Communication and social preferences: an experimental analysis

Antonio Cabrales, Francesco Feri, Piero Gottardi
and Miguel A. Meléndez-Jiménez*

January 2021

Abstract: This paper reports on experiments regarding cheap talk games where senders attempt deception when their interests are not in conflict with those of the receiver. The amount of miscommunication is higher than in previous experimental findings on cheap talk games in situations where senders' and receivers' interests are not in conflict. We obtain this even though, as in previous literature, some participants appear to feature a cost of lying. We argue our findings could be attributed to distributional preferences of senders who lie to avoid the receiver getting a higher payoff than herself.

Keywords: Experiments, Cheap talk, Deception, Conflicts of interest, Social preferences

JEL Classification: D83, C72, G14

* Antonio Cabrales, Dept. of Economics, Universidad Carlos III de Madrid; Francesco Feri, Dept. of Economics, Royal Holloway University of London and Università di Trieste; Piero Gottardi, Dept. of Economics, University of Essex; Miguel A. Meléndez-Jiménez, Dept. Teoría e Historia Económica, Universidad de Málaga. Miguel A. Meléndez-Jiménez acknowledges financial support from the Spanish Ministry of Science and Innovation through Project RTI2018-097620-B-I00 and from the Regional Government of Andalusia-FEDER through Projects UMA18-FEDERJA-243 and P18-FR-3840. All errors are our own.

1. Introduction

Communication is an important part of human interaction. It helps to solve coordination problems and it allows humans to use widely dispersed information to achieve personal and social goals. It is also a uniquely human activity (Nowak and Krakauer 1999, Berwick et al. 2013). Game theorists have also been interested in communication. Since the seminal work of Crawford and Sobel (1982), we learnt that under standard preferences there are sharp limits to the amount of information that can be credibly communicated among parties. Those limits arise from the extent of the alignment of preferences between the players. If the players' preferences are well aligned, in the most informative equilibrium all information from a sender can be passed on to the receiver. But if those preferences are completely misaligned, no information can be transmitted in equilibrium.

Yet a recent literature, prompted by experimental findings, has shown that individuals may have moral concerns, like an aversion to lying, in which case the limits to communication can be relaxed. This is particularly important in games where sender and receiver interests are misaligned. In those cases, standard theory would predict no communication is possible, but when senders have a psychological cost of lying, communication is at least partially restored.

We show in this paper that other factors affecting agents' preferences may create limits to communication. Specifically, we explore how distributional concerns can create impediments to communication in situations where standard theory predict they should not be present. The typical situation we have in mind – and examine - is one where the sender, by telling the truth and being believed, receives a benefit (or is not harmed) but the receiver receives an even bigger benefit. In that case, concerns for equality might tempt the sender to mislead the receiver.

This cannot happen in the standard cheap talk games that are commonly considered in the experimental literature following Crawford and Sobel (1982). In that framework, when preferences are aligned players have the same utility and there is no room for distributional concerns and no reason for deception. When preferences are not aligned, on the other hand, distributional concerns may arise but there is also a reason for lying even under standard preferences. Thus, in this benchmark situation, while the presence of moral concerns with regards to truth-telling generates a sharp prediction regarding the outcome, distributional concerns do not have an independent power to influence agents' behavior, and in particular induce deception, beyond what standard self-interest indicates. To allow for an independent role of distributional concerns, we need so to modify the experimental design.

We study a sender-receiver game where there is a state of nature known to the sender, but not to the receiver, and about which the sender can send a message to the receiver. The latter then chooses an action. This part would be standard in any sender-receiver game. But, in addition, the sender has a private type, known only to herself, while the receiver's type is common knowledge. These types and the state of nature determine whether or not a conflict of interest is present. In particular, if the sender's and the receiver's types happen to be the same and they also coincide with the state of nature, then their interests are misaligned. In all other situations, their interests are not misaligned.

However, while when the sender's type is equal to the state of nature her payoff is still affected by the action chosen by the receiver, in the complementary event where her type differs from the state of nature, her payoff is invariant with respect to the receiver's action. In that event, if the receiver chooses the action that is optimal for him in each state of nature, he gets a – significantly - higher payoff than the sender. This is a crucial difference with earlier designs. There is in fact no conflict of interest, but there is a distributional asymmetry, as the receiver does considerably better than the sender. This is the case that allows to separate a sender with distributional concerns from one without them. As we will argue below, the game described represents situations like one where sender and receiver are participants in the auction for an object and are a priori uncertain about its value for them.

We first provide a complete characterization of all the pure strategy equilibria of this game, under standard preferences. There is an equilibrium, which we label informative, where communication is maximal: the sender sends a message truthfully reporting the state of nature, except in the event we described where interests are misaligned. In that event the sender sends a false message, so as to deceive the receiver. As in all sender-receiver games, there is also a babbling equilibrium, where the sender's message is totally uninformative. Finally, there is another equilibrium, which we call opposite, in which the sender's message is only truthful when the types of sender and receiver differ. In comparing these equilibria it is of particular interest to point out that in the informative equilibrium the sender gets a lower payoff than the receiver while in the equilibrium labeled opposite the expected payoff difference between sender and receiver is lower.

The experimental results we obtain show that, as the previous experimental literature on sender-receiver games finds, subjects acting as senders sometimes tell the truth when there is a conflict of interest between sender and receiver (in contrast with the behavior prescribed in all the equilibria). But the crucial novelty is that we also find a remarkable amount of deception when the type of the sender differs from the state of nature, something not present in the previous experimental literature. To gain some explanation for these different findings, recall that in the designs based on Crawford and Sobel (1982) a situation with asymmetric payoffs and no conflict of interest was not present. In our game, when the sender's type differs from the state of nature, her payoff is not affected by the receiver's action, hence by telling the truth, and being believed, the sender does not harm her payoff, but clearly benefits the receiver. This finding survives a long period of learning, as our participants play for 40 periods. So, this is not simply the result of confusion, it seems deliberate. As a result of this higher level of deception, the average payoffs in the game are significantly lower than in the informative equilibrium.

As mentioned earlier, a candidate explanation for the observed behavior is that participants display sufficient concerns for equity to overcome the tendency to tell the truth that we also observe in the experiment. To test whether this is the case, we elicited the distributional preferences of the participants in our experiments using the method of Bartling et al. (2009). We then checked if there was a correlation between distributional preferences and the propensity for truth-telling. The results are quite clear, in that the combination of envious and non-pro-social attitudes leads subjects to lie when they would be expected to report the truth in the informative equilibrium. Interestingly, it is the combination of envy and non-pro-sociality that leads to this result, since envious and pro-social individuals are not more likely to lie. In other words, it is a status-seeking preference, rather than simple inequality aversion that is correlated with a propensity to lie to restore equity. On the other hand, a truth-telling behavior by the sender in situations where, in the informative equilibrium, she should send a deceptive report are not explained by social preferences. The main candidate for the explanation of this behavior, as in the previous literature, is some psychological cost of lying.

To better understand the behavior of experimental participants we also estimate a mixture model (à la Costa-Gomes, Crawford and Broseta 2001), where each sender is assumed to have a main "preferred" strategy and departures from it arise from random mistakes. This is important because in the previous discussion we have focused on individuals lying or telling the truth under different configurations of their state and the state of nature. However, we have not investigated their overall strategies in the game. By undertaking this analysis, we can get an idea of how the overall behavior of subjects relates to, or is different from, equilibrium strategies, and how it relates to distributional preferences.

We find the most parsimonious model categorizes the senders into four main strategies/types: two types that play, respectively, the informative and partially informative equilibria (but the latter one with a very high error rate), a type that never tells the truth, and another one that always tells the truth. This characterization is reminiscent of the theoretical model provided in Hurkens and Kartik (2009) that divides players into those that always tell the truth, and those that do so only when it is payoff maximizing. The most significant result from this analysis is that individuals who play the partially informative equilibrium are also likely to have non-prosocial and envious preferences. To be more precise, the individuals who are characterized as using more frequently the partially informative equilibrium strategy are also more likely to exhibit envious and non-prosocial preferences with the Bartling et al. (2009) test. This confirms, with a different kind of analysis the importance of concerns for equity to understand behavior in cheap-talk games, which is the main aim of this paper.

A caveat is needed regarding the analysis in this last part of the paper. We estimate a static model because not enough data are available to characterize a dynamic learning model, with individual level parameters.[1] But we observe individuals changing their behavior across repetitions of the game in ways that may suggest the presence of some learning. Possibly as a result of that, the estimates in this part are quite noisy and should then be taken as suggestive.

The paper is organized as follows. We first briefly describe the related literature. Then we present the design of the experiment and the analysis of the equilibria of the game we consider. The experimental results and their discussion follow. We conclude with some reflections for future research.

[1] Individual parameters are important to avoid biases, as shown for example in Wilcox (2006).

2. Literature

First, we should mention the seminal work of Crawford and Sobel (1982) on strategic information transmission, which studies how the alignment of preferences between sender and receiver affects information transmission (Sobel (2013) reviews the vast theoretical literature following that paper). As noted above, with respect to that paper (and the subsequent literature), we consider a richer game structure that allows for some novel results. In particular, the alignment of interests between senders and receivers is not commonly known, as it depends on the realized types of the sender and of the object. As explained earlier, this fact plus the asymmetry of payoffs in the states where the sender's type does not match the state imply that distributional preferences have distinct implications for agents' behavior from attitudes towards truth-telling.[2]

The experimental literature on information transmission has concentrated primarily on analyzing sender-receiver games à la Crawford and Sobel (1982). A first series of papers (e.g., Dickhaut, McCabe and Mukherji (1995), Blume et al. (1998, 2001), and Kawagoe and Takizawa (1999)) demonstrates that, when the interests of the sender and receiver are well aligned (the underlying game is one of common interest), play tends to converge to informative/separating equilibria. A more recent strand of the literature (see Sánchez-Pagés and Vorsatz (2007), Kawagoe and Takizawa (2005), Cai and Wang (2006), and Wang, Spezio and Camerer (2010)) finds evidence of a higher level of truthful communication than the most informative equilibrium would predict in games in which interests do not align well. This finding is then attributed to a truth-telling norm. While in our experiments we also find some evidence of aversion to lying, we also observe a substantial amount of deception/misinformation even when lying does not increase the senders' payoff but reduces that of the receivers. This tends to be the case for subjects who display (independently measured) non-pro-social or envious preferences.

Gneezy (2005) also explores deception by senders in situations where senders always benefit materially from the deception (although not very much, most of the times), while the receiver is harmed. Given the importance of this work and the connection to

[2] There is also a relevant theoretical literature that studies information transmission when agents may have a preference for telling the truth (see Kartik, Ottaviani and Squintani (2007) and Bolton, Freixas and Shapiro (2007)).

our paper, we shall discuss this at greater length in the main body of the paper when we comment on our results.

Maggian and Villeval (2016) study the connection between lying and social preferences by making participants (in their case young children) play a dictator game in which the dictator, by telling a lie or the truth to the experimenter, can vary the distribution between her and the receiver. They find that subjects playing the dictator often lie to avoid unequal distributions. The main difference with other papers we discussed, and with ours, is that this is not a game of communication, as the other participant is passive and the lie is directed to the experimenter.

Brandts and Charness (2003) study a situation where a player sends a cheap talk message about her intended action in a 2×2 game she plays with a receiver. The receiver is then allowed to inflict a costly punishment to the sender. They find that an action by the sender that leads to an unequal payoff induces the receiver to punish. However, the punishment is lower if the sender did not lie about his intended action. That is, a violation of trust induces more negative feelings regarding the sender than simple envy, thus establishing the separate importance of norm violation in addition to a concern for equity. We also find that both truth-telling and distributional concerns contribute to determine subjects' behavior, though the way in which these factors manifest themselves is different and so is the environment considered.

3. The game and experimental design

We now describe the game. In each round, each player is randomly assigned a color (either black or white, defining his type), with the assigned colors being i.i.d. There is also an i.i.d. random draw for the color of an object (defining the state of nature), which can be either black or white. Player 1 – the sender - is informed of the colors assigned to both agents and of the color of the object whereas player 2 – the receiver - is only informed of his assigned color. If the assigned color of a player coincides with the color of the object, we say that the player is *interested* (otherwise, *uninterested*).

Next, player 1 sends a message (or report) regarding the color of the object to player 2. The message can be either "the color is white" or "the color is black", i.e., player 1 can either send a truthful message or a false message. Player 2 observes the message sent by player 1 and chooses an action that can be either *left*, *center*, or *right*. The players' payoffs depend on whether they are interested in the object and on the choice of player 2, as indicated in Table 1.

Table 1. Payoffs in the game *Sender Receiver*

Player 2's choice	Left	Center	Right	Left	Center	Right
Player 1 not interested	20, 20	20, 60	20, 100	20, 160	20, 120	20, 50
Player 1 interested	20, 30	70, 90	120, 120	20, 60	70, 50	120, 40
	Player 2 not interested			*Player 2 interested*		

(in each cell: player 1's payoff, player 2's payoff)

The payoff structure resembles the conflict of interests that would arise in a hypothetical auction for the object, where player 1 (the informed player) makes a high bid when she is interested in the object and a low bid otherwise. Player 2's action *left* could be interpreted as a high bid in the same hypothetical auction, or more generally a choice to fight, action *center* as a medium bid or a choice to stay put, and action *right* as a low bid or a choice to concede. When player 2 is interested, action *left* is preferred by this player to *center*, which in turn is better than *right*. If he is not interested, the order is reversed.[3]

The game, however, also applies to a variety of other situations. The main features are that, when player 1 is not interested, her payoffs are independent of the action of player 2. In contrast, when player 1 is interested she prefers that player 2 chooses *right* rather than *center* and *center* rather than *left*. In this case the preferences of player 1 may be strictly aligned with those of player 2 (when player 2 is not interested), or strictly misaligned (when player 2 is also interested). Thus, depending on the state we may have a clear conflict or a perfect alignment of the interests of the two players, but also the case where there is still alignment but only because the sender does not care for the action of the receiver.

We want to stress this richness of our design. The sender has four information sets: in two of them (when she is interested) she has a strict interest that the receiver believes her messages, in the other two (when she is uninterested) she is indifferent. However, she could use her behavior in these last two information sets to shape the receiver's beliefs over the truthfulness of the messages she sends. We believe this characteristic is uniquely novel in our design.

These features characterize situations where each agent can act on the basis of his or her information, information has a partly rival nature and its value varies with the state,

[3] Moreover, player 2's payoff is strictly higher when her assigned color differs from the assigned color of player 1 than when they coincide.

as in the case of salesmen who are recommending brands to customers, but who may occasionally have incentives for selling a particular brand. These features are somewhat different from the ones present in the cheap talk games considered in most of the literature following Crawford and Sobel (1982), where the conflict of interest is known, independent of the state of nature.[4]

The analysis of the equilibria of the game described is developed in the online Appendix, where we characterize all the pure strategy (weak) perfect Bayesian equilibria. In the *informative equilibrium*, player 1 sends a message that informs the receiver of the true color of the object except in the event where she is interested in the object and player 2 is also interested. The optimal response of player 2 in this equilibrium is then to choose action *left* if the report says that she is interested (i.e., if it contains her assigned color) and action *right* otherwise. Thus player 2 trusts sufficiently the truthfulness of the message sent by player 1 to choose to fight when told he is interested and to concede when told he is not.

In addition to the informative equilibrium, there is also a *babbling equilibrium* in which player 1's message is uninformative (i.e., the sender uses the same rule to determine the content of the message regardless of what she learned about the players' types and the type of the object) and player 2 chooses then action *center* regardless of the content of the message received.[5]

Finally, there is another equilibrium in which player 1's message reports truthfully the color of the object when the assigned colors of players 1 and 2 differ, but reports instead the opposite of the true color of the object when the assigned colors of players 1 and 2 coincide.[6] Thus player 1's message says that player 2 is interested in the object when player 1 is not interested, and that player 2 is not interested when player 1 is interested. Hence the report always portrays the receiver's preferences for the object as being opposite to the ones of the sender and for such reason we refer to this as the *opposite*

[4] For an exception, see Le Quement (2016). One important difference with our setup is that he considers the case where the senders' bias is private information and is fixed across states of nature, whereas in our case the bias depends on the state. A more important difference is that there are many possible senders the receiver can consult (at a cost). Since some of the senders are unbiased, while others are biased in possibly different directions, the actions of senders are affected by competition among them and their private information. The equilibrium structure is then much more complicated than ours and the research objective, focused on the disciplining effect of competition on truthfulness, is quite different.

[5] The case in which player 2 responds using action *left* on the out-of-equilibrium path (and action *center* on the equilibrium path) also constitutes a babbling (weak) perfect Bayesian equilibrium that is outcome equivalent to the babbling equilibrium described in the text. See the online Appendix for details.

[6] Of course, associated with each one of the three equilibria there is another equilibrium, in which the colors reported by player 1 are exactly the opposite in every state. See the online Appendix for details.

equilibrium. Such report is truthful only in two out of four information states. The messages of the sender differ from those in the *informative* equilibrium only in the state where none of the players is interested in the object. In that event in the *informative* equilibrium the sender tells the truth, whereas in the *opposite* equilibrium the sender lies. So, clearly, in the *opposite* equilibrium less information is transmitted – though still some.

In this equilibrium player 2, even though he has less confidence in the truthfulness of the message sent by player 1, is still willing to follow a similar behavior as in the *informative* equilibrium: choose action *left* or action *center* if the report says that he is interested (both actions provide the same expected payoff to player 2) and action *right* otherwise. In fact, the information transmitted by the sender is the minimum one that induces the receiver to choose to concede when the report says he is not interested. This smaller amount of information transmitted lowers the payoffs of the receiver, but not the payoffs of the sender, something which is important for our analysis.

The equilibrium (expected) payoffs to players 1 and 2 are, respectively, 70 and 105 (in the *informative* equilibrium), 45 and 80 (in the *babbling* equilibrium), and 70 and 85 (in the *opposite* equilibrium).

Regarding the implementation of the experiment, participants were assigned to groups of four players and played 40 rounds of the game. In each round, subjects were randomly matched within their group, and in each pair, one subject was randomly assigned the role of player 1 (sender) with the other subject acting as player 2 (receiver).

After the 40 rounds of play, we elicited the subjects' attitudes towards risk and social preferences. We used the risk test proposed by Charness and Gneezy (2010). In particular, subjects had to decide how much of their endowment (5 euros) to invest in a risky asset and how much to keep. They earned 2.5 times the amount invested if the asset has a high yield (with prob. 0.5) and otherwise lose the entire amount invested.

Regarding social preferences, we used the approach proposed by Bartling et al. (2009) to identify pro-social and envious attitudes. This information then allows us to investigate possible rationales for the behavior observed in the sender-receiver game in terms of these attitudes. Each subject was asked to make four decisions corresponding to four dictator games. Each decision consists of a choice between *distribution 1* and *distribution 2*. The choice of a distribution determines a payoff for the player and a payoff for another player.[7] These payoffs are shown in Table 2.

Table 2. Dictator games for the elicitation of social preferences

Game *(All payoffs in euros)*	*Distribution 1* *self: other*	*Distribution 2* *self: other*
(I) Pro-sociality	2: 2	2: 1
(II) Costly pro-sociality	2: 2	3: 1
(III) Envy	2: 2	2: 4
(IV) Costly envy	2: 2	3: 5

Using the choices made in these games, we can classify subjects with respect to their pro-sociality and envy attitudes. Regarding pro-sociality (games I and II), subjects choosing distribution 1 in game I and distribution 2 in game II are classified as *weakly pro-social* and those choosing distribution 1 in both games are classified as *strongly pro-social*. Those who choose instead distribution 2 in both games are classified as *non-pro-social*. Regarding envy (games III and IV), subjects choosing distribution 1 in game III and distribution 2 in game IV are classified as *weakly envious*, while those choosing distribution 1 in both games are classified as *strongly envious*. In contrast, those choosing distribution 2 in both games are classified as *non-envious*.[8]

To explain the terms, a prosocial person prefers (perhaps weakly) to avoid having higher payoffs than others. An envious person prefers (perhaps weakly) to avoid having lower payoffs than others. It is indeed possible to be both pro-social and envious.

We ran three sessions at the laboratory of experimental economics of the University of Siena (LabSi). A total of 40 subjects participated in these sessions, recruited from the LabSi pool of human subjects, primarily consisting of undergraduate students from the University of Siena. The average duration of the sessions was 70 minutes (including the reading of instructions but excluding payment procedures). The experiment was computerized and conducted using the experimental software z-Tree (Fischbacher,

[7] Every subject acted as the decision maker in each of the four dictator games, with another subject randomly chosen to be the recipient, and one (randomly selected) choice of a decision maker in the four games was paid.

[8] Note that a subject choosing distribution 2 in game I and distribution 1 in game II would be hard to rationalize in terms of pro-social attitudes. Similarly, a subject choosing distribution 2 in game III and distribution 1 in game IV would be hard to rationalize in terms of envy attitudes. We do not find any of these choices in our sample.

2007). The experimental instructions, translated into English, are reported in the online Appendix.

4. Results

In Table 3 we report the behavior of senders. More specifically, we report the frequency of true messages, distinguishing the cases in which the senders and the receivers are, respectively, interested or not interested in the object. According to the informative equilibrium, the frequency should be 1 except for the cell in which both players 1 and 2 are interested, in which case it should be 0. We observe that the modal play coincides with the prediction of that equilibrium (the frequency of true messages is above 0.5 in the top cells and in the bottom-left cell of Table 3, and it is below 0.5 in the bottom-right cell), but there are significant deviations. These deviations are particularly evident in the top cells, in which approximately one-third of the uninterested senders (who are expected to report the truth since they are not interested) lie, and in the bottom-right cell, in which approximately one-third of the interested senders (who are expected to lie since they are interested and so is the receiver) report the truth.

Table 3. Player 1's behavior. Frequency of true messages

	All Rounds	Rounds 21-40	All Rounds	Rounds 21-40
Player 1 not interested	0.6315 (0.4834) [228]	0.6111 (0.4897) [108]	0.6497 (0.4782) [197]	0.6593 (0.4765) [91]
Player 1 interested	0.7419 (0.4387) [186]	0.8085 (0.3955) [94]	0.3439 (0.4762) [189]	0.3551 (0.4808) [107]
	Player 2 not interested		*Player 2 interested*	

(Std. Dev.), [number of observations]

In Table 4 we present the actions chosen by the receivers in response to the content of the message they received. Again, we observe that modal choices in each case correspond to the prescriptions of the informative equilibrium: the choice of *left* (that is, to fight) when the message says that the receiver is interested (*Color – Yes*, with a frequency slightly below 50%) and the choice of *right* (that is to concede) when the message says that he is not interested (*Color – No*, with a frequency slightly above 60%). Note the significant use of action center (between 25% and 40% of the observations).

Table 4. Player 2's behavior. Absolute frequencies of choices

	Left		Center		Right	
	All Rounds	*Rounds 21-40*	*All Rounds*	*Rounds 21-40*	*All Rounds*	*Rounds 21-40*
Color– No	65 (13.68)	31 (12.81)	109 (22.95)	62 (25.62)	**301 (63.37)**	**149 (61.57)**
Color – Yes	**161 (49.54)**	**76 (48.10)**	110 (33.85)	61 (38.61)	54 (16.62)	21 (13.29)

% over total in the row in brackets, theoretical predictions in bold

4.1 Social preferences

We investigate next what could explain the departures we observed from truth-telling, especially in cases where truth telling is consistent with agents' maximizing behavior. In particular, we explore to what extent the observed behavior of senders is related to features of their social preferences. To this end, we will use the choices made by subjects in the dictator games (as in Bartling et al. (2009)) described in Table 2. Subjects were asked to play such games at the end of each session.

In Table 5 we report the distribution of social preferences in our population of experimental subjects. We see that 42.5% of the subjects are classified as envious and 30% of them are classified as non-pro-social. We proceed then to examine to what extent these attitudes are associated with the deviations from the informative equilibrium we saw in Table 2.

Table 5. Distribution of social preferences

	Weakly Envious	*Strongly Envious*	*Non envious*	*Total*
Weakly pro-social	6 (25.00) [100]	5 (20.83) [45.45]	13 (54.16) [56.52]	24 (100) [60.00]
Strongly pro-social	0 (0) [0]	0 (0) [0]	4 (100) [17.39]	4 (100) [10.00]
Non pro-social	0 (0) [0]	6 (50.00) [54.54]	6 (50.00) [26.08]	12 (100) [30.00]
Total	6 (15.00) [100]	11 (27.50) [100]	23 (57.50) [100]	40 (100) [100]

(% over total row), [% over total column]

To this end, we perform a logit estimation of the probability that player 1 sends a true message. The explanatory variables are combinations of the social preference variables and dummies that determine whether the sender and the receiver are interested in the object. In particular, we define *Int1* (*Int2*) as a dummy that takes value 1 if player 1 (player 2) is interested in the object, i.e., if her assigned color and the color of the object coincide, and takes value 0 otherwise. We define then *NoInt1* = 1 – *Int1* and *NoInt2* = 1 – *Int2*. Similarly, we define *Prosoc* (*Env*) as a dummy that takes value 1 if player 1 is pro-social (envious), either weakly or strongly. Finally, we define *NoProsoc* = 1 – *Prosoc* and *NoEnv* = 1 – *Env*. We also include in the regression the variables *Round* (from 1 to 40) and *Risk*, which describes the choice made by the subject in the risk test.

In Table 6 we report the marginal effects of envious and pro-social attitudes on the probability of sending a truthful report (measured at round 20 and for the average risk aversion level). The full estimation is reported in Table 11 in the online Appendix.

Table 6. Marginal Effects of Envy and Pro-sociality (at round 20 and the average risk aversion level) on the probability of sending a true message

	Marginal effect of *Env* = 1 vs. *Env* = 0 (by values of *Prosoc, Int1* and *Int2*)		Marginal effect of *Prosoc* = 1 vs. *Prosoc* = 0 (by values of *Env*, *Int1* and *Int2*)	
	Prosoc = 0	*Prosoc* = 1	*Env* = 0	*Env* = 1
Int1 = 0 & *Int2* = 0	-0.2550** (0.1111)	-0.0883 (0.0795)	0.0392 (0.0905)	0.2058** (0.1017)
Int1 = 0 & *Int2* = 1	-0.3924*** (0.1175)	-0.0768 (0.0815)	-0.0419 (0.0928)	0.2736** (0.1098)
Int1 = 1 & *Int2* = 0	-0.4027*** (0.1103)	-0.1296* (0.0755)	-0.0557 (0.0766)	0.2173** (0.1102)
Int1 = 1 & *Int2* = 1	0.0036 (0.1133)	-0.0012 (0.0852)	0.1498 (0.0963)	0.1449 (0.1040)

More precisely, on the left-hand side of Table 6 we present the marginal effect of envy (i.e., of *Env* = 1 vs. *Env* = 0). This is measured separately for pro-social and non-pro-social subjects and for the different combinations of preferences for the object of the sender and the receiver (i.e., the four cells of Table 3). On the right-hand side of Table 6 we present the marginal effect of pro-sociality (i.e., of *Prosoc* = 1 vs. *Prosoc* = 0). This is measured separately for envious and non-envious subjects, again differentiating the various combinations of preferences for the object of the sender and the receiver.

The first implication we draw from the inspection of the marginal effects in Table 6 is that social preference attitudes are associated with deviations in the behavior of the sender from the informative equilibrium in the cases in which such equilibrium prescribes that senders report the truth, i.e., in the first three rows of Table 6. In these cases, the results on the left-hand side of the table (i.e., regarding the marginal effect of *Envy*) show that, for those subjects who are non-pro-social, the fact that they are also envious significantly reduces the probability of telling the truth. In contrast, envy does not have a significant effect for pro-social subjects.[9] Such finding is somewhat surprising. As shown above, the informative equilibrium, with maximal communication, features an asymmetric distribution of its gains, which favors the receiver. This may suggest the conjecture that the more limited level of communication observed in the experiment could at least partly be explained by envy traits in subjects, while non-pro-sociality should not play a role. In contrast, the results reported in Table 6 show that, while social preferences indeed contribute to explain the untruthful reports by senders in the absence of conflicts of interest, the relevant trait is not envy but envy together with non-pro-sociality.

This is an interesting observation that we could explain in the following way. Individuals who are *envious* and *pro-social* are likely to be *inequality averse*. That is, they have a preference for fair and equitable behavior. Individuals who are *envious* and *not pro-social* tend to have instead some form of *status-loving* preference. We may conjecture they are less likely to have moral concerns and so find not telling the truth less difficult than *inequality averse* individuals.

Similarly, the results on the right-hand side of the table (i.e., regarding the marginal effect of *Pro-sociality*) show that, for those subjects who are envious, the fact that they are also non-pro-social significantly reduces the probability of telling the truth. However, *Pro-sociality* does not have a significant effect for non-envious subjects. Thus, as explained earlier our results suggest that it is the combination of envious and non-pro-social attitudes that leads subjects to lie in situations in which they would be expected to report the truth. We believe this is an important insight of our study.[10]

Next, we turn our attention to situations in which the informative equilibrium prescribes the sender to lie: this happens when both the sender and the receiver are interested in the object. The results for this case, reported in the fourth line of Table 6,

[9] Only marginally at the 10% level if the sender is interested and the receiver is not.

[10] See also Morgan, Steiglitz and Reis (2003) for an analysis of the effect of social preferences on bidding behavior in auctions.

show that the marginal effects of envy and pro-sociality on the probability of sending a true message are not significant. In such a situation, the coefficient representing the marginal effect of envy is essentially 0 and the coefficient representing the marginal effect of pro-sociality is positive but not significant (p-values of 0.16 and 0.12 for non-envious and envious subjects, respectively). This is not surprising as in this case the main force affecting agents' behavior is the tension between pure self-interest and aversion to lying, while distributional preferences play a smaller role. Our findings are in line with those by Brandts and Charness (2003), who argue that truth-telling when not individually advantageous is often the product of a social norm against lying, rather than the result of distributional preferences.

It is also useful to discuss now in greater detail the relationship with the findings in Gneezy (2005). In the situation considered in that paper a sender must tell a receiver, who is uninformed about his payoffs, what is the receiver's most profitable action. The action of the receiver, in turn, affects both the sender and the receiver. There are three treatments. In all three treatments the sender is better off when the receiver takes action B, while the receiver is better off when action A is chosen. Thus, recommending B constitutes a deception that, if followed, is beneficial for the sender. The key difference between the treatments is given by the extent to which the sender and the receiver benefit from the two actions. In treatment 1 both agents benefit very little from the adoption of the action that is optimal for each of them. In treatment 2 the sender benefits very little from her preferred action, while the receiver benefits substantially from his optimal action. Finally, in treatment 3 both players benefit considerably from their preferred actions. The payoffs in the three treatments are summarized in Table 7.

An important feature of the environment considered in Gneezy (2005) is that the receiver does not know his payoffs nor those of the sender and hence ignores whether a conflict of interest exists or not, in fact we may conjecture is even unaware that a conflict of interest may exist. In the behavior observed in the experiment the receiver often follows the advice of the sender. This suggests that his priors regarding the presence of that conflict are typically not very large.

Table 7. Payoffs in Gneezy (2005)

Treatment	*Option*	*Payoff to*	
		Sender	*Receiver*
1	A	5	6
	B	6	5
2	A	5	15
	B	6	5
3	A	5	15
	B	15	5

Comparing Table 7 with Table 1 we see that the situation in treatment 1 in Gneezy (2005) is somewhat similar to the one arising in our game when both sender and receiver are not interested in the object. Treatment 2 is then close to the situation that obtains in our game when the sender is not interested in the object, but the receiver is. Finally, treatment 3 exhibits some analogies with the case when both sender and receiver are interested in the object.

To explore the interactions between distributional preferences and lying, Gneezy (2005) performs another experiment with a dictator game whose payoffs are the same as if the sender directly chose between options A and B in Table 7 (rather than making a recommendation to the receiver). The result is that the choice of options B is more common in this experiment (66% in treatment 1, 42% in treatment 2 and 91% in treatment 3, compared to 36% in treatment 1, 17% in treatment 2, and 52% in treatment 3 in the main experiment). This shows that having to lie to reach a favorable outcome for the sender is not acceptable for some people.[11,12] To further emphasize this point, Gneezy (2005) conducts a separate experiment, with different participants, asking them whether they think it is appropriate for a seller to hide a defect in a used car he is selling. He shows the majority of subjects consider it acceptable only if the seller's cost of fixing the damage is high, thereby showing that the acceptability of lying depends on the size of the payoff consequences.

This underlines a key difference between Gneezy (2005) and our experiment. His design aims to show that an aversion to lie mitigates the desire by some people to obtain

[11] Hurkens and Kartik (2009), using the set-up of Gneezy (2005), examine the relationship between lying aversion and preferences. They show that people can be categorized in two types, either they will never tell a lie, or they will tell a lie when it allows them to get a material benefit.

[12] Charness and Dufwenberg (2006) explore theoretically and experimentally whether the results in Gneezy (2005) can be explained by a norm that induces guilt in senders if they "let-down" receivers (i.e., senders believe that they harm receivers relative to what the latter believe they will receive). They construct a different game from Gneezy (2005) that allows them to separate the role of social preferences from that of an aversion to disappointing receivers.

higher payoffs (even at the expense of others). In contrast, we show that a desire to avoid earning less than receivers may induce senders to be more willing to tell a lie, even though this may end up lowering their material payoff. So, we show that envy can be a motivator for lying, whereas he shows that lying aversion can be a deterrent to pursue self-interest (something we also observe as some players tell the truth when they are interested in the object).

Furthermore, our analysis allows to isolate cleanly the effect of distributional preferences, which are measured directly, and to disentangle the envy and pro-sociality components in these preferences, while treatments considered by Gneezy focus on the degree of selfishness vs. altruism. This allows us to disentangle some subtle effects, like the one connecting envy and non-pro-sociality with propension to lie.

We should conclude this discussion by briefly commenting on the consequences of another significant difference between our experiment and his, given by the fact that in our case receivers are aware of the potential conflict of interest, and this is common knowledge, while in Gneezy (2005) there is always a conflict of interest, but receivers are not aware of it. When we compare the receivers' response in Gneezy (2005) to the one we observed, we see that the receivers' awareness of a possible conflict of interest makes them more wary about the sender's behavior. For example, 78% of receivers followed the recommendation of the sender in Gneezy (2005),[13] while in our case the receiver plays the action most preferred by the sender (the one that would be chosen in the *informative* equilibrium) between 50% and 60% of the time. This is true even though the frequency of lies that we observe in the three information sets where there is no conflict of interest are broadly comparable to one found by Gneezy and recalled above.[14] Hence we may be led to think that the possibility to understand the senders' incentives makes receivers less trustful and it is not even clear they are "deceived."

[13] In a set-up analogous to Gneezy (2005), Sutter (2009) analyzes the role for reporting choices of the sender's expectations regarding the receiver's response. He shows that senders sometimes expect their recommendation not to be followed. If they tell the truth and this leads to the sender's preferred outcome, senders' behavior, though truthful, should be called deceptive, that is, self-interested.

[14] Gneezy (2005) reports that senders recommend option B to the receiver (recall, this is the best for them but not for the receiver) is 36% in treatment 1, 17% in treatment 2, and 52% in treatment 3. In our experiment we saw in Table 3 that the frequency of lies in the situation that we argued is similar to treatment 1 (both not interested) is 37%, in the one similar to treatment 2 (sender not interested, receiver interested) is 35%, and in that similar to treatment 3 (both interested) is 66%.

4.2 Senders' strategies

In the previous section we have concentrated on the relationship between the probability a sender tells the truth and her social preferences, by considering independently her behavior at each different information set. To gain some understanding also of the overall strategy guiding senders' behavior we examine next the pattern of choices made by senders in the whole game.

To this end, in this section we investigate senders' choices building on the econometric model in Costa-Gomes, Crawford and Broseta (2001), suitably adapted to our framework. In this analysis, we assume that each sender can be classified into one out of several types, corresponding to the possible (pure) strategies of the sender.

More precisely, a strategy for the sender is a vector of four components $(c_{ni,ni}, c_{ni,i}, c_{i,ni}, c_{i,i}) \in \{0,1\}^4$, where $c_{ni,ni}$ is the sender's choice in the information set in which both the sender and the receiver are not interested in the object, $c_{ni,i}$ is the choice in the information set in which the sender is not interested and the receiver is interested, $c_{i,ni}$ represents the choice in the information set in which the sender is interested and the receiver is not, and $c_{i,i}$ represents the choice in the information set in which both the sender and the receiver are interested. For each information set, the choice $c = 1$ indicates that the sender sends a true message (i.e., the content of the message is the true color of the object), while the choice $c = 0$ indicates that the sender sends a false message (i.e., the content of the message is the opposite color of that of the object). Hence, we consider each strategy $(c_{ni,ni}, c_{ni,i}, c_{i,ni}, c_{i,i}) \in \{0,1\}^4$ to be a sender's type.[15] There are so 16 different types in total.

We assume that a type k sender chooses the actions prescribed by strategy k with probability $1 - \varepsilon_k$ and makes a mistake with the residual probability $\varepsilon_k \in [0,1]$, describing type k 's error rate. When she makes a mistake, she sends any of the two possible reports with equal probability.[16]

We consider a mixture model in which each sender's type is drawn from a common distribution over types and remains constant for all the periods in which a player acts as a sender. The estimates of this model provide us the probability distributions over types and the error rates. We estimate the model separately for the first 20 periods and then

[15] Technical details on the estimation can be found in the online Appendix.

[16] Mistakes can happen at all information nodes, with equal and independent probabilities across periods and subjects.

again for the last 20 periods, so as to be able to capture some evolution of the assigned types (due, for instance, to players' learning about the game.[17]

We find that the most parsimonious model [18] in the first 20 periods features the following three types. We have one type (which we denote T1) playing strategy (1,1,1,0), that is the one corresponding to the informative equilibrium, a second type (denoted T2) who plays strategy (1,1,1,1) which consists in always announcing the truth, and a third type (denoted T3) playing strategy (0,0,0,0) associated with never telling the truth. The most parsimonious model for the final 20 periods has also three types, but partly different: T1 and T2 are still present, as the associated strategies are played by a significant number of players, but T3 disappears and is replaced by another type (denoted T4), who plays strategy (0,1,1,0) corresponding to the *opposite* equilibrium. Thus, T1 and T4 are equilibrium types whereas T2 and T3 are not. T1 always tells the truth and type T4 never tells the truth.

Table 8 shows the estimated parameter values of this model. For each estimation of the variable *Probability* associated with each type the (two-sided) significance reported in the table refers to the probability of "being greater than 0". For the variable *Error rate* the (two-sided) significance reported in the table refers to the probability of "being smaller than 1". A way to understand this is that the error rate is smaller than one if the behavior of the sender is significantly different from that of an uninformative (babbling) strategy.

[17] An alternative modeling strategy to identify subjects' types was pursued by Cai and Wang (2006) and Wang, Spezio and Camerer (2010). They fit a level-*k* model, where level-0 for player 1 is truth-telling and for player 2 is to naively believe player 1. The behavior of players at all other levels is then strictly pinned down in their game by this choice of level 0 behavior. We did not pursue this route for two reasons. Using the same specification for level 0 does not uniquely pin down behavior in our game for other levels of $k > 1$ because of the insensitivity of player 1's payoff to player 2's action when player 1 is not interested. Furthermore, the observed behavior of player 1 when both players are interested is inconsistent with that kind of model.

[18] To select the most parsimonious model we apply the following iterated procedure. We begin by estimating this model with all (n = 16) types and compute the Akaike Information Criterion (AIC). Then, we estimate all models with n-1 types, choose the best one (according to the Loglikelihood) and compute AIC. If the best model with 15 types performs better of that with 16 types (according to the AIC), we continue estimating all models with 14 types, choose the best one and compute AIC. If the best model with 14 types performs better of that with 15 types, we continue estimating all models with 13 types. We stop this process when the best model with one type less does not perform better (according AIC) otherwise, we continue the process.

Table 8. Error-rate model

	Rounds 1-20		*Rounds 21-40*	
Type	*Probability*	*Error rate*	*Probability*	*Error rate*
T1 (1,1,1,0)	0.406*** (0.100)	0.225^^^ (0.068)	0.325*** (0.117)	0.179^^^ (0.115)
T2 (1,1,1,1)	0.345*** (0.126)	0.672 (0.236)	0.122 (0.105)	.252^^^ (0.263)
T3 (0,0,0,0)	0.249* (0.131)	0.586^^ (0.182)		
T4 (0,1,1,0)			0.553*** (0.140)	0.782 (0.155)

Standard errors within brackets
Significance: Probabilities (with respect to 0): ***, **, * at the 1%, 5%, 10% level, resp.
Error rates (with respect to 1): ^^^, ^^, ^ at the 1%, 5%, 10% level, resp.

The examination of the results in Table 8 suggests that a note of caution is due regarding the possibility of drawing clear-cut conclusions from the analysis of the mixture model. We see in fact that several of the types identified, like the truth telling type T2 (in the first half of the experiment), or type T4 associated with the opposite equilibrium (in the last half), feature rather high error rates (0.672 and 0.782, respectively), so they cannot be clearly distinguished from an uninformative (babbling) type.[19] This is most likely due to the restrictive nature of the model, which assumes that all departures from the adoption of a particular strategy should be imputed to an independently distributed noise term that puts equal weight on all other strategies. It seems quite plausible that individuals do not depart from a strategy by giving the same weight to all others. However, estimating a more complex learning model would be difficult given the amount of data we have.

With this proviso, it is still useful to combine the analysis of the mixture model with that of the social preferences of subjects to examine whether the different types we have identified with this model are related to social preference traits. To do this, we compute for each subject exhibiting one of the four possible social traits we considered in Section 4.1, the probability that the subject is of one of the three types identified in the model. In Table 9 we then report, for each social trait, the average of these probabilities across subjects in the first and second half of the experiment.

[19] In Table 12 in the online Appendix we report the results for some statistical tests aimed to determine whether, from the point of view of the receiver, the error rates of the different types are high enough to have the same implications as babbling strategies (i.e., to induce the same best response from the receiver as a babbling strategy). We find that this cannot be ruled out in the case of type T2 (in the first half of the experiment) and T4 (in the last half).

Table 9. Social preferences and types in the mixture model

Social Preferences	*Rounds*	*T1 (1,1,1,0)*	*T2 (1,1,1,1)*	*T3 (0,0,0,0)*	*T4 (0,1,1,0)*
Non envious & Prosocial	1-20	0.445	0.447	0.109	
	21-40	0.464	0.125		0.378
Envious & Non prosocial	1-20	0.135	0.330	0.535	
	21-40	0.140	0.002		0.858
Non envious & Non prosocial	1-20	0.781	0.126	0.0935	
	21-40	0.196	0.032		0.772
Envious & Prosocial	1-20	0.291	0.315	0.395	
	21-40	0.280	0.183		0.537
Aggregate	1-20	0.406	0.345	0.249	
	21-40	0.325	0.122		0.553

An immediate conclusion from the table is that social preferences are indeed related to the strategy choice. Subjects with different social preference traits have very different propensities to choose the different strategies. The non-envious and non-prosocial players, for example, mostly concentrate on type T1 (the strategy associated with the informative equilibrium) in the first 20 periods and then shift to type T4 (the strategy of the opposite equilibrium). Notice that a shift from the informative to the opposite equilibrium does not change the sender's utility, only the receiver's payoff is affected, which is consistent with the fact these subjects do not care about others' welfare. The modal type of envious and non-prosocial players (the *least nice* participants) is, in the first twenty periods, T3 (always lie) and, in the last twenty periods, T4 (the strategy of the opposite equilibrium), which leads to an outcome that is better for senders than T3, but worse for receivers, and thus could be attractive to those players that are *status seekers*.

Non-envious and pro-social players are more evenly divided across types T1 and T2 (always tell the truth) in the first 20 periods and across T1 and T4 in the last 20 periods. In general, they exhibit a significant tendency to use a more sophisticated strategy, entailing lies only when strictly beneficial to the sender, which is consistent with their social trait attributing importance to total social welfare and less to their own payoff. Something similar happens with the envious and pro-social players. These players, who are less interested in maximizing total surplus, and more on equity, have a somewhat higher tendency to use the strategy of the opposite equilibrium, which is more egalitarian than the informative equilibrium.

Overall, this analysis, in spite of the noisiness, which is likely attributable to the limitations of the model, points in a similar direction as the one we performed in section 4.1.

5. Conclusion

The traditional experimental economics literature on cheap talk has generally painted a positive view on the ability of communication to improve social welfare. Because of humans' moral tendency to avoid falsehood, individuals eschew lying even in circumstances where it would materially benefit them. This improves the chances of communication to coordinate social behavior. In this paper we introduce a counterpoint to that generally held belief. Many humans also exhibit a tendency to act so as to avoid inequality that goes against them. We show those individuals may be willing to overcome the aversion to lying if this could reduce a payoff gap with respect to others. This is true even when the lie does not benefit them.

We were able to find this because our experimental design, unlike the commonly used one, allows for the possibility that no material conflict of interest is present, but truth-telling – if believed - yields a material benefit for the receiver, not for the sender. We are also able to check that the individuals who in such situations choose deception are envious, because we measure independently the envy and pro-sociality of our experimental participants.

We believe our results open interesting new avenues of research. One could investigate whether our results also hold in the field, for tasks that are common in everyday life. Would a trader in the government bonds desk pass really valuable information about a stock to a colleague in the stocks desk who thereby might enjoy an important promotion? What would be the personal characteristics of such a bond trader (and the stock one) that would make that kind of communications more likely?

References

Alonso, R., Dessein, W., & Matouschek, N. (2008). When does coordination require centralization? *The American Economic Review*, *98*(1), 145-179.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review*, *53*(5), 941-973.

Bartling, B., Fehr, E., Maréchal, M. A., & Schunk, D. (2009). Egalitarianism and competitiveness. *The American Economic Review*, *99*(2), 93-98.

Berry, S. T. (1992). Estimation of a Model of Entry in the Airline Industry. *Econometrica*, 889-917.

Berwick, Robert C., Angela D. Friederici, Noam Chomsky, and Johan J. Bolhuis (2013). "Evolution, brain, and the nature of language." *Trends in cognitive sciences* 17.2: 89-98.

Blume, A., DeJong, D. V., Kim, Y. G., & Sprinkle, G. B. (1998). Experimental evidence on the evolution of meaning of messages in sender-receiver games. *American Economic Review*, 1323-1340.

Blume, A., DeJong, D. V., Kim, Y. G., & Sprinkle, G. B. (2001). Evolution of communication with partial common interest. *Games and Economic Behavior*, *37*(1), 79-120.

Bolton, P., Freixas, X., & Shapiro, J. (2007). Conflicts of interest, information provision, and competition in the financial services industry. *Journal of Financial Economics*, *85*(2), 297-330.

Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, *49*(1), 116-130.

Cabrales, A., & Gottardi, P. (2014). Markets for information: Of inefficient firewalls and efficient monopolies. *Games and Economic Behavior*, *83*(1), 24-44.

Cai, H., & Wang, J. T. Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, *56*(1), 7-36.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601.

Charness, G., & Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, *48*(1), 133-146.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*(5), 1193-1235.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431-1451.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1-38.

Dickhaut, J. W., McCabe, K. A., & Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, *6*(3), 389-403.

Erev, I., & Rapoport, A. (1998). Coordination, “magic,” and reinforcement learning in a market entry game. *Games and Economic Behavior*, *23*(2), 146-175.

Ericson, R., & Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, *62*(1), 53-82.

Fehr, E., Glätzle-Rützler, D., & Sutter, M. (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review*, *64*(C), 369-383.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171-178.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, *108*(5), 874-904.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, *95*(1), 384-394.

Haas-Wilson, D. (2001). Arrow and the information market failure in health care: The changing content and sources of health care information. *Journal of Health Politics, Policy and Law*, *26*(5), 1031-1044.

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, *12*(2), 180-192.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *The RAND Journal of Economics, 30*(2), 214-231.

López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, *16*(3), 233-247.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, *76*(4), 1359-1395.

Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory*, *134*(1), 93-116.

Kawagoe, T., & Takizawa, H. (1999). Instability of babbling equilibrium in cheap talk games. Saitama University.

Kawagoe, T., & Takizawa, H. (2005). Why lying pays: Truth bias in the communication with conflicting interests. *Available at SSRN 691641*.

Le Quement, Mark T. (2016). "The (human) sampler's curses." *American Economic Journal: Microeconomics* 8.4: 115-48.

Maggian, V., & Villeval, M.C. (2016). Social preferences and lying aversion in children. *Experimental Economics*, *19*, 663-685

Morgan, J., Steiglitz, K., & Reis, G. (2003). The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy*, *2*(1), Art. 5.

Nowak, Martin A., and David C. Krakauer (1999). "The evolution of language." *Proceedings of the National Academy of Sciences* 96.14: 8028-8033.

Sánchez-Pagés, S., & Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, *61*(1), 86-112.

Sobel, J. (2013). Giving and receiving advice, in *Advances in Economics and Econometrics*, D. Acemoglu, M. Arellano, and E. Dekel (eds.).

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, *119*(534), 47-60.

Wang, J. T. Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's Pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, *100*(3), 984-1007.

Wilcox, Nathaniel T. "Theories of learning in games and heterogeneity bias." *Econometrica* 74.5 (2006): 1271-1292.

ONLINE APPENDIX

A) EQUILIBRIUM ANALYSIS OF THE SENDER-RECEIVER GAME

In Table 10 below, we present the expected payoffs to players 1 and 2 associated with all of the pure strategy profiles and identify the Bayes-Nash equilibria (in grey), i.e., the candidates for perfect Bayesian equilibria.[20] In the rows of Table 10, we list all of the strategy profiles of player 1 and in the columns those of player 2. A strategy of player 1 (the sender) is a vector of four components $(c_{ni,ni}, c_{ni,i}, c_{i,ni}, c_{i,i}) \in \{0,1\}^4$, as defined in Section 4.1. A strategy of player 2 (the receiver) is a vector of two components $(c_{m\_ni}, c_{m\_i}) \in \{L, C, R\}^2$. The component c_{m_ni} represents player 2's choice in the information set in which player 1's message says that player 2 is not interested in the object (i.e., the color reported in the message does not coincide with player 2's assigned color); c_{m_i} represents player 2's choice in the information set in which player 1's message says that player 2 is interested in the object. For each information set, the choice $c = L$ indicates that player 2 chooses action *left*, the choice $c = C$ indicates that player 2 chooses action *center*, and the choice $c = R$ indicates that player 2 chooses action $right$.

Each cell of Table 10 contains a vector with the (*ex-ante*) expected payoffs to player 1 and player 2, associated with the respective strategies of players 1 and 2 indicated by row and column. The expected payoffs are computed using the payoffs contained in Table 1, taking into account that each of the four possible combinations of players 1 and 2 being interested/not interested in the object has an *ex ante* probability of 1/4.

In this table, we also mark in bold the best responses of players 1 and 2 and fill in grey those cells in which the strategies of players 1 and 2 are mutual best responses, i.e., the (pure strategy) Bayes-Nash equilibria.

We find that there are 10 Bayes-Nash equilibria that can be grouped into 3 classes (*informative* equilibria, *babbling* equilibria and *opposite* equilibria). The equilibria within each class are informationally equivalent and only differ: (i) in the use of colors, as each equilibrium has a reverse one, and (ii) in the case of the babbling and opposite equilibria, also in the choice of player 2 in one of his information sets, which can be either *left* or *center*.

[20] We then check which Bayes-Nash equilibria are perfect Bayesian equilibria.

Table 10. Bayes-Nash equilibria

	(L,L)	(L,C)	(L,R)	(C,L)	(C,C)	(C,R)	(R,L)	(R,C)	(R,R)
(1, 1, 1, 1)	(**20.0**, 67.5)	(32.5, 55.0)	(45.0, 35.0)	(32.5, 92.5)	(**45.0**, 80.0)	(57.5, 60.0)	(45.0, **110**)	(57.5, 97.5)	(**70.0**, 77.5)
(1, 1, 1, 0)	(**20.0**, 67.5)	(20.0, 57.5)	(20.0, 40.0)	(**45.0**, 90.0)	(**45.0**, 80.0)	(45.0, 62.5)	(**70.0**, **105**)	(**70.0**, 95.0)	(**70.0**, 77.5)
(1, 1, 0, 1)	(**20.0**, 67.5)	(**45.0**, 70.0)	(**70.0**, 57.5)	(20.0, 77.5)	(**45.0**, 80.0)	(**70.0**, 67.5)	(20.0, 87.5)	(45.0, **90.0**)	(**70.0**, 77.5)
(1, 0, 1, 1)	(**20.0**, 67.5)	(32.5, 65.0)	(45.0, 62.5)	(32.5, **82.5**)	(**45.0**, 80.0)	(57.5, 77.5)	(45.0, **82.5**)	(57.5, 80.0)	(**70.0**, 77.5)
(0, 1, 1, 1)	(**20.0**, 67.5)	(32.5, 65.0)	(45.0, 55.0)	(32.5, 82.5)	(**45.0**, 80.0)	(57.5, 70.0)	(45.0, **90.0**)	(57.5, 87.5)	(**70.0**, 77.5)
(1, 1, 0, 0)	(**20.0**, 67.5)	(32.5, 72.5)	(45.0, 62.5)	(32.5, 75.0)	(**45.0**, 80.0)	(57.5, 70.0)	(45.0, 82.5)	(57.5, **87.5**)	(**70.0**, 77.5)
(1, 0, 1, 0)	(**20.0**, 67.5)	(20.0, 67.5)	(20.0, 67.5)	(**45.0**, **80.0**)	(**45.0**, **80.0**)	(45.0, **80.0**)	(**70.0**, 77.5)	(**70.0**, 77.5)	(**70.0**, 77.5)
(0, 1, 1, 0)	(**20.0**, 67.5)	(20.0, 67.5)	(20.0, 60.0)	(**45.0**, 80.0)	(**45.0**, 80.0)	(45.0, 72.5)	(**70.0**, **85.0**)	(**70.0**, **85.0**)	(**70.0**, 77.5)
(1, 0, 0, 1)	(**20.0**, 67.5)	(**45.0**, 80.0)	(**70.0**, **85.0**)	(20.0, 67.5)	(**45.0**, 80.0)	(**70.0**, **85.0**)	(20.0, 60.0)	(45.0, 72.5)	(**70.0**, 77.5)
(0, 1, 0, 1)	(**20.0**, 67.5)	(**45.0**, **80.0**)	(**70.0**, 77.5)	(20.0, 67.5)	(**45.0**, **80.0**)	(**70.0**, 77.5)	(20.0, 67.5)	(45.0, **80.0**)	(**70.0**, 77.5)
(0, 0, 1, 1)	(**20.0**, 67.5)	(32.5, 75.0)	(45.0, 82.5)	(32.5, 72.5)	(**45.0**, 80.0)	(57.5, **87.5**)	(45.0, 62.5)	(57.5, 70.0)	(**70.0**, 77.5)
(1, 0, 0, 0)	(**20.0**, 67.5)	(32.5, 82.5)	(45.0, **90.0**)	(32.5, 65.0)	(**45.0**, 80.0)	(57.5, 87.5)	(45.0, 55.0)	(57.5, 70.0)	(**70.0**, 77.5)
(0, 1, 0, 0)	(**20.0**, 67.5)	(32.5, **82.5**)	(45.0, **82.5**)	(32.5, 65.0)	(**45.0**, 80.0)	(57.5, 80.0)	(45.0, 62.5)	(57.5, 77.5)	(**70.0**, 77.5)
(0, 0, 1, 0)	(**20.0**, 67.5)	(20.0, 77.5)	(20.0, 87.5)	(**45.0**, 70.0)	(**45.0**, 80.0)	(45.0, **90.0**)	(**70.0**, 57.5)	(**70.0**, 67.5)	(**70.0**, 77.5)
(0, 0, 0, 1)	(**20.0**, 67.5)	(**45.0**, 90.0)	(**70.0**, **105**)	(20.0, 57.5)	(**45.0**, 80.0)	(**70.0**, 95.0)	(20.0, 40.0)	(45.0, 62.5)	(**70.0**, 77.5)
(0, 0, 0, 0)	(**20.0**, 67.5)	(32.5, 92.5)	(45.0, **110**)	(32.5, 55.0)	(**45.0**, 80.0)	(57.5, 97.5)	(45.0, 35.0)	(57.5, 60.0)	(**70.0**, 77.5)

The equilibria for this game are as follows:

1.- *Informative equilibrium*:

- $((1,1,1,0),(R,L))$
- $((0,0,0,1),(L,R))$

The expected payoffs for players 1 and 2 are 70 and 105, respectively.

2.- *Babbling equilibrium*:

- $((1,0,1,0),(C,C\ or\ L))$
- $((0,1,0,1),(C\ or\ L,C))$

The expected payoffs for players 1 and 2 are 45 and 80, respectively.

3.- *Opposite equilibrium*:

- $((0,1,1,0),(R,L\ or\ C))$
- $((1,0,0,1),(L\ or\ C,R))$

The expected payoffs for players 1 and 2 are 70 and 85, respectively

We now check that all the Bayes-Nash equilibria we found in Table 9 also constitute (weak) perfect Bayesian equilibria, hereafter PBE.

1.- *Informative equilibrium*

As both profiles are informationally equivalent (they only differ in the use of colors), let us consider the profile $((1,1,1,0),(R,L))$.

1.i) *Player 2's beliefs*

If the message of player 1 says that player 2 is not interested in the object, then, given the strategy of player 1, the beliefs of player 2 derived from Bayes' rule assign equal probability (1/3) to the following three events: (i) player 2 is not interested in the object and player 1 is interested, (ii) neither player 2 nor player 1 is interested in the object, and (iii) both players 2 and 1 are interested in the object.

If the message of player 1 says that player 2 is interested in the object, then, given the strategy of player 1, the beliefs of player 2 assign probability 1 to the following event: player 2 is interested in the object, and player 1 is not interested.

1.ii) *Sequential rationality of player 2*

If the message of player 1 says that player 2 is not interested in the object, then, given the beliefs above, the expected payoffs of player 2 associated with the choices L, C and R are 110/3, 200/3 and 260/3, respectively (see Table 7). Thus, the choice $c_{m\_ni} = R$ prescribed by the strategy of player 2 is sequentially rational.

If the message of player 1 says that player 2 is interested in the object, then, given the beliefs above, the expected payoffs of player 2 associated with the choices L, C and R are 160, 120 and 50, respectively. Thus the choice $c_{m\_i} = L$ prescribed by the strategy of player 2 is sequentially rational.

1.iii) *Sequential rationality of player 1*

If player 1 is not interested in the object, his payoff to player 1 is 20, regardless the choices of players 1 and 2. Thus, the choices $c_{ni,ni} = 1$ and $c_{ni,i} = 1$ prescribed by the strategy of player 1 are sequentially rational.

If player 1 is interested in the object and player 2 is not interested, then, given the strategy of player 2, (R, L), the payoffs to player 1 associated with sending a true and false message are, respectively, 120 and 20. Thus the choice $c_{i,ni} = 1$ prescribed by the strategy of player 1 is sequentially rational.

If both player 1 and player 2 are interested in the object, then, given the strategy of player 2, (R, L), the payoffs to player 1 associated with sending a true and a false message are, respectively, 20 and 120 (see Table 1). Thus the choice $c_{i,i} = 0$ prescribed by the strategy of player 1 is sequentially rational.

Hence, $((1,1,1,0),(R,L))$ is a PBE, and therefore, $((0,0,0,1),(L,R))$ is also a PBE.

2.- *Babbling equilibrium*

Because we have two sets of profiles that are informationally equivalent (they only differ in the use of colors), let us focus on the profiles $\big((1,0,1,0),(C,C)\big)$ and $((1,0,1,0),(C,L))$

2.i) *Player 2's beliefs*

In this case, the message of player 1 is not correlated with the state of the world: it always says that player 2 is not interested. Thus, if the message of player 1 says that player 2 is not interested in the object, then the beliefs of player 2 using Bayes' rule assign equal probability (1/4) to each of the four possible events regarding whether players 1 and 2 are interested in the object.

If the message of player 1 says that player 2 is interested in the object (which does not happen on the equilibrium path), then beliefs cannot be determined by Bayes' rule and are specified below.

2.ii) *Sequential rationality of player 2*

If the message of player 1 says that player 2 is not interested in the object, then, given the beliefs above, the expected payoffs to player 2 associated with the choices L, C and R are 270/4, 320/4 and 310/4, respectively. Thus the choice $c_{m\_ni} = C$ prescribed by the strategy of player 2 is sequentially rational.

If the message of player 1 says that player 2 is interested in the object, then we can find beliefs such that both the choices (i) $c_{m\_i} = C$ and (ii) $c_{m\_i} = L$ are sequentially rational (the beliefs are free in this case). For instance, if the beliefs in this information set are the same as in the former one (i.e., all four possible events have the same probability), then the choice $c_{m\_i} = C$ is sequentially rational. Alternatively, if the beliefs in this information set assign probability 1 to the event in which player 2 is interested in the object and player 1 is not, then the choice $c_{m\_i} = L$ is sequentially rational.

2.iii) *Sequential rationality of player 1*

If player 1 is not interested in the object, then the payoff to player 1 is 20, regardless of the choices of players 1 and 2. Thus the choices $c_{ni,ni} = 1$ and $c_{ni,i} = 0$ prescribed by the strategy of player 1 are sequentially rational.

If player 1 is interested in the object and player 2 is not interested, then we have the following:

- Given the strategy of player 2 in the first equilibrium in this class, $\text{i.e.}, (C, C)$, the payoff of player 1 is 70 regardless of whether he sends a true or a false message. Thus, the choice $c_{i,ni} = 1$ prescribed by the strategy of player 1 is sequentially rational.
- Given the strategy of player 2 in the second equilibrium in this class, $\text{i.e.}, (C, L)$, the payoffs of player 1 associated with sending a true and a false message are 70 and 20, respectively. Thus, the choice $c_{i,ni} = 1$ prescribed by the strategy of player 1 is sequentially rational.

If both players 1 and 2 are interested in the object, then we have the following:

- Given the strategy of player 2 in the first equilibrium in this class, $\text{i.e.}, (C, C)$, the payoff of player 1 is 70 regardless of whether he sends a true or a false message. Thus, the choice $c_{i,ni} = 0$ prescribed by the strategy of player 1 is sequentially rational.
- Given the strategy of player 2 in the second equilibrium in this class, $\text{i.e.}, (C, L)$, the payoffs of player 1 associated with sending a true and a false message are,20 and 70, respectively. Thus, the choice $c_{i,ni} = 0$ prescribed by the strategy of player 1 is sequentially rational.

Hence, the profiles $\big((1,0,1,0), (C, L \text{ } or \text{ } C)\big)$ are PBE, and therefore, $((0,1,0,1), (L \text{ } or \text{ } C, C))$ also are PBE.

3.- *Opposite equilibrium*

As we have two sets of profiles that are informationally equivalent (they only differ in the use of colors), let us focus on the profiles $((0,1,1,0), (R, L))$ and $((0,1,1,0), (R, C))$

3.i) *Player 2's beliefs*

If the message of player 1 says that player 2 is not interested in the object, then, given the strategy of player 1, the beliefs derived from Bayes' rule of player 2 assign equal probability (1/2) to the following two events: (i) player 2 is not interested and player 1 is interested in the object, and (ii) both players 2 and 1 are interested in the object.

If the message of player 1 says that player 2 is interested in the object, then, given the strategy of player 1, the beliefs of player 2 assign equal probability (1/2) to the following two events: (i) neither player 2 nor player 1 is interested in the object, and (ii) player 2 is interested in the object and player 1 is not interested.

3.ii) *Sequential rationality of player 2*

If the message of player 1 says that player 2 is not interested in the object, then, given the beliefs above, the expected payoffs of player 2 associated with the choices L, C and R are 90/2, 140/2 and 160/2, respectively. Thus the choice $c_{m\_ni} = R$ prescribed by the strategy of player 2 is sequentially rational.

If the message of player 1 says that player 2 is interested in the object, then, given the beliefs above, the expected payoffs of player 2 associated with the choices L, C and R are 180/2, 180/2 and 150/2, respectively. Thus, the choices $c_{m\_i} = L$ and $c_{m\_i} = C$ are sequentially rational.

3.iii) *Sequential rationality of player 1*

If player 1 is not interested in the object, then his payoff is 20, regardless of the choices of players 1 and 2. Thus, the choices $c_{ni,ni} = 0$ and $c_{ni,i} = 1$ prescribed by the strategy of player 1 are sequentially rational.

If player 1 is interested in the object and player 2 is not interested, then we have the following:

- Given the strategy of player 2 in the first equilibrium in this class, i.e.(R, L)), the payoffs of player 1 associated with sending a true and a false message are 120 and 20, respectively. Thus, the choice $c_{i,ni} = 1$ prescribed by the strategy of player 1 is sequentially rational.
- Given the strategy of player 2 in the second equilibrium in this class, i.e., (R, C), the payoffs of player 1 associated with sending a true and a false message are 120 and 70, respectively. Thus, the choice $c_{i,ni} = 1$ prescribed by the strategy of player 1 is sequentially rational.

If both player 1 and player 2 are interested in the object, then we have the following:

- Given the strategy of player 2 in the first equilibrium in this class, $\text{i.e.}, (R, L)$, the payoffs of player 1 associated with sending a true and a false message are 20 and 120, respectively. Thus, the choice $c_{i,i} = 0$ prescribed by the strategy of player 1 is sequentially rational.

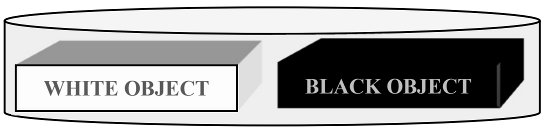
- Given the strategy of player 2 in the second equilibrium in this class, i.e., (R, C), the payoffs of player 1 associated with sending a true and a false message are 70 and 120, respectively. Thus, the choice $c_{i,i} = 0$ prescribed by the strategy of player 1 is sequentially rational.

Hence, $((0,1,1,0), (R, L\ or\ C))$ is a PBE, and therefore, $((1,0,0,1), (L\ or\ C, R))$ is also a PBE.

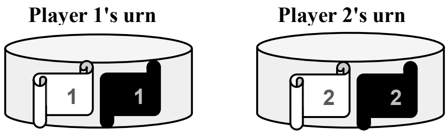
B) EXPERIMENTAL INSTRUCTIONS

The aim of this experiment is to study how individuals make decisions in certain contexts. The instructions are simple. If you follow them carefully you will earn a non-negligible amount of money in cash (euros) at the end of the experiment. During the experiment, your earnings will be in ECUs (experimental currency units). Individual payments will remain private, as nobody will know the other participants' payments. Any communication among you is strictly forbidden and will result in immediate exclusion from the experiment.

1. The experiment consists of 40 rounds. In each round, you will be randomly assigned to a group of 2 participants (including yourself). This group is determined randomly at the beginning of the round. Therefore, the group you are assigned to changes at each round. In this room, there are 4 participants (including yourself) who are potential members of your group. That is, at every round, your group is selected among these 4 participants, each of them being equally likely to be in your group. You will not know the identities of any of these participants. In each round, you will only interact with the other participant in your group, and your payoff will only depend on your choice and the choice of the other participant in your group.
2. In each round, one of the two participants in your group will have the role of player 1 and the other one will have the role of player 2. The roles will be randomly assigned, and both participants in a group are equally likely to have each role assigned. At the beginning of the round, each participant will be informed of his/her assigned role.
3. At the beginning of the round, the computer randomly draws one object from an (virtual) urn containing two objects: one white object and one black object. Each object is picked with equal probability (50%).



   The color of the object is revealed to player 1 but not to player 2 in your group.
4. At each round, each player is assigned a color. At the beginning of the round, the color assigned to each player is determined in the following way. There is one (virtual) urn for each player, containing two pieces of paper: one white and one black. The computer randomly (and independently) draws one piece of paper from each urn. In each urn, each piece of paper is picked with equal probability (50%). The piece of paper selected for each player determines that player's assigned color.



In each group, player 1 is informed both of his/her assigned color for the round and of the color assigned to player 2. Player 2 is only informed of his/her assigned color but not of the color assigned to player 1.

5. At each round, in each group, player 1 will be the first to make his/her decisions, knowing the color of the object drawn from the computer, his/her assigned color, and the assigned color of player 2. Player 1 has to decide what message to send to player 2 regarding the color of the object (which is unknown by player 2). The message can be either "*The object is white*" or "*The object is black*". Thus, the message can contain the true color or the false one.
6. Then, player 2, being informed of his/her assigned color (but neither of the color of the object nor of the color assigned to player 1), observes the content of the message sent by player 1 and decides which action to take: *Left*, *Center* or *Right*.
7. Round payoffs. At each round, the payoff to each player depends on whether the color of the object did or did not match his/her assigned color and on the action chosen by player 2:
   i. At each round, the payoff to player 1 is determined as follows.
      - If the color of the object is equal to the color assigned to player 1, then his/her payoff depends on the choice of player 2 in the following way:
        - 20 ECUs if player 2 has chosen *Left*
        - 70 ECUs if player 2 has chosen *Center*
        - 120 ECUs if player 2 has chosen *Right*
      - If the color of the object is different from the color assigned to player 1, then his/her payoff is 20 ECUs, regardless of the action chosen by player 2.

   ii. At each round, the payoff of player 2 is determined as follows.
      - If the color of the object is equal to the color assigned to player 2, then action *Left* provides him/her with a higher payoff than action *Center* or *Right*, and action *Center* provides him/her with a higher payoff than action *Right*.
      - If the color of the object is different from the color assigned to player 2, then action *Right* provides him/her with a higher payoff than action *Center* or *Left*, and action *Center* provides him/her with a higher payoff than action *Left*.

- The payoff of player 2 also depends on the correspondence between his/her assigned color and the color assigned to player 1: the payoff for player 2 when his/her assigned color is the same than the assigned color of player 1 is *lower* than in the case in which his/her assigned color is different from the color assigned to player 1.

The four tables below provide the payoffs of player 1 and player 2 in all possible situations:

- The top-left table corresponds to the cases in which both players have the same assigned color, which is different from the color of the object;
- The top-right table corresponds to the cases in which the color of the object is equal to the color assigned to player 2 but different from the color assigned to player 1;
- The bottom-left table corresponds to the cases in which the color of the object is equal to the color assigned to player 1 but different from the color assigned to player 2;
- The bottom-right table corresponds to the cases in which both players have the same assigned color, which is equal to the color of the object.

(Object: White Player 2: Black Player 1: Black) or (Object: Black Player 2: White Player 1: White)			
Action of player 2	Left	Center	Right
Payoff to player 1	20	20	20
Payoff to player 2	20	60	100

(Object: White Player 2: White Player 1: Black) or (Object: Black Player 2: Black Player 1: White)			
Action of player 2	Left	Center	Right
Payoff to player 1	20	20	20
Payoff to player 2	160	120	50

(Object: White Player 2: Black Player 1: White) or (Object: Black Player 2: White Player 1: Black)			
Action of player 2	Left	Center	Right
Payoff to player 1	20	70	120
Payoff to player 2	30	90	120

(Object: White Player 2: White Player 1: White) or (Object: Black Player 2: Black Player 1: Black)			
Action of player 2	Left	Center	Right
Payoff to player 1	20	70	120
Payoff to player 2	60	50	40

8. At the end of each round, prior to proceeding to the next round, all the players are informed about current and past rounds: assigned role, color of the object, color assigned to each player in the group, the message of player 1, the action chosen by player 2 and the payoff of each player.

9. Payments. At the end of the experiment, you will be paid the earnings that you obtained in 8 rounds (out of 40). These rounds will be randomly selected by the computer: 4 rounds will be selected from the rounds in which you were assigned the role of player 1 and the other 4 rounds will be selected from the rounds in which you were assigned the role of player 2. The earnings that you have obtained in the selected rounds will be converted into cash at the exchange rate of 40 ECUs = 1 euro and will be paid to you in private.

C) ADDITIONAL ECONOMETRIC ANALYSIS

Table 11. Determinants of the probability that player 1 sends a true message (logit estimation)

NoInt1 × NoInt2 × Prosoc × NoEnv	0.4020 (0.3575)
NoInt1 × NoInt2 × NoProsoc × Env	-0.8442** (0.4294)
NoInt1 × NoInt2 × NoProsoc × NoEnv	0.2168 (0.4534)
NoInt1 × Int2 × Prosoc × Env	0.0843 (0.3999)
NoInt1 × Int2 × Prosoc × NoEnv	0.4415 (0.3695)
NoInt1 × Int2 × NoProsoc × Env	-1.0487** (0.4714)
NoInt1 × Int2 × NoProsoc × NoEnv	0.6609 (0.5180)
Int1 × NoInt2 × Prosoc × Env	0.3446 (0.4069)
Int1 × NoInt2 × Prosoc × NoEnv	1.0806*** (0.4141)
Int1 × NoInt2 × NoProsoc × Env	-0.5787 (0.4610)
Int1 × NoInt2 × NoProsoc × NoEnv	1.5388** (0.6766)
Int1 × Int2 × Prosoc × Env	-0.9592** (0.3900)
Int1 × Int2 × Prosoc × NoEnv	-0.9540*** (0.3674)
Int1 × Int2 × NoProsoc × Env	-1.6444*** (0.5284)
Int1 × Int2 × NoProsoc × NoEnv	-1.6643*** (0.5149)
Round	0.0014 (0.0067)
Risk	-0.1276** (0.0560)
Constant	0.8704** (0.3578)

[Number of obs = 800], [***, **, * significant at the 1%, 5%, and 10% level, respectively]

D) TECHNICAL DETAILS OF THE MIXTURE MODEL ESTIMATION

For the estimation of the mixture model, let $i = 1, \dots, N$ index the different players and $k = 1, \dots, K$ index our types. We assume that a type-k player normally makes a type k decision, but in each period, he makes an error with probability $\varepsilon_k \in [0, 1]$, constituting type k's *error rate*, in which case he chooses to send a true or a false message with equal probability $\frac{1}{2}$. For a type-k player, the probability of a type k decision in any information set is then $1 - \frac{1}{2}\varepsilon_k$. Hence, the probability of a non-type-k decision is $\frac{\varepsilon_k}{2}$. We assume that errors are independently and identically distributed across periods and players and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)$.

The likelihood function can be constructed as follows. Let T_i denote the total number of periods in which player i acted as sender. Next, let x_{ik} denote the number of player i's decisions that equal type k's in periods in which he acts as a sender and $x_i = (x_{i1}, \dots, x_{iK})$, $x = (x_1, \dots, x_i, \dots, x_N)$. Let p_k denote the common probability that a player is of type k, $\sum_{k=1}^{K} p_k = 1$ and $p = (p_1, \dots, p_K)$. As each period has one type-k decision and one non-type-k decision, the probability of observing a particular sample with x_{ik} type-k decisions when player I is type k can be written as follows:

$$L_k^i(\varepsilon_k | x_{ik}) = \left[1 - \frac{1}{2}\varepsilon_k\right]^{x_{ik}} \left[\frac{1}{2}\varepsilon_k\right]^{T_i - x_{ik}}$$

Weighting the right-hand side by p_k, summing over k, taking logarithms, and summing over i yields the log-likelihood function for the entire sample:

$$lnL(p, \varepsilon | x) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} p_k L_k^i(\varepsilon_k | x_{ik})$$

This function is maximized by the EM algorithm.[21]

Note that, whatever the player's type, ε_k=1 would be equivalent to (uninformative) pooling, with an associated best response on the side of the receiver consisting on always playing action *center* (C), regardless the received message. Hence, for each type k, we can compute the threshold t_k such that, for each $\varepsilon_k \geq t_k$, from the perspective of the receiver, the sender's behavior is equivalent to pooling or, in other words, it induces a best response consisting on always playing action *cente*r. The thresholds associated to the types corresponding to the most parsimonious models

[21] As proposed in the seminal paper by Dempster, Laird and Rubin (1977).

reported in Table 8 (T1, T2, T3 and T4), together with the estimated error rates and the (one-sided) p-value for the error rate being lower than the threshold are displayed in Table 12.

Table 12. Thresholds for the error rates and tests of error rates with respect to threshold

		Rounds 1-20		*Rounds 21-40*	
Type	*Threshold (t_k)*	*Error rate*	*p-value*	*Error rate*	*p-value*
T1 (1,1,1,0)	12/13	0.225^^^ (0.068)	0.0000	0.179^^^ (0.115)	0.0000
T2 (1,1,1,1)	14/15	0.672 (0.236)	0.1368	.252^^^ (0.263)	0.0050
T3 (0,0,0,0)	14/15	0.586^^ (0.182)	0.0296		
T4 (0,1,1,0)	4/5			0.782 (0.155)	0.4538

Standard errors within brackets
Significance (with respect to t_k): ^^^, ^^, ^ at the 1%, 5%, 10% level, resp.