THE BENEFITS OF MISCOMMUNICATION

ARMIN RICK*

November 2013

ABSTRACT:

This paper investigates the effect of miscommunication on equilibrium information transmission, communication costs, and welfare. In particular, I analyze a persuasion mechanism between an informed sender and an uninformed receiver that nests a variety of prominent communication technologies such as cheap talk, verifiable messages, and costly signaling. Miscommunication can often increase welfare by improving the tradeoff between information transmission and communication costs. Such gains result from two sources, (i) an expansion of the set of communication strategy profiles that can induce a given distribution of posterior beliefs or (ii) an expansion of the set of distributions of beliefs that the sender may expect to induce by unused messages in equilibrium. I illustrate gains from miscommunication in a competitive version of the Spence signaling game. In particular, I characterize optimal miscommunication and derive conditions for which any equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication. This requires perfect information transmission to be sufficiently costly (or not implementable) on the one hand and partial information transmission to be sufficiently valuable on the other. If the latter is true, optimal miscommunication can even transform a signaling technology that makes all types of the sender worse off into one that makes everyone better off.

*Department of Economics, The University of Chicago, e-mail: arick@uchicago.edu.

I would like to thank Gary Becker, Alexander Frankel, Emir Kamenica, Roger Myerson, Derek Neal, Eric Nielsen, Phil Reny, Richard Van Weeden and the participants of various workshops for helpful comments and suggestions.
1. Introduction

Communication among economic agents is ubiquitous and communication efforts command a sizable share of economic resources. For instance, McCloskey and Klamer (1995) estimate this share to be as high as one quarter of GDP. Communication efforts arise in many situations of economic interest because a privately informed agent is affected by the decisions of a less informed decision maker and thus has an incentive to affect the informational environment of the latter. For instance, a buyer decides how much of a certain good to buy and the seller has private information about the good’s characteristics; a firm decides what job to assign and what wage to offer to a job applicant who has private information about his skill; or a politician considers regulations for a certain industry, and firms in this industry have better information about market characteristics. In these situations, communication efforts can often help to transmit valuable information and thus improve the quality of decision making in equilibrium. However, information transmission is often limited by the extent of preference alignment and/or communication efforts involve economic costs that reduce and may even dominate social gains from improved information transmission. The extent to which the performance of a communication technology may benefit from miscommunication or, equivalently, communication error, is the subject of this paper.

Most papers that study communication games between a privately informed sender and a decision-making receiver (see, e.g., Spence (1973, 1974); Grossman (1981); Milgrom (1981); Crawford and Sobel (1982); Kartik (2009)) assume that there is no miscommunication, i.e., the receiver can perfectly determine which message has been sent by the sender. There are a few relatively recent papers discussed in the next section that deviate from this assumption, but their analysis is restricted to very specific communication technologies and preference structures. This paper, in contrast, investigates the effect of miscommunication on equilibrium outcomes in a class of communication games referred to as persuasion mechanisms (Kamenica and Gentzkow (2011)) that nests a variety of ‘classic’ communication technologies considered in the literature such as cheap talk (Crawford and Sobel (1982)), verifiable messages (Grossman (1981); Milgrom (1981)), and costly signaling (Spence (1973, 1974)). In particular, a sender is assumed to have private information about his type and sends a (potentially costly) message to a receiver. The receiver observes a signal that is a potentially garbled version of this message and subsequently makes a decision that affects the welfare of both agents. Miscommunication occurs if the signal observed by the receiver is a garbled
version of the message that carries less information content. Beliefs are formed assuming that all agents are Bayesians and understand how the garbling mechanism works.

There are a variety of reasons why we should be interested in the effect of miscommunication. First, many communication technologies in the real world are noisy rather than error-free: For example, performance tests are noisy signals of a test taker’s skills and effort, and politicians communicate with voters through the media that restate and garble their messages. Given the amount of resources involved in communication efforts and the importance of information transmission for efficient decision making, it is thus important to ask if we should be concerned about the extent of miscommunication in a given situation or if miscommunication may in fact be beneficial. Second, if miscommunication can be beneficial, this raises a number of important questions for policy makers: Do some agents have an incentive to overinvest in the precision of a communication technology? Could a social planner improve equilibrium outcomes simply by limiting information flows, i.e., by introducing communication error? If so, what are optimal ways to limit information? And finally, how does the answer to these questions depend on preference alignment and the specifics of the communication technology?

In this paper, I argue that miscommunication can often help. In particular, it can improve equilibrium information transmission without raising communication costs and/or it can reduce communication costs without changing the quality of information transmission. In fact, optimal miscommunication may even transform a communication technology that makes all types of senders worse off compared to a situation without communication into one that makes all types of senders strictly better off—without hurting the receiver.

Such gains from miscommunication can arise from two sources. First, without miscommunication, a given strategy profile may fail to be an equilibrium strategy profile because the communication technology imposes restrictions on beliefs for unused messages that imply the existence of profitable deviations for some type(s) of the sender. By mixing unused messages with other messages, miscommunication can remove such profitable deviations. I refer to this source as gains from communication restrictions since it stems from removing access to certain combinations of communication costs and induced posteriors that could also be achieved by restricting the message space. Second, without miscommunication, any mixing of information as perceived by the receiver must be exactly mirrored by the sender’s strategy profile, i.e., the strategy profile must be as informative as the induced distribution of posteriors, in a sense I will make clear below. With appropriate miscommunication,
in contrast, a given distribution of posteriors can be induced by any strategy profile that is weakly more informative. As a result, miscommunication can expand the set of strategy profiles that implements a given distribution of posteriors in equilibrium. I thus refer to this source as gains from strategic expansion.

Each of these two sources of gains from miscommunication can materialize in two distinct ways: Either gains from miscommunication are associated with the implementation of a distribution of posteriors that is not implementable without miscommunication, or they are associated with a reduction in communication costs for a distribution of posteriors that is implementable without miscommunication. I refer to the former as decision-related gains and to the latter as cost-related gains.

**An Example**

To illustrate decision- and cost-related gains from strategic expansion, consider the following example: A patient observes symptoms of a health condition that can be either serious ($\theta = 1$) or harmless ($\theta = 0$). Since the symptoms are very similar in each case and the patient is inexperienced, she cannot judge her condition. However, it is common knowledge that the probability of a serious condition is $\pi_0 = \frac{1}{4}$. The patient attends a physician and has to decide whether or not to undergo a costly treatment. All agents are risk neutral and payoffs are as follows:

<table>
<thead>
<tr>
<th></th>
<th>health outcome if $\theta = 0$</th>
<th>health outcome if $\theta = 1$</th>
<th>cost to patient</th>
<th>physician’s profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>no treatment</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>treatment</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Note that the patient chooses the treatment if she believes that her condition is serious with probability $\pi > \frac{1}{2}$. Without further information, the patient hence decides not to be treated, which results in an expected payoff of $-1$ for the patient and 0 for the physician.

Now suppose that the physician is very experienced and can perfectly observe the patient’s condition. However, since the physician always has an interest to convince the patient that the treatment is necessary, cheap talk cannot credibly transmit any information. But suppose that the physician can administer a test at cost $c = \frac{3}{8}$ and reveal the test result to the patient. Further suppose that if the patient’s condition is serious, the test always generates a positive result, but if the condition is harmless, the test generates a false positive with probability $p = \frac{1}{2}$. Moreover, while the physician
cannot fake the test result, he can repeat the test as often as he wants and report only the last test result.\footnote{For simplicity, assume that submitting multiple positive test results does not reveal more information than a single positive test result since the patient cannot distinguish between the duplication of single test result and the results from two separate tests.} Now, there exists a number of outcome-equivalent informative equilibria. In each of these equilibria, the patient perceives that her condition is serious with probability $1/2$ if the test result is positive and chooses to be treated with probability $3/4$ in that case. Otherwise, she knows that her condition is harmless and chooses not to be treated. The physician administers the test with certainty if the patient’s condition is serious. If the patient’s condition is harmless, the physician generates a positive test result with probability $1/3$, e.g., he does not administer the test with probability $2/3$, but with probability $1/3$, he repeats the test until it shows a positive result.\footnote{There exists a number of outcome equivalent equilibrium strategies if $\theta = 0$, e.g., the physician may choose to administer the test with probability $1/2$ initially, but if it fails, he stops with probability $7/8$ and otherwise repeats it until it shows a positive result.} This equilibrium generates a payoff equal to 0 for the physician if the condition is harmless and of $3/8$ if the condition is serious. The patient still receives an expected payoff equal to $-1$.

Finally, suppose the test result is not directly revealed to the patient. Instead, the physician reveals the test result to a counselor. The counselor then communicates the test result to the patient, but with probability $\varepsilon = 1/4$, the counselor reports a positive test result even if the physician did not. Now, there exists an informative equilibrium in which the physician administers the test if and only if the condition is serious, and the patient chooses the treatment if and only if the counselor tells her that the test result was positive. In this equilibrium, the physician benefits from miscommunication in either state and receives an expected payoff equal to $1/4$ if the patient’s condition is harmless, and a payoff equal to $5/8$ if the patient’s condition is serious. The patient also strictly benefits from miscommunication and receives an expected payoff equal to $-7/8$.\footnote{She receives the treatment whenever her condition is serious and pays the treatment costs with probability $\frac{7}{16}$.} Note that in this equilibrium, communication costs are strictly smaller than without miscommunication, yet the test signal is strictly more informative. Thus, there are both decision- and cost-related gains from miscommunication. These gains stem from strategic expansion since a perfectly informative strategy profile is used to implement a partially informative distributions of posteriors.

The above example illustrates that the patient can extract more information from the physician simply by hiring a counselor that appropriately garbles the information provided by the physician. To improve information transmission, the counselor has to have no better information than the
patient and may in fact be completely ignorant. At the same time, the physician is also strictly better off in each state than without miscommunication. Similarly, employers may extract more information from their employees by using consultants as a garbling mechanism, and investors may improve their investment decisions by receiving less precise information about the messages sent by product developers. Also, politicians may effectively reveal more information to the public if their messages get garbled appropriately by the media.

How general are these results? After reviewing the relevant literature on communication games and communication error in section 2, section 3 shows that gains from miscommunication can occur in a wide variety of communication games and characterizes the sources of such gains. Sections 4, 5, and 6 apply these results to costly signaling, verifiable messages, and cheap talk. In particular, section 4 illustrates decision- and cost-related gains from strategic expansion in a competitive version of the Spence signaling game, characterizes optimal miscommunication, and provides conditions for which any equilibrium without miscommunication can be Pareto dominated by an equilibrium with optimal miscommunication. Section 5 illustrates decision-related gains from communication restrictions in the case of verifiable messages. Section 6 reviews decision-related gains from strategic expansion in cheap talk games. Section 7 concludes.

2. Literature Review

Ever since Akerlof’s (1970) analysis of the destructive effect of asymmetric information on market efficiency through adverse selection, economists have studied how economic agents may respond to informational asymmetries. Most relevant for this paper is the strand of the literature in which communication efforts originate from the more informed agent, i.e., a more informed sender sends a message to a less informed receiver before the latter makes a decision that affects the welfare of both. The formulation of these communication games differs along two important dimensions: First, the extent to which the sender’s and the receiver’s preferences over decisions vary differently with the sender’s type—sometimes referred to as the extent of preference alignment—and second, the characteristics of the communication technology, in particular the relationship between the sender’s type, available messages, and communication costs.

With respect to the latter, three prominent communication technologies considered in the literature are cheap talk (Crawford and Sobel (1982)), verifiable messages (Milgrom (1981); Grossman
(1981)) and *costly signaling* (Spence (1973, 1974, 2002); Riley (1975)). Among these communication technologies, cheap talk and verifiable messages represent two interesting extreme cases: If communication is cheap talk, the message space is arbitrarily large and each type of the sender has access to all messages at zero costs. Thus, any revelation of information requires that different types of the sender prefer different decisions. If agents communicate via verifiable messages, a message can be interpreted as a list of the sender’s types and the sender can send a given message at zero costs if and only if the message contains his true type; all other messages are infeasible. Therefore, the sender can always perfectly reveal his type at zero costs by sending the message that only contains his true type. Costly signaling represents an intermediate case: Communication costs vary differently across messages depending on the sender’s type, but each type can send any message at finite costs.

The extent to which each of these communication technologies can be used to transmit information in the absence of miscommunication crucially depends on the extent to which the sender’s and the receiver’s preferences over decisions vary differently with the sender’s type. For example, in the competitive version of the well-known Spence labor market signaling game (Spence (1974)) analyzed in section 4, the sender’s payoff only depends on his induced belief, but not on his own type, while the receiver’s payoff depends on both. In this case, the sender’s message may have informational content only if communication costs vary across messages and types of the sender. In particular, cheap talk cannot transmit any information in equilibrium, while verifiable messages can perfectly transmit information without imposing any communication costs. More generally, Spence derives conditions for which costly signaling can be used to transmit information in equilibrium but shows that if the costs of different messages vary little across types, communication efforts can lead to costly ‘rat races’ in the sense that potential gains from better information transmission are outweighed by the associated increase in communication costs.

Crawford and Sobel (1982), henceforth referred to as CS, suggest a different source of the message’s information content: If the receiver’s and the sender’s interest is at least partially aligned, then even cheap talk can result in informative communication. In particular, they show that there exist informative equilibria in which the type space is partitioned into different segments, and types in a given segment send the same message, whereas types in different segments send different messages. Since cheap talk does not involve communication costs, there is no risk of costly rat races, and from an ex ante perspective—i.e., prior to the revelation of private information—both the sender and the
receiver may benefit from informative cheap talk. However, the quality of information transmission is limited by the extent of preference alignment and unless alignment is perfect, cheap talk cannot implement a perfectly informative equilibrium. In this setting, introducing communication costs can improve the quality of equilibrium information transmission, but whether this also improves welfare is ambiguous. For instance, Kartik (2009) considers a hybrid of the CS-model and the Spence-model, a CS-model with lying costs. Lying costs result in a finer partition of the type space and thus better information transmission than in the cheap talk version of the CS-model, but whether this increases ex ante welfare is ambiguous.

Verifiable messages can often implement perfect information transmission at zero costs even if cheap talk only allows for very limited information transmission and costly signaling results in inefficient rat races. In particular, this is true for preferences structures considered in both the competitive Spence-model and the CS-model. However, even verifiable messages cannot always implement perfect information transmission. For example, if the sender’s preferences are not monotone in the sense that type $A$ prefers the receiver’s optimal decision for type $B$ while type $B$ prefers the receiver’s optimal decision for type $A$, a perfectly informative equilibrium may not exist even if it is ex ante efficient.

Surprisingly, the literature on noisy communication is relatively small and quite recent. To the best of my knowledge, Myerson (1991) is the first to mention that noise can improve cheap talk communication in a simple example that goes back Farrell (1988). In particular, Myerson models communication error as a carrier pigeon that delivers the sender’s message with probability less than 1. More recently, Blume et al. (2007) show that adding a specific type of communication error to the CS-model can improve welfare from an ex ante perspective since it allows for equilibria with a finer and more even partition of the type space. In particular, they show that if the type distribution is uniform and preferences are quadratic, there almost always exist sufficiently small levels of noise for which the sender reveals more information about her type than in any equilibrium without noise. Goltsman et al. (2009) show that the type of communication error considered by Blume et al. (2007) in fact maximizes ex ante welfare in the uniform-quadratic version of the CS-model. Similarly, Harbaugh and Rasmusen (2013) argue that coarse grades can transmit more information than precise grades.

---

4With some discrete probability, the message is transmitted without error; otherwise, it is drawn from an error distribution over the entire message space that is independent of the message sent and of the sender’s type.
The literature is even more limited in the context of costly signaling games. Hoppe, Moldovanu, and Sela (2009) consider matching with two-sided incomplete information and costly signaling and derive conditions for which it is efficient to shut down the signaling technology. However, they do not consider intermediate levels of noise. Fuchs and Skrzypacz (2013) consider a dynamic model of asset trading, in which the decision of when to trade can be interpreted as a costly signaling technology, and show that restricting people from trading the asset in some periods can improve market efficiency.

Kamenica and Gentzkow (2011) investigate the case in which the sender directly controls the mechanism through which private information is revealed to the receiver. The authors show how to use results from convex analysis to determine the mechanism—i.e., the transformation of a prior distribution of private information into a distribution of posterior beliefs—that is optimal from the sender’s (ex ante) perspective. This optimal mechanism determines an upper bound for the sender’s payoff that can be achieved with any communication device. In particular, their result implies that the sender may benefit from ‘persuasion’ for some prior distribution if and only if the sender’s expected payoff is not globally concave in the receiver’s belief. Similar to Kamenica and Gentzkow (2011), Ostrovsky and Schwarz (2010) argue that colleges can get better average placements for their students by making performance signals less informative, and Rayo and Segal (2010) show that partial information disclosure is optimal for the sender if the sender has private information on both her own and the receiver’s payoff. Moreover, in a recent paper, Chakraborty and Harbaugh (2013) illustrate that in a game between a buyer and a seller in which the seller has private information about the characteristics of the good and the buyer has private information about his tastes, mere puffery by the seller can be persuasive and benefit the seller.

To a lesser extent, this paper is also related to the strand of the literature in which the sender’s message or action not only affects information transmission but also has a direct effect on welfare such as the career concerns literature (see, e.g., Holmstrom (1999); Dewatripont and Tirole (1999), Lacker and Weinberg (1989); Maggi and Rodriguez-Clare (1995)), the political economy literature on voter selection of politicians (see, e.g., Ashworth, Bueno de Mesquita, and Friedenberg (2010)), and the literature on decision delegation. With respect to the latter, Prat (2005) shows that in a model in which a principal delegates a decision to a better informed expert and the expert’s precision of private information is type-specific, reputational concerns in combination with full transparency may lead to inefficient decision making and can prevent any transmission of information. Under
some assumptions, the quality of both decision making and information transmission can be improved if the principal can only observe the consequences of the agent’s action. However, Fox and Van Weelden (2012) point out that the conclusion that “transparency of consequences” is always beneficial while “transparency of action” is potentially harmful is not a general one.

Finally, this paper is related to the vast literature on mediation and optimal communication devices starting with the seminal work of Roger Myerson (1979; 1982; 1986). As shown by Myerson, the use of such devices can expand the set of outcomes that can be implemented in equilibrium. Communication error can be interpreted as a restricted form of mediation, in which the mediator is constrained to limiting public information about the players’ actions. In particular, the mediator cannot implement transfers between players. Analyzing this restricted form of mediation in communication games is interesting since first, the characteristics of communication games allow us to substantially simplify and concretize the analysis compared to general multi-stage games and second, the results show how restricting information can improve welfare in situations in which information transmission per se is socially valuable and the whole point of interaction between players is to affect the informational environment of other players.

3. Sources of Gains From Miscommunication

3.1. Basic Setup Without Miscommunication. Consider the following game between a privately informed agent $S$, call him sender, and an uninformed agent $R$, call her receiver: $S$ privately observes his type $\theta \in \Theta$ and sends a message $m \in M_\theta \subseteq M$ to $R$. $R$ observes $m$ (but not $\theta$), forms a posterior belief $\mu \in \Delta(\Theta)$, and makes a decision $d \in D$. Subsequently, payoffs are realized as follows: If $S$ is of type $\theta$, then $R$’s payoff from making decision $d$ is $u^R(d, \theta)$ and $S$’s payoff equals $u^S(d, \theta) - C(m|\theta)$. That is, the sender’s payoff is separable in a decision-related component, $u^S(d, \theta)$, and a communication cost component, $C(m|\theta)$. Sender and receiver are assumed to be Bayesians who share a common prior $\mu_0 \in \text{int}(\Delta(\Theta))$.

The game described above is a ‘persuasion mechanism’ with perfect private information as defined in the web appendix of Kamenica and Gentzkow (2011), which nests a variety of prominent communication games such as the Crawford-Sobel (1982) cheap talk game, Navin Kartik’s (2009) strategic

\footnote{Alternatively, $\theta$ can be interpreted as the state of the world.}

\footnote{Note that this formulation accounts for the possibility that some messages may not be available to all types as, for instance, in the case of verifiable messages.}

\footnote{$\Delta(X)$ denotes the set of all probability distributions on a set $X$.}
communication game with lying costs, the persuasion game with verifiable messages as in Grossman (1981) and Milgrom (1981), and Spence's (1973; 1974) signaling game.

For ease of exposition, I assume throughout this section that $\Theta$ is finite. Also, to make the game non-trivial, I assume that the receiver’s optimal strategy is not the same for all types of the sender and that the preferences of the sender and the receiver are not perfectly aligned, i.e., there exists at least one type $\theta$ that prefers a decision that the receiver considers optimal for $\theta' \neq \theta$, but not for $\theta$. Throughout the paper, the equilibrium concept employed is perfect Bayesian equilibrium.

3.2. Modeling Miscommunication. To allow for miscommunication, I assume that instead of observing the message $m$ directly, $R$ only observes a signal $y \in Y$ that is a potentially ‘garbled’ version of $m$, i.e., it results from a garbling mechanism

$$G : M \rightarrow \Delta(Y).$$

I further assume that $Y = \Delta(\Theta)$, so the size of $Y$ does not impose any restrictions on the garbling mechanism. A communication technology is thus formally defined by $M$, $C$, and $G$ and exhibits miscommunication if $G$ assigns positive probability to the same element of $Y$ for at least two messages in $M$. In particular, I use $G_0$ to denote communication without error.

3.3. Preliminary Considerations. Let $\sigma_\theta$ denote a communication strategy of a sender of type $\theta$ and let $\sigma$ denote the corresponding strategy profile. $\sigma$ is feasible if $\sigma_\theta$ is a probability distribution over $M_\theta$ for each $\theta \in \Theta$. Also, let $F \in \Delta(\Delta(\Theta))$ denote a distribution of posterior beliefs $\mu$, and let $\mu(\theta)$ denote the probability of type $\theta$ if the posterior belief is $\mu$.

$\sigma$ is said to induce $F$ given the garbling mechanism $G$, denoted $\langle \sigma, G \rangle = F$, if the distribution of posteriors inferred by the receiver is $F$ when the sender plays $\sigma$ and the garbling mechanism is $G$. Note that if $F$ can be induced by some $\sigma$, then $F$ is Bayes-consistent, i.e.,

$$\int_{\Delta(\Theta)} \mu dF(\mu) = \mu_0.$$

Now, consider which strategy profiles can induce a given distribution of posteriors for an appropriately chosen garbling mechanism. Note that without miscommunication, any given message can
only induce a single posterior, while with miscommunication, a single message can induce a non-degenerate distribution of posteriors. To describe the relationship between \( \sigma \) and inducible \( F \) more precisely, we need to introduce the concept of informativeness. Following Blackwell (1953), I say that a distribution of posteriors \( F \) is \textit{(weakly) more informative} than \( F' \), denoted \( F \succ F' \), if and only if for \textit{any} utility function \( u^R \), the receiver \textit{weakly} prefers \( F \) to \( F' \). If in addition, there exists \textit{some} \( u^R \) such that the receiver \textit{strictly} prefers \( F \) to \( F' \), I say that \( F \) is \textit{strictly more informative} than \( F' \), denoted \( F \succ F' \). For convenience, I will also say that a strategy profile \( \sigma' \) is \textit{(weakly) more informative} than another strategy profile \( \sigma \) or a distribution of posteriors \( F \) if \( \langle \sigma', \mathcal{G}_0 \rangle \succ \langle \sigma, \mathcal{G}_0 \rangle \) or \( \langle \sigma', \mathcal{G}_0 \rangle \succ F \), respectively. By Blackwell’s (1953) theorem\(^8\), there exists a garbling mechanism \( \mathcal{G} \) such that \( \sigma \) induces \( F \) if and only if \( \sigma \) is weakly more informative than \( F \). Without miscommunication, i.e., for \( \mathcal{G} = \mathcal{G}_0 \), \( \sigma \) induces \( F \) if and only if \( \sigma \) is as informative as \( F \). Thus, \textbf{miscommunication expands the set of strategy profiles than can induce a given distribution of posteriors.} \( \sigma \) is said to \textit{implement} \( F \) given the garbling mechanism \( \mathcal{G} \) if given \( \mathcal{G} \), there exists an \textit{equilibrium} strategy profile \( \sigma \) that induces \( F \). In particular, we say that \( F \) is implementable without miscommunication if there exists a strategy profile that implements \( F \) given \( \mathcal{G}_0 \); and \( F \) is implementable with miscommunication if there exists a (weakly more informative) strategy profile that implements \( F \) for \textit{some} \( \mathcal{G} \).

Note that in an equilibrium without miscommunication, the belief \( \mu_m \) that the sender may expect to induce with an unused message \( m \in M \setminus \text{Supp}(\sigma) \) must be \textit{reasonable} in the sense that \( \mu_m(\theta) = 0 \) if type \( \theta \) cannot send \( m \), i.e., if \( m \notin M_\theta \).\(^9\) In contrast, in an equilibrium with appropriate miscommunication, the sender may expect to induce \textit{any distribution} of posterior beliefs that are either (i) reasonable without miscommunication for \textit{some} unused message or (ii) induced by \textit{some} used message \( m \in \text{Supp}(\sigma) \). Thus, \textbf{miscommunication expands the set of distributions of beliefs that the sender may reasonably expect to induce by unused messages.}

3.4. \textbf{Gains from Miscommunication.} Consider a perfect Bayesian equilibrium of the game with garbling mechanism \( \mathcal{G} \). Note that the receiver’s set of optimal decisions depends on the signal

\(^8\)For a short proof, see Crémer (1982) or Leshno and Spector (1992).

\(^9\)The set of reasonable beliefs for unused messages may be further restricted by equilibrium refinements.
realization $y$ only through its information content, i.e., the induced posterior belief $\mu$. In particular, let $\delta(\mu)$ denote a distribution over optimal decisions given $\mu$. Then, we can define $U^S(\mu, \theta) \equiv \int_D u^S(d, \theta) d\delta(\mu)$ and $U^R(\mu, \theta) \equiv \int_D u^R(d, \theta) d\delta(\mu)$ to represent the sender’s and the receiver’s preference profiles over posterior beliefs, given the receiver’s optimizing behavior. Since the (unconditional) distribution of posteriors $F$ also determines the conditional distribution for any given type $\theta \in \Theta$, denoted $F_{\theta}$, it follows that $F$ determines decision-related payoff components for each (type of) agent. In particular, the receiver’s expected payoff equals

$$
\mathbb{E}[U^R(\mu, \theta)] = \int_{\Delta(\Theta)} \left( \sum_{\theta \in \Theta} U^R(\mu, \theta) \mu(\theta) \right) dF(\mu),
$$

and the decision-related component of the expected payoff of a sender of type $\theta$ equals

$$
\mathbb{E}[U^S(\mu, \theta) | \theta] = \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_\theta(\mu) = \int_{\Delta(\Theta)} U^S(\mu, \theta) \frac{\mu(\theta)}{\mu_0(\theta)} dF(\mu).
$$

Since the sender’s strategy profile determines the communication cost component of the sender’s payoff, the set of equilibrium payoff profiles for a given garbling mechanism $G$ is thus determined by the set of communication strategy profiles and induced distributions of posteriors that can occur in equilibrium. In particular, for a given set of Pareto-weights $\{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\}$, the welfare associated with a communication strategy profile $\sigma$ and a distribution of posteriors $F$ is given by

$$
W(\{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\}, F) = \alpha^R \int_{\Delta(\Theta)} \left( \sum_{\theta \in \Theta} U^R(\mu, \theta) \mu(\theta) \right) dF(\mu) + \sum_{\theta \in \Theta} \alpha^S_\theta \left( \int_{\Delta(\Theta)} U^S(\mu, \theta) \frac{\mu(\theta)}{\mu_0(\theta)} dF(\mu) - \int_M C(m|\theta) d\sigma_\theta(m) \right).
$$

I say that there are *gains from miscommunication* for the set of Pareto-weights $\{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\}$, if there exists an equilibrium with miscommunication that generates higher welfare than any equilibrium without miscommunication. To that end, note that the maximum welfare across equilibria without miscommunication is given by

$$
W(\{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\}, G_0) = \max_{\sigma \in F^I} W(\sigma, F | \{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\})
$$

s.t. $\sigma$ is as informative as $F$.

$$\sigma$$ implements $F$ given $G_0$.

---

10If the receiver’s optimal decision given $\mu$ is not unique and different optimal strategies result in different sets of equilibria, we will have to consider each distribution over optimal decisions in turn.
With miscommunication, maximum welfare is given by

$$W(\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}) = \max_{\sigma, F} W(\sigma, F|\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta})$$

s.t. \(\sigma\) is weakly more informative than \(F\).
\(\sigma\) implements \(F\) for some \(G\).

Comparing problems (3) and (4) reveals two sources of potential gains from miscommunication. The first source stems from the fact that miscommunication expands the set of distributions of beliefs that the sender can reasonably expect to induce with unused messages, and the second source stems from the fact that miscommunication expands the set of strategy profiles that can induce a given distribution of posteriors.

To illustrate the first source, consider the implementation of a pair \((\sigma, F)\) and suppose that \(F\) is as informative as \(\sigma\). Then, the set of incentive compatibility constraints that only involve used messages \(m \in \text{Supp}(\sigma)\) is exactly the same for both problems (3) and (4). Suppose that this set of incentive compatibility constraints holds. Then, implementation of \((\sigma, F)\) may still fail without miscommunication if for some unused message \(m \notin \text{Supp}(\sigma)\), any reasonable belief generates a profitable deviation for some type \(\theta \in \Theta\). In this case, however, \((\sigma, F)\) can still be implemented with miscommunication if no profitable deviation exists for some distribution of posteriors that are either (i) reasonable without miscommunication for some unused message or (ii) induced by some used message \(m \in \text{Supp}(\sigma)\). I refer to this source as gains from communication restrictions since such gains stem from removing profitable deviations to unused messages, which could also be achieved by restricting the message space.

To illustrate the second source, recall that without miscommunication, each message can only induce a single posterior belief. Thus, implementation of a distribution of posterior beliefs \(F\) requires that there exists a message for each belief \(\mu \in \text{Supp}(F)\) such that each \(\theta \in \Theta\) is indifferent between all messages that induce a belief \(\mu \in \text{Supp}(F_\theta)\) but at least weakly prefers such messages to messages that induce a belief \(\mu \in \text{Supp}(F)\). Such a set of messages may not exist for a given \(F\) or it may be associated with significant communication costs. With miscommunication, in contrast, a single message can induce a non-degenerate distribution of posteriors and a given \(F\) can be induced by any strategy profile that is more informative than \(F\). In particular, note that we can implement a Bayes-consistent \(F\) by a pure strategy profile if there exists a message \(m_\theta\) for each type \(\theta \in \Theta\) such
that if \( m_\theta \) induces \( F_\theta \), \( \theta \) prefers \( m_\theta \) to \( m_{\theta'} \) for all \( \{\theta, \theta'\} \subseteq \Theta \). The fact that miscommunication expands the set of strategy profiles that can induce a given distribution of posteriors can thus allow us to implement combinations of posterior distributions \( F \) and strictly more informative strategy profiles \( \sigma \) that are socially preferable to any implementable combination of \( F \) and \( \sigma \) that are equally informative. I thus refer to this source as gains from strategic expansion.

Proposition 1 below states that if there are gains from miscommunication, i.e., if \( \mathcal{W}(\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}) > \mathcal{W}(\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}|G_0) \), then they stem from one of these two sources.

**Proposition 1.** Suppose there are gains from miscommunication for some set of Pareto-weights \( \{\alpha^R, \{\alpha^S_\theta\}_{\theta \in \Theta}\} \) and \((\sigma^*, F^*)\) solves problem (4). Then, one of the following conditions is true:

1. \( \sigma^* \) is strictly more informative than \( F^* \).
2. \( \sigma^* \) is as informative as \( F^* \), but for some unused message \( m \in M \setminus \text{Supp}(\sigma^*) \), any reasonable belief without miscommunication implies a profitable deviation for some type \( \theta \in \Theta \). However, for each unused message \( m \in M \setminus \text{Supp}(\sigma^*) \), there exists a distribution of posteriors for which there is no profitable deviation for any \( \theta \in \Theta \) and each posterior in the support of that distribution is either (i) reasonable without miscommunication for some unused message or (ii) induced by some used message \( m \in \text{Supp}(\sigma^*) \).

**Proof.** If there are gains from miscommunication and \((\sigma^*, F^*)\) solves problem (4), then \( \sigma^* \) is weakly more informative than \( F^* \) but \((\sigma^*, F^*)\) does not solve problem (3). Hence, either \( \sigma^* \) is strictly more informative than \( F^* \) or, if \( \sigma^* \) is as informative as \( F^* \), there exists no profitable deviation to any unused message for any type \( \theta \in \Theta \) if and only if unused messages are garbled appropriately with other messages. This is true if and only if the conditions stated in (2.) hold. \( \Box \)

Next, note that if gains from miscommunication exist and \((\sigma^*, F^*)\) solves problem (4), then one of the following is true: Either \( F^* \) is not implementable without miscommunication, or \( F^* \) is implementable without miscommunication, but \( \sigma^* \) involves strictly lower communication costs\(^{12} \) than any \( \sigma \) that implements \( F^* \) without miscommunication. If the former is true, I refer to gains from miscommunication as decision-related. If the latter is true, I refer to them as cost-related.

Economists are typically interested in situations in which information transmission is socially valuable, i.e., in which more informative distributions of posteriors are ceteris paribus socially preferable.

\(^{11} \)Assume that unused messages provide no profitable deviation.

\(^{12} \)Measured by the communication cost component of social welfare, \( \sum_{\theta \in \Theta} \alpha^S_\theta \int_M C(m|\theta)d\sigma_\theta(m) \).
to less informative distributions. In this case, gains from miscommunication allow us to manage the tradeoff between information transmission and communication costs more efficiently.

The definition of gains from miscommunication outlined above compares the entire set of equilibria with and without miscommunication. Sometimes, we may be interested in the more narrowly defined question whether there are gains from miscommunication for a given equilibrium without miscommunication. Thus, if \( \sigma \) is an equilibrium strategy profile without miscommunication, I will say that there exist gains from miscommunication for the equilibrium associated with \( \sigma \) if

\[
W(\alpha_R, \{\alpha_S^\theta\}_{\theta \in \Theta}) > W(\sigma, \langle \sigma, G_0 \rangle | \alpha_R, \{\alpha_S^\theta\}_{\theta \in \Theta}).
\]

The next three sections illustrate gains from miscommunication in the case of costly signaling, verifiable messages, and cheap talk. If communication is cheap talk, gains from miscommunication are always decision-related and stem from strategic expansion. In the case of costly signaling, gains from miscommunication also stem from strategic expansion, but may be cost- or decision-related. In the case of verifiable messages, decision-related gains from communication restrictions can allow us to implement a perfectly informative equilibrium if and only if there is miscommunication.

4. MISCOMMUNICATION IN THE COMPETITIVE SPEHCE SIGNALING GAME

4.1. Basic Setup and Assumptions. To motivate the competitive Spence signaling game, consider an economy with a continuum of workers of mass 1 and free firm entry. The economy lasts for one period and all agents are risk neutral. Workers are endowed with skill level \( \theta \) that can be either low (\( L \)) or high (\( H \)). The fraction of high skill workers, \( \pi_0 \), is common knowledge. At date 0, a worker privately observes his individual skill level and sends a message \( m \in M \) with type-specific cost \( C(m|\theta) \). All firms observe the same potentially garbled signal \( y \in Y \) of the worker’s message and make a wage offer to the worker. The worker accepts and receives the highest wage offer. Subsequently, the hiring firm makes a decision \( d \in D \), e.g., an allocative or managerial decision regarding the worker, and the worker generates a finite net revenue of \( q(d, \theta) \) for the firm.\(^{13}\)

Note that since the firm makes decision \( d \) after the compensation of the worker has been determined, it will always choose the value of \( d \) that is optimal given its belief about the worker’s type. Also,\(^{13}\)

\(^{13}\)Here, it is assumed that the wage offer cannot condition on the ex post realized net revenue a worker generates for the firm. This assumption can make sense for a number of reasons: For instance, there might be a long lag between the hiring decision and the complete realization of output (e.g., research, managerial decisions, etc.), or a firm employs many workers and it is difficult for the firm to identify individual contributions to output (e.g., if output is a joint product). The wage determination process assumed here might therefore be a better assumption for young employees with short employment histories than for older employees.
because of free firm entry, there is perfect competition for workers among firms, and thus workers can claim the entire expected surplus from the match given the posterior belief induced by \( y \) and the associated optimal choice of \( d \). Thus, for a given posterior probability \( \pi \) that the worker is the high type, firms make zero expected profits and the worker’s wage is

\[
w(\pi) = \max_{d \in D} \pi q(d, H) + (1 - \pi) q(d, L).
\]

In the language of section 3, \( w \) determines the worker’s decision-related payoff component. Note that since \( w \) is the upper envelope of a family of linear functions, \( w \) is convex and absolutely continuous by construction. Also, since \( w \) is convex, it is semi-differentiable on \((0, 1)\) and differentiable at all but at most countably many points. Moreover, its left- and right-hand derivatives are monotone increasing functions and thus differentiable almost everywhere.

In what follows, I further assume that the high skill worker is always more productive than the low skill worker and that the firm’s decision is non-trivial. Formally,

(A1): \( q(d, H) > q(d, L) \) for any \( d \in D \) and \( \argmax_{d \in D} q(d, H) \neq \argmax_{d \in D} q(d, L) \).

(A1) implies that expected surplus is strictly increasing in the firm’s perception of the worker’s type and that perfect transmission of information about the worker’s type is socially valuable in the sense that it increases total surplus. To summarize, \( w \) is absolutely continuous, strictly increasing, convex and non-linear by construction. For ease of exposition, I assume that \( w \) is twice continuously differentiable everywhere.

Since workers can claim the entire expected surplus of a given match, the curvature of \( w \) captures what I will term the technological value of information (TVI). This value stems from the firm’s ability to make more suitable decisions in response to better information about the worker’s type. A natural local measure for the technological value of information is \( w^\prime \), which is simply the negative value of the well-known Arrow-Pratt measure of absolute risk aversion.\(^{14}\)

\(^{14}\)As an aside, note that if workers are risk averse rather than risk neutral and their preferences can be represented by \( u(w(\pi)) = C(m|\theta) \) for some (strictly) concave function \( u \), the subsequent analysis does not change substantively. The only difference is that the social value of information transmission is now determined by the curvature of \( u(w(\pi)) \) rather than just the curvature of \( w(\pi) \). As a result, the tradeoff between the worker’s risk aversion on the one hand and the technological value of information on the other will generally push optimal distributions of posteriors to be less informative than with risk neutrality.
**Definition.** The *technological value of information* of a twice differentiable function \( w : [0, 1] \to \mathbb{R} \) is defined at \( \pi \in (0, 1) \) as
\[
TVI(\pi|w) = \frac{w''(\pi)}{w'(\pi)}.
\]

Finally, I assume that the communication technology satisfies the following conditions:

**A2:** \( M = [m, \bar{m}] \) with \( \bar{m} > m \).

**A3:** \( C(m|\theta) \) is strictly increasing and continuous in \( m \) with \( C(m|\theta) = 0 \) and \( C(m_2|L) - C(m_1|L) > C(m_2|H) - C(m_1|H) \) for all \( m_2 > m_1 \).

In what follows, it will be useful to define \( c(x) \equiv C(C^{-1}(x|L)|H) \) as the value of the high type’s communication cost function that corresponds to the message \( m \) at which the low type’s communication cost equals \( x \). Assumptions (A2) and (A3) imply that \( x \in [0, C(\bar{m}|L)] \) and that \( c(x) \in [0, x) \) is strictly increasing and continuous in \( x \).

To summarize, we can characterize the labor market outlined above by the following competitive signaling game between a sender (the worker) and a receiver (the ‘market’), which is a special case of the persuasion mechanism outlined in section 3:

1. Nature draws the type of the sender \( \theta \in \Theta = \{L, H\} \) with \( Pr(\theta = H) = \pi_0 \in (0, 1) \).
2. The sender privately observes \( \theta \) and sends the message \( m \in M \).
3. The receiver observes a signal \( y \in Y \) that is generated by a garbling mechanism \( G : M \to \Delta(Y) \) and forms a posterior probability \( \pi \) that the worker is the high type.
4. The sender receives a payoff equal to \( U_S(\pi, \theta) = w(\pi) - C(m|\theta) \). The receiver’s expected payoff is zero.

As in section 3, the strategy of a sender of type \( \theta \), denoted by \( \sigma_\theta \), is a probability distribution over \( M \). In a slight abuse of notation, I will use \( \sigma_\theta = m \) to denote a pure strategy. \( \sigma = (\sigma_L, \sigma_H) \) is a (perfect Bayesian) equilibrium strategy profile of the competitive signaling game with garbling mechanism \( G \) if and only if for all \( \theta \in \{L, H\} \) and all \( m' \in \text{Supp}(\sigma_\theta) \),
\[
m' \in \arg\max_{m \in M} E[w|m, \sigma, G] - C(m|\theta),
\]
where \( E[w|m, \sigma, G] \) is the expected wage that follows from Bayes’ rule given the message \( m \), the strategy profile \( \sigma \), and the garbling mechanism \( G \). I will refer to an equilibrium as informative if there exists a set of signal realizations that occurs with strictly positive probability and induces
\[^{15}\text{(A2) and (A3) do not, however, imply anything about local convexity or concavity of } c(x)\text{.}\]
π ≠ π₀. An informative equilibrium is perfectly informative if the sender perfectly reveals his type; otherwise, it is partially informative.

Further, note that since there are only two types, the posterior probability that the sender is the high type, π, is a sufficient statistic for the belief μ = (1 − π, π). To simplify notation, I thus use F to denote a distribution of posterior probabilities π ∈ [0, 1] and express decision-related payoffs in terms of π and F. For further reference, note that if F has binary support {π₁, π₂}, Bayes-consistency implies that 1 ≥ π₂ > π₀ > π₁ ≥ 0. Moreover, given a prior π₀, the support {π₁, π₂} is a sufficient statistic for F since \( Pr(\pi = π₁) = \frac{π₂ − π₀}{π₂ − π₁} \) and \( Pr(\pi = π₂) = \frac{π₀ − π₁}{π₂ − π₁} \).

Finally, recall from section 3 that if there exists an equilibrium with distribution of posteriors F, then F is Bayes-consistent, i.e.,
\[
\int_0^1 πdF(π) = π₀.
\]

Also recall that the unconditional distribution of posteriors F implies conditional distributions for each type, denoted \( F_L \) and \( F_H \), respectively, which determine decision-related payoff components. In particular, the expected surplus claimed by a low type equals
\[
\int_0^1 w(π)dF_L(π) = \int_0^1 w(π) \frac{1 − π}{1 − π₀} dF(π)
\]
and the expected surplus claimed by a high type equals
\[
\int_0^1 w(π)dF_H(π) = \int_0^1 w(π) \frac{π}{π₀} dF(π).
\]

The remainder of this section proceeds as follows: I first characterize the set of equilibria without miscommunication and derive a set of sufficient conditions for which every equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication. To that end, I first characterize the garbling mechanism that efficiently implements a given distribution of posteriors. This characterization can be used to show that there are cost-related gains from strategic expansion for any partially informative equilibrium without miscommunication and assumptions (A1)-(A3) are sufficient for any partially informative equilibrium without miscommunication to be Pareto dominated by an equilibrium with miscommunication. I further show that any uninformative equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication if the technological value of information is sufficiently large at the prior. Also, if a perfectly informative equilibrium exists without miscommunication, it is Pareto dominated for a set of cost
functions that satisfy (A2) and (A3) and imply sufficiently high communication costs for the high type. The section concludes with a characterization of optimal miscommunication.

4.2. Equilibria Without Miscommunication. Proposition 2 characterizes the set of (perfect Bayesian) equilibria without miscommunication. If there are multiple equilibria that implement the same distribution of posteriors, I restrict attention to the equilibrium that minimizes communication costs.\textsuperscript{16}

**Proposition 2.** In any equilibrium without miscommunication, the support of the distribution of posteriors is of one of the following four types: \{\pi_0\}, \{\pi_1, 1\}, \{0, \pi_2\}, or \{0, \pi_3, 1\}.

- The (uninformative) best pooling equilibrium with support \{\pi_0\} always exists.
- If \(C(\bar{m}|H) \geq w(1) - w(0)\), then equilibria with support \{\pi_1, 1\} exist for \(0 < \pi_1 < \pi_0\). Otherwise, this type of equilibrium exists for \(\pi_1 \leq \pi_1 < \pi_0\), where \(C(\bar{m}|H) = w(1) - w(\pi_1)\).
- If \(C(\bar{m}|L) \geq w(1) - w(0)\), then equilibria with support \{0, \pi_2\} exist for \(\pi_0 < \pi_2 \leq 1\), which includes the best separating equilibrium. Otherwise, this type of equilibrium exists for \(\pi_0 < \pi_2 \leq \bar{\pi}_2\), where \(C(\bar{m}|L) = w(\pi_2) - w(0)\), which excludes perfectly informative equilibria.
- Equilibria with support \{0, \pi_3, 1\} and \(\pi_3 \in (0, 1)\) exist only if \(C(\bar{m}|L) > w(1) - w(0)\).

*Proof.* See Appendix. \(\square\)

4.3. Equilibria With Miscommunication. Proposition 3 below implies that for any set of cost functions that satisfy assumptions (A2) and (A3), miscommunication allows us to implement a much richer set of distributions of posteriors than without miscommunication. In fact, if \(C(\bar{m}|L) \geq w(1) - w(0)\), i.e., if the best separating equilibrium exists, then Bayes-consistency is the only restriction on the set of distributions that can be implemented with miscommunication. Otherwise, \(C(\bar{m}|L)\) limits the informativeness of implementable distributions. Proposition 3 further states that a garbling mechanism \(G\) implements a Bayes-consistent distribution \(F\) efficiently—in the sense that it minimizes communication costs for each type—only if it implements a pure strategy profile and the low type sends \(m\) with certainty. Note that this implies that there exist cost-related gains from strategic expansion for any partially informative equilibrium without miscommunication.

\textsuperscript{16}Note that for a given implementable distribution of posteriors, both the low and the high type’s communication costs can be minimized in the same equilibrium.
and any set of Pareto-weights. Finally, Proposition 3 suggests a simple garbling mechanism that efficiently implements a given distribution $F$. Since such a garbling mechanism minimizes communication costs for each type conditional on implementing $F$, I will henceforth refer to it as a cost-efficient garbling mechanism.

**Proposition 3.** There exists a garbling mechanism $G$ that can implement a distribution of posteriors $F$ if and only if $F$ is Bayes-consistent and $C(\bar{m}|L) \geq \int_{0}^{1} w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi)$, i.e., there exists a message that is sufficiently costly for the low type.

$G$ implements $F$ efficiently—in the sense that it minimizes communication costs for each type—if and only if it implements the pure strategy profile $(\sigma_L, \sigma_H) = (m, m^*_H)$, where

$$C(m^*_H|L) = \int_{0}^{1} w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi).$$

In particular, for $Y = [0, 1]$ and

$$G^*_F(m) = \begin{cases} F_L & \text{if } m \in [\underline{m}, m^*_H) \\ F_H & \text{if } m \in [m^*_H, \bar{m}] \end{cases},$$

the competitive Spence signaling game has an equilibrium with strategy profile $(\sigma_L, \sigma_H) = (m, m^*_H)$ and distribution of posteriors $F$.

**Proof.** See Appendix. □

### 4.4. Gains from Miscommunication

We can use Proposition 3 to derive a number of important properties of partially informative equilibria without miscommunication. In particular, any partially informative equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication.

**Theorem 1.** For any partially informative equilibrium of the competitive Spence signaling game without miscommunication, the following is true:

1. There exists a Pareto superior and weakly more informative equilibrium with miscommunication that involves 0 costs for the low type and strictly lower costs for the high type.

2. There exists a strictly more informative equilibrium with miscommunication that does not involve higher costs for any type.
3. The equilibrium is Pareto dominated by either the best pooling equilibrium or the most informative equilibrium without miscommunication and support \( \{0, \pi_2\} \), which is the best separating equilibrium if and only if \( C(\bar{m}|L) \geq w(1) - w(0) \).

Proof. See Appendix. □

Theorem 1 implies that only uninformative and perfectly informative equilibria may not be Pareto dominated by an equilibrium with miscommunication. Theorem 2 states a simple condition on the curvature of \( w \) at the prior \( \pi_0 \), which implies that for any uninformative equilibrium without miscommunication, there exists a Pareto superior equilibrium with miscommunication.

**Theorem 2.** Given any uninformative equilibrium of the competitive Spence signaling game without miscommunication, there exists a Pareto superior and strictly more informative equilibrium with miscommunication if \( TVI(\pi_0|w) = \frac{w''(\pi_0)}{w'(\pi_0)} > \frac{2}{1-\pi_0} \), i.e., if the technological value of information is sufficiently large at the prior.

Proof. See Appendix. □

Thus, if the technological value of information transmission is sufficiently large at the prior, theorems 1 and 2 imply that any equilibrium without miscommunication that is not perfectly informative is Pareto dominated by an equilibrium with miscommunication. In fact, if \( C(\bar{m}|L) < w(1) - w(0) \), this applies to any equilibrium without miscommunication since no perfectly informative equilibrium exists. Moreover, in this case, miscommunication can transform a signaling technology that can only make all types strictly worse off into a technology that makes all types strictly better off.

**Corollary 1.** If \( C(\bar{m}|L) < w(1) - w(0) \) and \( TVI(\pi_0|w) > \frac{2}{1-\pi_0} \), then for any equilibrium without miscommunication, there exists a Pareto superior equilibrium with miscommunication. Moreover, without miscommunication, giving agents access to the signaling technology makes all types strictly worse off if the technology is used to reveal information. In contrast, with appropriate miscommunication, giving agents access to the same signaling technology makes all types strictly better off.

Note that so far, we have not used any assumption about the communication technology other than (A2) and (A3). If \( C(\bar{m}|L) \geq w(1) - w(0) \), then a further assumption about cost functions is
sufficient to imply that any perfectly informative equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication.

**Theorem 3.** Given any perfectly informative equilibrium of the competitive Spence signaling game without miscommunication, there exists a Pareto superior equilibrium with miscommunication if and only if $c(w(1) - w(0))$ is sufficiently large, i.e., if and only if perfect information transmission is sufficiently costly to implement.

**Proof.** First, note that the low type is never worse off than in a perfectly informative equilibrium. Now, let $V_{H}^{max}$ denote the highest payoff that the high type can achieve in any equilibrium with miscommunication. Note that $V_{H}^{max} \geq w(\pi_0)$ since the best pooling equilibrium always exists.

Moreover, if $TVI(\pi_0|w) > \frac{2}{1-\pi_0}$, then by Theorem 2, there exists a partially informative equilibrium with miscommunication in which the high type is strictly better off than in the best pooling equilibrium and thus $V_{H}^{max} > w(\pi_0)$. Therefore, if $c(w(1) - w(0)) > w(1) - V_{H}^{max}$, then there exists an equilibrium with miscommunication that Pareto dominates any perfectly informative equilibrium. Since $w(1) - V_{H}^{max} \in (0, w(1) - w(\pi_0)]$, while assumptions (A2) and (A3) only imply that $c(w(1) - w(0)) < w(1) - w(0)$, there exists a wide range of cost functions that satisfy (A2) and (A3) as well as $c(w(1) - w(0)) > w(1) - V_{H}^{max}$.

It is important to note that the conditions in theorems 2 and 3 are never mutually exclusive since Theorem 2 only requires a condition on the curvature of $w$, and if this condition is satisfied, there exists a wide range of cost functions that satisfy (A2) and (A3) as well as the additional condition in Theorem 3. Thus, even if the best separating equilibrium exists, any equilibrium without miscommunication is Pareto dominated by an equilibrium with miscommunication if $TVI(\pi_0|w) > \frac{2}{1-\pi_0}$ and $c(w(1) - w(0))$ is sufficiently large, e.g., $c(w(1) - w(0)) > w(1) - w(\pi_0)$. In this case, we can state an analogous result to Corollary 1.

**Corollary 2.** If $C(\bar{m}|L) \geq w(1) - w(0)$, $TVI(\pi_0|w) > \frac{2}{1-\pi_0}$ and $c(w(1) - w(0))$ is sufficiently large, then for any equilibrium without miscommunication, there exists a Pareto superior equilibrium with miscommunication. In particular, if $c(w(1) - w(0)) > w(1) - w(\pi_0)$, then without miscommunication, giving agents access to the signaling technology makes all types strictly worse off if the technology is used to reveal information. In contrast, with appropriate miscommunication, giving agents access to the same signaling technology makes all types strictly better off.
4.5. Optimal Miscommunication. Proposition 3 states precisely which strategy profile efficiently implements a given distribution of posteriors $F$. Thus, given efficient implementation, $F$ becomes a sufficient statistic for both communication costs and decision-related payoffs. Therefore, we can state the problem of choosing an optimal cost-efficient garbling mechanism for an arbitrary Pareto-weight $\alpha \in [0, 1]$ associated with type $H$ as choosing the distribution of posteriors $F$ that maximizes

\[
(1 - \alpha) \int_0^1 w(\pi) \frac{1 - \pi}{1 - \pi_0} dF(\pi) + \alpha \left( \int_0^1 w(\pi) \frac{\pi}{\pi_0} dF(\pi) - c \left( \int_0^1 w(\pi) \left( \frac{1 - \pi}{1 - \pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi) \right) \right)
\]

subject to the constraints that $F$ is (i) Bayes-consistent, i.e.,

\[
\int_0^1 \pi dF(\pi) = \pi_0,
\]

and (ii) implementable, i.e.,

\[
\int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi) \leq C(\bar{m}|L).
\]

If $F^*$ solves the above problem for some $\alpha \in (0, 1)$, I refer to $F^*$ and the associated equilibrium as constrained Pareto efficient. If $\alpha$ equals 0 or 1, respectively, then constrained Pareto efficiency also requires that if type $L$ or $H$, respectively, is indifferent between multiple distributions of posteriors that solve the above problem, $F^*$ also maximizes the other type’s payoff among this set.

Appendix B shows how to solve this problem without imposing further restrictions on $c$ by decomposing it into two steps: The first step solves the problem for a given value of the gap between the expected surplus claimed by $L$ and $H$, $x = \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi)$, and derives the optimal distribution of posteriors conditional on $x$, denoted $F^*_x$. The second step uses $F^*_x$ to determine the optimal surplus gap $x^*$ that maximizes social welfare and infers the optimal distribution of posteriors $F^* = F^*_x$.

The problem is a lot simpler if we assume that the cost functions of low and high types are proportional, i.e., $c(x) = \gamma x$ for some $\gamma \in (0, 1)$. Also assume that $C(\bar{m}|L) \geq w(1) - w(0)$, in which case the best separating equilibrium exists. Then, we can restate the problem as follows:

\[
\max_{F \in \Delta([0, 1])} \int_0^1 w(\pi) \frac{1 - (1 - \psi)\pi}{1 - (1 - \psi)\pi_0} dF(\pi)
\]

s.t. $\int_0^1 \pi dF(\pi) = \pi_0$. 

(6)
where\(^{17}\)
\[
\psi = \frac{\alpha(1 - \gamma) - 1 - \pi_0}{1 - \alpha(1 - \gamma) - \pi_0}
\]
is the social planner’s marginal rate of substitution (MRS) of total surplus claimed by high types for total surplus claimed by low types given the Pareto-weight \(\alpha\), the cost parameter \(\gamma\), and the prior \(\pi_0\).\(^{18}\) The entire set of constrained Pareto efficient garbling mechanisms can thus be determined by solving problem (6) for \(\alpha \in [0, 1]\) or, equivalently, \(\psi \in [0, \frac{1-\gamma - \pi_0}{\gamma}]\).

The advantage of expressing the problem as in (6) is that it has a structure that is very similar to the problem of solving for the sender-optimal persuasion mechanism in Kamenica and Gentzkow (2011). The only difference is that the social value of inducing a posterior \(\pi\) is not simply the expected surplus \(w(\pi)\), but proportional to \(w(\pi)(1 - (1 - \psi)\pi) \equiv \nu(\pi|\psi, w)\). As a result, we can use the same technique as in Kamenica and Gentzkow (2011) to solve problem (6).

This works as follows: Following the terminology of Aumann and Maschler (1995), let \(\nu(\pi|\psi, w)\) denote the concavification of \(\nu(\pi|\psi, w)\), i.e.,
\[
\nu(\pi|\psi, w) = \sup \{z| (\pi, z) \in \text{co} (\nu(\pi|\psi, w))\},
\]
where \(\text{co}(f)\) denotes the convex hull of the graph of a function \(f\). Note that there exists a Bayes-consistent \(F\) such that \(\int_{0}^{1} \nu(\pi|\psi, w)dF(\pi) = z\) if and only if \((\pi_0, z) \in \text{co}(\nu(\pi_0|\psi, w))\) and thus \(\frac{\nu(\pi_0|\psi, w)}{1 - (1 - \psi)\pi_0}\) is the maximum social value that can be achieved by any Bayes-consistent distribution if the social MRS is \(\psi\). Since Bayes-consistency implies that \(\text{Supp}(F) \cap [0, \pi_0]\) and \(\text{Supp}(F) \cap [\pi_0, 1]\) are both non-empty, a Bayes-consistent \(F\) achieves this optimum if and only if for all \(\pi \in \text{Supp}(F)\),

\(^{17}\)Note that
\[
\int_{0}^{1} w(\pi) \left(1 - \alpha \right) \frac{1 - \pi}{1 - \pi_0} + \alpha \frac{\pi}{\pi_0} \right) dF(\pi) = \gamma \int_{0}^{1} w(\pi) \left(\frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi)
\]
\[
= \frac{1 - \alpha(1 - \gamma)}{1 - \pi_0} \int_{0}^{1} w(\pi) \left(1 - \left(1 - \frac{\alpha(1 - \gamma)}{1 - \alpha(1 - \gamma)} \frac{1 - \pi_0}{\pi_0} \right) \pi \right) dF(\pi)
\]
\[
= \int_{0}^{1} w(\pi) \frac{1 - (1 - \psi)\pi}{1 - (1 - \psi)\pi_0} dF(\pi),
\]
where the last line uses that
\[
\frac{1 - \pi_0}{1 - \alpha(1 - \gamma)} = \frac{(1 - \alpha(1 - \gamma))(1 - \pi_0) + \alpha(1 - \gamma)(1 - \pi_0)}{1 - \alpha(1 - \gamma)}
\]
\[
= 1 - \left(1 - \frac{\alpha(1 - \gamma)}{1 - \alpha(1 - \gamma)} \frac{1 - \pi_0}{\pi_0} \right) \pi_0
\]
\[
= 1 - (1 - \psi)\pi_0.
\]

\(^{18}\)Note that the total surplus claimed by high types is \(\pi_0 \int_{0}^{1} w(\pi) \frac{\pi}{\pi_0} dF(\pi) = \int_{0}^{1} w(\pi)\pi dF(\pi)\), and the total surplus claimed by low types is \((1 - \pi_0) \int_{0}^{1} w(\pi) \frac{1 - \pi}{\pi_0} dF(\pi) = \int_{0}^{1} w(\pi)(1 - \pi) dF(\pi)\).
the associated values of $\nu(\pi|\psi, w)$ all lie on the same line that goes through $(\pi_0, \hat{\nu}(\pi_0|\psi, w))$. This in turn implies that for any $\psi$, there exists an optimal distribution with at most binary support, and if an informative distribution is optimal, a binary support maximizes the social value and solves

$$\max_{\{\pi_1, \pi_2\} \in [0, \pi_0) \times (\pi_0, 1]} \frac{l(\pi_0|\pi_1, \pi_2, \psi)}{\pi_2 - \pi_1} = \nu(\pi_1) + (\pi_0 - \pi_1) \frac{\nu(\pi_2|\psi, w) - \nu(\pi_1|\psi, w)}{\pi_2 - \pi_1}.$$

Problem (6) can be solved graphically by plotting both $\nu(\pi|\psi, w)$ and its concavification $\hat{\nu}(\pi|\psi, w)$. In particular, if $\nu(\pi|\psi, w) = \hat{\nu}(\pi|\psi, w)$, then the best pooling equilibrium is optimal. If $\nu(\pi|\psi, w) < \hat{\nu}(\pi|\psi, w)$, then the least informative optimal binary support $\{\pi^*_1, \pi^*_2\}$ can be determined by the largest value of $\pi_1 \in [0, \pi_0)$ such that $\nu(\pi_1|\psi, w) = \hat{\nu}(\pi_1|\psi, w)$ and the smallest value of $\pi_2 \in (\pi_0, 1]$ such that $\nu(\pi_2|\psi, w) = \hat{\nu}(\pi_2|\psi, w)$. More generally, any Bayes-consistent distribution over posteriors $\pi$ for which $\nu(\pi|\psi, w) = l(\pi_0|\pi^*_1, \pi^*_2, \psi)$ is optimal.

Appendix B provides more information on how the optimal distribution of posteriors changes as we vary $\psi$ from 0 to $\frac{1-\gamma}{1-\pi}$. In particular, if $TVI(\pi|w) < \frac{2}{1-\pi}$ almost everywhere on $(0, 1)$, then $\nu(\pi|0, w)$ is globally strictly concave on $(0, 1)$ and thus the best pooling equilibrium is constrained Pareto efficient for any value of the prior $\pi_0 \in (0, 1)$. Otherwise, there exists a union of open intervals on $(0, 1)$ such that the best pooling equilibrium is Pareto dominated by an equilibrium with miscommunication if and only if the prior $\pi_0$ lies in this union of open intervals. In particular, this applies to all values of the prior in an open neighborhood around any $\pi$ for which $TVI(\pi|w) > \frac{2}{1-\pi}$, in which case $\nu(\pi|0, w)$ is locally strictly convex at $\pi$. Finally, Appendix B implies that there exists an equilibrium with miscommunication that is Pareto superior to the best separating equilibrium if

$$\gamma > \frac{1 - \pi_0}{1 - \pi_0(1 - \bar{\psi}(w))},$$

where

$$\bar{\psi}(w) = \sup_{\pi \in (0, 1)} \left\{ \frac{w(\pi) - w(0)}{\pi} / \frac{w(1) - w(\pi)}{1 - \pi} \right\} \in (0, 1).$$

### 5. Verifiable Messages and Decision-Related Gains From Communication Restrictions

In this section, I use the results of section 3 to analyze the relationship between the preference structure and gains from miscommunication if the communication technology is given by costlessly
verifiable messages. That is, \( M = \mathcal{P}(\Theta) \), and for a sender of type \( \theta \in \Theta \), the message \( m \) is feasible if and only if \( \theta \in m \), in which case \( C(m|\theta) = 0 \). Note that unlike cheap talk or costly signaling, this communication technology always allows each type \( \theta \in \Theta \) to perfectly reveal his type to the receiver at zero costs by sending the message \( m = \{\theta\} \). Given this property of verifiable messages, an interesting question is whether miscommunication can ever improve information transmission. The answer to this question depends on the preference structure. If preferences are monotone in the sense that we can order the sender’s types such that for each subset of \( \Theta \), there exists a type in that subset that prefers the perfect revelation of his type to being mixed with other types in that subset in any way, then a perfectly informative equilibrium is implementable at zero communication costs without miscommunication. If in addition, information transmission is socially valuable, miscommunication cannot improve welfare. In particular, this is true for preference structures considered in the CS-model or the competitive Spence-model and an ex-ante definition of social welfare.

However, this result does not hold in general: Consider the implementation of a Bayes-consistent distribution of posteriors \( F \). For simplicity, I restrict attention to distributions \( F \) such that for each message \( m \in M \), there exists at most one posterior belief \( \mu \in \text{Supp}(F) \) for which \( \text{Supp}(\mu) = m \). Let \( M_F = \{\text{Supp}(\mu) | \mu \in \text{Supp}(F)\} \) denote the set of supports of posterior beliefs that are in the support of \( F \) and let \( f \) denote the probability mass function associated with \( F \). Proposition 4 states that there are two conditions for the implementability of \( F \) without miscommunication: First, each type of the sender has to be indifferent between any posterior beliefs he induces, and second, for any message that corresponds to a support of a posterior that is not induced by \( F \), there must exist a reasonable belief such that no type for whom this message is feasible has an incentive to deviate.

**Proposition 4.** Without miscommunication, the communication game with verifiable messages has an equilibrium with Bayes-consistent distribution of posteriors \( F \) if and only if the following two conditions hold:

1. Every type \( \theta \in \Theta \) is indifferent between any \( \mu \in \text{Supp}(F_\theta) \).
2. For each subset of types \( m \subseteq \Theta \), there exists a probability distribution on \( m \), denoted \( \mu_m \), such that each \( \theta \in m \) at least weakly prefers \( F_\theta \) to \( \mu_m \).

**Proof.** I first show sufficiency. Consider a strategy profile \( \sigma \) such that for each type \( \theta \in \Theta \) and each belief \( \mu \in \text{Supp}(F) \), \( \sigma_\theta \) assigns probability \( \frac{\mu(\theta)}{\mu_\theta(\theta)} f(\mu) \) to the message \( m = \text{Supp}(\mu) \). Note

\[ f \] exists since \( \text{Supp}(F) \) is finite.
that $\text{Supp}(\sigma_\theta) = \{m \in M_F | \theta \in m\}$ and $\langle \sigma, G_0 \rangle = F$. Thus, $\sigma$ is feasible and induces $F$ without miscommunication. We are left to show that $\sigma$ is incentive compatible if conditions 1 and 2 hold. Consider an arbitrary $\theta \in \Theta$ and note that for any used message $m \in M_F$ either $m \in \text{Supp}(\sigma_\theta)$ or $m$ is not feasible for type $\theta$. Thus, condition 1 implies that incentive compatibility holds across all used messages $m \in M_F$. Moreover, condition 2 implies that for each unused message $m \in M \setminus M_F$, there exists a reasonable belief $\mu_m$ such that no type has an incentive to deviate from $\sigma_\theta$ to any unused message that is feasible for him. Therefore, $F$ is implementable.

Necessity is straightforward. Consider a type $\theta \in \Theta$ such that $\{\mu_1, \mu_2\} \subseteq \text{Supp}(F_\theta)$. If $U^S(\mu_1, \theta) \neq U^S(\mu_2, \theta)$, then $\theta$ is never willing to mix between any two messages that induce $\mu_1$ and $\mu_2$, respectively. Thus, $F$ is not implementable if condition 1 fails. Likewise, if condition 2 fails for some $m' \subseteq \Theta$, then for any $\sigma$ such that $\langle \sigma, G_0 \rangle = F$ and any reasonable belief for $m'$ given $\sigma$, there exists a type $\theta' \in m'$ such that $m' \notin \sigma_{\theta'}$ but $\theta'$ strictly prefers $m'$ to $\sigma_{\theta'}$. Thus, $F$ is not implementable if condition 2 fails.

In particular, Proposition 4 implies that implementation of a perfectly informative equilibrium may fail only due to restrictions on the set of reasonable beliefs for unused messages. This is formally stated in Corollary 3.

**Corollary 3.** Without miscommunication, the communication game with verifiable messages has a perfectly informative equilibrium if and only if for any subset of types $\Theta_1 \subseteq \Theta$, there exists a probability distribution over $\Theta_1$ such that each $\theta \in \Theta_1$ at least weakly prefers the perfect revelation of his type to this probability distribution.

Appropriate miscommunication relaxes both conditions in Proposition 4. In particular, it turns out to be without loss of generality to restrict attention to pure strategy profiles of the form $\sigma_\theta = \{\theta\}$. In this case, no type has to be indifferent between any posterior beliefs he induces, which can result in decision-related gains from strategic expansion. Moreover, appropriate miscommunication extends the set of distributions of beliefs that the sender may reasonably expect to induce by any unused message to an arbitrary distribution of posterior beliefs. This can result in decision-related gains from communication restrictions. Proposition 5 states these conditions formally.
Proposition 5. With appropriate miscommunication, the communication game with verifiable messages has an equilibrium with Bayes-consistent distribution of posteriors $F$ if and only if the following condition holds:

1. For every subset of types $m \subseteq \Theta$, there exists a probability distribution on $\Delta(\Theta)$, denoted $F_m$, such that each type $\theta \in m$ at least weakly prefers $F_\theta$ to $F_m$.

Proof. I first show sufficiency. Consider a pure strategy profile $\sigma$ such that $\sigma_\theta = \{\theta\}$. Obviously, $\sigma$ is feasible and since $\sigma$ is perfectly informative, there exists a garbling mechanism $G$ such that $\langle \sigma, G \rangle = F$. Moreover, among all used messages $m \in \text{Supp}\{\sigma\}$, the only feasible message for any type $\theta \in \Theta$ is $\{\theta\}$. Thus, incentive compatibility across the set of used messages holds by construction. Now consider an arbitrary unused message $m_u$, i.e., any non-singleton subset of $\Theta$. Since $m_u$ can be arbitrarily garbled with any other message without changing the induced distribution of posteriors $F$, we can choose $G$ such that $m_u$ induces an arbitrary distribution of posteriors $F_{m_u}$. Moreover, condition 1 implies that we can choose $F_{m_u}$ such that each type $\theta$ for whom $m_u$ is feasible, i.e., $\theta \in m_u$, at least weakly prefers $F_\theta$ to $F_{m_u}$ and thus weakly prefers $\sigma_\theta = \{\theta\}$ to $m_u$.

Necessity is straightforward. If condition 1 is violated for some $m' \subseteq \Theta$, then for any feasible strategy profile $\sigma$ and any distribution of beliefs that can be induced by $m'$, there exists a type $\theta' \in m'$ such that $\sigma_{\theta'} \neq m'$ but $\theta'$ strictly prefers $m'$ to $\sigma_{\theta'}$. □

Corollary 4. With appropriate miscommunication, the communication game with verifiable messages has a perfectly informative equilibrium if and only if for any subset of types $\Theta_1 \subseteq \Theta$, there exists a probability distribution over $\Delta(\Theta)$ such that each $\theta \in \Theta_1$ at least weakly prefers the perfect revelation of his type to this probability distribution.

Comparing corollaries 3 and 4 reveals that if preference are not monotone\textsuperscript{21}, then gains from communication restrictions can cause perfect information transmission to be implementable only if there is miscommunication. To illustrate such gains from communication restrictions, consider the following example.

\begin{footnotesize}
\textsuperscript{21}Recall that in the present context, I refer to preferences as monotone if there exists an ordering on $\Theta$ such that for each subset of $\Theta$, there exists a type in that subset that prefers the perfect revelation of his type to being mixed with other types in that subset in any way.
\end{footnotesize}
Example

Suppose the sender may be one of three different types with equal probability, i.e., $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and $\mu_0(\theta_i) = 1/3$ for $i \in \{1, 2, 3\}$. Also, the receiver’s choice set has three elements, $D = \{d_1, d_2, d_3\}$, and the payoff structure can be represented by the following matrices,

$$U^S = \begin{bmatrix} 2 & -1 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad U^R = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

where the element in the $i^{th}$ row and $j^{th}$ column refers to the payoff of $S$ and $R$, respectively, associated with decision $d_i$ and type $\theta_j$.

First, consider the case without miscommunication. Note that if $\theta = \theta_1$, the sender’s and the receiver’s preferences are perfectly aligned and both strictly prefer decision $d_1$. Thus, in any equilibrium without miscommunication, type $\theta_1$ perfectly reveals his type and induces decision $d_1$. There is, however, fundamental disagreement between the sender and the receiver if $\theta \in \{\theta_2, \theta_3\}$ since $\theta_2$ would like to be treated as $\theta_3$, while $\theta_3$ would like to be treated as $\theta_2$. What does this fundamental disagreement imply for equilibrium information transmission if there is no miscommunication? Note that if the receiver expects the sender to be either $\theta_2$ or $\theta_3$, he strictly prefers $d_3$ if $\mu(\theta_3) > 1/2$, but $d_2$ if $\mu(\theta_3) < 1/2$ and is indifferent between $d_2$ and $d_3$ if $\mu(\theta_3) = 1/2$. Now consider the message $m_{23} = \{\theta_2, \theta_3\}$. Since this message is infeasible for $\theta_1$, any reasonable belief must involve $\mu(\theta_1) = 0$.

It is now straightforward to verify that there exists no perfectly informative equilibrium: Note that if $\theta_2$ and $\theta_3$ perfectly reveal their type, both get a zero payoff. If $m_{23}$ is used by $\theta_2$, then it induces $d_2$ and thus $\theta_3$ has an incentive to deviate to this message. Likewise, if $m_{23}$ is used by $\theta_3$, then it induces $d_3$ and thus $\theta_2$ has an incentive to deviate to this message. Finally, if $m_{23}$ is an unused message, any reasonable belief generates a profitable deviation for either type $\theta_2$ or type $\theta_3$. Thus, a perfectly informative equilibrium is not implementable without miscommunication. Moreover, it is straightforward to verify that in any equilibrium without miscommunication, $\theta_1$ perfectly reveals his type and induces the receiver to choose $d_1$, but the receiver is unable to extract any information about whether the sender is $\theta_2$ or $\theta_3$ and mixes with equal probability over $d_2$ and $d_3$ in this case.

From an ex ante perspective, the sender gets an expected payoff equal to 1, while the receiver gets a payoff equal to 2. Thus, if the social planner weighs the sender’s and the receiver’s payoff equally, the ex ante welfare generated without miscommunication amounts to $3/2$. 

Now suppose there is miscommunication. In particular, suppose that all messages except for \( \{ \theta_2 \} \) and \( \{ \theta_3 \} \) are completely garbled such that the receiver can only distinguish three types of messages given by \( \{ \theta_2 \}, \{ \theta_3 \} \) and \( M \setminus \{ \{ \theta_2 \}, \{ \theta_3 \} \} \). Then, \( M \setminus \{ \{ \theta_2 \}, \{ \theta_3 \} \} \) induces \( \mu(\theta_1) \geq \frac{1}{3} \) and thus \( d_1 \).

But compared to that, both \( \theta_2 \) and \( \theta_3 \) prefer to perfectly reveal their type, which is their only feasible alternative. Thus, by completely garbling all messages other than \( \{ \theta_2 \} \) and \( \{ \theta_3 \} \), it is possible to implement a perfectly informative equilibrium. In this equilibrium, the sender gets an expected payoff equal to \( \frac{2}{3} \) from an ex ante perspective, while the receiver gets a payoff equal to \( \frac{8}{3} \). A social planner who weighs the sender’s and the receiver’s payoff equally achieves a total welfare of \( \frac{5}{3} > \frac{3}{2} \).

6. Cheap Talk and Decision-Related Gains From strategic expansion

If communication is cheap talk, i.e., if the message space \( M \) is arbitrarily large and all messages are costless, the discussion in section 3 simplifies considerably: Since all messages are available to all types of the sender at zero costs, there is no restriction on the set of reasonable beliefs for unused messages. This implies that with miscommunication, a Bayes-consistent distribution of posteriors \( F \) is implementable if and only if each type \( \theta \in \Theta \) at least weakly prefers his type-specific conditional distribution \( F_{\theta} \) to that of any other type. The argument is straightforward: Consider a pure strategy profile \( \sigma \) in which each type \( \theta \in \Theta \) chooses a distinct message \( m_{\theta} \in M \). Since \( \sigma \) is perfectly informative, there exists a garbling mechanism \( G \) such that \( \langle \sigma, G \rangle = F \) and thus \( \sigma \) induces \( F \) given \( G \). Moreover, \( \sigma \) implements \( F \) given \( G \) if and only if \( \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta}(\mu) \geq \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta'}(\mu) \) for any \( \{ \theta, \theta' \} \subseteq \Theta \).

Without miscommunication, on the other hand, a Bayes-consistent \( F \) is implementable if and only if for each \( \theta \in \Theta \), we have \( U^S(\mu, \theta) \geq U^S(\mu', \theta) \) for all \( \mu \in \text{Supp}(F_\theta) \) and all \( \mu' \in \text{Supp}(F) \). Obviously, this condition implies that \( \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta}(\mu) \geq \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta'}(\mu) \) for any \( \{ \theta, \theta' \} \subseteq \Theta \). Moreover, the two conditions are equivalent if and only if \( F_{\theta} \) is degenerate for each type. Thus, a Bayes-consistent distribution \( F \) is implementable with miscommunication but not without if and only if the following two conditions are satisfied

1. For some \( \theta \in \Theta \), there exist posterior beliefs \( \mu_1 \) and \( \mu_2 \) such that \( \{ \mu_1, \mu_2 \} \subseteq \text{Supp}(F_\theta) \) and \( U^S(\mu_1, \theta) \neq U^S(\mu_2, \theta) \).
2. \( \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta}(\mu) \geq \int_{\Delta(\Theta)} U^S(\mu, \theta) dF_{\theta'}(\mu) \) for any \( \{ \theta, \theta' \} \subseteq \Theta \).
Given these results, we can determine optimal miscommunication for a set of Pareto-weights \( \{ \alpha^R, \{ \alpha^S_\theta \}_{\theta \in \Theta} \} \) by restating problem (4) as follows:

\[
\max_F \int_{\Delta(\Theta)} \sum_{\theta \in \Theta} \left( \left( \alpha^R U^R(\mu, \theta) + \frac{\alpha^S_\theta}{\mu_0(\theta)} U^S(\mu, \theta) \right) \mu(\theta) \right) dF(\mu)
\]

s.t.

(i) \( F \) is Bayes-consistent.

(ii) \( \int_{\Delta(\Theta)} U^S(\mu, \theta) dF(\mu) \geq \int_{\Delta(\Theta)} U^S(\mu, \theta') dF(\mu') \) for any \( \{ \theta, \theta' \} \subseteq \Theta \).

Note that since there are no communication costs and there is no restriction on reasonable beliefs for unused messages, gains from miscommunication are always decision-related and stem from strategic expansion. The uniform-quadratic case of the CS-model discussed in Blume et al. (2007) and Goltsman et al. (2009) is an example for such gains: In this case, we have \( \Theta = M = [0, 1] \), \( D = \mathbb{R} \), \( u^R(d, \theta) = -(d - \theta)^2 \) and \( u^S(d, \theta) = -(d - (\theta + b))^2 \) for some bias level \( b > 0 \). Note that both the sender and the receiver strictly benefit from better information transmission ex ante. As demonstrated in Blume et al. (2007), a simple communication error structure can make both the sender and the receiver strictly better off for almost all bias levels \( b < 1/2 \). This error structure involves that with some sufficiently small probability \( \varepsilon > 0 \), the receiver observes a draw from a uniform distribution over \( M \), but otherwise perfectly observes \( m \). Goltsman et al. (2009) show that this simple error structure is in fact optimal in this specific version of the model.

7. Conclusion

I have examined gains from miscommunication in a persuasion mechanism that nests prominent communication technologies such as cheap talk, verifiable messages and costly signaling. A common feature of this class of communication games is that the set of equilibrium payoff profiles is entirely determined by the set of equilibrium combinations of communication strategy profiles on the one hand and induced distributions of posterior beliefs on the other hand. In particular, the former determine communication costs, while the latter determine decision-related payoffs.

Gains from miscommunication can arise from two sources. First, miscommunication expands the set of strategy profiles that can induce a given distribution of posteriors \( F \) to include strategy profiles that are strictly more informative than \( F \). Second, miscommunication can remove profitable deviations to unused messages since it expands the set of distributions of beliefs that the sender may
reasonably expect to induce by such unused messages. I refer to the former as gains from strategic expansion and to the latter as gains from communication restrictions. If information transmission is socially valuable, these two sources may allow us to manage the social tradeoff between information transmission and communication costs more efficiently by introducing communication error. In particular, optimal miscommunication can improve information transmission without increasing communication costs (decision-related gains) or reduce communication costs without affecting information transmission (cost-related gains).

I demonstrate such gains from miscommunication for different communication technologies. If communication is cheap talk, gains are always decision-related and stem from strategic expansion. In the case of verifiable messages, gains from communication restrictions can improve information transmission and welfare if the sender’s preferences are not monotone. In the competitive version of the Spence signaling game, gains from miscommunication stem from strategic expansion and can be both cost- and decision-related. For this game, I characterize optimal miscommunication and show that if the technological value of information is sufficiently large at the prior and perfect information transmission is either impossible or too costly to implement, then for any equilibrium without miscommunication, there exists a Pareto superior equilibrium with miscommunication.

While this paper provides some guidance on when miscommunication can be beneficial as well as how to characterize optimal miscommunication, there are a lot of open questions for future research. In particular, if miscommunication is socially optimal, then under what conditions would we expect optimal miscommunication to arise as an equilibrium outcome? In contrast, when may individual agents have an incentive to invest too much in the precision of the communication technology? Also, in which real world settings is it likely that observed miscommunication is beneficial and in which settings should we be concerned about the extent and type of miscommunication? Can we express the relationship between the primitives of the model and optimal miscommunication more generally than in this paper? Finally, this paper has limited the possibility of error to the transmission of the message from the sender to the receiver. Are there additional gains from introducing error at other stages, e.g., into the implementation of the decision made by the receiver or the revelation of private information to the sender? How does error at different stages interact? I plan to investigate these question in future research.


Appendix A. Proofs

A.1. Proof of Proposition 2.

Consider the competitive Spence signaling game without miscommunication. First, note that since the decision-related payoff component \( w(\pi) \) is identical for both types of the sender but differences in communication costs strictly increase in \( m \), there may exist at most one message that is in the support of both \( \sigma_L \) and \( \sigma_H \) in equilibrium. Moreover, while there may exist multiple equilibria that induce a given distribution of posteriors, communication costs are minimized for both types conditional on inducing a given distribution if and only if the low type sends \( m \) with strictly positive probability. Restricting attention to such equilibria, there are thus four sets of strategy profiles that may induce a cost-minimizing equilibrium:

1. Both types send \( m \) with certainty and induce the prior \( \pi_0 \), which results in a payoff equal to \( w(\pi_0) \) for both types. This is the best pooling equilibrium of the competitive Spence signaling game, which always exists.

2. The low type sends \( m_1 = m \) with certainty and the high type mixes between \( m \) and \( m_2 \in (m, \bar{m}] \). Such equilibria induce a distribution of posteriors with binary support \( \{\pi_1, \pi_2\} \), where \( \pi_1 \in (0, \pi_0) \), and generate a payoff of \( w(\pi_1) \) for both types. It is straightforward to verify that if \( C(\bar{m}|H) \geq w(1) - w(0) \), this type of equilibrium exist for all \( \pi_1 \in (0, \pi_0) \). Otherwise, this type of equilibrium exists for \( \pi_1 \in (\pi_1, \pi_0) \), where \( C(\bar{m}|H) = w(1) - w(\pi_1) \).

3. The low type mixes between \( m_1 = m \) and some \( m_2 \in [m, \bar{m}] \), while the high type sends \( m_2 \) with certainty. Such equilibria induce a binary support \( \{0, \pi_2\} \) with \( \pi_2 \in (\pi_0, 1] \) and induce a payoff equal to \( w(0) \) for the low type. It is straightforward to verify that if \( C(\bar{m}|L) \geq w(1) - w(0) \), then this type of equilibrium exists for all \( \pi_2 \in (\pi_0, 1] \). Otherwise, this type of equilibrium exists for \( \pi_2 \in (\pi_0, \bar{\pi}_2) \), where \( C(\bar{m}|L) = w(\bar{\pi}_2) - w(0) \).

4. The low type mixes between \( m_1 = m \) and \( m_2 \in (m, \bar{m}] \), while the high type mixes between \( m_2 \) and \( m_3 \in (m_2, \bar{m}] \). It is straightforward to verify that such equilibria may exist only if \( C(\bar{m}|L) > w(1) - w(0) \) and induce a support of the type \( \{0, \pi^*, 1\} \) for \( \pi^* \in (0, 1) \).

Suppose \( \mathcal{G} \) implements an equilibrium with strategy profile \((\sigma_L, \sigma_H)\) and distribution of posteriors \(F\). By construction, \(F\) is Bayes-consistent, and the high type receives an expected wage equal to \(\int_0^1 w(\pi) \frac{\pi}{\pi_0} dF(\pi)\), while the low type receives an expected wage equal to \(\int_0^1 w(\pi) \frac{1-\pi}{1-\pi_0} dF(\pi)\). Also, since type \(L\) at least weakly prefers \(\sigma_L\) to \(\sigma_H\), we must have

\[
\int_M C(m|L) d\sigma_H(m) - \int_M C(m|L) d\sigma_L(m) \geq \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1-\pi}{1-\pi_0} \right) dF(\pi).
\]

Since \(\int_M C(m|L) d\sigma_H(m) \leq C(\bar{m}|L)\) and \(\int_M C(m|L) d\sigma_L(m) \geq 0\), this implies

\[
C(\bar{m}|L) \geq \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1-\pi}{1-\pi_0} \right) dF(\pi).
\]

This concludes the proof of necessity of the conditions in the first statement. Now consider sufficiency. Suppose that \(F\) is Bayes-consistent and \(C(\bar{m}|L) \geq \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1-\pi}{1-\pi_0} \right) dF(\pi)\). Then, it is straightforward to verify that the garbling mechanism \(\mathcal{G}_F^*\) as defined in equation (5) can implement an equilibrium with pure strategy profile \((\sigma_L, \sigma_H) = (\bar{m}, m_H^*)\) and distribution of posteriors \(F\):

Since the induced distribution of the signal only depends on whether \(m \in [\bar{m}, m_H^*]\) or \(m \in [m_H^*, \bar{m}]\), the sender will never choose any message other than \(\bar{m}\) or \(m_H^*\). If the sender adheres to the proposed strategy profile, \(\bar{m}\) results in an expected wage of \(\int_0^1 w(\pi) \frac{\pi}{\pi_0} dF(\pi)\), while \(m_H^*\) results in an expected wage of \(\int_0^1 w(\pi) \frac{\pi}{\gamma_0} dF(\pi)\). Thus, \(L\) is just indifferent between both messages while \(H\) strictly prefers \(m_H^*\). Hence, no type has an incentive to deviate.

Finally, I show that if a garbling mechanism induces \(F\) through any equilibrium strategy profile other than \((\bar{m}, m_H^*)\), then it imposes strictly higher costs on \(H\) and weakly higher costs on \(L\) and is thus Pareto dominated by the equilibrium associated with the strategy profile \((\bar{m}, m_H^*)\), which can be implemented by \(\mathcal{G}_F^*\).

First, note that (A3) implies that if in equilibrium, \(L\) at least weakly prefers \(m_2\) to some \(m_1 < m_2\), then \(H\) strictly prefers \(m_2\) to \(m_1\). Likewise, if \(H\) at least weakly prefers \(m_1\) to some \(m_2 > m_1\), then \(L\) strictly prefers \(m_1\) to \(m_2\). It follows that in any equilibrium, \(\inf(\text{Supp}(\sigma_H)) \geq \sup(\text{Supp}(\sigma_L))\).

Now suppose given the garbling mechanism \(\mathcal{G}\), there exists an equilibrium in which the distribution of posteriors is \(F\) and \(H\) plays a non-pure strategy \(\sigma_H\). Let \(\bar{m}_H = \inf(\text{Supp}(\sigma_H))\) and \(w_H < \int_0^1 w(\pi) \frac{\pi}{\pi_0} dF(\pi)\) denote the equilibrium expected wage associated with \(\bar{m}_H\). Since \(H\) is indifferent
between all \( m \in \text{Supp}(\sigma_H) \), \( H \) is just as well off as if he sent \( m_H \) and received \( w_H \). Thus, for a given \( w_H \), the high type’s payoff is strictly larger the smaller \( m_H \). However, incentive compatibility determines a lower bound for \( m_H \), given by

\[
(7) \quad w_H - C(m_H | L) = \int_0^1 w(\pi) \frac{1 - \pi}{1 - \pi_0} dF(\pi) - \int_M C(m | L) d\sigma_L(m).
\]

Inspecting equation (7), first note that the lower bound for \( m_H \) is strictly increasing in the low type’s communication cost \( \int_M C(m | L) d\sigma_L(m) \). Thus, the best case scenario for the high type is that \( \sigma_L = m \), in which case the low type’s communication costs are globally minimized. Second, note that an increase in \( w_H \) by \( \varepsilon > 0 \) implies that \( C(m_H | L) \) has to increase by \( \varepsilon \) as well. By assumption (A3), however, \( C(m_H | H) \) increases by less then \( \varepsilon \). It is thus uniquely optimal for the high type if \( w_H \) is as large as possible. But conditional on implementing \( F \), \( w_H \) is maximized if and only if \( \sigma_H = m_H \), where \( m_H \) is uniquely determined by \( C(m_H | L) = \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi) \). □


First, consider a partially informative equilibrium without miscommunication and support \{\( \pi_1, 1 \)\} for some \( \pi_1 \in (0, \pi_0) \). Note that in this equilibrium, the high type must be indifferent between inducing the posteriors \( \pi_1 \) and 1 and thus both types receive an expected payoff equal to \( w(\pi_1) \). Since \( w(\pi_1) < w(\pi_0) \), this type of equilibrium is strictly Pareto dominated by the best pooling equilibrium. Moreover, Proposition 3 implies that there exists an equilibrium with miscommunication that (i) implements the same distribution of posteriors with support \{\( \pi_1, 1 \)\} and (ii) involves 0 costs for the low type but strictly smaller costs for the high type. Moreover, by Proposition 3, there exists a strictly more informative equilibrium with miscommunication and support \{\( \pi_1 - \varepsilon, 1 \)\} for some sufficiently small \( \varepsilon > 0 \) that involves 0 costs for the low type and costs for the high type that are not higher than in the equilibrium without miscommunication and support \{\( \pi_1, 1 \)\}. 


Next, consider a partially informative equilibrium without miscommunication and support \( \{0, \pi_2\} \) for some \( \pi_2 \in (\pi_0, 1) \). Note that in this equilibrium, the low type’s payoff equals \( w(0) \) and the high type’s payoff equals \( w(\pi_2) - c(w(\pi_2) - w(0)) \). By assumption (A3), the high type’s payoff is strictly increasing in \( \pi_2 \), and thus any equilibrium of this type is Pareto dominated by the most informative equilibrium without miscommunication and support \( \{0, \pi_2\} \). Obviously, the latter corresponds to the best separating equilibrium if and only if \( C(\bar{m}|L) \geq w(1) - w(0) \). Now compare the equilibrium without miscommunication and support \( \{0, \pi_2\} \) to the best equilibrium with miscommunication and the same support. By Proposition 3, the latter involves 0 costs for the low type (and thus a strictly higher payoff than \( w(0) \)) as well as strictly smaller costs for the high type. Hence, this equilibrium with miscommunication strictly Pareto dominates the corresponding equilibrium without miscommunication. Also, by Proposition 3, there exists a strictly more informative equilibrium with miscommunication and support \( \{0, \pi_2 + \varepsilon\} \) for some sufficiently small \( \varepsilon > 0 \) that involves a payoff strictly larger than \( w(0) \) as well as 0 costs for the low type and costs for the high type that are not higher than in the equilibrium without miscommunication and support \( \{0, \pi_2\} \).

Finally, consider a partially informative equilibrium with support \( \{0, \pi_3, 1\} \) for some \( \pi_3 \in (0, 1) \). By proposition 2, existence of such an equilibrium implies that \( C(\bar{m}|L) > w(1) - w(0) \) and thus the best separating equilibrium exists. Note that in the best separating equilibrium, the low type is as well off as in a partially informative equilibrium with support \( \{0, \pi_3, 1\} \), but assumption (A3) implies that the high type is strictly better off. Thus, any partially informative equilibrium of this type is Pareto dominated by the best separating equilibrium. By Proposition 3, we can implement the same distribution of posteriors in an equilibrium with miscommunication at 0 costs for the low type and strictly lower costs for the high type. Proposition 3 also implies that we can implement a strictly more informative equilibrium with miscommunication without increasing the costs for any type. □


We need to show that if \( TVI(\pi_0|w) = \frac{w''(\pi_0)}{w'(\pi_0)} > \frac{2}{1-\pi_0} \), then there exists an equilibrium with miscommunication that is Pareto superior to the best pooling equilibrium. First, note that if there exist an equilibrium with miscommunication in which the low type is strictly better off than in the best pooling equilibrium, then the high type is strictly better off as well. Next, note that if there are
no communication costs, the low type strictly prefers a Bayes-consistent distribution of posteriors $F$ to an uninformative distribution if and only if

$$\int_0^1 w(\pi)(1-\pi)dF(\pi) > w(\pi_0)(1-\pi_0). \tag{8}$$

Now, following the terminology of Aumann and Maschler (1995), let $\omega(\pi)$ denote the concavification of $w(\pi)(1-\pi)$, i.e.,

$$\omega(\pi) \equiv \sup \{z | (\pi, z) \in \text{co}(w(\pi)(1-\pi)) \},$$

where $\text{co}(f)$ denotes the convex hull of the graph of a function $f$. As shown in Kamenica and Gentzkow (2011), there exists a Bayes-consistent distribution that satisfies equation (8) if and only if $\omega(\pi_0) > w(\pi_0)(1-\pi_0)$. In particular, if $w(\pi)(1-\pi)$ is locally strictly convex at $\pi_0$, i.e., if

$$\text{TVI}(\pi_0|w) = \frac{w''(\pi_0)}{w'(\pi_0)} > \frac{2}{1-\pi_0},$$

then for any $\varepsilon > 0$ sufficiently small, the low type strictly prefers an informative distribution with binary support $\{\pi_0 - \varepsilon, \pi_0 + \varepsilon\}$ to an uninformative distribution. Proposition 3 implies that for $\varepsilon > 0$ sufficiently small, a distribution with binary support $\{\pi_0 - \varepsilon, \pi_0 + \varepsilon\}$ is implementable with miscommunication at zero cost for the low type and arbitrarily small cost for the high type.22 □

**Appendix B. General Characterization of Optimal Miscommunication**

Consider the problem stated at the beginning of section 4.5. To make further progress without imposing additional assumptions on cost functions, I decompose the problem into two steps. In the first step, I solve the problem for a given value of the gap between the expected surplus claimed by $L$ and $H$, $x = \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1-\pi}{1-\pi_0} \right) dF_{\pi}(\pi) \in [0, w(1) - w(0)]$, henceforth referred to as the surplus gap, and derive the optimal distribution of posteriors conditional on $x$, denoted $F^*_x$. Note that the size of the surplus gap $x$ is a measure of information transmission and thus the Lagrange multiplier associated with the constraint for $x$ can be interpreted as the shadow value of information transmission. In the second step, I use $F^*_x$ to determine the optimal surplus gap $x^* \in [0, \min \{w(1) - w(0), C(\bar{m}|L)\}]$ that maximizes social welfare and infer the optimal distribution of posteriors $F^* = F^*_x$.22

---

22Note that it is not sufficient to show that there exists some partially informative distribution $F$ that satisfies equation (8) since there only exists an equilibrium with miscommunication that implements $F$ if $C(\bar{m}|L) \geq \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1-\pi}{1-\pi_0} \right) dF(\pi)$. 

Step 1: Optimal Communication Error For a Given Surplus Gap

A distribution of posteriors optimally induces the surplus gap \( x \in (0, w(1) - w(0)) \) if and only if it solves the following problem:

\[
\max_{F \in \Delta([0,1])} \int_0^1 w(\pi) \left( (1 - \alpha) \frac{1 - \pi}{1 - \pi_0} + \alpha \frac{\pi}{\pi_0} \right) dF(\pi) \\
\text{subject to} \quad (i) \quad \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi) = x \\
(ii) \quad \int_0^1 \pi dF(\pi) = \pi_0.
\]

Letting \( \lambda \) denote the Lagrange multiplier for the first constraint, we can rewrite this problem as

\[
\max_{F \in \Delta([0,1])} \int_0^1 w(\pi) \left( (1 - \alpha) \frac{1 - \pi}{1 - \pi_0} + \alpha \frac{\pi}{\pi_0} \right) dF(\pi) + \lambda \left( x - \int_0^1 w(\pi) \left( \frac{\pi}{\pi_0} - \frac{1 - \pi}{1 - \pi_0} \right) dF(\pi) \right) \\
\text{subject to} \quad \int_0^1 \pi dF(\pi) = \pi_0,
\]

or, equivalently,

\[
(9) \quad \max_{F \in \Delta([0,1])} \int_0^1 w(\pi) \frac{1 - (1 - \psi) \pi}{1 - (1 - \psi) \pi_0} dF(\pi) + \lambda x \\
\text{subject to} \quad \int_0^1 \pi dF(\pi) = \pi_0,
\]

where

\[
\psi = \frac{\alpha - \lambda}{1 - (\alpha - \lambda)} / \frac{\pi_0}{1 - \pi_0}.
\]

Note that conditional on \( \psi \), the solution to this problem is the same as in the case of proportional cost functions discussed in section 4.5 if the social MRS equals \( \psi \). In what follows, I refer to a solution to problem 9 as \( \psi \)-optimal.

Our next goal is to describe how the set of \( \psi \)-optimal distributions depends on \( \psi \), \( w \), and \( \pi_0 \). In particular, we need to know when a \( \psi \)-optimal distribution can be uninformative, perfectly informative or partially informative, respectively. To that end, define the following sets:

\[
\Pi_U(\psi, w) = \{ \pi \in [0,1]| \nu(\pi|\psi, w) = \hat{\nu}(\pi|\psi, w) \} \\
\Pi_I(\psi, w) = \{ \pi \in [0,1]| \hat{\nu}(\pi|\psi, w) \text{ is linear in a neighborhood around } \pi \}
\]

Note that \( \Pi_U(\psi, w) \) is a closed set and that \( \{0,1\} \subseteq \Pi_U(\psi, w) \). \( \Pi_I(\psi, w) \) is empty if and only if \( \nu(\pi|\psi, w) \) is globally strictly concave; otherwise, \( \Pi_I(\psi, w) \) is a union of open intervals.
The results above imply that an uninformative distribution is \(\psi\)-optimal if and only if \(\pi_0 \in \Pi_U(\psi, w)\) and it is uniquely \(\psi\)-optimal if and only if \(\pi \in \Pi_U(\psi, w) \backslash \Pi_I(\psi, w)\). An informative distribution, on the other hand, is \(\psi\)-optimal if and only if \(\pi_0 \in \Pi_I(\psi, w)\). To describe the \(\psi\)-optimal distribution in this case, let \(t(\pi|\pi_0)\) denote the line that is tangent to \(\hat{\nu}(\pi|\psi, w)\) at \(\pi_0 \in \Pi_I(\psi, w)\) and let\(^{23}\)

\[
\Pi^*(\psi|\pi_0, w) = \begin{cases} 
\{\pi \in [0,1]|\nu(\pi|\psi, w) = t(\pi|\pi_0)\} & \text{if } \pi_0 \in \Pi_I(\psi, w), \\
\{\pi_0\} & \text{if } \pi_0 \notin \Pi_I(\psi, w).
\end{cases}
\]

Then, given a surplus function \(w\) and a prior \(\pi_0 \in (0,1)\), a Bayes-consistent \(F\) is \(\psi\)-optimal if and only if \(\text{Supp}(F) \subseteq \Pi^*(\psi|\pi_0, w)\). Finally, define

\[
\Pi^*_1(\psi|\pi_0, w) \equiv \Pi^*(\psi|\pi_0, w) \cap [0, \pi_0]
\]

and

\[
\Pi^*_2(\psi|\pi_0, w) \equiv \Pi^*(\psi|\pi_0, w) \cap [\pi_0, 1],
\]

and for \(i = 1, 2\), let

\[
\bar{\pi}^*_i(\psi|\pi_0, w) \equiv \min\{\Pi^*_i(\psi|\pi_0, w)\}
\]

\[
\tilde{\pi}^*_i(\psi|\pi_0, w) \equiv \max\{\Pi^*_i(\psi|\pi_0, w)\}.
\]

If \(\pi_0 \in \Pi_I(\psi, w) \backslash \Pi_U(\psi, w)\), then the least informative \(\psi\)-optimal distribution is determined by the binary support \(\{\bar{\pi}^*_1(\psi|\pi_0, w), \bar{\pi}^*_2(\psi|\pi_0, w)\}\) and the most informative \(\psi\)-optimal distribution is determined by the binary support \(\{\tilde{\pi}^*_1(\psi|\pi_0, w), \tilde{\pi}^*_2(\psi|\pi_0, w)\}\).

The above allows us to describe the set of \(\psi\)-optimal distributions of posteriors. Now, we are interested in how \(\Pi^*(\psi|\pi_0, w)\) depends on \(\psi\). Obviously, \(\Pi^*(\psi|\pi_0, w)\) depends crucially on the shape of \(\nu\), which is determined by \(w\) and \(\psi\). In particular, if \(\nu\) is locally strictly convex at \(\pi_0\), then an uninformative distribution cannot be \(\psi\)-optimal. Likewise, if an uninformative distribution is not \(\psi\)-optimal, then \(\nu\) must be strictly smaller than \(\hat{\nu}\) at \(\pi_0\), which in turn implies that \(\nu\) must be locally strictly convex for some \(\pi\) in an open interval around \(\pi_0\). Lemma 1 states the relationship between the curvature of \(\nu\) on the one hand and the technological value of information and \(\psi\) on the other hand.

\(^{23}\)The tangent exists since by construction, \(\hat{\nu}(\pi|\psi, w)\) is linear in a neighborhood around \(\pi_0\) for any \(\pi_0 \in \Pi_I(\psi, w)\).
Lemma 1. \( \nu \) is locally concave (convex) at \( \pi \) if and only if there exists an open neighborhood \( \mathcal{N}(\pi) \) around \( \pi \) such that for all \( x \in \mathcal{N}(\pi) \), we have

\[
TVI(x|w) (1 - (1 - \psi) x) \leq (\geq) \ 2(1 - \psi).
\]

Moreover, if

\[
TVI(\pi|w) (1 - (1 - \psi) \pi) \ < (>) \ 2(1 - \psi),
\]

then \( \nu \) is locally strictly concave (strictly convex) at \( \pi \in (0,1) \).

Proof. Consider an arbitrary belief \( \pi \in (0,1) \). \( \nu \) is locally concave if and only if for some open neighborhood \( \mathcal{N}(\pi) \) around \( \pi \in (0,1) \), we have \( \nu'(x)(y-x) \geq \nu(y) - \nu(x) \) for all \( x, y \in \mathcal{N}(\pi) \). Such a neighborhood exists if and only if \( \nu' \) is nonincreasing on some open neighborhood \( \mathcal{N}(\pi) \) around \( \pi \in (0,1) \), or equivalently,

\[
w''(x) (1 - (1 - \psi)x) \leq 2(1 - \psi)w'(x)
\]

for all \( x \in \mathcal{N}(\pi) \), which is equivalent to

\[
TVI(x|w) (1 - (1 - \psi)x) \leq 2(1 - \psi).
\]

Moreover, \( \nu \) is locally strictly concave at \( \pi \) if

\[
\nu''(\pi) < 0,
\]

or, equivalently,

\[
TVI(\pi|w) (1 - (1 - \psi)\pi) < 2(1 - \psi).
\]

The proof for local convexity and strict convexity is analogous. \( \square \)

Corollary 5. If \( \psi \leq 1 - \frac{1}{\pi} \) for some \( \pi \in (0,1) \), then \( \nu \) is locally strictly concave on \( (\pi', 1) \) for some \( \pi' < \pi \). If \( \psi = 0 \), \( \nu \) is globally strictly concave if and only if \( TVI(\pi|w) < \frac{2}{1 - \pi} \) almost everywhere on \( (0,1) \).

Proposition 6 uses the above results to describe how the set of \( \psi \)-optimal distributions of posteriors depends on the social MRS \( \psi \). In particular, it derives a lower and an upper bound for \( \psi \) such that a uninformative distribution is \( \psi \)-optimal if and only if \( \psi \) is smaller than the lower bound and a perfectly informative distribution is \( \psi \)-optimal if and only if \( \psi \) is larger than the upper bound. If \( \psi \)
lies strictly between these two bounds, then only a partially informative distribution is \( \psi \)-optimal, and the informativeness of the \( \psi \)-optimal distribution is strictly increasing in \( \psi \).

**Proposition 6.** A perfectly informative distribution of posteriors is \( \psi \)-optimal if and only if

\[
\psi \geq \sup_{\pi \in (0,1)} \left\{ \frac{w(\pi) - w(0)}{\pi} \right\} \equiv \bar{\psi}(w) \in (0,1).
\]

An uninformative distribution of posteriors is \( \psi \)-optimal if and only if

\[
\psi \leq \inf_{\pi \in [0,1]\setminus \pi_0} \left\{ \frac{1 - \pi_0 (\pi - w(\pi_0))}{\pi_0} \right\} \equiv \psi(w,\pi_0) \in \left(1 - \frac{1}{\pi_0}, \bar{\psi}(w)\right).
\]

If \( \psi \in (\bar{\psi}(w,\pi_0), \bar{\psi}(w)) \), then a Bayes-consistent and partially informative distribution of posteriors with support \( \text{Supp}(F) \subseteq \Pi^{*}(\psi|\pi_0, w) \) is \( \psi \)-optimal.

The correspondence \( \Pi^{*}_{1}(\psi|\pi_0, w) \) is upper hemicontinuous and decreasing in \( \psi \), while the correspondence \( \Pi^{*}_{2}(\psi|\pi_0, w) \) is upper hemicontinuous and increasing in \( \psi \). Moreover, \( \Pi^{*}_{1}(\psi|\pi_0, w) \) is strictly decreasing in the sense that for all \( \psi' > \psi \), \( \overline{\pi}^{*}_{1}(\psi'|\pi_0, w) > 0 \) implies \( \overline{\pi}^{*}_{1}(\psi'|\pi_0, w) < \overline{\pi}^{*}_{1}(\psi|\pi_0, w) \) and \( \overline{\pi}^{*}_{1}(\psi|\pi_0, w) = 0 \) implies \( \Pi^{*}_{1}(\psi|\pi_0, w) = \{0\} \). Also, \( \Pi^{*}_{2}(\psi|\pi_0, w) \) is strictly increasing in the sense that for all \( \psi' > \psi \), \( \overline{\pi}^{*}_{2}(\psi'|\pi_0, w) < 1 \) implies \( \overline{\pi}^{*}_{2}(\psi'|\pi_0, w) > \overline{\pi}^{*}_{2}(\psi|\pi_0, w) \) and \( \overline{\pi}^{*}_{2}(\psi|\pi_0, w) = 1 \) implies \( \Pi^{*}_{2}(\psi|\pi_0, w) = \{1\} \).

**Proof.** First, recall that there always exists a \( \psi \)-optimal distribution of posteriors with at most binary support. Now consider an arbitrary Bayes-consistent support \( \{\pi_1, \pi_2\} \in [0, \pi_0) \times (\pi_0, 1] \) and recall that if the social MRS is \( \psi \), the social value associated with that support is given by

\[
l(\pi_0|\pi_1, \pi_2, \psi) = \nu(\pi_1|\psi, w) + (\pi_0 - \pi_1) \frac{\nu(\pi_2|\psi, w) - \nu(\pi_1|\psi, w)}{\pi_2 - \pi_1}.
\]

The binary support \( \{\pi_1, \pi_2\} \) is \( \psi \)-optimal if and only if it maximizes the social value \( l(\pi_0|\pi_1, \pi_2, \psi) \) and thus if and only if \( l(\pi|\pi_1, \pi_2, \psi)) \geq \nu(\pi|\psi, w) \) for all \( \pi \in [0, 1] \).

In particular, this implies that the perfectly informative support \( \{0, 1\} \) is \( \psi \)-optimal if and only if \( l(\pi|0, 1, \psi) \geq \nu(\pi|\psi, w) \) for all \( \pi \in (0, 1) \). For a given value of \( \pi \), this is true if and only if

\[
l(\pi|0, 1, \psi) = \nu(0|\psi, w) + \pi (\nu(1|\psi, w) - \nu(0|\psi, w)) \geq \nu(\pi|\psi, w).
\]
or equivalently,

$$\psi \geq \frac{w(\pi) - w(0)}{\pi} \left/ \frac{1}{1 - \pi} \right. \ .$$

Thus, a perfectly informative signal solves problem (9) if and only

$$\psi \geq \sup_{\pi \in (0,1)} \left\{ \frac{w(\pi) - w(0)}{\pi} \left/ \frac{1}{1 - \pi} \right. \right\} \equiv \psi(w).$$

Note that \(\psi(w)\) is independent of the prior and \(\psi(w) \in (0,1)\) since \(w\) is strictly increasing, weakly convex and not linear.

In contrast, an uninformative support \(\{\pi_0\}\) is \(\psi\)-optimal if and only if \(\nu(\pi_0 | \psi, w) = \hat{\nu}(\pi_0 | \psi, w)\), or, equivalently, if and only if \(\nu\) is everywhere weakly smaller than the line that is tangent to \(\nu\) at \(\pi_0\), i.e., if and only if for all \(\pi \in [0,1]\), we have

$$\nu_t(\pi | w, \pi_0) \equiv \nu(\pi_0 | w, \pi_0) + \nu'(\pi_0 | w, \pi_0)(\pi - \pi_0) \geq \nu(\pi | w, \pi_0),$$

which is equivalent to

$$\psi \leq -\frac{1 - \pi_0}{\pi_0} \left( \frac{w(\pi) - w(\pi_0)}{w(\pi) - w(\pi_0)} \right) \frac{1}{1 - \pi_0} - \frac{w'(\pi_0)(\pi - \pi_0)}{w'(\pi_0)(\pi - \pi_0)} \equiv \psi(w, \pi_0 | \pi) \quad \forall \pi \neq \pi_0.$$  

It follows that in this case, an uninformative distribution of posteriors is \(\psi\)-optimal if and only if

$$\psi \leq \inf_{\pi \in [0,1] \setminus \pi_0} \left\{ -\frac{1 - \pi_0}{\pi_0} \left( \frac{w(\pi) - w(\pi_0)}{w(\pi) - w(\pi_0)} \right) \frac{1}{1 - \pi_0} - \frac{w'(\pi_0)(\pi - \pi_0)}{w'(\pi_0)(\pi - \pi_0)} \right\} \equiv \psi(w, \pi_0) \in (1 - \frac{1}{\pi_0}, \psi(w)].$$

Note that if \(\psi \leq 1 - \frac{1}{\pi_0}\), the \(\psi\)-optimal distribution must be uninformative. Why? If \(\psi \leq 1 - \frac{1}{\pi_0}\), then by Corollary 5, \(\nu\) is strictly concave above \(\pi_0\). Note that in this case, a necessary condition for a partially informative distribution to be \(\psi\)-optimal is that for some \(\pi_1 \in [0, \pi_0]\), \(\nu(\pi_1 | \psi, w)\) strictly exceeds the tangent to \(\nu\) to at \(\pi_0\), i.e.,

$$\nu(\pi_1 | \psi, w) > \nu(\pi_0 | \psi, w) - (\pi_0 - \pi_1)\nu'(\pi_0 | \psi, w).$$

However, this requires

$$(\pi_0 - \pi_1)w'(\pi_0)(1 - (1 - \psi)\pi_0) > (w(\pi_0) - w(\pi_1))(1 - (1 - \psi)\pi_1),$$
which is a contradiction since \((\pi_0 - \pi_1)w'(\pi_0) \geq w(\pi_0) - w(\pi_1) > 0\) as well as \(1 - (1 - \psi)\pi_0 \leq 0\) and \((1 - (1 - \psi)\pi_0) < (1 - (1 - \psi)\pi_1)\). Thus, if \(\psi \leq 1 - \frac{1}{\pi_0}\), the \(\psi\)-optimal distribution must be uninformative.

Finally, consider an arbitrary \(\psi^* \in (\psi(w, \pi_0), \tilde{\psi}(w))\) and suppose \(\{\pi_1^*, \pi_2^*\}\) is a \(\psi\)-optimal support given \(\psi^*\), which implies \(1 \geq \pi_2^* > \pi_0 > \pi_1^* \geq 0\). Then, we must have

\[
\nu(\pi|\psi^*, w) \leq l(\pi|\pi_1^*, \pi_2^*, \psi^*) \quad \forall \pi \in [0, 1].
\]

Now, consider \(\psi^{**} > \psi^*\) and let \(\{\pi_1^{**, \pi_2^{**}}\}\) denote a \(\psi\)-optimal binary support given \(\psi^{**}\). Note that

\[
\frac{\partial \nu(\pi|\psi, w)}{\partial \psi} = \pi w(\pi)
\]

is strictly increasing in \(\pi\). As a result, convexity of \(w\) implies that for any \(\pi_2 > \pi_1\), the change in the slope of \(l(\pi|\pi_1, \pi_2, \psi)\) that results from an increase in \(\psi\),

\[
\frac{\partial}{\partial \psi} \left( \frac{\nu(\pi_2|\psi, w) - \nu(\pi_1|\psi, w)}{\pi_2 - \pi_1} \right) = \pi_2 w(\pi_2) - \pi_1 w(\pi_1) = w(\pi_1) + \pi_2 \frac{w(\pi_2) - w(\pi_1)}{\pi_2 - \pi_1},
\]

is strictly increasing in both \(\pi_1\) and \(\pi_2\). In particular, this implies that

\[
\nu(\pi|\psi^{**}, w) < l(\pi|\pi_1^{**, \pi_2^{**}}, \psi^{**}) \quad \forall \pi \in (\pi_1^{**, \pi_2^{**}}).
\]

Hence, \(\pi_1^{**} \leq \pi_1^*\) as well as \(\pi_2^{**} \geq \pi_2^*\), which implies that the correspondence \(\Pi_1^*(\psi|\pi_0, w)\) is decreasing in \(\psi\), while the correspondence \(\Pi_2^*(\psi|\pi_0, w)\) is increasing in \(\psi\).

Next, note that if \(\pi_1^* > 0\) and \(\nu(\hat{\pi}_1|\psi^{**}, w) > l(\pi|\pi_1^{*, \pi_2^{**}}, \psi^{**})\) for some \(\hat{\pi}_1 < \pi_1^*\), then we must have \(\pi_1^{**} < \pi_1^*\). To see that this is true, suppose, by way of contradiction, that \(\pi_1^{**} = \pi_1^*\). Since \(\{\pi_1^*, \pi_2^*\}\) is \(\psi\)-optimal, we must have \(\nu(\pi_2^{**}|\psi^{**}, w) \geq l(\pi_0|\pi_1^{*, \pi_2^{**}}, \psi^{**})\). But then, \(\nu(\hat{\pi}_1|\psi^{**}, w) > l(\pi|\pi_1^{*, \pi_2^{**}}, \psi^{**})\) implies

\[
l(\pi_0|\hat{\pi}_1, \pi_2^{**}, \psi^{**}) > l(\pi_0|\pi_1^{*, \pi_2^{**}}, \psi^{**}),
\]

a contradiction to \(\pi_1^{**} = \pi_1^*\). By an analogous argument, if \(\pi_2^* < 1\) and \(\nu(\hat{\pi}_2|\psi^{**}, w) > l(\pi_0|\pi_1^{*, \pi_2^{**}}, \psi^{**})\) for some \(\hat{\pi}_2 > \pi_2^*\), then we must have \(\pi_2^{**} > \pi_2^*\).
Now suppose $\pi_1^* > 0$. Note that
\[
\frac{\partial \nu'(\pi_1^* | \psi, w)}{\partial \psi} = w(\pi_1^*) + \pi_1^* w'(\pi_1^*) < w(\pi_1^*) + \pi_2^* \frac{w(\pi_2^*) - w(\pi_1^*)}{\pi_2^* - \pi_1^*} = \frac{\partial \left( \nu(\pi_2^* | \psi, w) - \nu(\pi_1^* | \psi, w) \right)}{\partial \psi}.
\]
Thus, for any $\psi^{**} > \psi^*$, there exists a $\tilde{\pi}_1 \in [0, \pi_1^*)$ such that $\nu(\tilde{\pi} | \psi^{**}, w) > l(\pi_0 | \pi_1^*, \pi_2^*, \psi^{**})$, which implies that $\pi_1^{**} < \pi_1^*$.

An analogous argument holds for $\pi_2^{**}$ if $\pi_2^* < 1$. Finally, upper hemicontinuity of $\Pi_1^*(\psi | \pi_0, w)$ and $\Pi_2^*(\psi | \pi_0, w)$ follows straight from the fact that $\nu$ varies continuously in both $\pi$ and $\psi$. \[\square\]

Conditions (10) and (11) have nice intuitive interpretations: $\tilde{\psi}(w | \pi)$ is the ratio of the surplus loss incurred by low types relative to the surplus gain incurred by high types that results from replacing a signal that is degenerate at $\pi$ by a perfectly informative signal with support $\{0, 1\}$. $\bar{\psi}(w | \pi)$ can thus be interpreted as the marginal rate of transformation (MRT) of surplus claimed by high types for surplus claimed by low types that is associated with making a signal that is degenerate at $\pi$ perfectly informative. Stated differently, it is the price of a unit of the high type’s claim on surplus in terms of units of the low type’s claim on surplus associated with perfect information revelation.

Since $\psi$ is the social MRS between these two quantities, i.e., the social planner’s willingness to pay for a unit of the high type’s claim on surplus in terms of units of the low type’s claim on surplus, condition (10) states that a perfectly informative signal is $\psi$-optimal if and only if the social MRS at least weakly exceeds the MRT associated with perfect information revelation for all interior beliefs. It is worth noting that this MRT only depends on the curvature of $w$, but not on the prior $\pi_0$.

Similarly, $-\bar{\psi}(w, \pi_0 | \pi)$ is the ratio of the marginal change in the low type’s claim on surplus relative to the marginal change in the high type’s claim on surplus that results from replacing an uninformative signal with support $\{\pi_0\}$ by a binary signal that assigns an infinitesimal probability to a posterior $\pi \neq \pi_0$. $\bar{\psi}(w, \pi_0 | \pi)$ can thus be interpreted as the MRT of the high type’s claim on surplus for the low type’s claim on surplus that is associated with a marginal improvement in information transmission through a binary signal that assigns an infinitesimal probability to $\pi \neq \pi_0$. Stated differently, it is the price of a unit of the high type’s claim on surplus in terms of units of the low type’s claim on surplus associated with marginal information revelation. Thus, condition (11) states
that an uninformative signal is optimal if and only if the social MRS between these two quantities is smaller than the MRT associated with marginal information revelation for any posterior belief \( \pi \neq \pi_0 \).

To summarize, the results above allow us to identify the set of optimal distributions of posteriors for any given value of the social MRS \( \psi \in [\underline{\psi}(w, \pi_0), \bar{\psi}(w)] \) as the set of Bayes-consistent distributions over \( \Pi^*(\psi|\pi_0, w) \). Moreover, for any element of that set, we can determine decision-related payoffs and the surplus gap \( x \), which is a measure of information transmission. Given Proposition 6, we know that as \( \psi \) increases from \( \underline{\psi}(w, \pi_0) \) to \( \bar{\psi}(w) \), the optimal distribution becomes strictly more informative and thus the surplus gap strictly increases from 0 to \( w(1) - w(0) \). In particular, Proposition 6 implies that there exists a correspondence \( x_\psi \) from \( \psi \) to \( x \) that is upper hemicontinuous and strictly increasing on \( [\underline{\psi}(w, \pi_0), \bar{\psi}(w)] \). Moreover, for any \( x \in [0, w(1) - w(0)] \), there exists a unique \( \psi \) such that \( x \in x_\psi(\psi) \). Hence, we can determine an increasing and continuous function \( \psi_x : [0, w(1) - w(0)] \to [\underline{\psi}(w, \pi_0), \bar{\psi}(w)] \), \( x \mapsto \psi_x(x|w, \pi_0) \) such that for every \( x \in [0, w(1) - w(0)] \), there exists a Bayes-consistent distribution \( F_x^* \) over \( \Pi(\psi_x(x|w, \pi_0)|\pi_0, w) \) such that \( F_x^* \) is the optimal distribution of posteriors conditional on inducing the surplus gap \( x \).

**Step 2: Constrained Pareto Efficient Garbling Mechanisms**

We can now turn to the characterization of constrained Pareto efficient garbling mechanisms. To that end, recall that the solution to step 1 implies a positive relationship between the surplus gap \( x \) and the social MRS \( \psi \) that results from efficient implementation of \( x \). Moreover, the shadow value of information transmission associated with \( \psi \),

\[
\lambda = \alpha - \frac{\pi_0 \psi}{1 - \pi_0 (1 - \psi)},
\]

is strictly decreasing in \( \psi \) since

\[
\frac{\partial \lambda}{\partial \psi} = - \frac{(1 - \pi_0) \pi_0}{(1 - \pi_0 (1 - \psi))^2} < 0.
\]

It follows that for any value of the Pareto-weight \( \alpha \in [0, 1] \), the gross marginal benefit of improving information transmission in an optimal way as given by

\[
\lambda(x|\alpha, w, \pi_0) = \alpha - \frac{\pi_0 \psi_x(x|w, \pi_0)}{1 - \pi_0 (1 - \psi_x(x|w, \pi_0))}
\]
is decreasing in $x$. Hence, we can state the second step problem as

\[
\begin{align*}
\max_{x} & \quad \int_{0}^{x} \lambda(t|\alpha, w, \pi_0) dt - \alpha c(x) \\
\text{s.t.} & \quad 0 \leq x \leq \min\{C(\bar{m}|L), w(1) - w(0)\}.
\end{align*}
\]

Therefore, we can determine an optimal level of the surplus gap, $x^*(\alpha|\pi_0, w, c)$, as well as the associated optimal distribution of posteriors, $F^* = F^x_*$, by comparing the marginal value of information transmission $\lambda(x|\alpha, w, \pi_0)$ to the marginal social cost $\alpha c'(x)$. Moreover, the entire set of constrained Pareto efficient distribution of posteriors can be determined by varying $\alpha$ from 0 to 1. Note that since $\frac{\partial \lambda}{\partial \alpha} = 1 > c'(x) \in (0, 1)$, an increase in $\alpha$ results in a larger optimal surplus gap $x^*$ and thus induces a more informative distribution of posteriors. Hence, the largest solution to problem (12) for $\alpha = 0$ and the smallest solution for $\alpha = 1$, respectively, determine a lower and an upper bound for surplus gaps that are induced by constrained Pareto efficient garbling mechanisms.\(^{24}\) If the former strictly exceeds 0 and the latter is strictly smaller than $w(1) - w(0)$, then any constrained Pareto efficient garbling mechanism implements a partially informative equilibrium and thus each equilibrium without miscommunication is strictly Pareto dominated with miscommunication.

In particular, suppose that $C(\bar{m}|L) \geq w(1) - w(0)$, i.e., the best separating equilibrium exists. Note that if

\[
c'(w(1) - w(0)) > \lambda(w(1) - w(0)|\alpha = 1, w, \pi_0) \]

or, equivalently,

\[
\frac{\left(1 - c'(w(1) - w(0))\right)}{1 - (1 - c'(w(1) - w(0)))} / \frac{\pi_0}{1 - \pi_0} < \bar{\psi}(w),
\]

then marginal costs exceed the marginal benefit for sufficiently large levels of information transmission, i.e., for $x$ sufficiently close to $w(1) - w(0)$. Thus, condition (13) is sufficient for the best separating equilibrium not to be constrained Pareto efficient. Moreover, since $\lambda$ is decreasing, this condition is generically necessary and sufficient if $c$ is weakly convex\(^{25}\), i.e., if the high type's communication costs increase at least proportionally to the low type's communication costs. In this case, a

\(^{24}\)Recall that if $\alpha = 0$, the social planner maximizes the low type’s payoff. If there are multiple solutions to this problem that imply different values of the surplus $x$, the high type’s payoff is uniquely maximized by the largest value of $x$ among this set. Hence, only the largest solution to problem (12) for $\alpha = 0$ is constrained Pareto efficient. Analogously, if $\alpha = 1$, the social planner maximizes the high type’s payoff, and the low type’s payoff is uniquely maximized by the smallest solution to problem (12) for $\alpha = 1$.

\(^{25}\)In the non-generic case that $\lambda(x|\alpha = 1, w, \pi_0)$ is constant on some open interval directly below $w(1) - w(0)$, $c'(w(1) - w(0)) = \lambda(w(1) - w(0)|\alpha = 1, w, \pi_0)$ is necessary and sufficient.
distribution of posteriors $F^*_x$ is constrained Pareto efficient if and only if $x \in [\underline{x}(w, \pi_0), \bar{x}(w, \pi_0, c)]$, where $\underline{x}(w, \pi_0)$ is determined by the most informative Bayes-consistent distribution over $\Pi^*(0|\pi_0, w)$ and $\bar{x}(w, \pi_0, c)$ is determined by the smallest value of $x$ that solves
\[
\frac{1 - \pi_0}{1 - \pi_0(1 - \psi_x(x|w, \pi_0))} = c'(x).
\]