rho_0 \in [0, \underline{\rho}]$, a purely adversarial third party does not investigate the Sender so that the Receiver's payoff is zero. It follows that the Receiver strictly benefits from delegating to a purely adversarial Third Party over delegating to a purely Receiver-aligned Third Party if $\rho_0 \in (\underline{\rho}, 1)$. This is summarised in the corollary below.

Corollary 2.2. *For any $\rho_0 \in (\underline{\rho}, 1)$, the Receiver's 0-equilibrium payoff is strictly higher than his payoff in the ∞-equilibrium; i.e., the Receiver strictly prefers delegating investigation to a purely adversarial Third Party over a purely Receiver-aligned Third Party.*

Figure 2.5 shows how the Receiver's 0-equilibrium payoff (denoted V_R^0) varies with ρ_0 when $\mu_0 = 0.3$ and $\mu^* = 0.5$. Note that, for any value of $\rho_0 \in [0, 1]$, the Receiver's ∞-equilibrium payoff is zero. It shows that the Receiver strictly benefits from learning via a purely adversarial Third Party whenever $\rho_0 \in (\underline{\rho}, 1)$.

Figure 2.5: Receiver's delegation equilibrium payoff with purely adversarial Third Party, V_{R}^{0}.



Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, the figure shows (as a function of ρ_0) the Receiver's: 0-equilibrium payoff in the delegation game when the Third Party is purely adversarial, V_{R}^{0}; maximal payoff, $\overline{U}_{\mathrm{R}}$; commitment equilibrium payoff, V_{R}^{*}. The Receiver's no-commitment equilibrium and Sender-optimal full-reliability payoffs are zero, and μ_0 is his payoff when he can directly learn the state.

2.4.3 Optimal delegation

The next result establishes that the Receiver can benefit strictly more by delegating investigation to a λ-balanced Third Party with $\lambda > 0$. To see why delegating to such a Third Party can help, consider first the Third Party's continuation payoff:

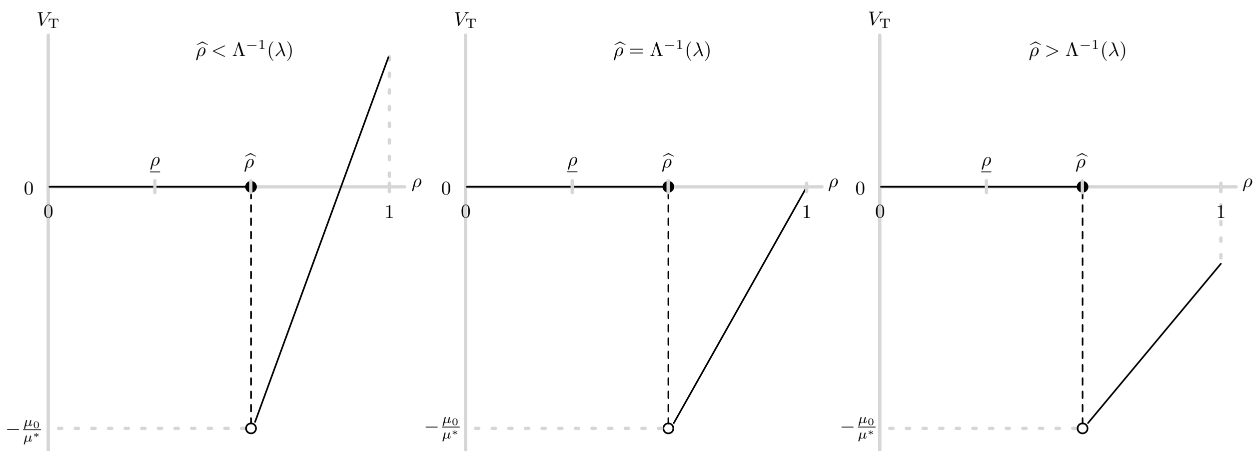
$$V_{\mathrm{T}}^{\lambda}\left(\widehat{\rho}, \iota\right) = \int_0^1 \mathbb{1}_{\{\rho > \widehat{\rho}\}} \left[-\left(1 - \frac{\rho}{\widehat{\rho}} \frac{\mu^* - \mu_0}{\mu^*}\right) + \lambda \frac{\mu^* - \mu_0}{\mu^*} \left(\frac{\rho}{\widehat{\rho}} - 1\right)\right] \mathrm{d}\iota\left(\rho\right).$$

Observe, in particular, that the sign of $V_{\mathrm{T}}^{\lambda}(\widehat{\rho}, \delta_1)$ (i.e., the Third Party's payoff when the Third Party believes that the Sender is reliable) depends on the weight λ on the Receiver's preference. Given any Sender's choice of an experiment $\widehat{\rho} \in [\underline{\rho}, 1]$, there exists a unique value of λ, denoted as $\Lambda(\widehat{\rho})$, such that $V_{\mathrm{T}}^{\Lambda(\widehat{\rho})}(\widehat{\rho}, \delta_1) = 0$. Moreover, because $\Lambda(\cdot)$ is strictly increasing,

$$V_{\mathrm{T}}^{\Lambda(\widehat{\rho})}\left(\widehat{\rho}, \delta_1\right) \geq 0 \Leftrightarrow \widehat{\rho} \leq \Lambda^{-1}\left(\lambda\right).$$

Figure 2.6 below shows the Third Party's continuation payoff as a function of ρ fixing $\iota(\widehat{\rho}) = \delta_\rho$.

Figure 2.6: Third Party's continuation payoff, $V_T(\widehat{\rho}, \delta_\rho)$ with $\widehat{\rho} \in (\underline{\rho}, 1)$.



The figure shows the Third Party's continuation payoff (as a function of ρ) when the Receiver's posterior belief that the Sender is reliable is degenerate at ρ, the Sender has chosen $\widehat{\rho} \in (\underline{\rho}, 1)$, and the Third Party's preference is such that (from left to right): $\widehat{\rho} < \Lambda^{-1}(\lambda)$, $\widehat{\rho} = \Lambda^{-1}(\lambda)$, and $\widehat{\rho} > \Lambda^{-1}(\lambda)$.

The Third Party's optimal investigation is given by ι that concavifies the function $V_T^\lambda(\widehat{\rho}, \delta_\rho)$. It is immediate from Figure 2.6 that: if $V_T^\lambda(\widehat{\rho}, \delta_1) \geq 0$, then a λ-balanced Third Party's optimal investigation is a full investigation, ι_1; and if $V_T^\lambda(\widehat{\rho}, \delta_1) \leq 0$, a λ-balanced Third Party's optimal investigation is $i^{\min}(\widehat{\rho})$. Moreover, if $V_T^\lambda(\widehat{\rho}, \delta_1) = 0$, a λ-balanced Third Party is indifferent between these two investigations. It follows that when $\Lambda^{-1}(\lambda) \leq \underline{\rho}$, a λ-balanced Third Party acts as a purely adversarial Third Party so that λ-equilibrium in such cases coincides with 0-equilibrium. To understand what happens when λ is sufficiently large so that $\Lambda^{-1}(\lambda) > \underline{\rho}$, notice first that the Sender's preferred experiment in the interval $(\underline{\rho}, \Lambda^{-1}(\lambda)]$ is $\Lambda^{-1}(\lambda)$—because a λ-balanced Third Party conducts a full investigation for any $\widehat{\rho} \in (\underline{\rho}, \Lambda^{-1}(\lambda)]$ so that the Sender's payoff is strictly increasing in $\widehat{\rho}$. However, the Sender would only choose $\Lambda^{-1}(\lambda)$ if her payoff, $V_S(\Lambda^{-1}(\lambda), \iota_1)$, is greater than her maxmin payoff V_S^{maxmin}, which is the case if and only if $\Lambda^{-1}(\lambda) \geq \widehat{\rho}^+$; otherwise, the Sender can choose $\widehat{\rho}^{\text{maxmin}}$ to obtain V_S^{maxmin}. These observations characterises λ-equilibrium in the delegation game. Define

$$\lambda^* := \Lambda\left(\max\left\{\underline{\rho}, \widehat{\rho}^+\right\}\right).$$

Theorem 2.3. *For any $\lambda \in \mathbb{R}_+ \cup \{\infty\}$, a pair $(\widehat{\rho}^\lambda, i^\lambda(\cdot))$ is a λ-equilibrium, where*

$$\widehat{\rho}^\lambda = \begin{cases} \Lambda^{-1}(\lambda) & \text{if } \lambda \geq \lambda^* \\ \widehat{\rho}^{\text{maxmin}} & \text{if } \lambda < \lambda^* \end{cases}, \quad i^\lambda(\widehat{\rho}) = \begin{cases} \iota_1 & \text{if } \lambda \geq \Lambda(\widehat{\rho}) \\ i^{\min}(\widehat{\rho}) & \text{if } \lambda < \Lambda(\widehat{\rho}) \end{cases}.$$

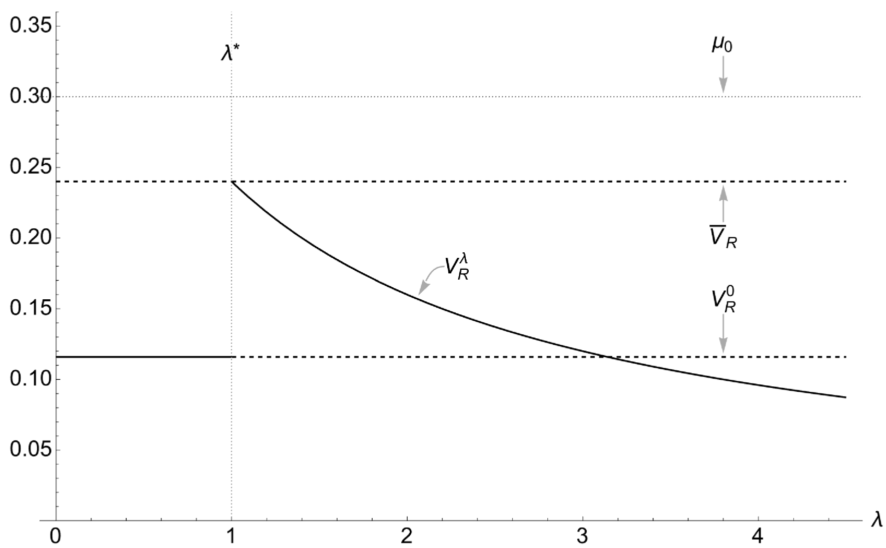
If $\lambda < \lambda^*$, λ-equilibrium coincides with 0-equilibrium (i.e., equilibrium with a purely-adversarial Third Party).

Now suppose that the Sender can choose the preference of the Third Party who will investigate the Sender's reliability; i.e., the Receiver can choose the weight λ in the Third Party's preference. An immediate observation is that, by delegating to a $\Lambda(\underline{\rho})$-balanced Third Party, the Receiver can obtain the commitment equilibrium payoff whenever the on-the-equilibrium-path investigation is a full investigation (i.e., when $\widehat{\rho}^+ \leq \underline{\rho}$). When $\widehat{\rho}^+ > \underline{\rho}$, a full investigation cannot be used to induce the Sender to choose the fully informative experiment, $\underline{\rho}$. Moreover, the partial investigation that a Receiver who can commit would have chosen (ι_{z^*}) is not optimal for any λ-balanced Third Party. Thus, the best the Receiver can do is to choose a $\Lambda(\widehat{\rho}^+)$-balanced Third Party to ensure that the Sender obtains her maxmin payoff when choosing $\widehat{\rho}^+$ (and the Third Party conducts a full investigation). These observations give the following result.

Corollary 2.3. *The Receiver's λ^*-equilibrium payoff is greater than in any other λ-equilibrium. Moreover, the Receiver's λ^*-equilibrium payoff coincides with the Receiver's commitment equilibrium payoff, which in turn coincides with the Receiver's payoff from the ideal outcome, whenever ρ_0 is sufficiently low.*

Figure 2.7 shows how the Receiver's λ-equilibrium payoff change as we vary λ when $\mu_0 = 0.3$, $\mu^* = 0.5$ and $\rho_0 = 0.8$. When the weight on the Receiver's payoff is small ($\lambda < \lambda^*$), a λ-balanced third party behaves as a purely adversarial Third Party so that the Receiver's payoff equal to his 0-equilibrium payoff. However, the Receiver's λ-equilibrium payoff "jumps" up at $\lambda = \lambda^*$ because the Third Party now finds it optimal to conduct a full investigation on the equilibrium path (instead of the punishing investigation). Moreover, given the parametric assumptions, the Receiver's λ^*-equilibrium payoff equals the maximal payoff, $\overline{V}_{\mathrm{R}}$, which, in turn, equals the Receiver's commitment equilibrium payoff. As λ increases beyond λ^*, the Receiver's payoff decreases and converges to zero (i.e., the no-commitment equilibrium payoff) as $\lambda \to \infty$.

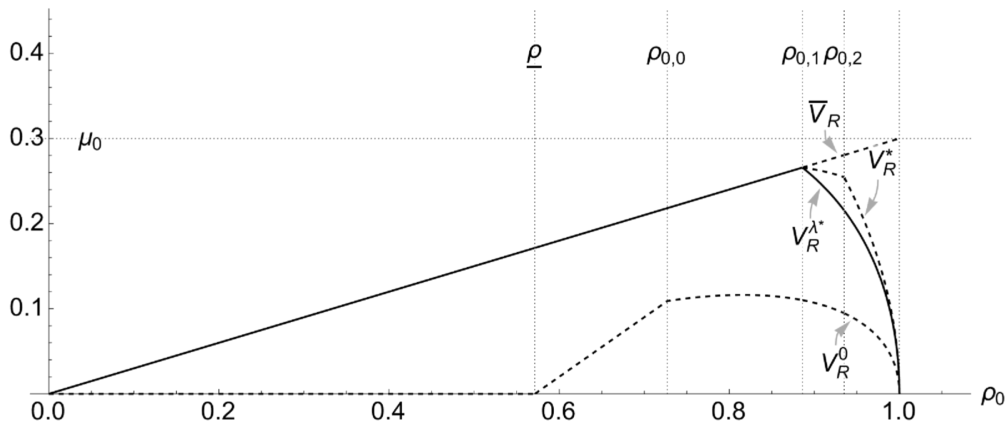
Figure 2.7: Receiver's delegation equilibrium payoff, V_{R}^{λ}.



Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, $\rho_0 = 0.8$, the figure shows (as a function of λ) the Receiver's: λ-equilibrium payoff in the delegation when the Third Party is λ-balanced; the Receiver's maximal payoff, which equals the Receiver's commitment, $\overline{V}_{\mathrm{R}} = V_{\mathrm{R}}^*$; 0-equilibrium payoff in the delegation game when the Third Party is purely adversarial, V_{R}^0. The Receiver's no-commitment equilibrium and Sender-optimal full-reliability payoffs are zero, and μ_0 is his payoff when he can directly learn the state.

Let us compare the Receiver's payoffs in λ^*-equilibrium with his commitment equilibrium payoffs as we vary ρ_0 when $\mu_0 = 0.3$ and $\mu^* = 0.5$. Figure 2.8 compares the Receiver's: λ^*-equilibrium payoff (denoted $V_R^{\lambda^*}$), 0-equilibrium payoff (V_R^0), commitment equilibrium payoff (V_R^*), and his maximal payoff ($\overline{V}_R$). The Receiver can obtain the commitment equilibrium payoff (that equals his maximal payoff) whenever $\rho_0 \leq \rho_{0,1}$ because this inequality implies $\widehat{\rho}^+ \leq \underline{\rho}$. However, when $\rho_0 > \rho_{0,1}$, the Receiver's payoff is lower in the λ^*-equilibrium because it is not sequentially rational for any λ-balanced Third Party to conduct a partial investigation of the form ι_z that a Receiver with commitment power would choose. The difference between V_R^* and $V_R^{\lambda^*}$ therefore represents the loss arising from the Third Party's inability to commit to ι_z on the equilibrium path. However, notice that the Receiver does better in λ^*-equilibrium than in 0-equilibrium (i.e., when the Third Party is purely adversarial). The difference here arises from the fact that a purely adversarial Third Party conducts a maximally punishing investigation even on the equilibrium path. This means that a purely adversarial Third Party can only induce the Sender to choose $\widehat{\rho}^{\text{maxmin}}$ whereas a λ^*-balanced Third Party can induce the Sender to choose a more informative experiment $\max\{\widehat{\rho}^+, \underline{\rho}\} \leq \widehat{\rho}^{\text{maxmin}}$ by not conducting a full investigation on the equilibrium path.

Figure 2.8: The Receiver's optimal delegation equilibrium payoff, $V_R^{\lambda^*}$.



Fixing $\mu_0 = 0.3$, $\mu^* = 0.5$, the figure shows (as a function of ρ_0) the Receiver's: λ^*-equilibrium payoff in the delegation game, $V_R^{\lambda^*}$; 0-equilibrium payoff in the delegation game when the Third Party is purely adversarial, V_R^0; maximal payoff, $\overline{V}_R$; commitment equilibrium payoff, V_R^*. The Receiver's no-commitment equilibrium and Sender-optimal full-reliability payoffs are zero, and μ_0 is his payoff when he can directly learn the state.

2.5 Discussions

In this section, I first discuss how my results shed light on the efficacy of cross-examination of witnesses, how audits can incentivise the auditee to provide more information in equilibrium, and benefits of *ad hominem* arguments in debates. I then discuss the robustness of the results with respect to certain alternative assumptions and provide some extensions.

2.5.1 Interpretations of the model

Cross-examinations Courts rely on witnesses to provide information about factual or technical matters concerning cases. A perennial worry, however, is that witnesses do not provide sufficient or reliable information. One prominent way courts deal with this concern is through cross-examinations in which a witness is interrogated by an attorney, usually from the opposing party, whose purported goal is to test the reliability of the witness (and consequently the reliability of the evidence provided by the witness). Cross-examination is an important feature of the US court system and has famously been described as the "greatest legal engine ever invented for the discovery of truth" by John Henry Wigmore, a leading legal scholar on US evidentiary law in the early 20th century.[32] Moreover, the US Supreme Court has held that the Confrontation Clause of the Sixth Amendment to the US Constitution provides a criminal defendant with the opportunity to cross-examine testimony that has been made against the defendant.[33] Similarly, in the US, parties can challenge the admissibility of expert evidence through a *Daubert challenge*.[34] The consequence of being found unreliable can range from partial exclusion to a full exclusion of the witness evidence (i.e., impeachment).[35] Between 2000 and 2021, there were 3,342 cases of Daubert challenges specifically against financial expert witnesses, and 43% of these challenges resulted in the partial or full exclusion of the expert (PricewaterhouseCoopers, 2022). The latter statistic, in particular, underscores the fact that outcomes of cross-examinations are not always predictable because they do not always lead to exclusions of the witness. It can also be the case that

[32]The quote is from an encyclopaedic survey of the development of the law of evidence published in 1904 by John Henry Wigmore. Known as *Wigmore on Evidence*, the US Supreme Court Justice Felix Frankfurter described the survey as being "unrivaled as the greatest treatise on any single subject of the law" (Frankfurter (1963)). Wigmore on Evidence was the dominant source of US evidentiary law until the codification of the Federal Rules of Evidence in 1975 (Friedman (2009)).

[33]*Melendez-Diaz v. Massachusetts*, 557 U.S. 305 (2009). US Supreme Court Justice Antonin Scalia wrote "[d]ispensing with confrontation because testimony is obviously reliable is akin to dispensing with jury trial because a defendant is obviously guilty. This is not what the Sixth Amendment prescribes."

[34]Coined after US Supreme Court case *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), a Daubert challenge is a type of motion to exclude expert witness testimony on the basis that it represents an unqualified evidence. *Daubert* altered the standard for determining the admissibility of scientific evidence in Federal courts that was previously based on *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923)) (some states continue to rely on the *Frye* or the "*Frye* plus" standard). Subsequently, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999) held that the *Daubert* standard applies to all expert testimony including those that are non-scientific.

[35]For example, the Manual of Model Criminal Jury Instructions and Manual of Model Civil Jury Instructions (2022) for the US Courts for the Ninth Circuit both state: "In deciding the facts in this case, you may have to decide which testimony to believe and which testimony not to believe. You may believe everything a witness says, or part of it, or none of it."

cross-examinations backfire and lead the court to believe the witness is more reliable than they had initially thought.

The features of cross-examination described above are consistent with how the Receiver in my model learns about the Sender's reliability using investigations. My results therefore highlight the role of cross-examination as not only a way to learn the reliability of witnesses, but also as a way for the court to obtain more information by inducing the parties to select more informative witnesses. In this light, my results have implications for the efficacy of cross-examination as an engine for the discovery of truth. Theorem 2.1 implies, perhaps surprisingly, that a cross-examination conducted by a judge may not help the court obtain additional information. Together with Theorem 2.2, they suggest that, for the judge to be an effective cross-examiner, he must paradoxically be able to commit to not discovering the truth about the witness' reliability. Theorem 2.2 tells us how the judge can optimally commit to being ignorant about the witness' reliability to obtain more information in equilibrium. Specifically, the theorem shows how the judge can optimally commit to not finding out that a witness is reliable as a way to punish the party for choosing uninformative witnesses. It also shows how the judge may optimally commit to not finding out that a witness is unreliable as a way to entice the party into choosing a more informative witness.

While Theorem 2.2 speaks to the potential benefits of the ability to commit to being ignorant, such commitments may not be possible in reality. To that end, Corollary 2.2 suggests delegating cross-examination to an adversarial third party—as is in fact done in the US and other jurisdictions—-allows the court to circumvent this commitment issue. Moreover, Theorem 2.3 and Corollary 2.3 suggest that there is a significant benefit in ensuring that the cross-examiner is not only adversarial but also cares about the discovery of truth. To this end, in the US, for example, prosecutors have a dual role as advocates seeking a conviction and as "ministers of justice" (Fisher, 1988). Moreover, the courts have also recognised that prosecutors have a special duty not to impede the truth (Gershman, 2001). These can be seen as ensuring that a prosecutor who cross-examines the defendant's witness is not only adversarial but also has a preference for the discovery of truth.

Interestingly, these obligations for prosecutors do not apply to defendants, because it would conflict with their right not to testify against themselves and the confidentiality of the lawyer-client relationship. However, there are still more aspects of legal systems that may assist in ensuring that the cross-examiner has a balanced preference. As studied by Shin (1998) and Dewatripont and Tirole (1999),[36] different jurisdictions have different norms as to the level of engagement by judges during trial procedures. For example, US courts place more weight on advocates relative to judges; consequently, judges tend to be relatively passive and act as "referees" to the parties. In particular, in the US, it is uncommon for judges to question witnesses. In contrast, in other jurisdictions (e.g., Germany), the court system is more inquisitorial (as opposed to adversarial) and the judges often direct the debates by asking questions. To the extent that a combination of the opposing party and an "inquisitorial judge" can be considered a balanced third party, Corollary 2.3 gives a reason to prefer an inquisitorial legal system over an adversarial system. Of course, the lesson from Theorem 2.1 applies—the judges must refrain from always finding out the truth as to the reliability of witnesses.

Note that the model predicts different kinds of cross-examinations depending on the cross-examiner's ability to commit or the weight the cross-examiner places on the court arriving at just decisions. For example, the ideal cross-examiner who can commit would not cross-examine the witness when the prior belief that the witness is reliable is sufficiently high. While this prediction appears reasonable, the model also predicts that such a cross-examiner would be willing to hide that the witness is unreliable, which is

[36] See also Timmerbeil (2003).

perhaps less reasonable. If we instead assume that the cross-examiner cannot commit and is strongly adversarial (i.e., $\lambda < \lambda^*$), then we should expect the cross-examiner to be less willing to provide evidence that suggests that the witness is reliable. In contrast, a cross-examiner who also cares more about the court arriving at just decisions (i.e., $\lambda \geq \lambda^*$) would be willing to provide such evidence. While we sometimes observe that prosecutors provide evidence that supports the defendant's innocence (which is about the state), there appear to be fewer instances in which the prosecutor provides evidence that supports the reliability of the defendant's witness. To the extent that my model, which is a significant simplification of the complex legal system in real life, has some empirical content, the discussion above suggests that reality might be most consistent with the case in which the cross-examiner does not have the ability to commit and is strongly adversarial.

Audit An audit of a piece of information such as financial statements or investment appraisals involve an examination of whether a particular method was followed to produce the information at hand. By interpreting the sender's choice of an experiment as a choice of such a method, an investigation into the sender's reliability can be thought of as a type of audit. Importantly, in my model, auditing is costless and it is not about the veracity of the sender's message but rather about the sender's reliability type. Theorem 2.1 suggests that an unfettered audit when conducted by the receiver (or by an auditor whose incentives are aligned with that of the receiver's) might not be beneficial because such an audit can result in the sender choosing a less informative method that negates the receiver's benefit from being able to identify unreliable information. Theorem 2.2 characterises how audits that are not always fully revealing can induce the sender to adopt a more informative method that provides the receiver with more information in equilibrium. Finally, Corollary 2.3 suggests how the receiver can implement such an "auditing strategy" by ensuring that the auditor balances his preference for the receiver and his antagonism toward the sender appropriately. I also note that delegating audits may also alleviate the coordination problem that might arise if a group of investors (as opposed to a single investor) is considering whether to buy the seller's asset.

***Ad hominem* arguments** As noted in the introduction, the realisations of investigations can be thought of as examples of *ad hominem* counter-arguments (meaning arguments "to the person") against the Sender's *ad rem* arguments (meaning arguments "to the point"). Under this interpretation, we can think of the Sender as being, for example, a politician making statements about a political issue, and the investigations as being about the politician (e.g., whether the politician is a flip-flopper) and not about the political issue itself. Such uses of *ad hominem* counter-arguments are prevalent in many political (or even non-political) debates.

In this light, we can interpret the results in this chapter as concerning: (i) the extent to which *ad hominem* arguments are effective as counter-arguments by looking at the effect of investigations on the Sender's payoff, and (ii) the extent to which *ad hominem* arguments are *productive* by looking at the effect of investigations on the Receiver's payoff. The results from the delegation game show that *ad hominem* counter-arguments can be both effective and productive. The latter is perhaps surprising given the oft-held view that *ad hominem* arguments are fallacious. In the political context, one can think of a politician's opposition or a media outlet that oppose the politician as examples of adversarial third parties that can provide information about the politician. My results demonstrate how they can in fact help the voters by induing the politician to speak more truthfully. Moreover, they also suggest that rules that prevent parties from making *ad hominem* arguments may, in fact, harm the receiver.

2.5.2 Extensions

Investigation strategies that can depend on the realisation of the Sender's experiment In some cases, an investigation is chosen after the Receiver has observed the Sender's message; e.g., a cross-examiner is able to read the witness statement before cross examining the witness. Formally, such a case corresponds to the investigation strategy being a function of both the Sender's choice of an experiment as well as the Sender's message. While this additional flexibility must weakly improve the Receiver to further improve his equilibrium payoff, it turns out the Receiver is no better off with this additional flexibility. In other words, allowing for the additional flexibility in the Receiver's investigation would not confer the Receiver any further advantage to benefit from strategic ignorance.

To see this, recall from Theorem 2.2 that for sufficiently low prior belief about the Sender's reliability (specifically $\rho_0 \leq \rho_{0,1}$), the Receiver can obtain his maximal payoff using investigations that do not depend on the realisation of experiments. It follows that the Receiver would not benefit from the additional flexibility for such sufficiently low ρ_0. Moreover, as already noted, the Sender's maxmin payoff is unaffected by this extra flexibility. These observations imply that the additional flexibility could (only) induce the Sender to choose an experiment that is not in the set of canonical experiments identified in Lemma 2.1 on the equilibrium path of the commitment equilibrium. Moreover, if this happens, it must be that both the Sender's and the Receiver's payoffs must be strictly greater than the case in which the Receiver's investigation can only depend on the experiment. For sufficiently high prior belief about the Sender's reliability (specifically, $\rho_0 \geq \rho_{0,2}$), the total surplus between the Sender and the Receiver is maximal (and equals $\frac{\mu_0}{\mu^*}$), meaning that it is not possible to improve both players' equilibrium payoffs simultaneously. For intermediate prior beliefs, $\rho_0 \in (\rho_{0,1}, \rho_{0,2})$, recall that the Receiver maximises his payoff subject to the Sender getting at least her maxmin payoff, and that the Sender is choosing the fully informative experiment. Thus, for the Receiver to obtain a higher payoff, it must be that he conducts a more informative investigation; however, doing so would necessarily result in lower payoffs for the Sender. Thus, it follows that the Receiver can do no better even for intermediate prior beliefs about reliability.

Finally, the additional flexibility in the Receiver's investigation would not change the payoffs that the Receiver can achieve by delegating investigations to a λ-balanced Third Party. This is because a λ-balanced Third Party either conducts a punishing or a full investigation, and the Sender's optimal experiments given such investigations are contained in the canonical set of experiments identified in Lemma 2.1.

Character witnesses The US Federal Rules of Evidence 608 states that: "A witness's credibility may be attacked or supported by testimony about the witness's reputation for having a character for truthfulness or untruthfulness, or by testimony in the form of an opinion about that character." We can think of a testimony in support of the witness' credibility as information provided by a purely Sender-aligned Third Party in the delegation model (i.e., a (λ_S, λ_R)-balanced Third Party with $\lambda_S > 0$ and $\lambda_R = 0$). When the prior belief about the reliability of a witness is low so that the Receiver would ignore the witness' testimony without any information (i.e., $\rho_0 < \underline{\rho}$), the Sender can use a purely Sender-aligned Third Party to its benefit.

To see this, a quick re-inspection of Figure 2.2 tell us that the Receiver-aligned Third Party's optimal investigation to the Sender's choice $\widehat{\rho}$ concavifies the Sender's payoff, so that an optimal investigation involves inducing posterior beliefs at 0 and $\widehat{\rho}$. Given this, the Sender's optimal choice of an experiment is the fully informative experiment $\widehat{\rho} = \underline{\rho}$, so that the Sender's (λ_S, λ_R)-equilibrium payoff is approximately $\frac{\mu_0}{\mu^*}\frac{\rho_0}{\underline{\rho}} > 0$. While the Sender benefits from a purely Receiver-aligned Third Party providing information about her reliability, the Receiver does not—the Receiver's payoff is approximately zero despite the Sender choosing a fully informative experiment.

Sender who observes her type before choosing an experiment Suppose the Sender observes her type $t \in T$ drawn according to ρ_0 before choosing her experiment. In this case, the Sender can potentially signal her type by choosing experiments according to her realised type. However, because the Sender cannot benefit from being identified as the unreliable type by the Receiver, the unreliable Sender would never choose an experiment that differs from the one that the reliable Sender chooses. By letting the off-equilibrium-path beliefs be such that the players treat any deviations as being uninformative, it is easy to show that there always exists an outcome-equivalent pooling equilibrium of the signalling version of the game. Thus, whether the sender knows her type before choosing an experiment would not affect the results.

Unreliable Sender who can observe the state I have assumed that the unreliable Sender does not observe the state before choosing a message. In some situations, it may be more reasonable to assume that the unreliable Sender observes the realisation of the state before choosing how to manipulate. For example, a product reviewer might be able to observe the quality of the product, θ, when deciding whether/how to (mis)communicate to the buyers about the quality for the manufacturer's benefit. When the unreliable Sender can observe θ before choosing m, she behaves exactly as the sender in Crawford and Sobel (1982).

If the unreliable Sender can observe the state, then her messaging strategy is now a mapping $\sigma : \Theta \to \Delta M$. The change allows for the possibility that an equilibrium exists in the game induced by some (ξ, ι) in which σ is informative; i.e., there exist $\xi \in \Xi$ and $\iota \in \mathcal{I}$ such that when the Sender makes action recommendations by sending messages m_0

and m_1, the unreliable Sender also sends both messages in a (ξ, ι)-equilibrium. Thus, it is no longer guaranteed that the unreliable Sender would only send messages in M_1^{ξ} (i.e., messages that would induce the Receiver to take action when the message is known to have been drawn from ξ). Moreover, because ξ and σ need not be Blackwell ordered, it is not clear whether the player's ex ante payoffs are higher or lower in equilibria in which the unreliable Sender sends messages only in M_1^{ξ} compared to equilibria in which the unreliable Sender sends messages from $\text{supp}(\xi) \backslash M_1^{\xi}$. This possibility did not arise when the unreliable Sender could not observe the state because ξ is always more informative than σ.

Nevertheless, for any (ξ, ι)-equilibrium in which σ is informative, there exists a slight perturbation of ι, $\iota' \in \mathcal{I}$, such that σ is not part of any (ξ, ι')-equilibrium. Consequently, even if the Sender is able to improve her payoff by choosing an informative σ for some ι, the Receiver can prevent such strategy from being part of an equilibrium. Thus, it follows that both the Sender and the Receiver's payoffs must be higher in any equilibrium in which σ is informative. But the same argument as in the case for when investigations can depend on the messages means that such equilibrium does not exist. In other words, whether the Sender can observe θ is not important for the results.

Receiver with limited commitment In some situations, it may not be possible for the Receiver to condition the investigation on the Sender's experiment. For example, this might be because the Receiver must commit to an investigation before the Sender chooses an experiment (e.g., a regulator committing to a rule that applies to all regulated entities) or because the Sender's experiment is unobservable (e.g., the experiment is the Sender's private communication strategy). One can model these situations by assuming that the Receiver must commit to a constant investigation strategy. Whether the Receiver benefits from this limited commitment depend on whether the Sender can observe the result of the investigation prior to choosing the experiment.

To see why, suppose that the Receiver has committed to an investigation $\iota \in \mathcal{I}$. Consider first the timing in which the Sender observes the result of the investigation before choosing an experiment. Then, every possible ρ in the support of ι induces a subgame in which the Receiver cannot investigate the Sender, and the prior belief that the Sender is reliable is ρ. As discussed above, the Receiver's payoff in such a game is always zero. It follows that the Receiver is unable to benefit from limited commitment under this timing.

Suppose now that the Sender does not observe the result of the investigation before choosing an experiment. Since the Receiver can always implement a constant investigation strategy in the commitment case, the Receiver's payoff under this timing must be weakly lower than the Receiver's commitment equilibrium payoff.[37] A pertinent question is thus whether the Receiver benefits from investigations when he must commit to a single investigation. Let us first consider what might happen if the Receiver conducts the type of on-the-equilibrium-path investigations that can arise in the commitment equilibrium: ι_z for some $z \in [\rho_0, 1]$ that induces posteriors at $\rho \in \{0, z\}$. Since the Receiver never takes action if $\rho = 0$, the Sender's optimal experiment given the Receiver commits to investigation ι_z is to choose $\widehat{\rho} = z$. But this implies that the Receiver's payoff is zero because $\widehat{\rho} = \max \operatorname{supp}(\iota)$. It follows that for the Receiver to benefit from committing to a single investigation ι, the Receiver must ensure that the Sender finds it optimal to choose $\widehat{\rho} < \max \operatorname{supp}(\iota)$. This raises the possibility that the Receiver can benefit from committing to an investigation with more than two elements in its support. Such an investigation allows the Receiver to exploit the strategic sender's incentive to provide more information to a receiver with (more) doubts about the

[37]In this case, the Receiver would not benefit from delegating investigations to a Third Party. This is because the Receiver can always mimic the investigation that a Third Party would have chosen without affecting the Sender's (subsequent) choice of an experiment.

sender's reliability by forcing the Sender to "guess" the doubt that the Receiver would have prior to deciding on the action. Importantly, it is still possible to show that the Receiver can benefit from committing to a single investigation.[38]

2.6 Conclusion

In this chapter, I study how a receiver can obtain more information from a sender by strategically resolving his doubts about the sender's reliability by conducting investigations about the sender. I show that the receiver's optimal investigation strategy uses ignorance in two ways: as a punishment off the equilibrium path and as a reward on the equilibrium path. Because the receiver prefers to find out the sender's reliability, the receiver uses ignorance on the equilibrium path only when he must. The receiver's penchant for finding out the sender's reliability also means the optimal investigation strategy is not sequentially rational for the receiver. I therefore study the extent to which the receiver can implement commitment outcomes by delegating investigations to a third party. In particular, I show that the receiver can obtain more information from the sender by delegating to an adversarial third party who also cares about the receiver's payoff. Moreover, I find that the receiver can obtain the optimal payoff by delegating when the initial level of doubt about the sender's reliability is high. I explain how my results suggest that the efficacy of cross-examination is affected by the cross-examiner's incentives and their ability to commit to not discovering the truth about the witness' reliability. Outside of the court context, the results imply that investors might not benefit from unfettered audits and that investors can benefit from ensuring that auditors have some adversarial incentives toward the party whose information is being audited.

[38]For example, if $\rho_0 > \underline{\rho}$ is sufficiently low, the Receiver can benefit by committing to $\iota \in \mathcal{I}$ such that $\text{supp}(\iota) = \{\underline{\rho} + \epsilon, 1\}$ for some small $\epsilon > 0$. One can also construct examples in which the Receiver's payoff is strictly higher with investigations that have strictly more than two elements in their support.

Finally, I also explain how my results imply that *ad hominem* counter-arguments can be effective, and perhaps surprisingly, how they can also be productive by inducing the speaker to be more truthful.

2.A Appendix

2.A.1 Proof of Theorem 2.1

The proof of Theorem 2.1 does not in fact require the simplification in Lemma 2.1. To proceed, I first establish that given any Sender's experiment and the unreliable Sender's strategy, a more informative investigation results in a mean-preserving spread of induced beliefs.

Lemma 2.3. *Fix any $\xi, \sigma \in \Xi$ and $\iota, \iota' \in \mathcal{I}$ such that ι is a mean-preserving spread of ι'. Then, distribution of posterior beliefs about the state induced by (ξ, σ, ι) is a mean-preserving spread of that induced by (ξ, σ, ι').*

Proof. Let N be the message space about the Sender's reliability that is sufficiently rich and let $\eta, \eta' : T \to \Delta N$ be the signal. It is well known that for any $\tilde{\iota} \in \mathcal{I}$, there exists such an $\tilde{\eta} : T \to \Delta N$. Let η' be a garbling of η; i.e., there exists $g : N \to \Delta N$ such that

$$\eta'(n|t) = \sum_{n} g(n'|n)\, \eta(n|t).$$

Given $\xi, \sigma \in \Xi$ and η, the Receiver's joint posterior belief about the state and the type is given by

$$\nu(\theta, r|m, n) = \frac{\xi(m|\theta)\, \eta(n|r)\, \rho_0 \mu_0(\theta)}{\sum_{\tilde{\theta}} \left[\xi(m|\tilde{\theta})\, \eta(n|r)\, \rho_0 + \sigma(m|\tilde{\theta})\, \eta(n|u)\, (1-\rho_0)\right] \mu_0(\tilde{\theta})},$$
$$\nu(\theta, u|m, n) = \frac{\sigma(m|\theta)\, \eta(n|u)\, (1-\rho_0)\, \mu_0(\theta)}{\sum_{\tilde{\theta}} \left[\xi(m|\tilde{\theta})\, \eta(n|r)\, \rho_0 + \sigma(m|\tilde{\theta})\, \eta(n|u)\, (1-\rho_0)\right] \mu_0(\tilde{\theta})}.$$

Thus, the Receiver's marginal belief about the state given any (m, n) in the support is

$$\mu(\theta|m, n) := \mathrm{marg}_{\Theta} \nu(\theta, r|m, n) = \frac{\left[\xi(m|\theta)\, \eta(n|r)\, \rho_0 + \sigma(m|\theta)\, \eta(n|u)\, (1-\rho_0)\right] \mu_0(\theta)}{\sum_{\tilde{\theta}} \left[\xi(m|\tilde{\theta})\, \eta(n|r)\, \rho_0 + \sigma(m|\tilde{\theta})\, \eta(n|u)\, (1-\rho_0)\right] \mu_0(\tilde{\theta})}.$$

Hence, beliefs about the state are updated after observing (m, n) as if the pair was drawn according to signal structure $\pi^{\xi,\sigma,\eta} \in \Xi$ such that

$$\pi^{\xi,\sigma,\eta}(m, n|\theta) := \xi(m|\theta)\, \eta(n|r)\, \rho_0 + \sigma(m|\theta)\, \eta(n|u)\, (1 - \rho_0)\,.$$

Using the fact that η' is a garbling of η,

$$\begin{aligned}
&\pi^{\xi,\sigma,\eta'}\left(m, n'|\theta, t\right)\\
&= \xi(m|\theta)\left(\sum_n g\left(n'|n\right) \eta(n|r)\right)\rho_0 + \sigma(m|\theta)\left(\sum_n g\left(n'|n\right)\eta(n|u)\right)(1-\rho_0)\\
&= \sum_n g\left(n'|n\right) \underbrace{\left[\xi(m|\theta)\,\eta(n|r)\,\rho_0 + \sigma(m|\theta)\,\eta(n|u)\,(1-\rho_0)\right]}_{=\pi^{\xi,\sigma,\eta}(m,n|\theta)}.
\end{aligned}$$

If we let $f : M \times N \to \Delta(M \times N)$ be

$$f\left(m', n'|m, n\right) := \mathbb{1}_{\{m'=m\}}\, g\left(n'|n\right)$$

we realise that $\pi^{\xi,\sigma,\eta'}$ is a garbling of $\pi^{\xi,\sigma,\eta}(m, n|\theta)$ via f. ■

Theorem 2.1 follows almost immediately from the previous lemma.

Proof of Theorem 2.1. The previous lemma, together with Blackwell's Theorem, implies that the sequentially rational investigation for the Receiver is always fully revealing in any no-commitment equilibrium; i.e., $i(\cdot) = \bar{\imath} := \rho_0\delta_1 + (1-\rho_0)\delta_0$. By condition (i),

$$\mu(1|\xi', \bar{\imath}, 0, \cdot) = \mu_0,\ \mu(1|\xi', \bar{\imath}, 1, \cdot) = \mu^{\xi}(1|\cdot)$$

Moreover, condition (iii) is moot because the unreliable Sender's payoff is always zero.

Moreover,

$$V_S\left(\xi', \bar{\iota}, \sigma, \alpha\right) = \rho_0 \sum_{m\in M}\sum_{\theta\in\Theta} \alpha\left(1|\xi, \bar{\iota}, 1, m\right) \xi\left(m|\theta\right) \mu_0\left(\theta\right).$$

Observe that the Sender's problem given above is equivalent to the Sender's problem in the case when $\rho_0 = 1$ except for the coefficient ρ_0 in Sender's payoff. Thus, the Sender-optimal fully-reliability experiment, $\widehat{\rho} = 1$, is optimal for the Sender. ■

2.A.2 Proof of Lemma 2.1

Theorem 2.1 implies that it is without loss to consider $(\xi, \overline{\sigma})$ as specified in Lemma 2.1. Hence, the proof focuses on the commitment case. Recall that given any $(\xi, \iota) \in \Xi \times \mathcal{I}$, a tuple (σ, α, μ) is a (ξ, ι)-equilibrium if is a PBE of the game induced by (ξ, ι). Fix some $(\xi, \iota) \in \Xi \times \mathcal{I}$. In this part of the proof, I suppress (ξ, ι) in the notation for brevity. Define the unreliable and reliable Sender's interim payoff from sending message m given (σ, α), respectively, as follows:

$$V_u\left(m|\sigma, \alpha\right) := \int \alpha\left(1|\rho, m\right) \frac{1-\rho}{1-\rho_0} \mathrm{d}\iota\left(\rho\right),$$
$$V_r\left(m|\sigma, \alpha\right) := \int \alpha\left(1|\rho, m\right) \frac{\rho}{\rho_0} \mathrm{d}\iota\left(\rho\right),$$

where $\frac{1-\rho}{1-\rho_0}\iota(\rho)$ (resp. $\frac{\rho}{\rho_0}\iota(\rho)$) is the probability that posterior belief ρ is induced when the Sender is unreliable (resp. reliable). These two interim payoffs combine to give the Sender's ex ante payoff in this (ξ, ι)-equilibrium:

$$V_S\left(\sigma, \alpha\right) = \rho_0 \sum_{m\in M} V_r\left(m|\sigma, \alpha\right) \xi\left(m\right) + \left(1-\rho_0\right) \int V_u\left(m|\sigma, \alpha\right) \sum_{m\in M} \sigma\left(m\right),$$

where $\xi(m) = \sum_{\Theta} \xi(m|\theta)\mu_0(\theta)$. Define

$$x^{\xi}(\cdot) := \xi(\cdot|1) - \xi(\cdot|0)\frac{1-\mu_0}{\mu_0}\frac{\mu^*}{1-\mu^*} \in \Delta M$$
$$M_1^{\xi} := \left\{m \in \operatorname{supp}(\xi) : x^{\xi}(m) \geq 0\right\},$$
$$M_0^{\xi} := \left\{m \in \operatorname{supp}(\xi) : x^{\xi}(m) < 0\right\}.$$

Because, $\mu^{\xi}(m) \geq \mu^* \Leftrightarrow x^{\xi}(m) \geq 0$ given any $m \in \operatorname{supp}(\xi)$, M_a^{ξ} represent the set of messages that can induce the Receiver to choose action $a \in A$ when the Receiver's belief about the Sender's reliability is $\rho = 1$. Define

$$\overline{\rho}(m|\xi,\sigma) := \frac{\frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma(m)}{\frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma(m) + x^{\xi}(m)}.$$

Then, for any $m_1 \in M_1^{\xi}$,

$$\mu(1|\rho, m_1) \geq \mu^* \Leftrightarrow \rho \geq \overline{\rho}(m|\xi,\sigma);$$

and for any $m_0 \in M_0^{\xi}$, $\mu(1|\rho, m_0) < \mu^*$ for any $\rho \in [0,1]$. It follows that, for any $m_1 \in M_1^{\xi}$ and $m_0 \in M_0^{\xi}$,

$$V_u(m_1|\sigma,\alpha) = \int \left[\mathbb{1}_{(\overline{\rho}(m_1|\xi,\sigma),1]} + \mathbb{1}_{\overline{\rho}(m_1|\xi,\sigma)}\alpha(1|\overline{\rho}(m_1|\xi,\sigma), m_1)\right]\frac{1-\rho}{1-\rho_0}\mathrm{d}\iota(\rho), \tag{2.7}$$
$$V_r(m_1|\sigma,\alpha) = \int \left[\mathbb{1}_{(\overline{\rho}(m_1|\xi,\sigma),1]} + \mathbb{1}_{\overline{\rho}(m_1|\xi,\sigma)}\alpha(1|\overline{\rho}(m_1|\xi,\sigma), m_1)\right]\frac{\rho}{\rho_0}\mathrm{d}\iota(\rho), \tag{2.8}$$
$$V_u(m_0|\sigma,\alpha) = V_r(m_0|\sigma,\alpha) = 0.$$

The following lemma shows that pooling messages do not affect equilibrium payoffs.

Lemma 2.4. *Fix $(\xi,\iota) \in \Xi \times \mathcal{I}$. Suppose (σ,α,μ) is a (ξ,ι)-equilibrium with strictly positive*

Sender ex ante payoff. There exists $\tilde{m}_1 \in M_1^{\xi}$ *and* $\tilde{m}_0 \in M_0^{\xi}$*, with*

$$\tilde{\xi}\left(m|\cdot\right) := \begin{cases} \xi\left(M_1^{\xi}|\cdot\right) & \text{if } m = \tilde{m}_1 \\ \xi\left(M_0^{\xi}|\cdot\right) & \text{if } m = \tilde{m}_0 \\ 0 & \text{otherwise} \end{cases}, \ \sigma^*\left(m|\cdot\right) := \mathbb{1}_{\{m=\tilde{m}_1\}},$$

such that $(\sigma^*, \alpha^*, \mu^*)$ *is a* $(\tilde{\xi}, \iota)$*-equilibrium is with the same payoffs for the players.*

Proof. That Sender's ex ante payoff is strictly positive implies that the unreliable Sender's payoff must be strictly positive, $M_1^{\xi}, M_0^{\xi} \neq \varnothing$, and $\text{supp}(\sigma) \subseteq M_1^{\xi}$. By IC,

$$V_u\left(m\right) = V_u\left(m'\right) \geq V_u\left(m''\right) \ \forall m, m' \in \text{supp}\left(\sigma\right) \ \forall m'' \in M_1^{\xi} \backslash \text{supp}\left(\sigma\right),$$

where

$$V_u\left(m\right) \in \begin{cases} \{1\} & \text{if } x^{\xi}\left(m\right) > 0 \\ [0,1] & \text{if } x^{\xi}\left(m\right) = 0 \\ \{0\} & \text{if } x^{\xi}\left(m\right) < 0 \end{cases} \cdot$$

Let $\overline{\rho}_m = \overline{\rho}(m|\xi, \sigma)$ and define

$$\overline{\rho}_{\min} := \min_{m \in \text{supp}(\sigma) \cap M_1^{\xi}} \overline{\rho}_m, \ \overline{\rho}_{\max} := \max_{m \in \text{supp}(\sigma) \cap M_1^{\xi}} \overline{\rho}_m.$$

Let $m_{\min}$ and $m_{\max}$be such that $\overline{\rho}_{\min} = \overline{\rho}_{m_{\min}}$ and $\overline{\rho}_{\max} = \overline{\rho}_{m_{\max}}$.

First, suppose that $\text{supp}(\sigma) = M_1^{\xi}$. Observe that pooling messages in $\text{supp}(\sigma) \cap M_1^{\xi}$

results in a threshold reliability belief that is a weighted average:

$$\overline{\rho}_{\text{supp}(\sigma)\cap M_1^\xi} = \frac{\sum_{m\in \text{supp}(\sigma)\cap M_1^\xi} \frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma\left(m\right)}{\sum_{m\in \text{supp}(\sigma)\cap M_1^\xi} \frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma\left(m\right) + x^\xi\left(m\right)}$$
$$= \sum_{\tilde{m}\in \text{supp}(\sigma)\cap M_1^\xi} \frac{\frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma\left(\tilde{m}\right) + x^\xi\left(\tilde{m}\right)}{\sum_{m\in \text{supp}(\sigma)\cap M_1^\xi} \frac{\mu^*-\mu_0}{\mu_0(1-\mu^*)}\sigma\left(m\right) + x^\xi\left(m\right)}\overline{\rho}\left(\tilde{m}|\sigma,\alpha\right).$$

Hence, pooling weights on $\text{supp}(\sigma)\cap M_1^\xi$ to some $\tilde{m}_1 \in \text{supp}(\sigma)\cap M_1^\xi$ results in a threshold such that

$$\tilde{\rho} := \overline{\rho}\left(\text{supp}\left(\sigma\right)\cap M_1^\xi|\xi,\sigma\right) \in \left[\overline{\rho}_{\min}, \overline{\rho}_{\max}\right].$$

If $\overline{\rho}_{\min} < \overline{\rho}_{\max}$, $\tilde{\rho} \in (\overline{\rho}_{\min}, \overline{\rho}_{\max})$ and

$$\iota\left(\left(\overline{\rho}_{\min}, \overline{\rho}_{\max}\right)\right) = 0$$

because the unreliable Sender's incentive compatibility requires, for any $m \in \text{supp}(\sigma)$,

$$\begin{aligned} 0 &= V_u\left(\overline{m}_{\min}\right) - V_u\left(m\right) \\ &= \int \left(\left[\mathbb{1}_{(\overline{\rho}_{\min},1]} + \mathbb{1}_{\overline{\rho}_{\min}}\alpha\left(1|\overline{\rho}_{\min}, m_{\min}\right)\right] - \left[\mathbb{1}_{(\overline{\rho}_m,1]} + \mathbb{1}_{\overline{\rho}_m}\alpha\left(1|\overline{\rho}_m, m\right)\right]\right)\frac{1-\rho}{1-\rho_0}\mathrm{d}\iota\left(\rho\right) \\ &= \int \left(\left[\mathbb{1}_{(\overline{\rho}_{\min},\overline{\rho}_m)} + \mathbb{1}_{\overline{\rho}_{\min}}\alpha\left(1|\overline{\rho}_{\min}, m_{\min}\right) + \mathbb{1}_{\overline{\rho}_m}\left[1-\alpha\left(1|\overline{\rho}_m, m\right)\right]\right]\right)\frac{1-\rho}{1-\rho_0}\mathrm{d}\iota\left(\rho\right). \end{aligned}$$

Thus, pooling messages in $\text{supp}(\sigma)\cap M_1^\xi$ to $\tilde{m}_1$ would not alter the players' payoffs.

Suppose now there exists $m'' \in M_1^\xi\backslash\text{supp}(\sigma)$ such that $x^\xi(m'') > 0$. Then, by biased Sender's incentive compatibility, it must be that $V_u(m) = 1$ for all $m \in \text{supp}(\sigma)$. By the argument above, pooling messages in $\text{supp}(\sigma)\cap M_1^\xi$ would not affect Sender's incentives. Moreover, putting weights on $M_1^\xi\backslash\text{supp}(\sigma)$ to $\tilde{m}_1$ can only lower $\tilde{\rho}$, which, in turn, can only weakly improve the payoffs. Since V_u is already maximal at one, it follows that pooling messages would not alter the players' payoff. Finally, suppose there exists $m'' \in M_1^{\xi} \backslash \text{supp}(\sigma)$ such that $x^{\xi}(m'') = 0$. Because pooling m'' would not alter $\tilde{\rho}$, players' payoff remain unchanged. To ensure that $(\sigma^*, \alpha^*, \mu^*)$ is a $(\tilde{\xi}, \iota)$-equilibrium, we can specify off-path μ^* to equal μ_0 and ensure that α^* is optimal for the Receiver given μ^*. ■

Remark 2.1. Importantly, given any (σ, α, μ) is a (ξ, ι)-equilibrium, the payoffs from a $(\tilde{\xi}, \iota)$-equilibrium derived using the lemma does not depend on σ since it only matters in choosing $\tilde{m}_1$.

I now show that pooling messages using the lemma above would not affect the choice of an investigation.

Lemma 2.5. *Fix $(\xi, \iota) \in \Xi \times \mathcal{I}$. Suppose (σ, α, μ) is a (ξ, ι)-equilibrium and let $(\sigma^*, \alpha^*, \mu^*)$ be a $(\tilde{\xi}, \iota)$-equilibrium derived via the previous lemma. Suppose there exists $\tilde{\iota}$ such that players' $(\tilde{\xi}, \tilde{\iota})$-equilibrium payoffs are different from their payoff under $(\sigma^*, \alpha^*, \mu^*)$. Then, there exists a $(\xi, \tilde{\iota})$-equilibrium in which the players' payoff are the same as in the $(\tilde{\xi}, \tilde{\iota})$-equilibrium.*

Proof. By the previous lemma,

$$V_S(\xi, \iota, \sigma, \alpha) = V_S\left(\tilde{\xi}, \iota, \sigma^*, \alpha^*\right),\ V_R(\xi, \iota, \sigma, \alpha) = V_R\left(\tilde{\xi}, \iota, \sigma^*, \alpha^*\right).$$

Suppose there exists $\tilde{\iota} \in \mathcal{I}$ and a tuple $(\tilde{\sigma}, \tilde{\alpha}, \tilde{\mu})$ that is a $(\tilde{\xi}, \tilde{\iota})$-equilibrium and

$$V_R\left(\tilde{\xi}, \tilde{\iota}, \tilde{\sigma}, \tilde{\alpha}\right) \neq V_R\left(\tilde{\xi}, \iota, \sigma^*, \alpha^*\right) \text{ or } V_S\left(\tilde{\xi}, \tilde{\iota}, \tilde{\sigma}, \tilde{\alpha}\right) \neq V_S\left(\tilde{\xi}, \iota, \sigma^*, \alpha^*\right).$$

The goal is to construct a tuple $(\widehat{\sigma}, \widehat{\alpha}, \widehat{\mu})$ that is a $(\xi, \tilde{\iota})$-equilibrium such that

$$V_j\left(\xi, \tilde{\iota}, \widehat{\sigma}, \widehat{\alpha}, \widehat{\mu}\right) = V_j\left(\tilde{\xi}, \tilde{\iota}, \tilde{\sigma}, \tilde{\alpha}\right)\ \forall j \in \{S, R\}.$$

Let

$$\widehat{\sigma}\left(m_1\right) := \frac{x^{\xi}\left(m_1\right)}{x^{\xi}\left(M_1^{\xi}\right)},\ \widehat{\alpha}\left(1|\cdot,\widehat{\sigma}\right) = \widehat{\alpha}\left(1|\cdot,\sigma^*\right).$$

The former implies that

$$\overline{\rho}_{m_1}\left(\xi,\hat{\sigma}\right) = \overline{\rho}_{\tilde{m}_1}\left(\tilde{\xi},\sigma^*\right)\ \forall m_1 \in M_1^{\xi}.$$

Then,

$$\begin{aligned}
&V_{\mathrm{S}}\left(\xi,\tilde{\iota},\hat{\sigma},\widehat{\alpha}\right)\\
&= \sum_{m_1 \in M_1^{\xi}} \int \left[\mathbb{1}_{\left(\overline{\rho}_{m_1}(\xi,\hat{\sigma}),1\right]} + \mathbb{1}_{\overline{\rho}_{m_1}(\xi,\hat{\sigma})}\widehat{\alpha}\left(1|\overline{\rho}_{m_1}\left(\xi,\widehat{\sigma}\right),\widehat{\sigma}\right)\right]\left[\xi\left(m_1\right)\rho + (1-\rho)\right]\mathrm{d}\tilde{\iota}\left(\rho\right)\\
&= \int \left[\mathbb{1}_{\left(\overline{\rho}_{\tilde{m}_1}(\tilde{\xi},\sigma^*),1\right]} + \mathbb{1}_{\overline{\rho}_{\tilde{m}_1}(\tilde{\xi},\sigma^*)}\alpha^*\left(1|\overline{\rho}_{\tilde{m}_1}\left(\tilde{\xi},\sigma^*\right),\sigma^*\right)\right]\left[\xi\left(M_1^{\xi}\right)\rho + (1-\rho)\right]\mathrm{d}\tilde{\iota}\left(\rho\right)\\
&= V_{\mathrm{S}}\left(\tilde{\xi},\tilde{\iota},\tilde{\sigma},\tilde{\alpha}\right) = V_{\mathrm{S}}\left(\tilde{\xi},\tilde{\iota},\sigma^*,\alpha^*\right)
\end{aligned}$$

and

$$V_{\mathrm{R}}\left(\xi,\tilde{\iota},\hat{\sigma},\widehat{\alpha}\right) = V_{\mathrm{R}}\left(\tilde{\xi},\tilde{\iota},\sigma^*,\alpha^*\right).$$

Thus, players' payoffs would be the same as in $(\xi,\tilde{\iota},\hat{\sigma},\widehat{\alpha})$ and $(\tilde{\xi},\tilde{\iota},\sigma^*,\alpha^*)$ (for appropriately specified belief maps). ■

Let us now prove Lemma 2.1.

Lemma 2.1. *Fix some $m_0, m_1 \in M$. Any commitment and no-commitment equilibrium payoffs are achievable via some $(\xi,\overline{\sigma}) \in \Xi^2$ pair such that* $\operatorname{supp}(\xi) = \{m_0, m_1\}$*, $\xi(m_1|1) = 1$, $\xi(m_1|0) \leq 1-\underline{\rho}$, and $\overline{\sigma}(m_1|\cdot) = 1$.*

Proof. By the previous two lemmata, given any (ξ,ι)-equilibrium, there exists a payoff equivalent $(\tilde{\xi},\iota)$-equilibrium. Moreover, if there exists a $(\xi,\tilde{\iota})$-equilibrium with different payoffs, there exists $(\tilde{\xi}, \tilde{\iota})$-equilibrium that obtains the same payoffs. Thus, it is without loss to focus on the equivalence class of Ξ given by $\tilde{\xi}$, which, in turn, implies that we can focus on $\text{supp}(\xi) = \{m_0, m_1\}$ for some $m_0, m_1 \in M$ and σ such that $|\text{supp}(\sigma)| = 1$.

Recall $\underline{\rho} = \frac{\mu^* - \mu_0}{\mu^*(1-\mu_0)}$. By the previous Lemma, it is without loss to focus on $\xi \in \Xi$ with supports $\{m_0, m_1\}$ and in which $\overline{\sigma}$ always sends m_1. In such equilibrium, for any $\iota \in \mathcal{I}$,

$$V_S(\xi, \iota, \overline{\sigma}) = \int \left[\mathbb{1}_{\left(\overline{\rho}_{m_1}, 1\right]} + \mathbb{1}_{\overline{\rho}_{m_1}} \alpha \left(1 | \overline{\rho}_{m_1}, \overline{\sigma}\right) \right] \left[\rho \xi (m_1) + (1-\rho)\right] \mathrm{d}\iota(\rho),$$

where

$$\overline{\rho}_{m_1} = \frac{\frac{1-\mu_0}{\mu_0} \frac{\mu^*}{1-\mu^*} - 1}{\frac{1-\mu_0}{\mu_0} \frac{\mu^*}{1-\mu^*} - 1 + \xi(m_1|1) - \xi(m_1|0) \frac{1-\mu_0}{\mu_0} \frac{\mu^*}{1-\mu^*}}.$$

Since $\{m_1\} = M_1^{\xi}$, it must be that

$$x^{\xi}(m_1) > 0 \Leftrightarrow \xi(m_1|1) > \xi(m_1|0) \frac{1-\mu_0}{\mu_0} \frac{\mu^*}{1-\mu^*}.$$

Observe that $\xi(m_1) = \xi(m_1|1)\mu_0 + \xi(m_1|0)(1-\mu_0)$ is increasing in $\xi(m_1|1)$ and $\overline{\rho}_{m_1}$ is decreasing in $\xi(m_1|1)$, so that Sender's payoff is increasing in $\xi(m_1|1)$ and, moreover, larger $\xi(m_1|1)$ relaxes the constraint on $\xi(m_1|0)$. It follows that $\xi(m_1|1) = 1$. Finally, given $m_1 \in M_1^{\xi}$, it must be the $\text{supp}(\sigma) = m_1$. ■

2.A.3 Proof of Theorem 2.2

I first characterise the Sender's maxmin experiment.

Lemma 2.2. *The Sender's maxmin experiment is given by*

$$\widehat{\rho}^{maxmin} = \max \left\{ \underline{\rho}, \left(1 + \sqrt{\frac{\mu_0}{\mu^* - \mu_0} \frac{1-\rho_0}{\rho_0}}\right)^{-1} \right\} \in \left[\underline{\rho}, \rho_0\right).$$

Proof. First, observe that $V_S(\cdot) \in [0,1]$. Kamenica and Gentzkow (2011) implies that the problem is equivalent to concavifying $-V_S(\widehat{\rho}, \delta_\rho)$. In particular, it is without loss to assume that $-V_S(\widehat{\rho}, \delta_\rho)$ is upper semicontinuous (i.e., the Receiver breaks ties by not taking action).

Suppose $\rho_0 < \underline{\rho}$. Then, setting $\iota = \delta_{\rho_0}$ ensures that $V_S(\widehat{\rho}, \delta_{\rho_0}) = 0$ for any $\widehat{\rho} \in [\underline{\rho}, 1]$. Hence, $V_S^{\text{maxmin}} = 0$. Suppose instead that $\rho_0 \geq \underline{\rho}$. If $\widehat{\rho} \geq \rho_0$, $\iota = \delta_{\rho_0}$ ensures that Sender's payoff is zero. If $\widehat{\rho} \in [\underline{\rho}, \rho_0)$, then $\supseteq_{\widehat{\rho}}$ such that $\text{supp}(\supseteq_{\widehat{\rho}}) = \{\widehat{\rho}, 1\}$ and the Sender's payoff in this case is strictly positive $(1 - \frac{1}{\widehat{\rho}} \frac{\mu^* - \mu_0}{\mu^*}) \supseteq_{\widehat{\rho}}(1)$. Given this, the Sender's problem is to choose

$$\max_{\widehat{\rho} \in [\underline{\rho}, \rho_0)} \left(1 - \frac{\mu^* - \mu_0}{\mu^*} \frac{1}{\widehat{\rho}}\right) \frac{\rho_0 - \widehat{\rho}}{1 - \widehat{\rho}},$$

which is solved by $\widehat{\rho}^{\text{maxmin}}$ given in the lemma. ■

Remark 2.2. $\widehat{\rho}^{\text{maxmin}} = \underline{\rho}$ if $\rho_0 \leq \rho_{0,0} := \frac{\mu^* - \mu_0}{\mu^*} \frac{1}{1 - (2 - \mu^*)\mu_0}$. $\widehat{\rho}^{\text{maxmin}}$ is strictly increasing in $\rho_0 \in (\rho_{0,1}, 1]$ while V_S^{maxmin} is also strictly increasing in $\rho_0 \in [\underline{\rho}, 1]$. Moreover, $\lim_{\rho_0 \to 1} \widehat{\rho}^{\text{maxmin}} = 1$ and $\lim_{\rho_0 \to 1} V_S^{\text{maxmin}} = \frac{\mu_0}{\mu^*}$.

Recall that

$$\iota_z = \begin{cases} \frac{\rho_0}{z} \delta_z + \frac{\rho_0}{z - \rho_0} \delta_0 & \text{if } z \in (\rho_0, 1] \\ \delta_{r_0} & \text{if } z = \rho_0 \end{cases}.$$

Thus, ι_z is an investigation that induces posterior belief about the Sender's reliability (i) at 0 and at z if $z > \rho_0$ or (ii) at ρ_0 if $z = \rho_0$. Thus, $\iota_1 = \rho_0 \delta_1 + (1 - \rho_0)\delta_0$ is a full investigation while $\iota_{\rho_0} = \delta_{\rho_0}$ is a fully uninformative investigation.

Theorem 2.2. $(\widehat{\rho}^*, i^{\min}_{\widehat{\rho}^*,\iota_{z^*}}(\cdot))$ *is a commitment equilibrium, where*

$$\widehat{\rho}^* = \max\left\{\underline{\rho}, \frac{\mu^* - \mu_0}{\mu^*}\frac{\rho_0}{1 - V_{\mathrm{S}}^{\mathrm{maxmin}}}\right\},$$
$$z^* = \max\left\{\min\left\{\frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1-\mu_0)\,\rho_0}, 1\right\}, \rho_0\right\}.$$

The players' commitment equilibrium payoffs are

$$V_{\mathrm{R}}\left(\widehat{\rho}^*, \iota_{z^*}\right) = \min\left\{\overline{V}_{\mathrm{R}}, \frac{\mu_0}{\mu^*}\left[(1-\mu_0)\,\rho_0 + V_{\mathrm{S}}^{\mathrm{maxmin}}\right] - V_{\mathrm{S}}^{\mathrm{maxmin}}, \frac{\mu_0}{\mu^*} - V_{\mathrm{S}}^{\mathrm{maxmin}}\right\}$$
$$V_{\mathrm{S}}\left(\widehat{\rho}^*, \iota_{z^*}\right) = \max\left\{V_{\mathrm{S}}\left(\underline{\rho}, \iota_1\right), V_{\mathrm{S}}^{\mathrm{maxmin}}\right\}.$$

In particular, both payoffs are strictly positive for any $\rho_0 \in (0,1)$.

Proof. Let us start by observing the following: (i) fixing $\widehat{\rho}$, the Receiver's optimal investigation is fully informative; and (ii) fixing ι, the Receiver's payoff is decreasing in $\widehat{\rho}$ (i.e., increasing in the informativeness of the Sender's experiment). Thus, the Receiver's maximal payoff is given by $\overline{V}_{\mathrm{R}} := V_{\mathrm{R}}(\underline{\rho}, \iota_1)$.

Suppose $\widehat{\rho}^{\mathrm{maxmin}} = \underline{\rho}$ which is the case when

$$\rho_0 \leq \rho_{0,0} \equiv \frac{\mu^* - \mu_0}{\mu^*}\frac{1}{1 - (2-\mu^*)\,\mu_0}.$$

Suppose the Receiver follows $i^{\min}(\cdot)$ for all $\widehat{\rho} \neq \underline{\rho}$ but chooses to fully investigate when $\widehat{\rho} = \underline{\rho}$. Then, because $i^{\min}(\underline{\rho}) \neq \iota_1$, the Sender's payoff from choosing $\underline{\rho}$ is greater than $V_{\mathrm{S}}^{\mathrm{maxmin}}$. Hence, the Receiver can obtain $\overline{V}_{\mathrm{R}}$ via such $i^*(\cdot)$.

Now suppose that $\widehat{\rho}^{\mathrm{maxmin}} > \underline{\rho}$. Define $\widehat{\rho}^+$ as

$$\widehat{\rho}^+ := \min\left\{\widehat{\rho}' \in \left[\underline{\rho}, 1\right] : V_{\mathrm{S}}\left(\widehat{\rho}', \iota_1\right) \geq V_{\mathrm{S}}^{\mathrm{maxmin}}\right\} = \frac{\mu^* - \mu_0}{\mu^*}\frac{\rho_0}{\rho_0 - V_{\mathrm{S}}^{\mathrm{maxmin}}},$$

which is strictly increasing in ρ_0 (given $\widehat{\rho}^{\text{maxmin}} > \underline{\rho}$). Notice also that because $\lim_{\rho_0 \to 1} \widehat{\rho}^+ = 1$ and

$$\lim_{\rho_0 \to \rho_{0,0}} \widehat{\rho}^+ \to \frac{\mu^* - \mu_0}{\mu^*} \frac{\rho_0}{\rho_0 - \frac{\mu^*(1-\mu_0)\rho_0 - (\mu^* - \mu_0)}{1-\mu^*}} < \underline{\rho}.$$

Thus, there exists a unique prior about Sender's type $\rho_{0,1} \in (\rho_{0,1}, 1)$ such that $\widehat{\rho}^+ = \underline{\rho}$ given by

$$\rho_{0,1} = \frac{1}{1 + \frac{\mu_0}{\mu^* - \mu_0} \left(1 - \sqrt{\mu^*}\right)^2}.$$

This means that for any $\rho_0 \in (\rho_{0,0}, \rho_{0,1}]$, the Receiver can induce the Sender to choose $\widehat{\rho} = \underline{\rho}$ with a full investigation.

Suppose now that $\rho_0 > \rho_{0,1}$ so that $\widehat{\rho}^+ > \underline{\rho}$. The Receiver must decide whether to induce $\widehat{\rho} \in [\underline{\rho}, \widehat{\rho}^+)$ by choosing some partially revealing investigation. Note that $\widehat{\rho}^+ < \widehat{\rho}^{\text{maxmin}} < \rho_0$. Given any $\widehat{\rho} \in [\underline{\rho}, \rho_0]$, both the Sender and the Receiver's payoffs are the same for any $\iota \in \mathcal{I}$ such that $\text{supp}(\iota) \subseteq (\widehat{\rho}, 1]$. In particular, the payoffs are the same as when the Receiver chooses ι_{ρ_0} (i.e., does not investigate the Sender); i.e., $V_j(\widehat{\rho}, \iota_{\rho_0})$. Observe that the Receiver can approximately induce this payoff by choosing ι_z and z arbitrarily close to ρ_0 and the Receiver will be strictly better off. In fact, to induce a payoff for the Sender between $V_{\text{S}}(\widehat{\rho}, \iota_{\rho_0})$ and $V_{\text{S}}(\widehat{\rho}, \iota_1)$, the Receiver should choose a partial investigation in $\{\iota_z\}_{z \in [\rho_0, 1]}$. Thus, the Receiver's problem can be written as

$$\max_{\widehat{\rho} \in [\underline{\rho}, \widehat{\rho}^+], z \in [\rho_0, 1]} V_{\text{R}}\left(\widehat{\rho}, \iota_z\right) \text{ s.t. } V_{\text{S}}\left(\widehat{\rho}, \iota_z\right) \geq V_{\text{S}}^{\text{maxmin}}.$$

Substituting the functional forms yields

$$\max_{\widehat{\rho} \in [\underline{\rho}, \widehat{\rho}^+], z \in [\rho_0, 1]} \frac{\mu^* - \mu_0}{\mu^*} \left(\frac{z}{\widehat{\rho}} - 1\right) \frac{\rho_0}{z} \text{ s.t. } \widehat{\rho} \geq \frac{\rho_0 \frac{\mu^* - \mu_0}{\mu^*}}{\frac{\rho_0}{z} - V_{\text{S}}^{\text{maxmin}}}.$$

Because the objective is strictly decreasing in $\widehat{\rho}$, at the optimal either $\widehat{\rho} = \underline{\rho}$ or the constraint must bind; i.e.,

$$\widehat{\rho}^* = \max\left\{\underline{\rho}, \frac{\rho_0 \frac{\mu^* - \mu_0}{\mu^*}}{\frac{\rho_0}{z^*} - V_{\mathrm{S}}^{\mathrm{maxmin}}}\right\}.$$

Moreover, if $\widehat{\rho}^* = \underline{\rho}$, then $z^* = 1$ because the objective is strictly increasing in z. Note that

$$V_{\mathrm{R}}\left(\frac{\rho_0 \frac{\mu^* - \mu_0}{\mu^*}}{\frac{\rho_0}{\rho} - V_{\mathrm{S}}^{\mathrm{maxmin}}}, \iota_z\right) = \frac{\rho_0}{z}\frac{\mu_0}{\mu^*} - V_{\mathrm{S}}^{\mathrm{maxmin}}$$

is decreasing in z. Hence, if $\widehat{\rho}^* \geq \underline{\rho}$, the receiver chooses the smallest possible $z \in [\rho_0, 1]$ such that

$$\frac{\rho_0 \frac{\mu^* - \mu_0}{\mu^*}}{\frac{\rho_0}{z} - V_{\mathrm{S}}^{\mathrm{maxmin}}} \geq \underline{\rho} \Leftrightarrow z \geq \frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1 - \mu_0)\,\rho_0} > \rho_{0,1},$$

where the right-hand side is strictly decreasing in ρ_0. Since

$$\lim_{\rho_0 \to 1} \frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1 - \mu_0)\,\rho_0} = \frac{\mu_0}{\mu^* + \mu_0\,(1 - \mu_0)} < 1,$$

there exists $\rho_{0,2} \in (\rho_{0,1}, 1)$ such that

$$\frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1 - \mu_0)\,\rho_0} \leq \rho_0 \; \forall \rho_0 \geq \rho_{0,2}.$$

In fact,

$$\rho_{0,2} = \frac{\mu_0}{\mu^*\,(2 - \mu_0) - 2\sqrt{\mu^*\,(1 - \mu_0)\,(\mu^* - \mu_0)}}.$$

Therefore,

$$z^* = \max\left\{\frac{\rho_0}{V_{\mathrm{S}}^{\mathrm{maxmin}} + (1 - \mu_0)\,\rho_0}, \rho_0\right\}.$$

Observe that

$$\frac{\rho_0}{V_S^{\text{maxmin}} + (1-\mu_0)\,\rho_0} > 1 \Leftrightarrow \mu_0\rho_0 > V_S^{\text{maxmin}} \Leftrightarrow \rho_0 \geq \rho_{0,1}.$$

Hence, z^* for any $\rho_0 \in [0,1]$ can be defined as in the statement of the proposition. ∎

Remark 2.3. The proof identifies the following regions.

- ▷ $\rho_0 \in [0, \rho_{0,0}]$, where $\rho_{0,0} \in (\underline{\rho}, 1)$: $\widehat{\rho}^* = \widehat{\rho}^{\text{maxmin}} = \underline{\rho}$; $z^* = 1$; $V_R = \overline{V}_R$; $V_S = V_S(\underline{\rho}, \iota_1) > V_S^{\text{maxmin}} \equiv V_S(\underline{\rho}, \iota^{\min}(\underline{\rho}))$.
- ▷ $\rho_0 \in (\rho_{0,0}, \rho_{0,1})$: $\widehat{\rho}^* = \underline{\rho} < \widehat{\rho}^{\text{maxmin}}$; $z^* = 1$; $V_R = \overline{V}_R$; $V_S = V_S(\underline{\rho}, \iota_1) > V_S^{\text{maxmin}}$.
- ▷ $\rho_0 \in [\rho_{0,1}, \rho_{0,2}]$: $\widehat{\rho}^* = \frac{\rho_0 \frac{\mu^*-\mu_0}{\mu^*}}{\frac{\rho_0}{z^*} - V_S^{\text{maxmin}}} = \underline{\rho}$; $z^* = \frac{\rho_0}{V_S^{\text{maxmin}} + (1-\mu_0)\rho_0}$, $V_R(\widehat{\rho}^*, \iota_{z^*}) = \frac{\mu_0}{\mu^*}(1-\mu_0)\,\rho_0 - \frac{\mu^*-\mu_0}{\mu^*} V_S^{\text{maxmin}} < \overline{V}_R$ and $V_S = V_S^{\text{maxmin}}$.
- ▷ $\rho_0 \in (\rho_{0,2}, 1]$: $\widehat{\rho}^* = \frac{\rho_0 \frac{\mu^*-\mu_0}{\mu^*}}{1 - V_S^{\text{maxmin}}}$, $z^* = \rho_0$, $V_R(\frac{\rho_0 \frac{\mu^*-\mu_0}{\mu^*}}{1-V_S^{\text{maxmin}}}, \delta_{\rho_0}) = \frac{\mu_0}{\mu^*} - V_S^{\text{maxmin}} < \overline{V}_R$ and $V_S = V_S^{\text{maxmin}}$.

2.A.4 Proof of Theorem 2.3

Theorem 2.3. *A pair ($\widehat{\rho}^\lambda, i^\lambda(\cdot)$) is a λ-equilibrium, where*

$$\widehat{\rho}^\lambda = \begin{cases} \Lambda^{-1}(\lambda) & \text{if } \lambda \geq \lambda^* \\ \widehat{\rho}^{\text{maxmin}} & \text{if } \lambda < \lambda^* \end{cases}, \; i^\lambda(\widehat{\rho}) = \begin{cases} \iota_1 & \text{if } \lambda \geq \Lambda(\widehat{\rho}) \\ i^{\min}(\widehat{\rho}) & \text{if } \lambda < \Lambda(\widehat{\rho}) \end{cases}.$$

If $\lambda \in \lambda^*$, λ-equilibrium coincides with 0-equilibrium.

Proof. First suppose that $\rho_0 \geq \underline{\rho}$. Let $\Lambda(\widehat{\rho})$ denote the normalised weight on the Receiver's preferences such that $V_{\mathrm{T}}^{\lambda}(\widehat{\rho},\delta_1)=0$. Then,

$$U_{\mathrm{T}}^{\lambda}\left(\widehat{\rho},\delta_1\right)\geq 0\Leftrightarrow\lambda\geq\Lambda\left(\widehat{\rho}\right)\equiv\frac{\widehat{\rho}\frac{\mu^*}{\mu^*-\mu_0}-1}{1-\widehat{\rho}}$$

Because $\Lambda'(\cdot)>0$,$\Lambda^{-1}(\cdot)$ is well defined and given by

$$V_{\mathrm{T}}^{\lambda}\left(\widehat{\rho},\delta_1\right)\geq 0\Leftrightarrow\widehat{\rho}\leq\Lambda^{-1}\left(\lambda\right)=\frac{1+\lambda}{\frac{\mu^*}{\mu^*-\mu_0}+\lambda}\in\left[\left(1-\mu_0\right)\underline{\rho},1\right).$$

Suppose we fix $\lambda\in(0,\infty)$. If $\Lambda^{-1}(\lambda)\leq\underline{\rho}$; i.e.,

$$\Lambda^{-1}\left(\lambda\right)=\frac{1+\lambda}{\frac{\mu^*}{\mu^*-\mu_0}+\lambda}\leq\underline{\rho}\Leftrightarrow\lambda\leq\frac{\mu^*}{1-\mu^*}.$$

Then, the Third Party's optimal investigation against any $\widehat{\rho}\in[\underline{\rho},1]$ is $i^{\min}(\widehat{\rho})$ thus the λ-equilibrium payoff coincides with 0-equilibrium.

Now suppose that $\Lambda^{-1}(\lambda)>\underline{\rho}$. Then, the Third Party conducts a full investigation upon seeing any $\widehat{\rho}\in[\underline{\rho},\Lambda^{-1}(\lambda)]$ and $i^{\min}(\widehat{\rho})$ for any $(\Lambda^{-1}(\lambda),1]$. Recalling that the Sender's payoff conditional on a full investigation is strictly increasing in $\widehat{\rho}$, the Sender's most preferred threshold in the interval $[\underline{\rho},\Lambda^{-1}(\lambda)]$ is $\Lambda^{-1}(\lambda)$.

Moreover, if $\Lambda^{-1}(\lambda)>\widehat{\rho}^{\mathrm{maxmin}}$, the Sender in fact chooses $\Lambda^{-1}(\lambda)$ because choosing any $\widehat{\rho}>\Lambda^{-1}(\lambda)$ results in a payoff that is weakly lower than $V_{\mathrm{S}}^{\mathrm{maxmin}}$ and choosing $\Lambda^{-1}(\lambda)$ results in payoff at least $V_{\mathrm{S}}^{\mathrm{maxmin}}$ (when the Sender chooses $\Lambda^{-1}(\lambda)$, the Third Party does not punish the Sender and the Sender benefits from a less informative experiment than $\widehat{\rho}^{\mathrm{maxmin}}$). Recall that for $\rho_0\in[\underline{\rho},\rho_{0,0}]$, $\widehat{\rho}^{\mathrm{maxmin}}=\underline{\rho}$ and so $\Lambda^{-1}(\lambda)>\widehat{\rho}^{\mathrm{maxmin}}$ holds if $\lambda>\frac{\mu^*}{1-\mu^*}$. When $\rho_0\in[\rho_{0,0},1)$, then

$\widehat{\rho}^{\text{maxmin}} = (1 + \sqrt{\frac{\mu_0}{\mu^* - \mu_0} \frac{1-\rho_0}{\rho_0}})^{-1}$ and

$$\Lambda^{-1}(\lambda) \geq \widehat{\rho}^{\text{maxmin}} = \frac{1}{1 + \sqrt{\frac{\mu_0}{\mu^* - \mu_0} \frac{1-\rho_0}{\rho_0}}} \Leftrightarrow \lambda > \sqrt{\frac{\mu_0}{\mu^* - \mu_0} \frac{\rho_0}{1-\rho_0}} - 1 > \frac{\mu^*}{1-\mu^*} > 0.$$

Now suppose that $\Lambda^{-1}(\lambda) \in [\underline{\rho}, \widehat{\rho}^{\text{maxmin}}]$ when $\rho_0 \in [\rho_{0,0}, 1)$. Then, the Sender would either choose $\Lambda^{-1}(\lambda)$ to get a payoff associated with $V_S(\Lambda^{-1}(\lambda), \iota_1)$ or choose $\widehat{\rho}^{\text{maxmin}}$ and get a payoff V_S^{maxmin}. Thus, the Sender chooses the former if and only if

$$\Lambda^{-1}(\lambda) \geq \widehat{\rho}^{+} \Leftrightarrow \lambda > \frac{V_S^{\text{maxmin}}}{\frac{\mu_0}{\mu^*}\rho_0 - V_S^{\text{maxmin}}}.$$

Because $\Lambda^{-1}(\lambda) \geq \underline{\rho}$ and $\widehat{\rho}^{+} \leq \underline{\rho}$ if $\rho_0 \in [\rho_{0,0}, \rho_{0,1}]$, if ρ_0 lies in this interval, the inequality above holds; i.e., Sender chooses $\Lambda^{-1}(\lambda)$. If $\rho_0 \in (\rho_{0,1}, 1]$, then $\widehat{\rho}^{+} > \underline{\rho}$ and Sender chooses $\Lambda^{-1}(\lambda)$ if and only if $\Lambda^{-1}(\lambda) \geq \widehat{\rho}^{+}$. But because $\widehat{\rho}^{+} < \widehat{\rho}^{\text{maxmin}}$, for any fixed ρ_0, there exists λ large enough so that $\Lambda^{-1}(\lambda) > \widehat{\rho}^{+}$ and for λ small enough $\Lambda^{-1}(\lambda) < \widehat{\rho}^{+}$. ■

Remark 2.4. The proof identifies the following regions.

▷ If $\lambda \leq \frac{\mu^*}{1-\mu^*}$ ($\Leftrightarrow \Lambda^{-1}(\lambda) \leq \underline{\rho}$), the λ-balanced Third Party behaves as an adversarial Third Party. The Sender chooses $\widehat{\rho}^{\text{maxmin}}$ and gets V_S^{maxmin}.

▷ If $\lambda > \frac{\mu^*}{1-\mu^*}$ ($\Leftrightarrow \Lambda^{-1}(\lambda) > \underline{\rho}$).

  ▷ If $\rho_0 \in [\underline{\rho}, \rho_{0,0}]$, then $\widehat{\rho}^{\text{maxmin}} = \underline{\rho}$ and the Sender chooses $\Lambda^{-1}(\lambda)$, the Third Party fully investigates, ι_1. In this case, $\Lambda^{-1}(\lambda) > \underline{\rho} > \widehat{\rho}^{+}$.

  ▷ If $\rho_0 \in (\rho_{0,0}, 1)$, then $\widehat{\rho}^{\text{maxmin}} \in (\underline{\rho}, \rho_0)$.

    • If $\lambda \in \left(\frac{\mu^*}{1-\mu^*}, \sqrt{\frac{\mu_0}{\mu^* - \mu_0} \frac{\rho_0}{1-\rho_0}} - 1\right)$ ($\Leftrightarrow \Lambda^{-1}(\lambda) \in [\underline{\rho}, \widehat{\rho}^{\text{maxmin}}]$).

      ◦ The Sender chooses $\widehat{\rho}^{\text{maxmin}}$ if $\Lambda^{-1}(\lambda) < \widehat{\rho}^{+}$, and gets V_S^{maxmin}.

- ◦ The Sender chooses $\Lambda^{-1}(\lambda)$ if $\Lambda^{-1}(\lambda) \geq \widehat{\rho}^{+}$, the Third Party fully investigates, ι_1.
- If $\lambda > \sqrt{\frac{\mu_0}{\mu^*-\mu_0}\frac{\rho_0}{1-\rho_0}} - 1$ ($\Leftrightarrow \Lambda^{-1}(\lambda) > \widehat{\rho}^{\text{maxmin}} > \widehat{\rho}^{+}$), the Sender chooses $\Lambda^{-1}(\lambda)$, the Third Party fully investigates, ι_1.

Corollary 2.3. *The Receiver's λ^*-equilibrium payoff is greater than in any other λ-equilibrium.*

Proof. Let us consider the Sender's choice of an experiment against a λ-balanced Third Party, denoted $\widehat{\rho}(\lambda)$. If $\widehat{\rho} \leq \Lambda^{-1}(\lambda)$, then the Third Party behaves as if it were Receiver-aligned and fully reveals, in which case the Sender's payoff is

$$V_{\mathrm{S}}\left(\widehat{\rho}, \rho_0\delta_1 + (1-\rho_0)\,\delta_0\right) = \left(1 - \frac{1}{\widehat{\rho}}\frac{\mu^*-\mu_0}{\mu^*}\right)\rho_0 > 0.$$

Because the payoff is increasing in $\widehat{\rho}$, Sender prefers $\widehat{\rho} = \Lambda^{-1}(\lambda)$ among all $\widehat{\rho} \leq \Lambda^{-1}(\lambda)$. If, instead, $\widehat{\rho} > \Lambda^{-1}(\lambda)$, then the Third Party behaves as if it were adversarial and chooses $\iota = \delta_{\rho_0}$ if $\widehat{\rho} \geq \rho_0$ and $\iota = \frac{1-\rho_0}{1-\widehat{\rho}}\delta_{\widehat{\rho}} + \frac{\rho_0-\widehat{\rho}}{1-\widehat{\rho}}\delta_1$ if $\widehat{\rho} < \rho_0$. If $\widehat{\rho}^{\text{maxmin}} > \Lambda^{-1}(\lambda)$, then the Sender prefers $\widehat{\rho}^{\text{maxmin}}$ among all $\widehat{\rho} > \Lambda^{-1}(\lambda)$, and the Sender prefers $\Lambda^{-1}(\lambda)$ over $\widehat{\rho}^{\text{maxmin}}$ (and therefore all $\widehat{\rho} \leq \Lambda^{-1}(\lambda)$) if

$$\left(1 - \frac{1}{\Lambda^{-1}(\lambda)}\frac{\mu^*-\mu_0}{\mu^*}\right)\rho_0 \geq V_{\mathrm{S}}^{\text{maxmin}} \Leftrightarrow \Lambda^{-1}(\lambda) \geq \frac{\mu^*-\mu_0}{\mu^*}\frac{\rho_0}{\rho_0 - V_{\mathrm{S}}^{\text{maxmin}}}$$

where the right-hand side is less than $\widehat{\rho}^{\text{maxmin}}$. Alternatively, the Sender prefers $\widehat{\rho}^{\text{maxmin}}$ if

$$\Lambda^{-1}(\lambda) \leq \widehat{\rho}^{+}$$

which also implies that $\Lambda^{-1}(\lambda) \leq \widehat{\rho}^{\text{maxmin}}$. Let $\lambda^* = \widehat{\rho}^{+}$ be such that the inequality above holds with equality. Then,

$$\Lambda^{-1}\left(\lambda^{*}\right) \leq \frac{\mu^{*}-\mu_{0}}{\mu^{*}} \frac{\rho_{0}}{\rho_{0}-V_{\mathrm{S}}^{\mathrm{maxmin}}} \; \forall \lambda \leq \lambda^{*}$$

and so Sender's optimal choice against a λ-balanced Third Party and the third party's best responses are:

$$\widehat{\rho}\left(\lambda\right)=\begin{cases}\widehat{\rho}^{\mathrm{maxmin}} & \text{if } \lambda \leq \lambda^{*} \\ \max\{\Lambda^{-1}(\lambda), \underline{\rho}\} & \text{if } \lambda > \lambda^{*}\end{cases},$$

$$i\left(\widehat{\rho}\left(\lambda\right)\right)=\begin{cases}\frac{1-\rho_{0}}{1-\widehat{\rho}^{\mathrm{maxmin}}}\delta_{\widehat{\rho}^{\mathrm{maxmin}}}+\frac{\rho_{0}-\widehat{\rho}^{\mathrm{maxmin}}}{1-\widehat{\rho}^{\mathrm{maxmin}}}\delta_{1} & \text{if } \lambda \leq \lambda^{*} \\ (1-\rho_{0})\delta_{0}+\rho_{0}\delta_{1} & \text{if } \lambda > \lambda^{*}\end{cases}.$$

Because $\Lambda^{-1}(\lambda) \leq \widehat{\rho}^{\mathrm{maxmin}}$, $\Lambda^{-1}(\lambda^{*}) < \Lambda^{-1}(\lambda)$ for all $\lambda > \lambda^{*}$ the Receiver's payoff is greatest when $\lambda = \lambda^{*}$. Moreover, recall that $\Lambda^{-1}\left(\lambda^{*}\right) = \widehat{\rho}^{+}$ is exactly the condition that defined the experiment that the Receiver would induce in the commitment equilibrium (together with the feasibility constraint that $\Lambda^{-1}(\lambda^{*}) \geq \underline{\rho}$), and there and in this case too, a fully informative investigation is conducted. Thus, the two payoffs must coincide.

Finally, observe that when $\rho_{0} < \underline{\rho}$, since $V_{\mathrm{S}}^{\mathrm{maxmin}} = 0$ in this case, λ^{*}-balanced Third Party with $\Lambda^{-1}(\lambda^{*}) = \underline{\rho}$ induces the same payoff to the Receiver as in the no-commitment equilibrium. ■

Chapter 3

Knowing the informed player's payoffs and simple play in repeated games

We study a particular class of two-player, undiscounted, infinitely repeated game in which only one player is informed about the state of the world, and players observe only each other's actions in each stage of the repeated game.[1] Specifically, we focus on the class of these games in which the informed player's preferences are state independent. The absence of discounting implies that the informed player's actions in any initial finite stages of the game can be interpreted as costless signals. Thus, one can think of the informed and the uniformed players as the sender and the receiver in communication games, respectively, despite the players not having explicit access to costless signals. In this light, our focus on state-independent preferences for the informed player is akin to the assumption made in the previous chapters that the sender's incentives in communication games are transparent; e.g., a seller always wants a buyer to purchase a product.

Our main result is that, in the class of games that we study, any equilibrium payoffs

[1]This chapter is based on a joint work with Elliot Lipnowski and Doron Ravid with the same title.

for the informed player can be obtained as an equilibrium in which the informed player's actions depend on the state only in an initial finite number of stages. We therefore show that, to characterise the informed player's equilibrium payoffs, one can restrict attention to equilibria in which the informed player only uses an initial finite number of stages to communicate information about the state to the uninformed player with no additional information being conveyed for the rest of the game. We also give examples that demonstrate that the result does not extend to equilibrium payoff vectors (i.e., payoffs of both informed and uniformed players) or to the uninformed player's equilibrium payoffs.

Aumann and Maschler (1966) and Aumann, Maschler and Stearns (1968) were the first to study the two-player, undiscounted, infinitely repeated game with a lack of information on one side.[2] In particular, Aumann, Maschler and Stearns (1968) characterise payoff vectors of a particular class of equilibria, which we call *AMS equilibria,* whereby the informed player's actions depend on the state only in an initial finite number of stages. Although simple, AMS equilibria entail a loss of generality. Hart (1985) provides a full characterisation of all equilibrium payoff vectors using bimartingale processes—a type of martingale process studied in detail in Aumann and Hart (1986) that represents much more sophisticated communication and coordination between the players. Lipnowski and Ravid (2020), *inter alia*, show that when an informed sender has state-independent preferences, the sender does not benefit from extra rounds of pre-play communication. We also prove our results using by appealing to the same connection using results from Hart (1985), Aumann and Hart (1986), and Lipnowski and Ravid (2020).

[2]See the preface of Aumann and Maschler (1995) and the foreword in Mertens, Sorin and Zamir (2015) for historiographical accounts.

Related literature We contribute to the rich literature on repeated games with incomplete information that began with Aumann and Maschler (1966). Forges (1990) shows that tools developed by Aumann, Maschler and Stearns (1968) and Hart (1985) in the context of repeated games can also be applied to communication games. Forges (2020) provides a survey of the literature and explains the connections between repeated games and communication games.

While we focus on the case in which the informed player's preferences are state independent, Shalev (1994) considers the "known-own payoff" case in which the *uninformed* player's payoff is state independent.[3] He shows the payoff vectors attainable in fully revealing AMS equilibria characterise all possible equilibrium payoff vectors. Cripps and Thomas (2003) introduce discounting to the known-own payoff case and show that, when the informed player is arbitrarily patient relative to the uniformed player, the characterisation of the informed player's payoffs are essentially the same as in the undiscounted case. They also study the case with equal discount rates for the players. Their result, and subsequent generalisations by Pęski (2008; 2014), demonstrate that the set of equilibrium payoff vectors in the undiscounted case are typically a strict subset of the limit set of equilibrium payoff vectors in the discounted case with equal discount rates.

Mertens and Zamir (1971) study the case in which there is a lack of information on both sides. Results in this direction are covered in depth by Mertens, Sorin and Zamir (2015). Our model is a two-player repeated game with undiscounted utility, one-sided incomplete information, and observable actions, as studied by Aumann, Maschler and Stearns (1968) and Hart (1985), specialised to the case in which the informed player has state-independent preferences.

[3]A salient feature of the known-own payoff case is that the uninformed player always knows his own payoff, whereas in the general case, the uninformed player might never learn his own payoff.

3.1 Model

There are two players: one *informed* (player 1) and one *uninformed* (player 2). The game begins with a realisation of a payoff-relevant random state θ from a finite set Θ (with at least two elements) according to a full-support distribution $\mu_0 \in \Delta\Theta$.[4] Then, player 1 observes the realisation of θ, and the players subsequently play the stage game infinitely many times. In each period $t \in \mathbb{N}$, each player $j \in \{1,2\}$ chooses an action from a finite set A_j (with at least two elements) simultaneously, and the stage payoff is given by $u_1 : A \to \mathbb{R}$ and $u_2 : A \times \Theta \to \mathbb{R}$ for players 1 and 2, respectively, where $A := A_1 \times A_2$. At the end of each period, players observe the period's chosen action profile but not the resulting payoffs.

Let $\sigma_1 : \mathcal{H} \times \Theta \to \Delta A_1$ denote player 1's strategy and let $\sigma_2 : \mathcal{H} \to \Delta A_2$ denote player 2's strategy, where $\mathcal{H} := \bigcup_{t=0}^{\infty} A^t$ is the set of public histories. A strategy profile $\sigma = (\sigma_1, \sigma_2)$ and a belief $\mu \in \Delta\Theta$ induce a unique probability measure $\mathbb{P}_{\sigma,\mu}$ on $\Omega := A^\infty \times \Theta$; let $\mathbb{E}_{\sigma,\mu}$ denote the expectation operator with respect to $\mathbb{P}_{\sigma,\mu}$.[5] Define expectations of players 1 and 2's payoffs up to and including stage t, respectively, as

$$v_1^t(\sigma) := \mathbb{E}_{\sigma,\mu_0}\left[\frac{1}{t}\sum_{s=1}^{t} u_1(a_t)\right],\ v_2^t(\sigma) := \mathbb{E}_{\sigma,\mu_0}\left[\frac{1}{t}\sum_{s=1}^{t} u_2(a_t,\theta)\right]. \tag{3.1}$$

[4]We adopt the following notational conventions throughout the chapter. Given a finite set X, let ΔX denote the set of all probability measures over X. Given a real-valued function f on some convex space Z, let vex f denote the convexification of f (i.e., pointwise largest convex function on Z that does not exceed f). Given a correspondence $V : X \rightrightarrows Y$, let $\mathrm{gr}(V)$ denote its graph $\{(y,x) :\ x \in X,\ y \in V(x)\}$.

[5]For each $t \in \mathbb{N}$, let $\mathscr{H}^t$ be the finite algebra generated by the discrete algebra on A^t, and let $\mathscr{H}^\infty$ denote the product σ-algebra on A^∞. Then, $\mathbb{P}_{\sigma,\mu}$ is a probability measure on the measurable space $(\Omega, \mathscr{H}^\infty \otimes 2^\Theta)$, which is uniquely defined by the Kolmogorov extension theorem.

Following Aumann and Maschler (1968),[6] a strategy profile σ is an *equilibrium* if:[7]

$$\liminf_{t\to\infty} v_1^t(\sigma) \geq \limsup_{t\to\infty} \sup_{\sigma_1'} v_1^t(\sigma_1', \sigma_2) \text{ and } \liminf_{t\to\infty} v_2^t(\sigma) \geq \limsup_{t\to\infty} \sup_{\sigma_2'} v_2^t(\sigma_1, \sigma_2'). \tag{3.2}$$

The payoffs for players 1 and 2 associated with an equilibrium σ are their respective limit payoffs:[8]

$$v_1(\sigma) := \lim_{t\to\infty} v_1^t(\sigma) \text{ and } v_2(\sigma) := \lim_{t\to\infty} v_2^t(\sigma). \tag{3.3}$$

A vector $s = (s_1, s_2) \in \mathbb{R}^2$ is an *equilibrium payoff vector* if an equilibrium σ exists such that $s = (v_1(\sigma), v_2(\sigma))$. For $j \in \{1, 2\}$, $s_j \in \mathbb{R}$ is an *equilibrium Pj-payoff* if an equilibrium σ exists for which $s_j = v_j(\sigma)$.

3.2 Informed player's equilibrium payoff

Aumann, Maschler and Stearns (1968) characterise equilibrium payoff vectors that can be induced by a simple strategy profiles in which player 1's behaviour in the initial stages is used to communicate information about θ to player 2, and the players subsequently coordinate their actions with no further information being revealed. We call an equilibrium σ an *AMS equilibrium* if some $\ell \in \mathbb{N}$ exists such that (i) σ_1 does not condition on θ for any on-path history after stage ℓ; and (ii) players ignore player 2's behaviour in the first ℓ stages.[9] We refer to an equilibrium payoff vector associated with

[6]Robert J Aumann, Michael B Maschler and Richard E Stearns (1968) and Hart (1985) refer to σ that satisfies (3.2) as a uniform equilibrium. The associated payoffs $(v_1(\sigma), v_2(\sigma))$ of a (uniform) equilibrium σ (see (3.3)) can be achieved in an ϵ-equilibrium of some finitely repeated game; that is, for any $\epsilon > 0$, there exists $T_\epsilon \in \mathbb{N}$ such that, for all $t > T_\epsilon$, $v_1^t(\sigma', \sigma_2) \leq v_1(\sigma) + \epsilon$ for all σ_1' and $v_2^t(\sigma_1, \sigma_2') \leq v_2(\sigma) + \epsilon$ for all σ_2'. Hart (1985) defines an equilibrium in which T_ϵ may depend on σ and shows the set of payoffs of such equilibria and the set of payoffs of the equilibrium defined above coincide (Proposition 2.1.4).

[7]Unlike under more general payoff environments (e.g., Simon, Spież and Toruńczyk, 1995; Shalev, 1994), equilibrium existence is immediate; e.g., players could employ, i.i.d. across histories, a mixed-strategy equilibrium from the complete-information static game with payoffs u_1 and $\int_\Theta u_2(\cdot, \theta)\mathrm{d}\mu_0(\theta)$.

[8]The limits exist given (3.2).

[9]Formally, an equilibrium σ is an AMS equilibrium if, for any $t \in \mathbb{Z}_+$ with $t \geq \ell$, any pair of public

an AMS equilibrium as an *AMS-equilibrium payoff vector*. Finally, for $j \in \{1,2\}$, $s_j \in \mathbb{R}$ is an *AMS-equilibrium Pj-payoff* if some AMS equilibrium σ exists such that $s_j = v_j(\sigma)$.

Aumann, Maschler and Stearns (1968) provide examples of equilibrium payoff vectors that cannot be AMS-equilibrium payoff vectors. Hart (1985) subsequently provides a characterisation of all equilibrium payoffs via strategy profiles that allow players to engage in more sophisticated communication and coordination than in Aumann, Maschler and Stearns (1968). Our main result is that, when player 1's preferences are state independent, the additional sophistication allowed under Hart (1985) is unnecessary from player 1's perspective. In particular, simple communication-coordination strategy profiles that induce AMS equilibria are sufficient. We formally state the result below.

Proposition 3.1. *Any equilibrium P*1*-payoff is an AMS-equilibrium P*1*-payoff.*

The proof applies Hart's (1985) characterisation of equilibrium payoff vectors via bimartingales, showing player 1's payoff induced by such a martingale satisfies the conditions for an AMS equilibrium when player 1's payoff is state independent.

We first introduce a correspondence yielding the set of payoffs that are feasible and individually rational given any belief $\mu \in \Delta\Theta$ as defined by Hart (1985). Letting $\overline{u}$ be a bound on the players' possible payoff magnitudes and $\mathcal{R} := [-\overline{u}, \overline{u}] \subseteq \mathbb{R}$, define

$$F^* : \Delta\Theta \rightrightarrows \mathcal{R}^2 \tag{3.4}$$

$$\mu \mapsto \left\{(s_1, s_2) \in F(\mu) : \ s_1 \geq \underline{u}_1,\ s_2 \geq \operatorname{vex} \underline{u}_2(\mu)\right\},$$

histories $h, h' \in A^t$, and any pair of states $\theta, \theta' \in \Theta$, we have (i) $\sigma_1(h,\theta) = \sigma_1(h,\theta')$, and (ii) if h and h' differ only in the first ℓ periods of player 2's play, then $\sigma_1(h,\theta) = \sigma_1(h',\theta)$ and $\sigma_2(h) = \sigma_2(h')$.

where

$$F(\mu) : \Delta\Theta \rightrightarrows \mathcal{R}^2$$
$$\mu \mapsto \operatorname{co}\left\{ \left(u_1\left(a\right), \int_\Theta u_2\left(a, \cdot\right) \mathrm{d}\mu \right) : \, a \in A \right\}$$

is the correspondence that gives the set of feasible expected payoffs in the one-stage game from using a correlated state-independent strategy given a prior belief $\mu \in \Delta\Theta$; $\underline{u}_1 \in \mathbb{R}$ and $\underline{u}_2 : \Delta\Theta \to \mathbb{R}$ are the minmax values for players 1 and 2, respectively, in the one-stage game in which neither player observes the realisation of θ with a common prior belief $\mu \in \Delta\Theta$; that is

$$\underline{u}_1 := \min_{\alpha_2 \in \Delta A_2} \max_{\alpha_1 \in \Delta A_1} \int_A u_1 \mathrm{d}\left(\alpha_1 \otimes \alpha_2\right),$$
$$\underline{u}_2\left(\mu\right) := \min_{\alpha_1 \in \Delta A_1} \max_{\alpha_2 \in \Delta A_2} \int_{A \times \Theta} u_2 \mathrm{d}\left(\alpha_1 \otimes \alpha_2 \otimes \mu\right).$$

Define $F_1^* : \Delta\Theta \rightrightarrows \mathcal{R}$ as the projection of F^* to player 1's payoffs; that is

$$F_1^*\left(\mu\right) := \left\{ s_1 \in \mathcal{R} : \exists s_2 \in \mathcal{R} \text{ with } \left(s_1, s_2\right) \in F^*\left(\mu\right) \right\}.$$

We first observe that F_1^* is a Kakutani correspondence (i.e., nonempty-valued, compact-valued, convex-valued, and upper hemicontinuous).

Lemma 3.1. *The correspondence F_1^* is a Kakutani correspondence.*

Proof. Let $\hat{F} : \Delta\Theta \rightrightarrows \mathcal{R}^2$ denote the correspondence $\mu \mapsto \{(u_1(a), \int_\Theta u_2(a, \cdot)\mathrm{d}\mu) : a \in A\}$. Observe that $\hat{F}$ is closed- and bounded-valued. Thus, $\hat{F}$ is compact-valued, and because its graph is closed (being a union of the graphs of finitely many continuous functions), $\hat{F}$ is upper hemicontinuous. As a convex hull of the real-valued correspondence $\hat{F}$, F is convex- and compact-valued and upper hemicontinuous. Define $\tilde{F} : \Delta\Theta \rightrightarrows \mathcal{R}^2$ as the individual rationality correspondence; that is, $\mu \mapsto \{(s_1, s_2) \in \mathcal{R}^2 : s_1 \geq \underline{u}_1, s_2 \geq \operatorname{vex} \underline{u}_2(\mu)\}$. Observe that $\tilde{F}$ is convex-, closed-, and bounded-valued (taking values in a bounded set $\mathcal{R}^2$). Moreover, $\tilde{F}$ is upper hemicontinuous because $\operatorname{vex} \underline{u}_2$ is lower semicontinuous. Because $\hat{F}$ is convex-valued with a closed graph, whereas $\tilde{F}$ is convex-valued with a compact graph, their intersection is convex-valued with a compact graph. Hence, F^* is convex-valued, compact-valued, and upper hemicontinuous. For any $\mu \in \Delta\Theta$, $F^*(\mu)$ contains the payoff vector associated with player 1 playing the minmax mixed strategy for u_1 and player 2 playing a minmax mixed strategy for $\int_\Theta u_2(\cdot, \theta) \mathrm{d}\mu(\theta)$, and so F^* is nonempty-valued. Therefore, F^* is a Kakutani correspondence. Because F_1^* is a projection of F^* to the first coordinate, which is a continuous transformation, F_1^* is a Kakutani correspondence. ■

The remainder of the proof centres around the following lemma that follows immediately from the Main Theorem in Hart (1985). To state it, define a *bimartingale* as a $\mathcal{R} \times \Delta\Theta$-valued martingale, $\{(s^t, \mu^t)\}_{t=1}^{\infty}$, on some filtered probability space such that, for all $t \in \mathbb{N}$, either $s^{t+1} = s^t$ almost surely or $\mu^{t+1} = \mu^t$ almost surely. We say the bimartingale has *initial value* (s, μ) if $(s^1, \mu^1) = (s, \mu)$ almost surely. Given a measurable subset Z of $\mathcal{R} \times \Delta\Theta$, we say the bimartingale has *terminal values in* Z if the almost-sure limit of the martingale is contained in Z almost surely.

Lemma 3.2. *If s_1 is an equilibrium $P1$-payoff, then some bimartingale exists with initial value (s_I, μ_0) and terminal values in* $\operatorname{gr}(F_1^*)$.

It remains to show the bimartingale given by Lemma 3.2 gives an AMS-equilibrium $P1$-payoff. The lemma below, which follows from a definition of AMS equilibrium as pointed out by Hart (1985),[10] gives a sufficient condition for a payoff $s_1 \in \mathcal{R}$ to be an AMS-equilibrium $P1$-payoff. Let $\mathcal{I}(\mu_0) = \{\tau \in \Delta\Delta\Theta : \int_{\Delta\Theta} \mu \mathrm{d}\tau(\mu) = \mu_0\}$ denote the set

[10]See the second paragraph of section 6 in Hart (1985).

of all Bayes plausible distribution over posterior beliefs.

Lemma 3.3. *Let $s_1 \in \mathcal{R}$. Suppose some $\tau \in \mathcal{I}(\mu_0)$ with finite support exists such that*

$$\tau\left(\{\mu \in \Delta\Theta : s_1 \in F_1^*\left(\mu\right)\}\right) = 1. \tag{3.5}$$

Then, s_1 is an AMS-equilibrium $P1$-payoff.

Proof. Let $\tau \in \mathcal{I}(\mu_0)$ satisfy the premise of the lemma and let $\{\mu(m)\}_{m=1}^{|\text{supp}(\tau)|} = \text{supp}(\tau)$. For each m, there exists $s_2(m) \in \mathcal{R}$ such that $(s_1, s_2(m), \mu(m)) \in \text{gr}(F^*)$. Since $\tau \in \mathcal{I}(\mu_0)$, $\sum_m p(\mu(m))\mu(m) = \mu_0$ and $\sum_m p(\mu(m))s_1 = s_1$. By the result from the second paragraph of of section 6 in Hart (1985), $(s_1, \sum_m p(\mu(m))s_2(m))$ is an AMS-equilibrium payoff vector; i.e., s_1 is an AMS-equilibrium $P1$-payoff. ■

The following lemma, essentially proved as part of Proposition 4 in Lipnowski and Ravid (2020), serves as a link between the previous two lemmata.

Lemma 3.4. *Suppose $V : \Delta\Theta \rightrightarrows \mathbb{R}$ is a Kakutani correspondence. If some bimartingale exists with initial value (s_1, μ_0) and terminal values in $\text{gr}(V)$, then some $\tau \in \mathcal{I}(\mu_0)$ exists such that*

$$\tau\left(\{\mu \in \Delta\Theta : s_1 \in V\left(\mu\right)\}\right) = 1. \tag{3.6}$$

Proof. We prove the contrapositive statement. Consider $s_1 \in \mathcal{R}$ such that no $\tau \in \mathcal{I}(\mu_0)$ exists for which (3.6) holds. In the proof of Proposition 4, Lipnowski and Ravid (2020) construct a continuous and biconvex function $B : \mathcal{R} \times \Delta\Theta \to \mathbb{R}$ that separates (s_1, μ_0) from $\text{gr}(V)$; i.e., $B|_{\text{gr}(V)} = 0$ and $B(s_1, \mu_0) > 0$. By Theorem 4.7 in Aumann and Hart (1986), it follows that there does not exist a bimartingale process with initial value (s_1, μ_0) and terminal values in $\text{gr}(V)$. ■

We are now ready to prove our main result.

Proof of Proposition 3.1. Let $s_1 \in \mathcal{R}$ be an equilibrium $P1$-payoff. Lemma 3.2 delivers a bimartingale with initial value (s_1, μ_0) and terminal values in $\text{gr}(F_1^*)$. By Lemma 3.4 and 3.1, some $\tau \in \mathcal{I}(\mu_0)$ exists that satisfies (3.5). Because Θ is finite, Carathéodory's theorem implies τ may be chosen to ensure $\text{supp}(\tau)$ is finite. Hence, s_1 is an AMS-equilibrium $P1$-payoff by Lemma 3.3. ■

3.3 Discussions

Proposition 3.1 characterises the informed player's equilibrium payoffs, and does not extend to characterising equilibrium payoff vectors. To demonstrate this fact, we adapting two examples from the cheap-talk literature.[11] The first example, adapted from Aumann and Hart (2003), demonstrates some equilibrium payoff vector exists that is not an AMS-equilibrium payoff vector. Nevertheless, the set of equilibrium Pj-payoffs in this example coincides with the set of AMS-equilibrium Pj-payoffs for each player j. In Example 2, adapted from Lipnowski and Ravid (2020), we show an equilibrium $P2$-payoff exists that is not attainable in an AMS equilibrium. Taken together, the examples demonstrate that whereas assuming player 1's preferences are state independent simplifies the characterisation of equilibrium $P1$-payoffs, the same simplification does not apply to the set of attainable equilibrium payoff vectors or the set of equilibrium $P2$-payoffs.

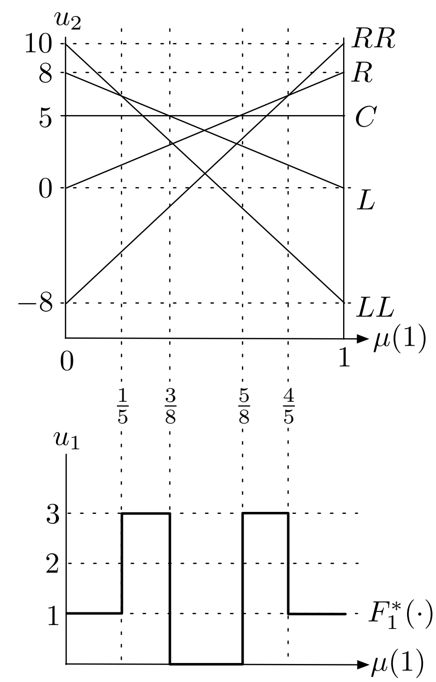
Example 3.1 (Example 2.6 in Aumann and Hart, 2003)**.** Two possible states, $\Theta := \{0, 1\}$, are equally likely, and player 2 has five possible actions, $A_2 := \{LL, L, C, R, RR\}$. The figure below shows the payoffs associated with each action under each state, player 2's

[11] In the cheap-talk version of the examples, the informed player has no "action" to take but can send a cheap-talk message to the uninformed player, whose action is payoff relevant. Because our environment has no explicit communication technology, we adapt the original examples by allowing the informed player to have at least two payoff-irrelevant actions.

best response as a function of his belief that the state is B, $\mu(1)$, and player 1's value correspondence, F_1^*, as a function of $\mu(1)$.

Figure 3.1: Example 2.6 from Aumann and Hart (2003).

$u_1(\cdot\|\theta), u_2(\cdot\|\theta)$	LL	L	C	R	RR
$\theta = 0$	$1, 10$	$3, 8$	$0, 5$	$3, 0$	$1, -8$
$\theta = 1$	$1, -8$	$3, 0$	$0, 5$	$3, 8$	$1, 10$



As explained in Aumann and Hart (2003), $(2, 8)$ is an equilibrium payoff vector of this game. One can achieve this payoff, for example, by first performing a jointly controlled lottery with equal probabilities,[12] and depending on the outcome of the jointly controlled lottery, player 1 communicates the state either fully yielding a payoff of $(1, 10)$,[13] or partially so that player 2's posterior belief, $\mu(1)$, is either $\frac{1}{4}$ or $\frac{3}{4}$ yielding a payoff of $(3, 6)$.[14]

We now argue $(2, 8)$ cannot be an AMS-equilibrium payoff vector. First, observe that

[12]For example, player 1 chooses first-stage action uniformly and player 2 chooses first-stage action uniformly among $\{LL, RR\}$. Then, player 1 fully reveals if and only if $a_1^1 = 1$ or $a_2^1 = LL$, and partially reveals (in the manner specified below) if $a_1^1 = 0$ and $a_2^1 = RR$.

[13]For example, player 1 chooses $a_1^2 = \theta$ if and only if the state is $\theta \in \Theta$.

[14]For example, player 1 chooses $a_1^2 = 1$ with probability $\frac{1}{4}$ if the state is 1 and with probability $\frac{3}{4}$ if the state is 0.

jointly controlled lotteries cannot be part of any AMS-equilibrium strategy, because players do not ignore player 2's action in responding to the lottery. To induce a payoff of 2 for player 1 in an AMS equilibrium, the posterior belief for player 2 must therefore always be one of $\frac{1}{5}$, $\frac{3}{8}$, $\frac{5}{8}$, or $\frac{4}{5}$. However, with such beliefs, player 2's expected payoffs are strictly below 8. Hence, it follows that $(2,8)$ cannot be an AMS-equilibrium payoff vector.

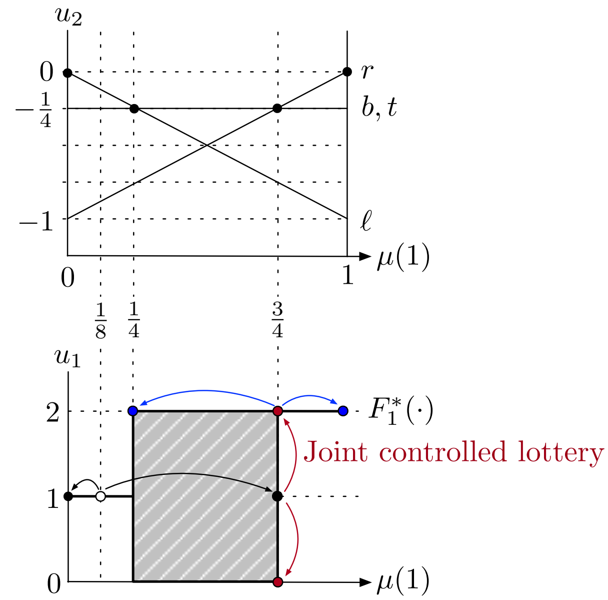
However, let us observe that the set of equilibrium $P2$-payoffs and the set of AMS-equilibrium $P2$-payoffs coincide. First, observe that in any equilibrium, player 2's payoff must lie in $[5,10]$. Lemma 3.3 together with the fact that player 1's value correspondence is symmetric means any $s_1 \in [0,3]$ is an AMS-equilibrium $P1$-payoff, and any such s_1 can be achieved by inducing a symmetric posterior belief around $\frac{1}{2}$. Because such distribution over posterior beliefs can induce any expected payoff for player 2 in $[5,10]$, it follows that any equilibrium $P2$-payoff is an AMS-equilibrium $P2$-payoff.

We now demonstrate that the last observation, namely, that the set of equilibrium $P2$-payoffs and the set of AMS-equilibrium $P2$-payoffs coincide, does not hold generally.

Example 3.2 (Appendix C.3 in Lipnowski and Ravid, 2020)**.** Two states, $\Theta := \{0,1\}$, are possible, and the prior belief is that the state is 1 with probability $\frac{1}{8}$. Player 2 has four possible actions, $A_2 := \{\ell, b, t, r\}$. The figure below shows the payoffs associated with each action under each state, player 2's best response as a function of his belief that the state is 1, $\mu(1)$, and player 1's value correspondence, F_1^*, as a function of $\mu(1)$.

Figure 3.2: Example from Appendix C.3 in Lipnowski and Ravid (2020).

$u_1(\cdot\vert\theta), u_2(\cdot\vert\theta)$	ℓ	b	t	r
0	$1, 0$	$0, -\frac{1}{4}$	$2, -\frac{1}{4}$	$2, -1$
1	$1, -1$	$0, -\frac{1}{4}$	$2, -\frac{1}{4}$	$2, 0$



By Lemma 3.3, because $F_1^*(\mu) = \{1\}$ for any $\mu(1) \leq \frac{1}{8}$, player 1's payoff must be 1 in any AMS equilibrium. Moreover, in any AMS equilibrium, player 2's maximum payoff is $-\frac{1}{24}$, corresponding to a distribution over beliefs $\mu(1) = 0$ with probability $\frac{5}{6}$ and $\mu(1) = \frac{3}{4}$ with probability $\frac{1}{6}$.[15] However, as explained in Lipnowski and Ravid (2020), players may perform a jointly controlled lottery (with equal probabilities) following realisation of posterior belief $\frac{3}{4}$ (red arrows), and player 1 could further communicate (upper set of blue arrows) so that player 2's payoff will be supported by the solid dots as shown in the figure, which must yield a strictly higher payoff than $-\frac{1}{24}$. Such splits (called diconvexifications) are allowed under Hart (1985)'s characterisation, and it follows that the resulting player-2 payoff is an equilibrium $P2$-payoff. Thus, some equilibrium $P2$-payoff exists that is not an AMS-equilibrium $P2$-payoff.

In the case in which, instead, the *uninformed* player (player 2) has state-independent preferences, Shalev (1994) shows every equilibrium payoff vector is a fully revealing

[15]For example, player 1 chooses $a_1^1 = 1$ with probability 1 if the state is 1 and with probability $\frac{1}{21}$ if the state is 0.

AMS-equilibrium payoff vector. In the case in which the informed player (player 1) has state-independent preferences, however, fully revealing AMS-equilibrium $P1$-payoffs do not characterise equilibrium $P1$-payoffs. For instance, in Example 3.1, attainable payoffs for player 1 (e.g., 5) exist that are unattainable with a fully revealing AMS equilibrium; and Example 3.2 has the feature that no fully revealing AMS equilibrium exists; hence, none of the nonempty set of attainable payoffs for player 1 can be attained in a fully revealing AMS equilibrium.

3.4 Conclusion

In this chapter, we study a two-player, undiscounted, infinitely repeated game in which only one player is informed about the state of the world, and players observe only each other's actions in each stage of the repeated game. Aumann, Maschler and Stearns (1968) characterised payoff vectors that can be attained in equilibrium with strategies that use only the initial finite stages for communication. Hart (1985) provides a full characterisation of equilibrium payoff vectors in such games including those that require infinite stages of communication. We specialise the model by assuming that the informed player's payoff is state independent and show that finite stages of communication are sufficient to characterise the informed player's equilibrium payoff. In other words, the class of equilibrium characterised by Aumann, Maschler and Stearns (1968) is sufficient to characterise the informed player's equilibrium payoff when the sender's motives are transparent. We also provide two examples demonstrating that finite stages of communication are not sufficient to capture the entire set of equilibrium payoff vectors or or the uninformed player's equilibrium payoffs even with state-independent preferences for the informed player.

Bibliography

Arieli, Itai, and Yakov Babichenko. 2019. "Private Bayesian persuasion." *Journal of Economic Theory*, 182: 185–217.

Aumann, Robert J, and Michael B Maschler. 1966. "Game theoretic aspects of gradual disarmament." *Report of the US Arms Control and Disarmament Agency*, 80: 1–55.

Aumann, Robert J, and Michael B Maschler. 1968. "Repeated games of incomplete information: The zero-sum extensive case." *Mathematica, Inc*, 37–116.

Aumann, Robert J, and Michael B Maschler. 1995. *Repeated Games with Incomplete Information.* MIT Press.

Aumann, Robert J, and Sergiu Hart. 1986. "Bi-convexity and bi-martingales." *Israel Journal of Mathematics*, 54(2): 159–180.

Aumann, Robert J, and Sergiu Hart. 2003. "Long Cheap Talk." *Econometrica*, 71(6): 1619–1660.

Aumann, Robert J, Michael B Maschler, and Richard E Stearns. 1968. "Repeated games of incomplete information: An approach to the non-zero-sum case." *Report of the US Arms Control and Disarmament Agency ST-143*, 117–216.

Bachracht, Yoram, Reshef Meir, Kyomin Jung, and Pushmeet KohIit. 2010. "Coalitional structure generation in Skill Games." *Proceedings of the National Conference on Artificial Intelligence*, 2: 703–708.

Balbuzanov, Ivan. 2019. "Lies and consequences." *International Journal of Game Theory*, 48: 1203–1240.

Battaglini, Marco. 2002. "Multiple Referrals and Multidimensional Cheap Talk." *Econometrica*, 70: 1379–1401.

Bergemann, Dirk, and Stephen Morris. 2019. "Information Design: A Unified Perspective." *Journal of Economic Literature*, 57(1): 44–95.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *The Annals of Mathematical Statistics*, 24: 265–272.

Border, Kim C, and Joel Sobel. 1987. "Samurai Accountant: A Theory of Auditing and Plunder." *The Review of Economic Studies*, 54: 525.

Chakraborty, Archishman, and Rick Harbaugh. 2010. "Persuasion by Cheap Talk." *American Economic Review*, 100(5): 2361–2382.

Che, Yeon Koo, and Navin Kartik. 2009. "Opinions as incentives." *Journal of Political Economy*, 117: 815–860.

Crawford, Vincent P, and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451.

Cripps, Martin W, and Jonathan P Thomas. 2003. "Some Asymptotic Results in Discounted Repeated Games of One-Sided Incomplete Information." *Mathematics of Operations Research*, 28(3): 433–462.

Dewatripont, Mathias, and Jean Tirole. 1999. "Advocates." *Journal of Political Economy*, 107: 1–39.

Dworczak, Piotr, and Alessandro Pavan. 2022. "Preparing for the Worst but Hoping for the Best: Robust (Bayesian) Persuasion." *Econometrica*, 90: 2017–2051.

Dziuda, Wioletta, and Christian Salas. 2018. "Communication with Detectable Deceit." *SSRN Electronic Journal*.

Ederer, Florian, and Weicheng Min. 2022. "Bayesian Persuasion with Lie Detection." National Bureau of Economic Research Working Paper 30065.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap Talk with Two Audiences." *American Economic Review*, 79(5): 1214–1223.

Fershtman, Chaim, and Kenneth L Judd. 1987. "Equilibrium Incentives in Oligopoly." *American Economic Review*, 77: 927–940.

Fisher, Stanley Z. 1988. "In Search of the Virtuous Prosecutor: A Conceptual Framework." *American Journal of Criminal Law*, 197.

Forges, Françoise. 1990. "Equilibria with Communication in a Job Market Example." *The Quarterly Journal of Economics*, 105(2): 375–398.

Forges, Françoise. 2020. "Games with Incomplete Information: From Repetition to Cheap Talk and Persuasion." *Annals of Economics and Statistics*, 137: 3–30.

Frankfurter, Felix. 1963. "John Henry Wigmore: A Centennial Tribute." *Northwestern University Law Review*, 58(4).

Frechette, Guillaume R, Alessandro Lizzeri, and Jacopo Perego. 2019. "Rules and Commitment in Communication: An Experimental Analysis." *SSRN Electronic Journal*.

Friedman, Richard D. 2009. *The Yale Biographical Dictionary of American Law.* Yale University Press.

Gentzkow, Matthew, and Emir Kamenica. 2017*a*. "Bayesian persuasion with multiple senders and rich signal spaces." *Games and Economic Behavior*, 104: 411–429.

Gentzkow, Matthew, and Emir Kamenica. 2017*b*. "Competition in persuasion." *Review of Economic Studies*, 84: 300–322.

Gerardi, Dino, and Leeat Yariv. 2008. "Costly expertise." *American Economic Review*, 98: 187–193.

Gershman, Bennett L. 2001. "The Prosecutor's Duty to Truth." *Gerogetown Journal of Legal Ethics*, 14: 309–354.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information avoidance." *Journal of Economic Literature*, 55: 96–135.

Goltsman, Maria, and Gregory Pavlov. 2011. "How to talk to multiple audiences." *Games and Economic Behavior*, 72(1): 100–122.

Hart, Sergiu. 1985. "Nonzero-Sum Two-Person Repeated Games with Incomplete Information." *Mathematics of Operations Research*, 10(1): 117–153.

Ivanov, Maxim. 2010. "Informational control and organizational design." *Journal of Economic Theory*, 145: 721–751.

Ivanov, Maxim, and Alex Sam. 2022. "Cheap talk with private signal structures." *Games and Economic Behavior*, 132: 288–304.

Kamenica, Emir. 2019. "Bayesian Persuasion and Information Design." *Annual Review of Economics*, 11: 249–272.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review*, 101(6): 2590–2615.

Krähmer, Daniel. 2021. "Information Design and Strategic Communication." *American Economic Review: Insights*, 3: 51–66.

Levkun, Aleksandr. 2022. "Communication with Strategic Fact-Checking." *Working Paper*.

Lichtig, Avi. 2020. "Adversarial Disclosure." *Working Paper*.

Lipnowski, Elliot, and Doron Ravid. 2020. "Cheap Talk With Transparent Motives." *Econometrica*, 88(4): 1631–1660.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2022. "Persuasion via Weak Institutions." *Journal of Political Economy*, 130(10): 2705–2730.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva. 2020. "On information design in games." *Journal of Political Economy*, 128(4): 1370–1404.

McAdams, David. 2012. "Strategic ignorance in a second-price auction." *Economics Letters*, 114: 83–85.

Mertens, Jean-Francois, and Shmuel Zamir. 1971. "The value of two-person zero-sum repeated games with lack of information on both sides." *International Journal of Game Theory*, 1: 39–64.

Mertens, Jean-François, Sylvain Sorin, and Shmuel Zamir. 2015. *Repeated Games. Econometric Society Monographs*, Cambridge University Press.

Min, Daehong. 2021. "Bayesian persuasion under partial commitment." *Economic Theory*, 72: 743–764.

Mookherjee, Dilip, and Ivan Png. 1989. "Optimal Auditing, Insurance, and Redistribution." *The Quarterly Journal of Economics*, 104: 399.

Ohta, Naoki, Atsushi Iwasaki, Makoto Yokoo, Kohki Maruono, Vincent Conitzer, and Tuomas Sandholm. 2006. "A compact representation scheme for coalitional games in open anonymous environments." *Proceedings of the National Conference on Artificial Intelligence*, 1(1994): 697–702.

Onuchic, Paula. 2022. "Advisors with Hidden Motives." *Working Paper*.

Özdogan, Ayça. 2016. "A Survey of strategic communication and persuasion." *Bogazici Journal*, 30.

Pęski, Marcin. 2008. "Repeated games with incomplete information on one side." *Theoretical Economics*, 3(1): 29–84.

Pęski, Marcin. 2014. "Repeated games with incomplete information and discounting." *Theoretical Economics*, 9(3): 651–694.

Posner, Richard A. 1999. "An Economic Approach to the Law of Evidence." *Law Review*, 51: 1477–1546.

PricewaterhouseCoopers. 2022. "Daubert challenges to financial expert."

Rahwan, Talal, Tomasz P Michalak, Michael Wooldridge, and Nicholas R Jennings. 2015. "Coalition structure generation: A survey." *Artificial Intelligence*, 229: 139–174.

Roesler, Anne-Katrin, and Balázs Szentes. 2017. "Buyer-Optimal Learning and Monopoly Pricing." *American Economic Review*, 107: 2072–2080.

Rogoff, Kenneth. 1985. “The Optimal Degree of Commitment to an Intermediate Monetary Target.” *The Quarterly Journal of Economics*, 100: 1169–1189.

Sadakane, Hitoshi, and Yin Chi Tam. 2022. “Cheap talk and Lie detection.” *Working Paper*.

Salcedo, Bruno. 2019. “Persuading part of an audience.”

Sandholm, Tuomas. 1999. “An algorithm for optimal winner determination in combinatorial auctions.” *IJCAI International Joint Conference on Artificial Intelligence*, 1: 542–547.

Sandholm, Tuomas, Kate Larson, Martin Andersson, Onn Shehory, and Fernando Tohmé. 1999. “Coalition structure generation with worst case guarantees.” *Artificial Intelligence*, 111(1-2): 209–238.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* Harvard University Press.

Schnakenberg, Keith E. 2015. “Expert advice to a voting body.” *Journal of Economic Theory*, 160: 102–113.

Schnakenberg, Keith E. 2017. “Informational Lobbying and Legislative Voting.” *American Journal of Political Science*, 61(1): 129–145.

Shalev, Jonathan. 1994. “Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Known-Own Payoffs.” *Games and Economic Behavior*, 7(2): 246–259.

Shin, Hyun Song. 1998. “Adversarial and Inquisitorial Procedures in Arbitration.” *The RAND Journal of Economics*, 29: 378.

Simon, Robert S., Stanisław Spież, and Henryk Toruńczyk. 1995. “The existence of equilibria in certain games, separation for families of convex functions and a theorem of Borsuk-Ulam type.” *Israel Journal of Mathematics*, 92(1–3): 1–21.

Sklivas, Steven D. 1987. “The Strategic Choice of Managerial Incentives.” *The RAND Journal of Economics*, 18: 452.

Sobel, Joel. 2013. “Giving and Receiving Advice.” *Advances in Economics and Econometrics: Tenth World Congress*, , ed. Daron Acemoglu, Manuel Arellano and Eddie Dekel Vol. 1, Chapter 10, 305–341. Cambridge University Press.

Spence, Michael. 1973. “Job Market Signaling.” *The Quarterly Journal of Economics*, 87: 355.

Taylor, Curtis R, and Huseyin Yildirim. 2011. “Subjective performance and the value of blind evaluation.”

Timmerbeil, Sven. 2003. “The Role of Expert Witnesses in German and U.S. Civil Litigation.” *Annual Survey of International and Comparative Law*, 9(1): 163–187.

Townsend, Robert M. 1979. "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory*, 21: 265–293.

Vickers, John. 1985. "Delegation and the Theory of the Firm." *The Economic Journal*, 95: 138.

Ye, Minlei. 2021. "Theory of Auditing Economics: A Review of Analytical Auditing Research." *SSRN Electronic Journal*.