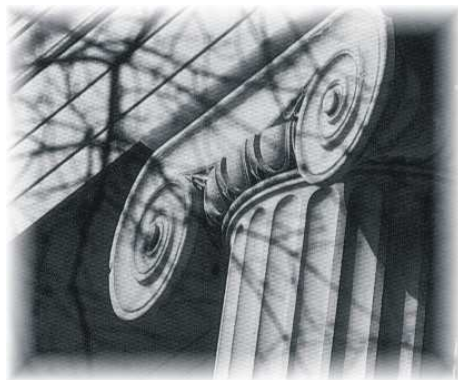


THE UNIVERSITY OF ALABAMA

ECONOMICS, FINANCE AND LEGAL STUDIES

WORKING PAPER SERIES



Working Paper No. 09-09-02

Strategic Truth and Deception

Paan Jindapon and Carlos Oyarzun

September 10, 2009

This paper can be downloaded without charge at:
Department of Economics, Finance and Legal Studies Working Paper Series Index:
www.cba.ua.edu

Culverhouse College of Commerce and Business Administration
The University of Alabama

Strategic Truth and Deception

Abstract. We study strategic communication in a sender-receiver game in which the sender sends a message about the observed quality of the good to the receiver who may accept or reject the good without knowing the true quality or the sender's type. The game has infinitely many perfect Bayesian Nash equilibria. An equilibrium refinement identifies a unique class of equilibria that are outcome-equivalent to the equilibrium in which the neutral sender always tells the truth and the biased sender adopts a feigning strategy to disguise himself by not fully exaggerating about the quality of the good.

Paan Jindapon, University of Alabama, Email: pjindapo@cba.ua.edu

Carlos Oyarzun, University of Alicante, Email: oyarzun@merlin.fae.ua.es

“If you wish to strengthen a lie, mix a little truth in with it.”—Zohar

1 Introduction

Making a decision under imperfect information is very common. A decision maker usually asks a third party or an expert to provide some information or a recommendation. For example, an individual is considering buying a new car to replace the car that she currently owns. While a car dealer recommends a new vehicle whose quality is unknown to her, she knows exactly the quality of her car. She cannot directly observe the quality of the new car or the car dealer’s incentives. How should she make a decision based on the information that she has on her car and the dealer’s recommendation on the new car? Another example is the problem of internal promotion versus external hiring. Why do some firms promote an internal person despite the fact that an outside candidate seems to have a stronger record? A recruiter often uses letters of recommendation to decide whether she should hire an outside candidate without observing first-hand the candidate’s quality. However, the individuals who write the letters may have conflicts of interest with the recruiter; some recommenders may be biased and want the job candidates they recommend to be hired even though the candidates are not qualified for the job or have no qualifications whatsoever. Some recommenders may be honest, and some may want to contribute to a successful hiring process where only qualified candidates are hired or the best applicant is selected. Most of the time, the recruiters are unlikely to know the incentives of the person who wrote a given recommendation letter.

In order to analyze communication in these situations, we develop a model of strategic information transmission from a sender to a receiver. The receiver (she) has to choose one of two goods, X and Y , which are assumed to be equally priced. The quality of each good is random and exogenously determined. The sender (he) observes the quality of X but not the quality of Y , while the receiver observes the quality of Y but not the quality X . The receiver aims to buy the good that has higher quality and her decision is based on the observed quality of Y and the message about the quality of X . We assume that

the receiver does not know the sender's incentives. On the one hand, the sender may have aligned objectives with the receiver, i.e., the X seller may want the buyer to buy X only if its quality is higher than Y , the buyer's outside option. This may occur when the seller has moral concerns or other-regarding preferences.¹ In such a case, we say that the sender is *neutral*. On the other hand, the sender may be *biased*, i.e., he does not care about the quality of the good chosen by the receiver; his goal is to send a message to the receiver so that X is chosen. There is no cost for the sender to send a message, and the quality of X is not verifiable at the time the receiver decides which good to buy.

The timing of the game is as follows. First, the sender observes his type and the quality of X , and the receiver observes the quality of Y . Then, the sender strategically sends a message to the receiver about the quality of X . Upon receiving the message, the receiver updates her beliefs about the type of the sender and the quality of X , and makes a decision to buy either X or Y . The payoff of the receiver is the quality of the good that she has chosen. If the sender is neutral, his payoff is the same as the receiver's. If the sender is biased, his payoff is strictly higher when the receiver chooses X than when the receiver chooses Y . The game can be applied to various situations even though it is based on one sender and one receiver. For example, this game is equivalent to a two-sender game in which the first sender strategically sends a message about X , and the second sender is a non-strategic player who always reports the true quality about Y . Alternatively, Y can be interpreted as a random quality aspiration which is the receiver's private information.

Several questions arise regarding how agents communicate in these situations. Since the receiver does not know the type of the sender, it is unclear how much information the neutral sender can convey. At the same time, the biased sender may be able to exploit the fact that the receiver believes that the sender might be neutral. Yet, it is not clear to what extent the biased sender may take advantage of this situation. Our analysis aims to answer these and other related questions.

¹In his experiment, Gneezy (2005) provides evidence that senders do care about receivers' payoffs and are less likely to engage in deceptive communication when the detriment they may cause to the receivers is greater.

Before stating our results, we compare our framework to those previously analyzed in the literature. The study of sender-receiver games with costless communication (cheap talk) was initiated by Crawford and Sobel (1982). They study information transmission from a better-informed sender to a receiver who later takes an action, within a continuous space, that affects the welfare of both. We deviate from Crawford and Sobel’s framework in two aspects. First, inspired by the literature of persuasion games (for example, Milgrom, 1981; Milgrom and Roberts, 1986; Fishman and Hagerty, 1990; Glazer and Rubinstein, 2004), we assume that the action that the biased sender wants the receiver to choose is common knowledge.² In our model, the biased sender wants the receiver to choose X , regardless of the state of the world. Hence, he has a strict incentive to induce an expected value of the quality of X as high as possible. Second, we assume that the receiver cannot observe the sender’s incentives, which can be either fully biased or perfectly aligned. These assumptions are more suitable descriptions for a number of problems in economics, including the examples described at the beginning of this paper and the problems described in persuasion games. Therefore, our analysis provides additional insights for the analysis of these problems. Other studies that consider uncertainty in the sender’s incentives and inspire our framework include dynamic models by Sobel (1985), Benabou and Laroque (1992), and Morris (2001). In Sobel (1985), the neutral sender is not a strategic player and will always mechanically report the truth. Sobel finds that the biased sender has an incentive to behave as if he is a neutral sender to build credibility and increase his future opportunity. Benabou and Laroque (1992) extend Sobel’s model by introducing noisy signals about states of nature which allow the biased sender to hide his own incentives repeatedly. Morris (2001) extends Benabou and Laroque’s model by allowing the neutral sender to be a strategic player. Morris finds that increased reputation concerns provide an incentive for the neutral sender to lie so that his truthful message has a larger impact on the receiver’s decision in later periods of the game.

These three dynamic cheap-talk games have two common features: (i) the authors assume binary state and message spaces in order to make their dynamic models tractable in

²For a further discussion of the relation between the results provided here and the literature on persuasion games, see Section 5.

repeated game settings, and (ii) the authors find that reputation can be strategically built through repeated information transmission. In these papers, building reputation provides strong incentives for message distortion regardless of the sender's incentives. What would be other incentives of lying if a cheap-talk game is not repeated? Morgan and Stocken (2003) analyze a model of stock recommendations in which a biased stock analyst can manipulate stock markets through strategic announcements. Their model is based on Benabou and Laroque's (1992) single-period game, but assumes continuous state and message spaces. A key difference between Morgan and Stocken's model and ours is that, in Morgan and Stocken's, the expected utility of the biased sender is not monotonically increasing in the induced expected value of the state. This occurs because, even though the biased stock analyst has an incentive to induce high market prices, he also has an incentive to provide valuable information to the public. This implies that there are some states of the world in which the induced stock price which is optimal for the biased sender is smaller than the upper bound of the state space. In contrast, the biased sender in our model always wants to induce an expected quality as high as possible, regardless the observed quality. Furthermore, Morgan and Stocken (2003) focus on pure reporting strategies, while in our analysis, as explained below, mixed revealing strategies play an important role.³

We study the perfect Bayesian Nash equilibria of the game. As it is standard in games of strategic information transmission, our game allows for infinitely many perfect Bayesian Nash equilibria. There exists a unique equilibrium in which the neutral sender always reveals the true quality of the good. We call this equilibrium *truth-telling*. In this equilibrium, the biased sender pretends to be neutral by not fully exaggerating, or even understating the quality he observed. In this case, we say that the biased sender uses a *feigning* strategy. The rationale for feigning strategies is fairly intuitive: if the biased sender always claimed that he observed the highest quality, then, upon receiving his message, the receiver would have a strong belief that the sender was biased and the message was not informative. In

³In an interesting extension to Morgan and Stocken's (2003) framework, Li and Madarász (2008) show that there are conditions under which both sender and receiver may be better off when the conflict of interest between them is private information than when it is common knowledge.

order to avoid this, the biased sender has to choose a message within a range of quality so that the chosen message does not lead to too much suspicion. All the messages the biased sender may send must induce the same expected quality from the point of view of the receiver; otherwise, there would be a possibility of arbitrage for the biased sender which is not allowed in equilibrium. For the same reason, this expected quality is the maximum quality that can be induced in each equilibrium.

We find that a higher probability that the sender is biased leads to a more intense feigning strategy. Formally, an increase in the probability that the sender is biased leads to a leftward probabilistic shift in the distribution of the messages sent by the biased sender. In other words, the initial distribution of messages sent by the biased sender first-order stochastically dominates the distribution of messages when the probability that the sender is biased has increased. This result follows from the fact that a higher probability that the sender is biased causes the receiver to be more suspicious about the type of the sender upon receiving a high quality message. Therefore, to compensate this effect, the biased sender has to assign less probability density on the high quality messages and, therefore, spread the density to the left. In terms of exaggeration, the more likely that the sender is biased, the less inflated the messages sent by the biased sender in the truth-telling equilibrium are. Furthermore, we show that the maximum expected quality induced in equilibrium is smaller when the probability that the sender is biased is higher.

Other equilibria which are outcome-equivalent to the truth-telling equilibrium also arise.⁴ These equilibria use different communication codes, but the amount of transmitted information is the same. In particular, we formalize the concept of *upholding equilibrium*, in which the biased sender sends the highest message and so does the neutral sender who has observed a quality which is higher than the maximum expected quality induced in the truth-telling equilibrium. The neutral sender who observes a quality which is lower than the maximum expected quality induced in the truth-telling equilibrium sends a message equal to the quality he has observed. This equilibrium has the intuitive interpretation that

⁴Two equilibria are outcome-equivalent if, for a given type of sender and realization of quality, they induce the same action of the receiver.

the biased sender always inflates his message as much as possible and so does the neutral sender who has observed a quality that is high enough. In other words, this equilibrium pools the biased sender with the neutral sender who has observed high quality. In this equilibrium, the codes used by the biased sender and the neutral sender who has observed high quality are inflating, nonetheless, it is not a deception by the neutral sender.⁵ This profile of revealing strategies seems to be consistent with the pattern of communication through letters of recommendation observed in highly competitive job markets. In such markets, if most recommendation letters are highly inflated, then a message which is not clearly fully supportive may be interpreted by the receiver as a negative signal. In this case even the sender whose objectives are aligned with those of the receiver may exaggerate about the quality. We call this strategy of the neutral sender an *upholding strategy* since the sender tries to support a person (or a good) that is worth supporting.

We provide a few examples of equilibria with coarser information transmission. Equilibria in which less information is transmitted are usually regarded as unlikely to be descriptive of actual behavior. Furthermore, experimental evidence shows that, in the lab, the amount of information transmission is consistent with the equilibrium predictions of the most informative equilibria.⁶ Altogether, this has motivated the search for equilibrium refinement criteria for information transmission games (see Rabin, 1990; Matthews, Okuno-Fujiwara, and Postlewaite, 1991; Farrell, 1993; Conlon, 1997; and Chen, Kartik, and Sobel, 2008). We adapt one of the refinement criteria introduced by Matthews *et al.* (1991), called equilibrium-announcement proofness, to the framework of this paper.⁷ We show that a

⁵There is a range of unused messages in this equilibrium. In order to avoid incentives to deviate, out-of-equilibrium beliefs of the receiver have to be specified so that departing from this revealing strategy would harm every player.

⁶Indeed, Dickhaut, McCabe, and Mukherji (1995), and Cai and Wang (2006) show that experimental subjects playing a version of Crawford and Sobel's model tend to overcommunicate, i.e., the correlation between states, messages, and actions are consistent with the most informative equilibrium or even higher. Sanchez-Pages and Vorsatz (2007) provide experimental evidence which suggests that the overcommunication phenomenon occurs because some experimental subjects have non-strategic concerns for telling the truth.

⁷Matthews *et al.* (1991) introduced three criteria for refinements, namely announcement proofness, strong

perfect Bayesian equilibrium is equilibrium-announcement proof if and only if it is outcome equivalent to the truth-telling equilibrium.

Most papers in the literature, including Crawford and Sobel (1982) and Morgan and Stocken (2003), identify partitional equilibria in which a sender sends an interval of messages and analyze the relationship between the degree of preference misalignment and the size of the interval. Crawford and Sobel (1982) show that a “babbling” equilibrium, in which no information is conveyed, always exists and the maximum amount of information conveyed in a partitional equilibrium depends on the degree of misalignment in the agents’ preferences. In the truth-telling equilibrium of our model, since the incentives of the neutral sender and the receiver are the same, the neutral sender communicates with full accuracy any observed quality which is lower than the expected quality induced by the biased sender. However, details about the levels of quality which are higher than the expected quality induced by the biased sender cannot be conveyed because all the messages claiming have observed such quality levels induce the same expected quality (equal to the expected quality induced by the biased sender). This result is more extreme than Morgan and Stocken’s (2003) where equilibrium prices are not fully responsive to good news when there is uncertainty about the sender’s incentives. In our model, the induced expected quality is not responsive *at all* as a consequence of the feigning strategy used by the biased sender. This strategy makes all the induced expected qualities to be exactly the same when the neutral sender observes a high quality. The feigning strategies that arise in the truth-telling equilibrium has no parallel in Morgan and Stocken (2003) because they do not consider mixed strategies of the biased sender. These strategies do not arise in Sobel (1985), Benabou and Laroque (1992), or Morris (2001) because the state and message spaces are binary in those models.

Olszewski (2004) analyzes a model in which the sender may be honest (mechanically reveals the truth) or may have reputation concerns of being honest and possibly other incentives. In his model, the receiver also has private information about the true state of

announcement proofness, and weak announcement proofness. Conlon (1997) shows that announcement proofness does not imply weak announcement proofness and suggests renaming this criterion as equilibrium-announcement proofness. Here we adopt the name suggested by Conlon (1997).

the world. Olszewski shows that if the sender has only reputation concerns, there is an equilibrium in which he always reveals his private information. This is the only equilibrium of the game when each signal he may observe fares differently with respect to the private information of the receiver. As a result, the receiver always observes a message exactly equal to the private information of the sender. This contrasts with the coarse communication for high levels of quality that we obtain in our model. Yet, truth-telling is no longer an equilibrium in his analysis when the sender, besides reputation concerns, has other interests even if those incentives are perfectly aligned with those of the receiver.

We provide a formal description of the game in the next section. In Section 3, we provide an analysis of perfect Bayesian equilibria. In Section 4, we show that every equilibrium-announcement proof equilibrium is outcome-equivalent to the truth-telling equilibrium. We conclude and discuss our results in Section 5.

2 The Model

Consider a model of a sender, a receiver, and two goods whose qualities are represented by two independent random variables, X and Y , which take values in the interval $[0, 1]$. We assume that X and Y are absolutely continuous. Let the realization, distribution function, support, and density function of X be denoted by x , F , \bar{F} , and f , respectively. Similarly, the realization, distribution function, support, and density function of Y are denoted by y , F_Y , \bar{F}_Y , and f_Y . We assume that $F(0) = F_Y(0) = 0$, $F(1) = F_Y(1) = 1$, and $f(x), f_Y(y) > 0$ for all $x, y \in [0, 1]$. Hence $E[X], E[Y] \in (0, 1)$. The receiver's utility (u_R) equals to the quality of the chosen good. Let $p \in [0, 1]$ be the probability that the receiver choose X . Therefore, given p , x , and y , the receiver's expected utility is

$$E[u_R(x, y)] = px + (1 - p)y. \tag{1}$$

We assume that the receiver can observe y , but not x , and the sender can observe x , but not y . After the sender observes x , he will send a message about the observed quality to the receiver. This message is not verifiable, and the incentives of the sender are not

observable by the receiver. Let S be a random variable, with realization s , which represents the sender's type. We assume that the sender takes on one of the two types, neutral ($s = 0$) and biased ($s = 1$) and the probability that the sender is biased is $\beta \in (0, 1)$. Furthermore, X , Y , and S are assumed to be mutually independent. If the sender is neutral, his utility (u_S) is the same as the utility of the receiver. In other words, the neutral sender's incentives are perfectly aligned with the receiver's. If the sender is biased his utility depends only on the receiver's choice: his utility is equal to 1 if the receiver chooses X and 0 if the receiver chooses Y . Since the biased sender does not care about how the receiver benefits from the good's quality, for simplicity, we focus our analysis on the biased sender's strategies that are not conditional on the quality he has observed.⁸ Therefore, the sender's expected utility can be written as

$$E[u_S(s, x, y)] = sp + (1 - s)[px + (1 - p)y]. \quad (2)$$

F , F_Y , and β are assumed to be common knowledge. The game proceeds as follows. Nature chooses x , y , and s . The sender observes x and s , and the receiver observes y . The sender sends a message m about x to the receiver. The receiver then forms a conditional expectation of S and X given m and chooses p to maximize his expected utility. Since p chosen by the receiver depends on m and y , we denote the strategy of the receiver by $p(m, y)$ and the expected utility of the receiver (upon observing m and y) can be written as

$$E[u_R(X, y)|m] = p(m, y)E[X|m] + (1 - p(m, y))y, \quad (3)$$

where $E[X|m]$ is the conditional expected value of X , conditional on observing a message m . The neutral sender chooses a message and sends it to the receiver so that the receiver makes a choice that maximizes her (and also his) expected utility. Even though the neutral sender cannot observe y , he knows that the receiver can. His choice of message determines $E[S|m]$, $E[X|m]$ and consequently affects p chosen by the receiver. If $E[X|m]$ is greater than y , the receiver will choose $p = 1$; if $E[X|m]$ is less than y , the receiver will choose

⁸Indeed, allowing the biased sender revealing strategy to depend on the observed quality enlarges the set of equilibria. However, this extension of the analysis does not provide further insights as biased senders, in equilibrium, always induce the same expected quality in order to avoid arbitrage.

$p = 0$; and, if $E[X|m] = y$, then p can take any value in $[0, 1]$. Since F_Y is absolutely continuous, the optimal choice of p for the receiver can be in $(0, 1)$ with probability zero. To avoid trivial technicalities, we assume that the receiver chooses $p = 1$ if $E[X|m] = y$. The biased sender aims to maximize the probability that the receiver chooses X , thus he will choose a message m that maximizes $E[X|m]$. Let the distribution function of messages sent by the neutral sender who has observed quality x be denoted by $F_{n|x}$ and the distribution function of messages sent by the biased sender be denoted by F_b . Now we define formally an equilibrium for this game.

Definition 1 *A perfect Bayesian equilibrium of the game is a triplet of (i) an action rule $p^*(m, y)$ for the receiver, (ii) a revealing strategy for the neutral sender defined by the family of distributions $(F_{n|x})_{x \in [0, 1]}$ with supports $\bar{F}_{n|x} \subseteq [0, 1]$ for all $x \in [0, 1]$, and (iii) a revealing strategy for the biased sender defined by the distribution F_b , with support $\bar{F}_b \subseteq [0, 1]$, such that:*

(a)

$$p^*(m, y) \in \arg \max_{p \in \{0, 1\}} \{pE[X|m] + (1 - p)y\} \quad (4)$$

for all $m \in [0, 1]$, and $y \in [0, 1]$,

(b)

$$m^* \in \arg \max_{m \in [0, 1]} E[p^*(m, Y)x + (1 - p^*(m, Y))Y] \quad (5)$$

for all $m^* \in \bar{F}_{n|x}$ and $x \in [0, 1]$.

(c)

$$m^* \in \arg \max_{m \in [0, 1]} E[X|m] \quad (6)$$

for all $m^* \in \bar{F}_b$, and beliefs are updated according to Bayes' rule whenever possible.

For each perfect Bayesian equilibrium there is a function which maps the realization of X , Y , and S to the choice of the receiver induced in the equilibrium. We call this function the *outcome-function* and two equilibria are said to be *outcome-equivalent* if their outcome-functions are the same.

We end this section by providing an example of equilibrium in a game in which both X and Y are uniformly distributed in $[0, 1]$. Consider a revealing profile in which the neutral

sender always sends a message truthfully, i.e., if $S = 0$, then $m = x$ for all $x \in [0, 1]$. Hence, the marginal density of the neutral sender's messages, denoted by $f_n(m)$, is also uniform, that is, $f_n(m) = f(m) = 1$ for all $m \in [0, 1]$. Let the biased sender's revealing strategy be given by the density function $f_b(m)$. A preliminary observation is that the revealing strategy of the biased sender must be atomless because $f_n(m)$ is atomless. If $f_b(m)$ has an atom at $m_0 \in [0, 1]$ and the receiver receives message m_0 , she will believe that the sender's type is biased and induce $E[X|m] = E[X]$. Since $E[X|m] = E[E[X|S, m]|m]$ where $E[X|s, m]$ is the expected value of X conditional on observing m and s , then $E[X|s, m] = (1 - s)m + sE[X]$, and $E[E[X|S, m]|m] = E[(1 - S)m + SE[X]|m] = (1 - E[S|m])m + E[S|m]E[X]$. Therefore, we have

$$E[X|m] = (1 - E[S|m])m + E[S|m]E[X]. \quad (7)$$

Using Bayes' rule we find $\Pr\{S = 1|m\} = \beta f_b(m) / [(1 - \beta)f(m) + \beta f_b(m)]$. Since $E[S|m] = \Pr\{S = 1|m\}$, then

$$E[X|m] = \frac{(1 - \beta)f(m)}{(1 - \beta)f(m) + \beta f_b(m)}m + \frac{\beta f_b(m)}{(1 - \beta)f(m) + \beta f_b(m)}E[X]. \quad (8)$$

In equilibrium, any message sent by a biased sender induces the same expected from the receiver's point of view. We denote this expected quality by c . Since X has the standard uniform distribution, then

$$c = \frac{(1 - \beta)}{(1 - \beta) + \beta f_b(m)}m + \frac{\beta f_b(m)}{(1 - \beta) + \beta f_b(m)}\frac{1}{2} \quad (9)$$

for any m sent by the biased sender. Thus

$$f_b(m) = \frac{(1 - \beta)}{\beta} \frac{m - c}{c - \frac{1}{2}}. \quad (10)$$

If $c < \frac{1}{2}$, then the support of the biased sender's strategy is bounded above by c because $f_b(m) \geq 0$. This could not be an equilibrium since, biased senders would have an incentive to deviate and reveal a message $m > c$, which would induce a higher expected quality for the receiver. Consider, then, $c > \frac{1}{2}$. Since $f_b(m) \geq 0$, we conclude that the support of the biased sender's revealing strategy is bounded below by c . Furthermore, any $m > c$ must be in the support; otherwise, sending m not in the support would induce an expected value for the

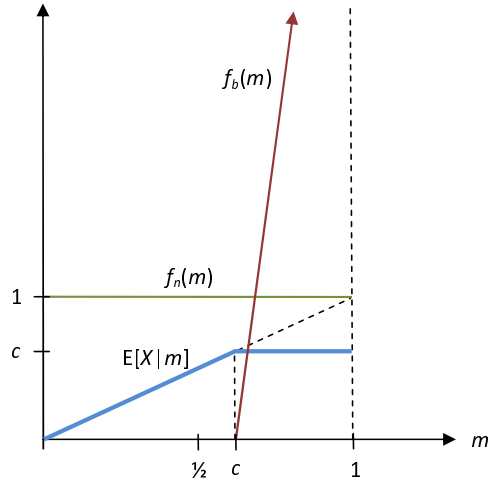


Figure 1: Revealing strategies and conditional expected quality of X when X and Y are uniform random variables.

receiver $E[X|m] > c$ and the biased sender would have an incentive to deviate. Therefore $\int_c^1 f_b(m)dm = 1$; or, equivalently,

$$\int_c^1 \frac{(1-\beta)m-c}{\beta} \frac{1}{c-\frac{1}{2}} dm = 1. \quad (11)$$

Rearranging (11) yields $c = \frac{1}{1+\sqrt{\beta}}$. Thus, the density of the revealing strategy for biased senders is given by

$$f_b(m) = \begin{cases} \frac{2}{\beta}(1+\sqrt{\beta})[(1+\sqrt{\beta})m-1], & \text{if } \frac{1}{1+\sqrt{\beta}} < m \leq 1; \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Figure 1 shows the marginal density functions of the messages sent by each type of sender, and the conditional expected quality for the receiver upon receiving a message m . Since X has the standard uniform distribution and the neutral sender always reports truthfully, the marginal density of messages sent by the neutral sender is given by $f_n(m) = 1$. The density function of messages sent by the biased sender is derived from (12). The conditional expectation given m is drawn in bold; $E[X|m] = c$ for $m > c$, otherwise, $E[X|m] = m$. We find that the biased sender chooses $f_b(m)$ so that any $m \in \bar{F}_b$ leads to highest expected quality. When $f(x)$ is uniform, the biased sender sends $m = 1$ with highest density and

no density for any $m < c$. For any $m \in (c, 1)$, the density increases with m at a constant rate. We can see that in order to convince the receiver to expect the highest quality, the biased sender cannot fully exaggerate about the good's quality by always sending $m = 1$. He assigns a positive density to some other messages that are lower than 1 but not lower than c . If the biased sender assigns too much density to some high message, it will raise too much suspicion for the receiver to believe that the received message is the true quality of X . The strategy that the biased sender uses to disguise himself as a neutral sender, $f_b(m)$ according to (12), is a feigning strategy. Finally notice that no neutral sender has incentives to deviate since those who observe $x \leq c$ induce an expected quality exactly equal to the one they observe and those who observe $x > c$ induce the expected quality c which is the highest expected quality they can induce.

3 Perfect Bayesian equilibria

In this section we analyze the perfect Bayesian equilibria of this game. First, we provide a lemma which shows that, in equilibrium, biased senders always induce an expected quality which is greater than the prior expected quality of the good X . Furthermore, all messages that may be sent by the biased sender induce the same expected quality which corresponds to the maximum expected quality that can be induced in such equilibrium.

Lemma 1 *Let $(p^*(m, y), (F_{n|x})_{x \in [0,1]}, F_b)$ be a perfect Bayesian equilibrium. There exists $c \geq E[X]$ such that (i) $c = E[X|m]$ for all $m \in \bar{F}_b$ and (ii) $c \geq E[X|m]$ for all $m \notin \bar{F}_b$.*

Proof. Conditions (i) and (ii) are non-arbitrage conditions which must hold in equilibrium so the biased sender does not deviate. Let M be the random variable which represents the messages sent in the equilibrium. To prove that $c \geq E[X]$, we argue by contradiction.

First notice that $\sup_{\substack{m \in \\ x \in [0,1]} \bar{F}_{n|x}} E[X|S = 0, m] \geq E[X]$. This follows from the fact that if $\sup_{\substack{m \in \\ x \in [0,1]} \bar{F}_{n|x}} E[X|S = 0, m] < E[X]$, then $E[X] = E[X|S = 0] = E[E[X|S = 0, M]|S = 0] < E[X]$ which is a contradiction.

Now, for all $m \in \bar{F}_b$, we have

$$\begin{aligned}
c &= E[X|m] \\
&= E[E[X|S, m]|m] \\
&= E[(1 - S)E[X|S = 0, m] + SE[X|S = 1, m]|m] \\
&= (1 - E[S|m])E[X|S = 0, m] + E[S|m]E[X].
\end{aligned}$$

Suppose $c < E[X]$, then it follows that $E[X|S = 0, m] < c < E[X]$ for all $m \in \bar{F}_b$. But, since $\sup_{m \in \bigcup_{x \in [0,1]} \bar{F}_{n|x}} E[X|S = 0, m] \geq E[X]$, there exists $m' \in \left(\bigcup_{x \in [0,1]} \bar{F}_{n|x} \right) \setminus \bar{F}_b$ such that $E[X|m', s = 0] > c$ and therefore the biased sender would deviate by sending m' instead of $m \in \bar{F}_b$. Therefore $c \geq E[X]$. ■

In Section 3.1, we turn our attention to the truth-telling equilibrium. We show that there is a unique equilibrium in which the neutral sender always sends a truthful message. In such an equilibrium, the biased sender uses a feigning strategy which mixes his messages within a range bounded below by the maximum quality induced in the equilibrium. The size of the message range (or the support of the revealing strategy) used by the biased sender is increasing in the probability that a sender is biased. Thus, when the probability that a sender is biased is larger, the support includes lower quality levels, and hence, the message is less inflated. Consequently, in a truth-telling equilibrium, the maximum induced expected quality is decreasing on the probability that the sender is biased. In Sections 3.2 and 3.3, we provide equilibria in which the sender and the receiver use different codes and equilibria that involve less information transmission, respectively.

3.1 Truth-telling equilibrium

In a truth-telling equilibrium, the neutral sender always reports the observed quality, i.e., if the sender is neutral, then $m = x$ for all $x \in [0, 1]$. In this case we say that the neutral sender adopts a *truth-telling* strategy. We find that there is a unique truth-telling equilibrium.

Theorem 1 *There exists a unique truth-telling equilibrium.*

Proof. First we provide a revealing strategy for the biased sender characterized by a revealing strategy $f_b(m)$ such that the neutral sender's truth-telling strategy and $f_b(m)$ induce a perfect Bayesian equilibrium.

Step 1: Bayesian updating and computation of $E[X|m]$.

Given a truth-telling strategy for the neutral sender and a revealing strategy with density $f_b(m)$ for the biased sender, then the expected quality of the good X is given by

$$\begin{aligned} E[X|m] &= E[E[X|S, m]|m] \\ &= E[(1-S)m + SE[X]|m] \\ &= m(1 - E[S|m]) + E[S|m]E[X]. \end{aligned}$$

$E[S|m] = \Pr\{S = 1|m\}$, thus, using Bayes' rule, we obtain

$$\Pr\{S = 1|m\} = \frac{\beta f_b(m)}{(1 - \beta)f(m) + \beta f_b(m)}.$$

Therefore,

$$E[X|m] = \frac{(1 - \beta)f(m)}{(1 - \beta)f(m) + \beta f_b(m)}m + \frac{\beta f_b(m)}{(1 - \beta)f(m) + \beta f_b(m)}E[X].$$

Step 2: The biased sender's feigning strategy.

In equilibrium all the messages sent by the biased sender induce the same expected quality from the point of view of the receiver. Denote this expected quality by c . Then

$$c = \frac{(1 - \beta)f(m)}{(1 - \beta)f(m) + \beta f_b(m)}m + \frac{\beta f_b(m)}{(1 - \beta)f(m) + \beta f_b(m)}E[X].$$

Solving for $f_b(m)$ we obtain that the density for the revealing function of biased advisors is given by

$$f_b(m) = \frac{1 - \beta}{\beta} \left(\frac{m - c}{c - E[X]} \right) f(m). \quad (13)$$

Since the marginal distribution of the neutral sender's message has the support $[0, 1]$, then Lemma 1 implies $\bar{F}_b := [c, 1]$.

Step 3: Existence and uniqueness of c .

The revealing strategy of the biased sender satisfies $\int_{[c,1]} f_b(m)dm = 1$, thus,

$$\int_{[c,1]} \frac{1 - \beta}{\beta} \left(\frac{m - c}{c - E[X]} \right) f(m)dm = 1,$$

or equivalently,

$$E[X] + \frac{1-\beta}{\beta} \left[\int_c^1 (m-c)f(m)dm \right] - c = 0. \quad (14)$$

Define the function $t(z) : [0, 1] \rightarrow \mathbb{R}$ by

$$t(z) := E[X] + \frac{1-\beta}{\beta} \left[\int_z^1 (m-z)f(m)dm \right] - z.$$

Notice that $t(0) = \frac{E[X]}{\beta} > 0$, $t(1) = E[X] - 1 < 0$. By the Leibniz rule, dt/dz can be obtained as

$$\frac{dt}{dz} = -\frac{1-\beta}{\beta} \int_z^1 f(m)dm - 1.$$

Therefore, the intermediate value theorem guarantees the existence of $z^* \in (0, 1)$ such that $t(z^*) = 0$. Since $dt/dz < 0$ on $(0, 1)$, $t(z) = 0$ only at $z = z^*$.

In order to finish the proof we need to show that the neutral sender would not deviate from the truth-telling strategy. Since $f_b(m) = 0$ for all $m \leq c$, for $x \leq c$, we have $E[X|m] = m$ and $m = x$, thus $E[X|m] = x$ and therefore the neutral sender does not deviate. Since $f_b(m) > 0$ for $m > c$, then $E[X|m] = c$ for all $m > c$; thus for $x > c$, we have $E[X|m] = c$. Since $c = \sup_{m \in [0,1]} E[X|m]$, the neutral sender who has observed $x > c$ does not deviate. ■

Theorem 1 shows that the truth-telling equilibrium is unique, i.e, given that the neutral sender always tells the truth, the revealing strategy chosen by the biased sender ($f_b(m)$) that yields a perfect Bayesian equilibrium is unique. Next we show some characteristics of $f_b(m)$: (i) $f_b(m)$ first-order stochastically dominates (FOSD) $f(m)$, and (ii) If $\beta' > \beta$, then $f_b(m)$ FOSD $f'_b(m)$, the biased sender's mixed strategy given β' . Furthermore, if c' is the expected value of quality upon receiving high messages given β' , then $c' < c$.

Remark 1 $f_b(m)$ FOSD $f(m)$.

Proof. From (13), we know that $f_b(m) \geq f(m)$ when

$$\frac{1-\beta}{\beta} \left(\frac{m-c}{c-E[X]} \right) \geq 1. \quad (15)$$

This condition is equivalent to $m \geq m^*$, where

$$m^* := \frac{c - \beta E[X]}{1 - \beta}. \quad (16)$$

Therefore in \bar{F}_b , $f_b(m)$ crosses $f(m)$ only once. ■

Remark 2 If $\beta' > \beta$, then $f_b(m)$ FOSD $f'_b(m)$ and $c' < c$.

Proof. From (14), call the LHS, $G(\beta, c)$. We know that $G(\beta, c) = 0$. Using IFT, we have

$$\frac{dc}{d\beta} = -\frac{G_\beta}{G_c} = -\frac{\int_c^1 (m-c)f(m)dm}{\beta(1-\beta)\int_c^1 f(m)dm + \beta^2} < 0. \quad (17)$$

Therefore, if $\beta' > \beta$, then $c' < c$. It follows that, for $m \in [c', c]$, $f'_b(m) \geq f_b(m)$. For $m \in (c, 1]$, let $g(m) := f'_b(m)/f_b(m)$ and notice that

$$g(m) = \frac{(1-\beta')\beta(c-E[X])(m-c')}{\beta'(1-\beta)(c'-E[X])(m-c)}. \quad (18)$$

Therefore

$$\frac{dg}{dm} = \frac{(1-\beta')\beta(c-E[X])(c'-c)}{\beta'(1-\beta)(c'-E[X])(m-c)^2} < 0$$

and thus $f'_b(m)$ and $f_b(m)$ cross only once. So $f_b(m)$ FOSD $f'_b(m)$. ■

3.2 Upholding equilibria

To derive a truth-telling equilibrium, we first assume that the neutral sender always tells the truth, and then we find a corresponding feigning strategy chosen by the biased sender that constitutes a perfect Bayesian equilibrium. In this section, we identify another class of equilibria where the biased sender always fully exaggerates, i.e. $\bar{F}_b = \{1\}$, and there is a revealing strategy for the neutral sender which constitutes an equilibrium. Here we denote revealing strategies for neutral senders which are deterministic by the function $\mu_n : [0, 1] \rightarrow [0, 1]$. For example, the truth-telling revealing strategy can be represented by the function $\mu_n(x) = x$, for all $x \in [0, 1]$. In particular, we focus on revealing strategies for the neutral sender who reveals the true quality up to a certain quality (threshold), and for qualities higher than the threshold, the neutral sender uses an inflated language by sending the message $m = 1$. In other words, for high enough qualities, neutral senders pool together all the qualities, therefore inducing the highest possible quality which is possible to induce in equilibria with this kind of revealing strategies.

Definition 2 *The neutral sender's strategy is upholding if there exists $\omega \in (0, 1)$ such that*

$$\mu_n(x) = \begin{cases} 1, & \text{if } x \geq \omega; \\ x & \text{otherwise.} \end{cases}$$

An equilibrium is upholding if μ_n is upholding.

Note that this definition is not limited to sending the message $m = 1$ when $x \geq \omega$. The definition can be generalized to any message that is greater than ω or any randomization in an interval that is a subset of $[\omega, 1]$. We assume that $E[X|m] = E[X]$ for all m which is not in $\left(\bigcup_{x \in [0,1]} \bar{F}_{n|x} \right) \cup \bar{F}_b$.

Theorem 2 *There exists a unique upholding equilibrium in which $\bar{F}_b = \{1\}$.*

Proof. Let $c := E[X|m = 1]$. Then, in equilibrium, $\mu_n(x) = 1$ for all $x > c$, otherwise $E[X|m = x] = x > c$ and the biased sender would deviate. Then

$$c = (1 - E[S|m = 1]) \frac{\int_c^1 x f(x) dx}{\int_c^1 f(x) dx} + E[S|m = 1] E[X]$$

Using Bayes' rule,

$$E[S|m = 1] = \frac{\beta}{\beta + (1 - \beta) \int_c^1 f(x) dx},$$

and, therefore,

$$c = \frac{(1 - \beta) \int_c^1 x f(x) dx}{\beta + (1 - \beta) \int_c^1 f(x) dx} + \frac{\beta}{\beta + (1 - \beta) \int_c^1 f(x) dx} E[X].$$

The last expression can be shown to be equivalent to (14). It follows that c uniquely exists.

■

In this case, the neutral sender whose objectives are aligned with those of the receiver may use an inflated language about the quality in order to avoid the receiver being misled by off-equilibrium-path beliefs. That is why this strategy of the neutral sender is said to be upholding. It is important to notice that this revealing strategy for the neutral sender, while being inflated, is not deceptive. The neutral sender, when sending the message $m = 1$, knows that the receiver does not interpret this message literally; in fact the expected value

induced by this message is less than the quality observed by the neutral sender. Indeed, this equilibrium is outcome-equivalent to the truth-telling equilibrium, yet only deterministic revealing strategies are used by both neutral and biased senders.

3.3 Less informative equilibria

In most cheap-talk models, it is easy to verify that a babbling equilibrium and other equilibria which allow for less information transmission than the truth-telling equilibrium exist. Here we provide an example. We revisit the example in Section 2 where both X and Y are uniformly distributed. Consider a strategy profile $(p^*(m, y), (F_{n|x})_{x \in [0,1]}, F_b)$ which is partitioned: (i) the neutral sender who has observed a quality above a certain level \hat{c} , to be specified below, randomizes his messages uniformly over the interval $[\hat{c}, 1]$, independently of the observed quality, (ii) the neutral sender who has observed a quality below \hat{c} randomizes his messages uniformly over the interval $[0, \hat{c}]$, independently of the observed quality, and (iii) the biased sender randomizes his messages uniformly over the interval $[\hat{c}, 1]$. We find that, to the receiver, messages in $[0, \hat{c}]$ induce an expected quality of $\frac{\hat{c}}{2}$, while messages in $[\hat{c}, 1]$ induce an expected quality denoted by $\gamma(\hat{c})$, which is a function of \hat{c} . Specifically, $\gamma(\hat{c}) := E[X|m]$ for $m \in [\hat{c}, 1]$. We can write

$$\begin{aligned} \gamma(\hat{c}) &= \frac{(1 - \beta) \int_{\hat{c}}^1 x f(x) dx + \frac{1}{2} \beta}{(1 - \beta) \int_{\hat{c}}^1 f(x) dx + \beta} \\ &= \frac{(1 - \beta) \int_{\hat{c}}^1 x dx + \frac{1}{2} \beta}{(1 - \beta) \int_{\hat{c}}^1 1 dx + \beta} \\ &= \frac{1(1 - \beta)(1 - \hat{c}^2) + \beta}{2(1 - \beta)(1 - \hat{c}) + \beta}. \end{aligned} \tag{19}$$

If the neutral sender who has observed $x > \hat{c}$ reveals a message in $[\hat{c}, 1]$ and the neutral sender who has observed $x < \hat{c}$ reveals a message in $[0, \hat{c}]$, it follows that the former prefers inducing $\gamma(\hat{c})$ while the latter prefers inducing $\frac{\hat{c}}{2}$. Figure 2 shows the induced expected qualities in both intervals for every possible bipartition with cutoff \hat{c} . We find that $\gamma(\hat{c}) > \frac{\hat{c}}{2}$ for all $\hat{c} \in (0, 1)$. Since Y is uniformly distributed, it is intuitive to formalize that, for the neutral sender who has observed \hat{c} to be indifferent between randomizing in each interval,

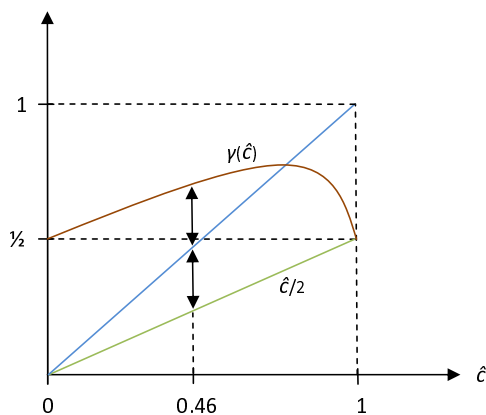


Figure 2: Revealing strategies and conditional expected quality of X under standard uniform distribution

it is necessary and sufficient that $\gamma(\hat{c}) - \hat{c} = \hat{c} - \frac{\hat{c}}{2}$, or equivalently, $\gamma(\hat{c}) = \frac{3}{2}\hat{c}$. Thus, using (19), we find that \hat{c} satisfies

$$\frac{1}{2} \frac{(1-\beta)(1-\hat{c}^2) + \beta}{(1-\beta)(1-\hat{c}) + \beta} = \frac{3}{2}\hat{c}.$$

This expression has only one root in $[0, 1]$, which is given by $\hat{c} = \frac{3 - \sqrt{9 - 8(1-\beta)}}{4(1-\beta)}$. For example, if $\beta = 0.1$, then $\hat{c} \approx 0.46$. Thus, in this equilibrium, the neutral sender who has observed a quality greater than 0.46 and the biased sender randomizes his messages uniformly in the interval $[0.46, 1]$ and each of these messages induces an expected quality of 0.69. On the other hand the neutral sender who has observed a quality lower than 0.46 randomizes his messages uniformly in the interval $[0, 0.46]$ and each of these messages induces an expected quality of 0.23.

The example with two pools of messages can be easily generalized to $N + 1$ pools of messages, for $N = 1, 2, \dots$, where the biased sender pools with the neutral sender who has observed the highest quality.⁹ In other words, information transmission can be arbitrarily detailed within the interval of qualities corresponding to the first N pools of the partition,

⁹In this case, the highest pool of senders are those who observe qualities in the interval $[\hat{c}, 1]$ with $\hat{c} = \frac{3 - \frac{N-1}{N} - \sqrt{(3 - \frac{N-1}{N})^2 - 4(1-\beta)(2 - \frac{N-1}{N})}}{2(1-\beta)(2 - \frac{N-1}{N})}$. This expression for \hat{c} can be obtained by identifying the quality \hat{c} which leaves a neutral sender indifferent between pooling in the group of neutral senders who observe highest qualities and inducing an expected quality of $\gamma(\hat{c})$, and pooling in the group of neutral senders who

with the highest pool bounded above by \hat{c} . As $N \rightarrow \infty$, $\hat{c} \nearrow c = \frac{1}{1+\sqrt{\beta}}$, and the expected value induced by each type of sender, for all $x \in [0, 1]$, converges to the induced expected value of the truth-telling equilibrium.

We do not attempt to characterize all the perfect Bayesian equilibria of this game. Several refinement criteria in the literature suggest that the notion of perfect Bayesian equilibrium allows for too many uninformative equilibria which are implausible. In the next section, we analyze a refinement criterion and identify the equilibria that satisfy it. In this characterization, the truth-telling equilibrium described above plays an important role.

4 Equilibrium announcements

The multiplicity of equilibria makes it difficult to use the perfect Bayesian equilibrium concept to provide predictions about how people behave. Moreover, equilibria which predict little or no information transmission at all, such as the babbling equilibrium, often are regarded as unlikely to provide a useful description of players' messages and actions (see Matthews *et al.*, 1991; Farrell, 1993; Cai and Wang, 2006). Some of the refinement criteria studied in the literature help select equilibria, especially those which involve more information transmission. Matthews *et al.* (1991) provide one of such criteria, namely equilibrium-announcement proofness (EAP). In this section, we adapt the EAP concept to the game analyzed in this paper. Matthews *et al.* (1991) motivate its construction by imposing a number of rationality restrictions. This motivation captures and carries further the notion of neologism proofness behind Farrell's (1993) refinement criterion. Furthermore, Conlon (1997) shows that if there is a Pareto dominant equilibrium, then it will be selected by EAP. It turns out that this refinement criterion is particularly tractable in our model: as we will see below, it is easy to show that the truth-telling equilibrium is EAP. Furthermore, we show that any EAP equilibrium is outcome equivalent to the truth-telling equilibrium.

observe the second group of highest qualities and inducing an expected quality of $\frac{\hat{c} + \frac{N-1}{N}\hat{c}}{2}$. Therefore the expression for \hat{c} provided above corresponds to the root of $\gamma(\hat{c}) - \hat{c} = \hat{c} - \frac{\hat{c} + \frac{N-1}{N}\hat{c}}{2}$. The other N pools of the partition are equally sized and cover all the range $[0, \hat{c}]$.

Finally, the equilibrium selected by EAP, i.e., the truth-telling equilibrium, also satisfies an adapted version of the seemingly less restrictive NITS criterion studied by Chen *et al.* (2008).

The concept of equilibrium-announcement proofness relies on the assumption that senders can make an argument which may convince the receiver to do something different from what she is prescribed to play on the equilibrium path given the type of the sender and observed quality. Making such an argument obliges the sender to describe a full profile of messages and announcements which every type of sender would do. This profile, called *announcement strategy*, distinguishes between the senders who announce this profile and those who do not, and follow the equilibrium revealing prescription. The senders who announce this profile are called deviants. Along with the profile, a deviant sender reveals a specific message which may be interpreted as a claim about the quality he has observed. The whole set of information, including both the specific message and the described announcement strategy is called an *announcement*. If the announcement is compelling to the receiver, the expected quality is computed consistently with the announcement strategy and the Bayes rule. The following definition formalizes this concept.

Definition 3 An announcement strategy is a function $\delta : D \rightarrow \Delta([0, 1])$ with $D \subseteq (\{0, 1\} \times [0, 1])$. An announcement is a pair $\langle a, \delta \rangle$ where δ is an announcement strategy and $a \in \delta(D)$ with $\delta(D) := \bigcup_{(s,x) \in D} \overline{\delta(s, x)}$ and $\overline{\delta(s, x)}$ is the support of $\delta(s, x)$. An announcement $\langle a, \delta \rangle$ is believed if

$$E[X|\langle a, \delta \rangle] = \frac{(1 - \beta) \int_{(0,x) \in D} \delta(a|0, x) f(x) x dx + \beta \delta(a|1) E[X]}{(1 - \beta) \int_{(0,x) \in D} \delta(a|0, x) f(x) dx + \beta \delta(a|1)},$$

where $\delta(a|0, x)$ is the probability distribution of the announcement a conditional on the type of sender being $s = 0$ and the quality of the good being x ; and $\delta(a|1)$ is the probability of the announcement a conditional on the type of sender being $s = 1$.

In order to be compelling to the receiver, announcement strategies (and the corresponding announcements) have to satisfy some credibility criteria. Matthews *et al.* (1991) suggest a credibility criterion which forces the receiver to have rational beliefs in the sense that she,

upon hearing an announcement, realizes that the sender knows the class of announcement strategies that are believable, and that he could have obtained the equilibrium payoffs by sticking to the equilibrium profile. Therefore, the information deviant senders want to convey must induce actions that these senders prefer to those in the equilibrium path or those actions induced by other credible announcement strategies. Lastly, in order to avoid deviations from this announcement strategy, there must exist some announcement strategy for the non-deviants such that the whole profile of announcements form a perfect Bayesian equilibrium in which the specific set of messages sent by the deviants and the non-deviants are disjoint.

Altogether, credibility may be summarized in Conditions 1-5 below (in these conditions, $p^*(m, y)$ and $E[X|m]$ refer to those of the equilibrium which is being tested against the EAP criterion). The first condition (C1) requires that all the deviant senders weakly prefer the outcome of the announcement (if believed) to the outcome of the equilibrium and, at least one of them, strictly prefers the outcome of the announcement (if believed) to the outcome of the equilibrium.

Condition 1 *Neutral senders:*

Let $p(\langle a, \delta \rangle, y) \in \arg \max_{p \in \{0,1\}} \{pE[X|\langle a, \delta \rangle] + (1-p)y\}$. Then

$$E[p(\langle a, \delta \rangle, Y)x + (1-p(\langle a, \delta \rangle, Y))Y] \geq \max_{m \in [0,1]} E[p^*(m, Y)x + (1-p^*(m, Y))Y] \quad (20)$$

for all $(0, x) \in D$ and $a \in \delta(0, x)$ with strict inequality for some $(0, x) \in D$ and $a \in \overline{\delta(0, x)}$.

Biased senders:

$$E[X|\langle a, \delta \rangle] \geq \max_{m \in [0,1]} E[X|m] \quad (21)$$

for all $a \in \overline{\delta(1, x)}$ and $(1, x) \in D$.

And either (20) or (21) is satisfied with strict inequality for some $(s, x) \in D$ and $a \in \overline{\delta(s, x)}$.

The second condition (C2) requires that all the non-deviant senders weakly prefer the outcome of the equilibrium to the outcome of the announcement if believed by the receiver.

Condition 2 *Neutral senders:*

$$E[p(\langle a, \delta \rangle, Y)x + (1 - p(\langle a, \delta \rangle, Y))Y] \leq \max_{m \in [0,1]} E[p^*(m, Y)x + (1 - p^*(m, Y))Y]$$

for all $(0, x) \in D^c$ and $a \in \delta(D)$.¹⁰

Biased senders:

$$E[X|\langle a, \delta \rangle] \leq \max_{m \in [0,1]} E[X|m] \text{ for all } (1, x) \in D^c \text{ and } a \in \delta(D).$$

The third condition (C3) requires that within deviant types, each of them prefers sending messages according to the announcement strategy prescription for that type to sending messages according to the announcement strategy prescription for another deviant type.

Condition 3 *Neutral senders:*

$$E[p(\langle a, \delta \rangle, Y)x + (1 - p(\langle a, \delta \rangle, Y))Y] \geq E[p(\langle \hat{a}, \delta \rangle, Y)x + (1 - p(\langle \hat{a}, \delta \rangle, Y))Y]$$

for all $(0, x) \in D$, $a \in \overline{\delta(0, x)}$, and $\hat{a} \in \delta(D) \setminus \overline{\delta(0, x)}$.

Biased senders: $E[X|\langle a, \delta \rangle] \geq E[X|\langle \hat{a}, \delta \rangle]$ for all $(1, x) \in D$, $a \in \overline{\delta(1, x)}$, and $\hat{a} \in \delta(D) \setminus \overline{\delta(1, x)}$.

The fourth condition (C4) requires that there exists a revealing profile for the non-deviant senders such that, along with the revealing strategies of the deviant senders, this profile forms a perfect Bayesian equilibrium in which the set of messages sent by the deviant senders and the non-deviant senders are disjoint.

Condition 4 *If the set of non-deviants is not empty, there is an announcement strategy $\delta^* : D^c \rightarrow \Delta([0, 1])$ such that $\delta(D) \cap \delta^*(D^c) = \emptyset$ and the profile*

$$(p^*(m, y), (\delta(0, x))_{(0,x) \in D} \cup (\delta^*(0, x))_{(0,x) \in D^c}, (\delta(1, x))_{(1,x) \in D} \cup (\delta^*(1, x))_{(1,x) \in D^c}),$$

where the receiver holds Bayesian beliefs wherever it is possible, is a perfect Bayesian equilibrium.

¹⁰The complement of D is defined with respect to the type space $\{0, 1\} \times [0, 1]$, i.e., $D^c := (\{0, 1\} \times [0, 1]) \setminus D$.

The last condition (C5) requires that no deviant sender prefers any other announcement strategy in which he is also a deviant sender and C1, C2, C3, and C4 are satisfied.

Condition 5 *If the announcement strategy δ' (with domain D') also satisfies C1-C4 with respect to the equilibrium $(p^*(m, y), (F_{n|x})_{x \in [0,1]}, F_b)$, then*

Neutral senders:

$$E[p(\langle a, \delta \rangle, Y)x + (1 - p(\langle a, \delta \rangle, Y))Y] \geq E[p(\langle a', \delta' \rangle, Y)x + (1 - p(\langle a', \delta' \rangle, Y))Y]$$

for all $(0, x) \in D \cap D'$, $a \in \overline{\delta(0, x)}$, and $a' \in \overline{\delta'(0, x)}$

Biased senders:

$E[X|\langle a, \delta \rangle] \geq E[X|\langle a', \delta' \rangle]$ for all $(1, x) \in D \cap D'$, $a \in \overline{\delta(1, x)}$, and $a' \in \overline{\delta'(1, x)}$.

Definition 4 *An announcement strategy δ (with domain D) and the corresponding announcements $\langle a, \delta \rangle$ are credible related to the equilibrium $(p^*(m, y), (F_{n|x})_{x \in [0,1]}, F_b)$ if they satisfy C1, C2, C3, C4, and C5.*

A perfect Bayesian equilibrium is EAP if there is no announcement satisfying C1, C2, C3, C4, and C5.

Definition 5 *An equilibrium $(p^*(m, y), (F_{n|x})_{x \in [0,1]}, F_b)$ is equilibrium-announcement proof (EAP) if there is no credible announcement strategy related to it.*

This following result establishes that the truth-telling equilibrium is EAP. Furthermore, every EAP equilibrium is outcome-equivalent to the truth-telling equilibrium.

Theorem 3 *The truth-telling equilibrium is EAP. A perfect Bayesian equilibrium is EAP if and only if it is outcome-equivalent to the truth-telling equilibrium.*

The proof of this result relies on the fact that a credible announcement strategy requires some senders to be better off than in the truth-telling equilibrium by making an announcement. However, the proof reveals that it is not possible to find a profile such that a sender can have a higher expected utility than in the truth-telling equilibrium. This occurs because the neutral sender who has observed quality $x \leq c$ already maximizes his utility in

the truth-telling equilibrium by inducing $E[X|m] = x$, and it is impossible for the neutral sender who has observed $x > c$ and the biased sender, to raise the induced expected quality beyond c . The second part of the result follows from the fact that the same argument applies to every equilibrium which is outcome equivalent to the truth-telling equilibrium. Furthermore, for any equilibrium which is not outcome equivalent to the truth-telling equilibrium, a revealing profile corresponding to the truth-telling equilibrium can be announced and such an announcement satisfies C1–C5.

Proof. First we prove that the truth-telling equilibrium is EAP. From the proof of the theorem that shows existence and uniqueness of the truth-telling equilibrium, and, in particular, from the expression for $f_b(m)$, we know that $c > E[X]$. For every $x \leq c$, $E[X|m] = x$. Therefore, the neutral sender with $x \leq c$ reaches his maximum utility in the truth-telling equilibrium. Suppose there exists a credible announcement strategy δ which satisfies C1–C5. Since every type $(0, x)$ such that $x \leq c$ is reaching his maximum expected payoff, we conclude that the announcement strategy δ must provide a strictly higher payoff either to a type $(0, x)$ such that $x > c$, or a strictly higher payoff to a type $(1, x)$. In either case, this implies that there is a non empty set of announcements $\langle a, \delta \rangle$ such that $E[X|\langle a, \delta \rangle] > c$. Let $c' := \sup_{a' \in \delta(D)} E[X|\langle a', \delta \rangle]$ and

$$\bar{A} := \{a \in \delta(D) : E[X|\langle a, \delta \rangle] = \sup_{a' \in \delta(D)} E[X|\langle a', \delta \rangle]\}.$$

C3 imposes that every type $(0, x)$ such that $x \geq c'$ and every type $(1, x)$ makes an announcement $\langle a, \delta \rangle$ such that $a \in \bar{A}$. Suppose only types $(0, x)$ such that $x \geq c'$ and every type $(1, x)$ make announcements in \bar{A} . Then

$$E[X|\langle a, \delta \rangle] = \frac{(1 - \beta) \int_{c'}^1 f(x)xdx + \beta E[X]}{(1 - \beta) \int_{c'}^1 f(x)dx + \beta}$$

for all $a \in \bar{A}$. Recall that in the truth-telling equilibrium we have $c = \frac{(1 - \beta) \int_c^1 f(x)xdx + \beta E[X]}{(1 - \beta) \int_c^1 f(x)dx + \beta}$, which follows from (14). Now consider the function $\gamma : [0, 1] \rightarrow \mathbb{R}$ given by

$$\gamma(z) = \frac{(1 - \beta) \int_z^1 f(x)xdx + \beta E[X]}{(1 - \beta) \int_z^1 f(x)dx + \beta}.$$

The first derivative of $\gamma(z)$ is given by

$$\gamma'(z) = \frac{(1 - \beta)f(z) \left[(1 - \beta) \int_z^1 f(x)(x - z)dx + \beta(E[X] - z) \right]}{\left[(1 - \beta) \int_z^1 f(x)dx + \beta \right]^2},$$

thus the sign of $\gamma'(z)$ is the same as the sign of $(1 - \beta) \int_z^1 f(x)(x - z)dx + \beta(E[X] - z)$. From the proof of the existence and uniqueness of the truth-telling equilibrium, $(1 - \beta) \int_z^1 f(x)(x - z)dx + \beta(E[X] - z) < 0$ for all $z > c$. It follows that $c' < c$, which is a contradiction. So far we have assumed that types $(0, x)$ make announcements $\langle a, \delta \rangle$ with $a \in \bar{A}$ only if $x \geq c'$. If types $(0, x)$ with $x < c'$ too make announcements $\langle a, \delta \rangle$ with $a \in \bar{A}$, then we would have

$$E[X|\langle a, \delta \rangle] < \frac{(1 - \beta) \int_{c'}^1 f(x)xdx + \beta E[X]}{(1 - \beta) \int_{c'}^1 f(x)dx + \beta}$$

for $a \in \bar{A}$ and this would lead to the contradiction $E[X|\langle a, \delta \rangle] < c$ as well. Thus no credible announcement strategy can be made related to the truth-telling equilibrium. The same arguments show that every equilibrium which is outcome-equivalent to the truth-telling equilibrium is EAP.

Now, we prove that any equilibrium which is not outcome-equivalent to the truth-telling equilibrium is not EAP. In such an equilibrium, an announcement strategy δ in which every neutral sender who observes x announces $\langle x, \delta \rangle$ with probability one and biased senders make announcements $\langle m, \delta \rangle$ according to the density $f_b(m)$ described in (13) with c given by the root of (14) is credible related to such an equilibrium. It is straightforward checking that such an announcement strategy satisfies C1, C2, C3, and C5. Finally C4 is readily verified since the set of non-deviants is empty. Thus, such an equilibrium cannot be EAP.

■

The previous result suggests that the truth-telling equilibrium stands out from other equilibria—beyond the fact of restricting the neutral senders to behave according to a natural focal point and ubiquitous social norm. This equilibrium is selected according to the equilibrium-announcement proofness criterion and any equilibrium which satisfies this criterion is outcome equivalent to the truth-telling equilibrium. For example, this refinement criterion sets aside the truth-telling equilibrium from all the partitional equilibria described

in the uniform example of Section 3.3. None of these equilibria is EAP as they are not outcome equivalent to the truth-telling equilibrium. However, the upholding equilibrium we described in Section 3.2 is outcome equivalent with the truth-telling and thus is EAP.

As mentioned above, the literature offers several other refinement criteria for cheap talk games. Although less formally, we now discuss how the truth-telling equilibrium fares with a few of these criteria. Farrell (1993) introduces a refinement criterion based on the concept of *neologisms*. Neologisms are statements senders may make to identify their self, or themselves in the case a group of senders takes part in a given neologism. Credibility of a neologism requires the senders who sends a neologism to prefer the expected quality induced by the neologism to the quality which would be induced in the equilibrium (see Matthews *et al.*, 1991, page 255). However, as shown in the proof of the Theorem 3, it is not possible for any sender in the truth-telling equilibrium to change the expected quality induced by this equilibrium to another induced expected quality which is preferred to this sender. For the neutral sender who has observed quality $x \leq c$, this is impossible because he is inducing an expected quality which is exactly equal to the observed quality, and that is exactly his objective. For the neutral sender who has observed quality $x > c$, it is also impossible to increase the expected induced quality, because of the same argument in the proof of the Theorem 3. Thus, the truth-telling equilibrium is neologism proof as well.

We now analyze the implications of the refinement criterion of Chen *et al.* (2008) for our equilibrium analysis. Chen *et al.* (2008) propose a criterion called NITS (for no incentives to separate) to select among equilibria in the model of Crawford and Sobel (1982). The equilibria that satisfy it are ones that satisfy a simple condition in the equilibrium payoffs. In the context of our model, this condition is translated to the requirement that the sender who has observed the lowest quality is at least as well-off in the equilibrium as when he can accurately induce the observed quality. Chen *et al.* (2008) show that NITS selects among equilibria and under certain conditions, the only equilibrium which satisfies NITS is the most informative equilibrium. In the truth-telling equilibrium of our game, there is no incentive to separate because the neutral sender who observes the lowest quality, in equilibrium, induces an expected quality equal to the lowest quality, and the biased sender

does not have an incentives to separate either. Therefore, the truth-telling equilibrium satisfies NITS. Finally, our example of the partitional equilibria in Section 3.3 allows us to illustrate further the selecting power of NITS. In particular, none of the partitional equilibria with $N = 1, 2, \dots$ partitions satisfies NITS. In any of these equilibria, the neutral sender who observes $x = 0$ would prefer the receiver to expect the quality to be equal to zero. However in all those equilibria, the neutral sender who observes $x = 0$ induces an expected quality greater than zero. Thus, none of these equilibria satisfy NITS.

5 Summary and Discussion

Our main results are Lemma 1 and Theorem 2. Lemma 1 establishes that, in any perfect Bayesian equilibrium, all the messages sent by the biased sender induce the same expected quality from the point of view of the receiver. This expected quality is the highest expected quality which may be induced by any message in that equilibrium. This result follows immediately from the fact that the biased sender only cares about maximising the expected quality he induces. Perhaps more importantly, this expected quality is greater than the unconditional expected quality and strictly less than the maximum quality. This implies that there is an upper bound to the quality that can be communicated in equilibrium.

We also provide different equilibria of the game. In one of these equilibria (Theorem 2), the neutral sender reveals the observed quality truthfully, i.e., his message is always the quality he has observed. It may not be surprising that such a result emerges in a pure common interest strategic situation, yet this is much less obvious in a context where the sender is suspected to be biased. Theorem 2 establishes that truth-telling for neutral senders arises in equilibrium, even though the receiver *knows* that the sender may be biased. This, in our analysis, occurs only because the preferences of the neutral sender and the receiver are perfectly aligned and without assuming intrinsic preferences for truth-telling (or cost of lying). We also show that in this equilibrium the messages sent by the biased sender cannot be fully inflated, which means that he does not always report the maximum quality. Instead, the biased sender uses a feigning strategy by spreading out the range of messages

he may send. However, the support of this mixed strategy is bounded below by the expected quality induced in the equilibrium, which is strictly higher than the unconditional expected quality.

The other class of equilibria we analyze is characterized by inflated messages from the biased sender, inflated messages from the neutral sender who observed a quality level above a certain threshold, and truthful messages from the neutral sender who observed a quality level below the threshold. This result is consistent with observed patterns of communication transmission in recommendation letters of highly competitive job-markets. In such markets, if most recommendation letters are written very generously, then a message which is not clearly fully supportive may be interpreted by the receiver as a negative signal. In our model, even the neutral sender, whose objectives are perfectly aligned with those of the receiver, may exaggerate about the quality of the good in order to avoid an undesired choice of the receiver. This incentive to exaggerate occurs only when the observed quality of the recommended good is high enough. We call such strategy upholding because the neutral recommender uses inflated language to induce the recruiter to choose the recommended person. However, the messages sent by the neutral sender are not deceptive as the quality they observe when they exaggerate is greater than the induced expected quality.

It is worth comparing our results to those in the literature of costly talk (see Ottaviani and Squintani, 2006; Kartik, Ottaviani, and Squintani, 2007). Kartik *et al.* (2007) study a general framework in which a biased sender sends a message to a group of audience consisted naive and strategic receivers. They show that a fully separating equilibrium exists when the state and message spaces are unbounded above. They also show that strategic receivers are always able to infer exactly the state of the world, despite the inflated language used by the sender, while naive receivers are deceived as they believe the received message is the true state of the world. In contrast, we show that when there is a sender whose type is unobservable, it is possible to have an equilibrium where the neutral sender always reveals the truth. However, this equilibrium fails to be fully separating as the potential presence of a biased sender makes high messages unreliable. Therefore, in our setting, all high qualities induce the same expected quality in equilibrium.

In a related work, Ottaviani and Squintani (2006) analyze a model similar to the one in Kartik *et al.* (2007), but with a bounded state space. Similar to our paper, they find that fully revealing is possible for low states of the world. For the top range of the state space, they identify a partitioned equilibrium, while we show that information transmission on the top range of qualities is totally uninformative. It is important to notice a fundamental difference between these two models with naive receivers and our model which follows directly from the assumptions. In the aforementioned papers, since the preferences of the senders are common knowledge, only naive receivers are victims of deception. In our analysis, even though the receiver is fully strategic, she is expected to be deceived by the biased sender because, in equilibrium, the biased sender can always persuade the receiver to believe that the expected quality given his message is higher than the unconditional expected quality.

The fact that the neutral sender in our model is a strategic player, i.e., he does not necessarily have to be honest, is also assumed by Morris (2001). However, he focuses on the dynamics of information transmission and the sender's reputation. Morris finds that both sender types have a short-term incentive to deceive when they are concerned about reputation. In this paper, neutral senders do not have an incentive to deceive the receiver since there are no benefits from gaining reputation. In the upholding equilibrium in which messages by the neutral senders are inflated, the receiver is not deceived by the neutral sender. In fact, in such equilibrium, given the set of equilibrium beliefs, the neutral sender would have misled the receiver if he revealed the true quality.

Our description of the preferences of the biased sender resembles the description of the sender in persuasion games. In the literature on persuasion, usually it is common knowledge that the sender wants to persuade the receiver to make a certain choice and the message may be fully or partially verifiable. Dziuda (2008) considers conditions in which revealing information that, at face value, does not seem to favor the interests of the biased sender. In our cheap talk model, the biased sender can not credibly reveal any information. However, the feigning strategies we obtain play a similar role by raising the posterior belief of the receiver that the sender may be neutral.

There is a close connection between Theorem 1 in our paper and the proposition in

Wolinsky (2003). He considers a sender-receiver game where the receiver must decide (within a finite set) the amount of money to invest in a given project. The sender can be either fully biased to the left or fully biased to the right, i.e., he may prefer the receiver to minimize or maximize her investment. The sender's private information is his type of bias and the state of the world, which corresponds to the amount of money to invest that would be optimal for the receiver. In his model, the receiver can realize that the sender is lying only if the message of the sender is greater than the true state. This leads to the result that every equilibrium of his model is equivalent to an equilibrium in which the right-biased sender always reveals truthfully the observed state of the world, and the left biased sender uses a mixed strategy in the bottom range of messages. This result is analogous to our truth-telling equilibrium and feigning strategies. In contrast to our results, the verifiability assumption rules out equilibria with less information transmission such as the babbling equilibria or those we described in Section 3.3.

Finally, it seems natural to extend our model to a sender-receiver game with multiple senders. Consider a situation where there is one position to be filled and n applicants have applied. Each applicant's quality is independent from one another's, observed by the applicant's professor, and unobserved by the recruiter. If each professor sends a letter of recommendation to the recruiter to report the quality of the applicant, there will be n senders, each sends a message about the quality of his student to the recruiter. The analysis of such a game would be analogous to the analysis of the game described in this paper.

Acknowledgments

We thank Gary Charness, John Conlon, Brit Grosskopf, Bill Neilson, Paul Pecorino, Rajiv Sarin, and seminar participants at University of Alabama, University of Alicante, University of Mississippi, University of Tennessee, and Texas A&M University for valuable comments. We also thank Phillip Stocken for providing us with unpublished parts of his work with John Morgan. Paan Jindapon acknowledges financial support from the Culverhouse College of Commerce and Business Administration (Summer 2008). Carlos Oyarzun acknowledges

financial support from the Ministerio de Ciencia y Tecnologia, Spain, (SEJ2007-62656 and JC2008-00289) and hospitality from the Department of Economics at Texas A&M University.

References

- [1] Benabou, R. and Laroque, G. "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, Vol. 107 (1992), pp. 921-958.
- [2] Cai, H.B. and Wang, J.T. "Overcommunication in strategic information transmission games", *Games and Economic Behavior*, 56 (2006), 7-36.
- [3] Chakraborty, A. and Harbaugh, R. "Comparative Cheap Talk," *Journal of Economic Theory*. Vol. 132 (2007), pp. 70-94.
- [4] Chen, Y., N. Kartik and J. Sobel (2008): "Selecting Cheap-Talk Equilibria," *Econometrica*, Vol. 76, No. 1, 117-136.
- [5] Conlon, J.R. "Coordination and Weak Announcement Proofness: Two Comments on Refining Cheap-Talk Equilibria," *Journal of Economic Theory*. Vol. 72 (1997), pp. 220-224.
- [6] Crawford, V.P. and Sobel, J. "Strategic Information Transmission," *Econometrica* Vol. 50 (1982), pp. 1431-1451.
- [7] Dickhaut, J., McCabe, K. and Mukherji, A. "An experimental study of strategic information transmission," *Economic Theory* 6, (1995), pp. 389-403.
- [8] Dziuda, W. "Strategic Argumentation," typescript, MEDS, Northwestern University, 2008.
- [9] Farrell, J. "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior* Vol. 5 (1993), pp. 514-531.

- [10] Fishman, M. J. and K. M. Hagerty “The Optimal Amount of Discretion to Allow in Disclosures,” *Quarterly Journal of Economics* 105, (1990) 427–444.
- [11] Glazer, J. and Rubinstein, A. “On Optimal Rule of Persuasion,” *Econometrica*, Vol. 72 (2004), pp.1715-1736.
- [12] Gneezy, U. “Deception: the Role of Consequences” *American Economic Review* 95, (2005) 384–394.
- [13] Kartik, N., Ottaviani, M. and Squintani, F. “Credulity, Lies, and Costly Talk,” *Journal of Economic Theory*, Vol. 134 (2007), pp.93-116.
- [14] Li, M. and Madarász, K. “When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests,” *Journal of Economic Theory* Vol. 139 (2008), pp. 47-74.
- [15] Matthews, S.A., Okuno-Fujiwara, M., and Postlewaite, A. “Refining Cheap-Talk Equilibria,” *Journal of Economic Theory*. Vol. 55 (1991), pp. 247-273.
- [16] Milgrom, P. “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics* Vol. 12 (1981), pp. 380-391.
- [17] Milgrom, P. and Roberts, J. “Relying on the Information of Interested Parties,” *RAND Journal of Economics* Vol. 17 (1986), pp. 18-32.
- [18] Morgan, J. and Stocken, P. “An Analysis of Stock Recommendation,” *RAND Journal of Economics* Vol. 34 (2003), pp. 183-203.
- [19] Morris, S. “Political Correctness,” *Journal of Political Economy* Vol. 109 (2001), pp. 231-265.
- [20] Olszewski, W. “Informal Communication,” *Journal of Economic Theory* Vol. 117 (2004), pp. 180-200.
- [21] Ottaviani, M. and Squintani, F. “Naive Audience and Communication Bias.” *International Journal of Game Theory* Vol. 35 (2006), pp. 129-150.

- [22] Rabin, M. "Communication Between Rational Agents," *Journal of Economic Theory* 51 (1990), 144-170.
- [23] Sanchez-Pages, S. and Vorsatz, M. "An Experimental Study of Truth-Telling in a Sender-Receiver Game," *Games and Economic Behavior* Vol. 61 (2007), pp. 86-112.
- [24] Sobel, J. "A Theory of Credibility," *Review of Economic Studies* Vol. 52 (1985), pp. 557-573.
- [25] Wolinsky, A. "Information Transmission when the Sender's Preferences are Uncertain," *Games and Economic Behavior* Vol.42 (2003), pp. 319-326.