

# RULES AND COMMITMENT IN COMMUNICATION

Guillaume Fréchette    Alessandro Lizzeri    Jacopo Perego  
New York University    New York University    Yale University

January 31, 2018

**Preliminary and Incomplete.**

## ABSTRACT

We investigate models of cheap talk, disclosure of verifiable information, and Bayesian persuasion, in a unified experimental framework. We examine how commitment affects the informativeness of communication and how this depends on the verifiability of information. We show that both senders and receivers react to commitment. Results move in the direction predicted by the model, but there are some revealing ways in which they do not correspond to the exact theoretical predictions. In particular, there is considerable heterogeneity among senders and although, on average the sender's behavior is closer to the theory's prediction under unverifiable information; a higher fraction of subjects send equilibrium type messages under verifiable information. Nonetheless, under verifiable information many senders have difficulty using commitment to reduce informativeness to their advantage, but instead they introduce noise that does not benefit them. On the receiver side, rules have an important effect and their behavior is closer to optimal under verifiable information than under unverifiable information.

We thank Andreas Blume, Santiago Oliveros and Salvatore Nunnari for useful comments.

# 1 Introduction

The goal of this paper is to study how *rules* and *commitment* affect the amount of information that can be transmitted between a sender and a receiver who have conflicting interests. The structure of our experimental design allows us to jointly analyze models that share an underlying structure regarding preferences and information, but that are distinguished either by the rules governing communication, i.e., whether the sender can lie about the content of her information, or by the extent to which the sender can commit to her communication plan. These models generate different predictions regarding both the communication strategies that are used and the amount of information that is revealed in equilibrium. With minimal differences between treatments, our common structure ranges from models of cheap talk (Crawford and Sobel (1982)), to models of disclosure (Grossman (1981), Jovanovic (1982), Okuno-Fujiwara et al. (1990)), to models of Bayesian persuasion (Kamenica and Gentzkow (2011)), as well as intermediate cases between these extremes. Hence, we span a considerable portion of the models of strategic information revelation that have been discussed in the literature in the last decades, and we experimentally study novel dimensions of the sender-receiver interaction.

Regarding the rules governing communication, we consider two cases that highlight the following contrast. In the first scenario, senders' messages are *unverifiable*, i.e. senders can freely misreport their private information. In the second scenario, instead, information is *verifiable*, i.e. while information can be hidden, it cannot be misreported.<sup>1</sup> The first scenario corresponds to models of cheap talk (e.g., Crawford and Sobel (1982)), while the latter corresponds to models of disclosure with verifiable information (e.g., Grossman (1981), Milgrom (1981), Okuno-Fujiwara et al. (1990)). We interact these two scenarios with treatments in which we vary the degree of commitment available to the sender. The way we model commitment, is as follows. Before learning the true state, the sender publicly selects an information structure. After observing the state, with probability  $(1 - \rho)$  the sender is then given the opportunity to secretly revise her plan of action. The higher is  $\rho$ , the higher the probability that the sender will not be able to revise her strategy, hence the higher the extent to which she is committed to her initial plan of actions. In the limit case in which  $\rho = 1$  the sender has full commitment and outcomes converge to a Bayesian persuasion

---

<sup>1</sup>The Sender misreports her private information when she sends messages that are *false*. A message is false if none of the statements it contains is true. E.g., provided that the ball is blue, message “the ball is red or black” is false. We will formalize this in Section 2.

model (Kamenica and Gentzkow (2011)).<sup>2</sup>

Novel comparative statics arise from this set up. We focus our attention on the effect that the degree of commitment has on the amount of information that is revealed in equilibrium and how this depend on the verifiability of messages. While in any specific environment and application it is hard to know (and measure) the exact extent of commitment available to an agent, it is natural to think that this can vary and may depend on observable correlates such as the protocols and the frequency of communication. Thus, it seems important to study how communication varies with the extent of commitment.

When messages are unverifiable, the sender seeks commitment because it allows her to credibly tell the truth, at least some of the time. That is, commitment *increases* the amount of information that is revealed in equilibrium. In contrast, in verifiable environments, i.e. when any informative message needs to correspond to the state, the sender still seeks commitment, but for the opposite reason: to credibly hide information, at least some of the time. That is, commitment *decreases* the amount of information that is revealed in equilibrium. We also show that, when senders have full commitment, verifiability has no impact on the amount of information that is communicated. That is, equilibrium informativeness is exactly the same regardless of the verifiability of messages. However, the senders implement this outcome with very different strategies. Thus, the model generates a rich set of predictions that we then bring to the laboratory.

We depart from the previous experimental literature on information transmission in several ways. First, we innovate by conducting an analysis *across* a variety of models. Of course, when performing such an exercise, it is crucial to make sure that all sources of variations coming from seemingly unimportant details of the design are reduced to a minimum, so that differences in outcomes in the data can be imputed to differences in the treatments. In order to do this, we take advantage of our theoretical framework, thanks to which we are able to design an experiment that allows us to move from one model to another by simply changing one of the two parameters, namely the degree of commitment on the sender's part and the verifiability of messages. An additional advantage of considering all these treatments under the same umbrella is that it provides discipline on the explanations that can be used to

---

<sup>2</sup>The (un)verifiability of communications could be modeled in more detail. For instance, verifiable messages could be the equilibrium result of the Sender's aversion to lying: when lying is cheap, communication is unverifiable; when the cost of lying is sufficiently high, communication becomes verifiable. Alternatively, verifiability could be modeled as access, with some probability, to an authority that verifies the truthfulness of the report. When that probability is zero, communication is unverifiable, and when it is one, communication is verifiable.

rationalize potential deviations from theoretical predictions.

A second way in which we depart from the previous experimental literature on cheap talk is that we do not investigate the relationship between the informativeness of communication and the degree of preference alignment between the sender and the receiver. Rather than preference alignment, we focus on the effect of commitment on the amount of information that is revealed in equilibrium.<sup>3</sup>

A third element of novelty in our design is the treatment under full commitment. As discussed above, this treatment coincides with a model of Bayesian persuasion as introduced in [Kamenica and Gentzkow \(2011\)](#). This model has become influential in the recent theoretical literature, e.g. [Gentzkow and Kamenica \(2014\)](#), [Alonso and Camara \(2016\)](#), [Gentzkow and Kamenica \(2016\)](#), etc. Evaluating how the degree of commitment affects outcomes is one way to experimentally evaluate the model of Bayesian persuasion.

Our main results are the following. We first show that subjects understand the power of commitment: senders figure out how to exploit commitment and receivers how to react to it. We show that senders understand commitment by contrasting their behavior in the commitment stage to their behavior in the revision stage in a treatment with partial commitment. The theory predicts that in the case of unverifiable information the sender should reveal more information in the commitment stage than in the revision stage, and that this ranking should be reversed when information is verifiable. The data supports this prediction of the theory. The receiver should understand that information conveyed in the commitment stage is more meaningful when the level of commitment is higher, and this is what we find in the data: receivers are more responsive to information in higher commitment treatments. Second, we investigate whether the senders use commitment in a way that is consistent with Bayesian persuasion. We do this by considering treatments in which we vary the receivers' payoffs so as to change what we call the persuasion threshold, i.e., the minimal amount of information required to persuade the receivers to choose the action desired by the sender. The theory predicts that the sender conveys more information when the persuasion threshold is higher, and this is consistent with what we find in the data. We then consider how commitment affects overall equilibrium informativeness by comparing different levels of commitment. We find that subjects behave in ways that are consistent with theory, that is, informativeness decreases with commitment in the verifiable treatments and increases with commitment in the unverifiable treatments. However, quantitatively there are significant departures from the

---

<sup>3</sup>In doing this, we are also addressing recent theoretical contributions on persuasion under partial commitment, such as [Min \(2017\)](#).

theory: informativeness does not rise enough in the unverifiable treatments and it does not drop enough in the verifiable treatments. Furthermore, commitment seems to work better for the sender when communication is unverifiable. In the limiting case of full commitment,  $\rho = 1$ , a lot more information is revealed when communication is verifiable than when it is not, despite the fact that, in theory, the equilibrium amount of communication is the same in the two treatments. It appears that this is partly due to the fact that Senders find it harder to understand how to strategically hide *good news* than to lie about *bad* ones. We also find that, in the unverifiable treatments with a substantial amount of commitment, receivers are excessively skeptical. This is partly in contrast with prior literature on cheap talk, e.g. [Cai and Wang \(2006\)](#). From a policy perspective, this is a novel justification for making it harder for senders to misreport their information.

Models with unverifiable communication have been used to study a variety of phenomena, including lobbying [Austen-Smith \(1993\)](#), [Battaglini \(2002\)](#); the relation between legislative committees and a legislature, as in e.g., [Gilligan and Krehbiel \(1989\)](#), [Gilligan and Krehbiel \(1987\)](#); the production of evidence to a jury ([Kamenica and Gentzkow \(2011\)](#), [Alonso and Camara \(2016\)](#)). Models of disclosure of verifiable information have been used to study the disclosure of quality by a privately informed seller, for instance, via warranties,<sup>4</sup> of the contents of financial statements by a firm,<sup>5</sup> and in many other contexts. [Dranove and Jin \(2010\)](#) survey the literature on product quality and the disclosure of information.

There are a number of experimental papers on cheap talk. [Blume et al. \(2017\)](#) provides a survey of the experimental literature on communication. [Dickhaut et al. \(1995\)](#) is the first experimental paper to test the central prediction of Crawford and Sobel that more preference alignment between the sender and the receiver should result in more information transmission. Their main result is consistent with this prediction. [Forsythe et al. \(1999\)](#) add a cheap talk communication stage to an adverse selection environment with the feature that the theory predicts no trade and that communication does not help. In the experiment, in contrast, communication leads to additional trade, partly because receivers are too credulous. [Blume et al. \(1998\)](#) study a richer environment and compare behavior when messages have preassigned meanings with behavior when meaning needs to emerge. Among other findings, they confirm that, as in [Forsythe et al. \(1999\)](#), receivers are gullible. [Cai and Wang \(2006\)](#) find that Senders are overly truthful and they also find that receivers are overly trusting, relative to the predictions of the cheap talk model. They also study information revelation

---

<sup>4</sup>E.g., [Grossman \(1981\)](#).

<sup>5</sup>See for instance, [Verrecchia \(1983\)](#), [Dye \(1985\)](#) and [Galor \(1985\)](#).

as players’ preferences become more aligned: consistently with the theory, they find that the amount of information transmission increases with the degree of preference alignment. They then discuss how to reconcile the departures from the predictions of the cheap talk model via a model of cognitive hierarchy and via quantal response equilibrium.<sup>6</sup>

Conversely, experiments on the disclosure of verifiable information typically find that there is under-revelation of information when compared with the theoretical predictions. For instance, [Jin et al. \(2016\)](#) find that receivers are insufficiently skeptical when senders do not provide any information. This in turn leads Senders to underprovide information, thereby undermining the unraveling argument.<sup>7</sup> There are also some papers that study information unraveling with field data. In particular, [Mathios \(2000\)](#) studies the impact of a law requiring nutrition labels for salad dressings. He shows that, prior to mandatory disclosure, low-fat salad dressings posted labels, while a range of high-fat salad dressings chose not to disclose. Mandatory disclosure was followed by reductions in sales for the highest fat dressings. These results are in conflict with the predictions of the unraveling result from the literature on verifiable communication. [Jin and Leslie \(2003\)](#) study the consequences of mandatory hygiene grade cards in restaurants. They show that hygiene cards lead to increases in hygiene scores, that demand becomes more responsive to hygiene, and to lower food borne illness hospitalizations.

## 2 Theory

In the next Section we describe the experimental design and the equilibrium outcomes that obtain in our specific parametrization. In the present Section we describe the framework we consider and the theoretical predictions of the model. In the Appendix we show that the predictions extend to a more general environment.

The state space  $\Theta$  is binary:  $\Theta = \{\theta_L, \theta_H\}$ . There is a common prior  $\mu_0$  that denotes the probability that the state is  $\theta_H$ . There are two players: a sender and a receiver. The sender has information, the receiver has the ability to act. Communication takes place with the sender sending messages in an attempt to influence the action chosen by the receiver. The game has three stages: the sender is active in the first two, while the receiver is active in the third. The receiver chooses actions in a binary set  $A = \{a_L, a_H\}$ . The receiver’s preferences

---

<sup>6</sup>See also [Sánchez-Pagés and Vorsatz \(2007\)](#); [Wang et al. \(2010\)](#).

<sup>7</sup>See also [Forsythe et al. \(1989\)](#); [King and Wallin \(1991\)](#); [Dickhaut et al. \(2003\)](#); [Forsythe et al. \(1999\)](#); [Benndorf et al. \(2015\)](#); [Hagenbach et al. \(2014\)](#);

are given by the following utility function:

$$u(a_L, \theta_L) = u(a_H, \theta_H) = 0, \quad u(a_L, \theta_H) = -(1 - q), \quad u(a_H, \theta_L) = -q.$$

Thus, the receiver wishes to match the actions to the state, and the relative cost of the mistakes in the two states is parametrized by  $q$ . Given these parameters, a Bayesian receiver would choose action  $a_H$  whenever her posterior that the state is  $\theta_H$  is larger than  $q$ . Thus, we call  $q$  the *persuasion threshold*.

We assume that the prior  $\mu$  that the state is high is such that  $\mu < q$  so that, absent any information, the receiver would choose  $a_L$ . Without this assumption the sender would have no reason to attempt to communicate with the receiver.

The sender's preferences are given by  $v(a) := I(a = a_H)$ , i.e., the sender receives a positive payoff only if the receiver chooses  $a_H$ . Thus, the sender would like the receiver to choose the high action.

The sender sends messages to the receiver from a subset of  $M = \{\theta_L, \theta_H, n\}$ . The interpretation of these possible messages is that two of the messages represent statements about the states, while the last one is a "noise" message that has no relation to the state. In the experiment we call  $n$  "no message." Allowing for this message is important in the case of verifiable information. Let  $M^\theta \subseteq M$  be the set of messages that the sender can use in state  $\theta$ . We say that information is *unverifiable* if  $M^\theta = M$  for all  $\theta$ . We say that information is *verifiable* if  $M^\theta = \{\theta, n\}$  for all  $\theta$ . Thus, when information is unverifiable, the sender is allowed to send any message, including any lie about the content of her information. When information is verifiable, the sender cannot lie about her type but can choose not to report the type by choosing the noise message  $n$ .

Messages are generated by the sender in two possible stages, a *commitment stage* in which she publicly chooses an information structure before observing the state, and a *revision stage* in which, after observing the state, she can secretly revise her strategy. The revision stage only occurs with probability  $1 - \rho$  that is exogenous and common knowledge. Thus,  $\rho$  is a measure of the sender's commitment: for high values of  $\rho$  the choice made at the commitment stage is likely to be the relevant one; for low values of  $\rho$  the sender is very likely to have the opportunity to revise the message, so that the choice at the revision stage is likely to be more relevant.

1. **Commitment Stage.** In this stage, the sender chooses an information structure,

which is given by a commitment strategy:  $\pi_C : \Theta \rightarrow \Delta(M^\theta)$  that defines the probability that the sender will send a specific message for any given realization of her information.

2. **Revision Stage.** The sender learns the state  $\theta \in \Theta$ . She now has the chance to revise her initial strategy  $\pi_C$ , by specifying a new probability distribution  $\pi_R : \Theta \rightarrow \Delta(M^\theta)$ .
3. **Guessing Stage.** The receiver chooses  $a : M \times \Pi_c \rightarrow A$ . In words, the receiver observes the commitment strategy  $\pi_C$  but does not observe the revision strategy  $\pi_R$ . The receiver also observes a message  $m$  which is generated with probability  $\rho$  from  $\pi_C$  and  $(1 - \rho)$  from  $\pi_R$ . The receiver makes her best guess about  $\theta$ , given each possible message  $m \in M$  she might receive.

One possible interpretation of the game is that the revision game is always available but the sender has a type that determines whether she will take advantage of the opportunity to revise the strategy. The parameter  $\rho$  is then the probability that the sender is not this opportunistic type.

This game includes interesting classic models as special extreme cases. When  $\rho = 0$  and information is unverifiable, this is a cheap talk model. In the binary model, in this case, the unique equilibrium outcome involves no information transmission. When  $\rho = 0$  and information is verifiable, this is a model of disclosure. The unique equilibrium in this case involves full information transmission. When  $\rho = 1$  and information is unverifiable, this is a model of Bayesian persuasion. The unique equilibrium outcome involves partial information transmission.

A key variable in our analysis is the amount of information that is transmitted from the sender to the receiver. We measure the amount of informativeness of communication in two ways: the correlation between the state and the action, and the dispersion of posteriors by state. In our discussion of the data these two measures are both useful and highlight different aspects of the phenomena of interest. However, our main theoretical results are independent of the specific measure of informativeness. When we say that there is a positive amount of information in equilibrium we mean that the equilibrium messages are correlated with the state. When we say that there is less than maximal information, we mean that this correlation is less than one.

Our first result considers a given level of commitment  $\rho$  and compares the informativeness of communication between the commitment and the revision stage.



**Proposition 1** (i) *There is a  $\hat{\rho}$  such that, if  $\rho > \hat{\rho}$ , in equilibrium, there is positive information in unverifiable treatments and less than maximal information in verifiable treatments.* (ii) *Consider  $\rho$  such that  $\hat{\rho} < \rho < 1$ . Under verifiable information, more information is transmitted at the revision stage than in the commitment stage; under unverifiable information less information is transmitted at the revision stage than in the commitment stage.*

This result highlights the tension between the commitment and the revision stage, and how this tension manifests itself differently under the different verifiability scenarios. In order to understand this result it is useful to consider the extreme cases: when  $\rho = 0$ , full disclosure takes place under verifiable information and no information is communicated under unverifiable information. When  $\rho = 1$ , it turns out that both scenarios lead to the same outcome of Bayesian persuasion and partial information revelation (see Proposition below).

The intuition for Proposition 1 is the following. Under both verifiable and unverifiable information, the sender would like to commit to persuade the receiver to choose the high action as often as possible, and this requires partial information revelation. However, at the revision stage, the sender is unable to resist the temptation to use her opportunity to “cheat” and manipulate information in her favor. Under verifiable information this opportunity implies full information disclosure in the revision stage; under unverifiable information, it implies always stating the same favorable signal regardless of the state. When  $\rho$  is high, some commitment is possible because the revision stage cannot completely undo the positive effect of the commitment stage, explaining part (i) of the result. However, the revision stage partially undoes the information communicated in the commitment stage, and this undoing goes in opposite directions in verifiable relative to unverifiable information, explaining part (ii) of this result. Part (ii) also says that the role of commitment in facilitating information transmission is completely reversed when we move from non verifiable to verifiable information.

Our next result describes how equilibrium informativeness changes with the degree of commitment, and how this depends on the verifiability of messages.

**Proposition 2** (i) *When messages are unverifiable, increasing commitment increases informativeness.* (ii) *When messages are verifiable, increasing commitment decreases informativeness.* (iii) *When  $\rho = 1$ , the equilibrium outcome is independent of the mode of communication: under full commitment verifiability does not matter.*

Thus, increasing commitment affects equilibrium informativeness in opposite ways depending on whether communication is verifiable or not, but the verifiability of messages has no effect on the amount of information transmitted in equilibrium when the sender has full commitment.

We now consider another interesting comparative statics result that can be tested in our environment. This keeps fixed the degree of commitment and discusses how the equilibrium changes as we vary the persuasion threshold parameter  $q$  that measures the required posterior to induce the sender to choose  $a_H$ .

**Proposition 3** *For any  $\rho > 0$ , for both cases of verifiable and unverifiable messages, as the persuasion threshold  $q$  increases, the strategy of the sender becomes more informative.*

As the persuasion threshold increases, the sender must reveal more information in order to induce the receiver to choose the high action. Of course, for low values of  $\rho$ , in the unverifiable scenario, no information can be credibly transmitted in equilibrium for any value of  $q$  because the revision stage is too dominant in the game, and the revision stage is never informative in unverifiable scenarios. However, for  $\rho$  high enough, increasing  $q$  will strictly increase the amount of information transmitted.

## 3 Design

### 3.1 Basic structure

We now describe the specific parametrization in our experimental design.

An urn contains three balls, two blue ( $B$ ) and one red ( $R$ ). The color of the ball represents the state of nature, i.e.  $\theta \in \Theta := \{R, B\}$ . One ball is drawn at random from the urn and, therefore, the *prior* probability that the ball is red is given by  $\mu_0(\theta = R) = \frac{1}{3}$ . Since the state space is binary, we denote the posterior beliefs by  $\mu \in [0, 1]$  as the probability that the receiver attributes to the ball being red. The set of messages in the unverifiable case is  $M = \{r, b, n\}$ : red, blue and no message.<sup>8</sup> Thus,  $r$ , and  $b$  are the type-messages and  $n$  is the noise-message. In verifiable treatments  $M^r = \{r, n\}$ ,  $M^b = \{b, n\}$ : under verifiable information it is not possible to send  $m = r$  when  $\theta = B$  or  $m = b$  when  $\theta = R$ . No restriction is imposed on the probability of  $m = n$ . The two players, sender and receiver, are assumed to have preferences over money that are analogous to those introduced in the

---

<sup>8</sup>See Appendix D for more details about the language used in the laboratory.

previous section, up to a normalization. The receiver wins \$2 if she correctly guesses the state, and wins nothing otherwise. The sender wins \$2 if the receiver guesses  $a = R$  and nothing otherwise. Payoffs are summarized in Table 1.

	$\theta = R$		$\theta = B$	
$a = R$	Receiver \$2	Sender \$2	Receiver \$0	Sender \$2
$a = B$	Receiver \$0	Sender \$0	Receiver \$2	Sender \$0

Table 1: Payoffs

Each round is divided into three stages, following the structure of the game introduced in Section 2: Commitment Stage, Revision Stage, and Guessing Stage. The sender is active in the first two stages, while the receiver is active only in the last stage.

It is common knowledge that the actual message sent to the receiver will be generated from the choice made in the Commitment Stage with probability  $\rho$ , and from the choice made in the Revision Stage with probability  $(1 - \rho)$ . In treatments with  $\rho = 1$ , the Revision Stage is not relevant, and therefore is not included in the game description.

Figure 1 contains a screen shot of the Commitment Stage. This highlights the fact that, in our design, implementing the choice of an information structure is straightforward and easy to visualize: senders simply move sliders on the screen. Moving the slider changes the color of each bar in a way that reflects the chosen probabilities of each message. The implied probabilities of their current choice are updated in real-time in the cells above the sliders. Screens for the other stages are provided in Appendix C. Each round is concluded by a feedback screen in which all the relevant information, with the exception of the choice of  $\pi_2$  from the revision stage, become common knowledge. In particular, subjects are told the state, the message, the guess, the realized payoffs, the stage (Commitment or Revision) from which the message was generated, the sender’s Commitment strategy, and the receiver’s Guessing strategy.

Our initial set of treatments constitute a  $2 \times 3$  factorial between-subjects design. In treatments with verifiable communication, the interface prevents senders from assigning positive probability to a red message when the ball is blue or a blue message when the ball is red.

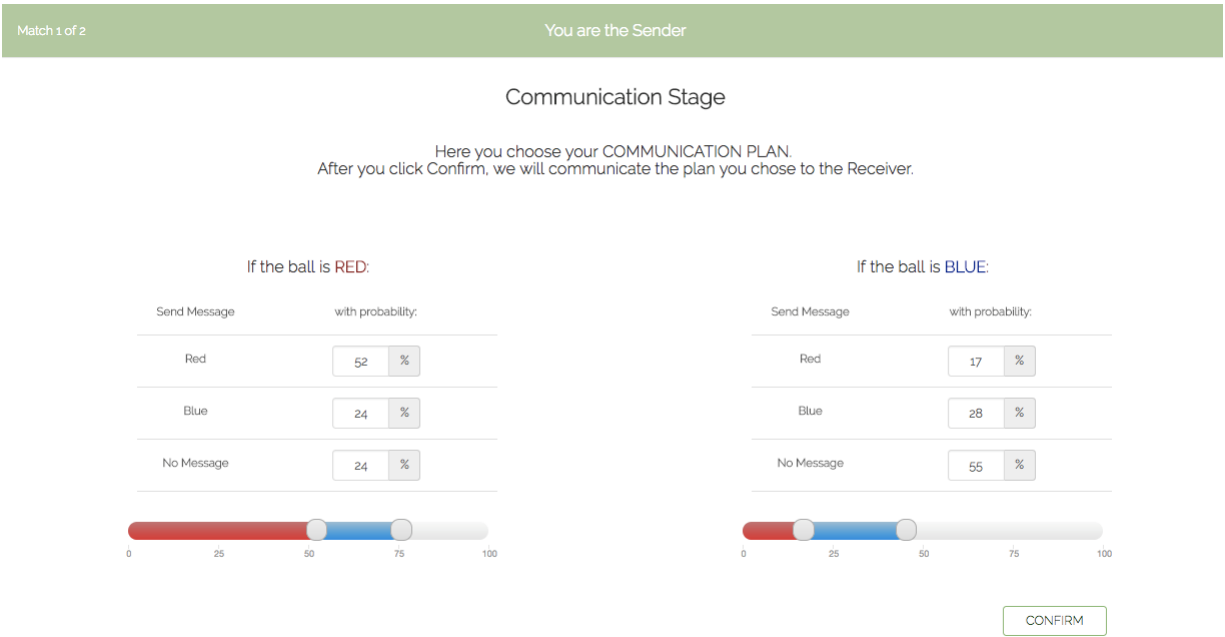


Figure 1: Design - Commitment Stage.

The interfaces are identical in all other respects. For both verifiable and unverifiable treatments, we conduct three variations generating a total of six treatments:  $V20$ , verifiable communication with  $\rho = 0.2$ ;  $V80$ , verifiable communication with  $\rho = 0.8$ ;  $V100$ , verifiable communication with  $\rho = 1$ ;  $U20$ , unverifiable communication with  $\rho = 0.2$ ;  $U80$ , unverifiable communication with  $\rho = 0.8$ ; and  $U100$ , unverifiable communication with  $\rho = 1$ . Bayesian persuasion as in [Kamenica and Gentzkow \(2011\)](#) is implemented in treatment  $U100$ . The treatment  $V100$  also implements the outcome of Bayesian persuasion. Cheap talk and disclosure of verifiable information would correspond to treatments  $U0$  and  $V0$ , respectively. Treatments with  $\rho = 0.2$  and  $\rho = 0.8$  still provide us with predictions in the spirit of cheap talk and disclosure, but allow us to vary the commitment parameter  $\rho$  while maintaining a cohesive experimental design. Treatments with  $\rho = 1$  are included to explore the extreme case where the predictions of verifiable and unverifiable exactly coincide, allowing for a tight test of the implications of verifiable communication. We chose not to perform a treatment for the cases of  $\rho = 0$  for two reasons. First, these cases have been studied in previous experiments. Second, in our environment the key predictions of the cases of  $\rho = 0.2$  in terms of information transmission are the same as in the case of  $\rho = 0$ . Furthermore, as will be shown later, behavior in treatments with  $\rho = 0.2$  is in line with prior experimental evidence on cheap talk and disclosure games. [Figure 2](#) and [Table 2](#) summarize the theoretical

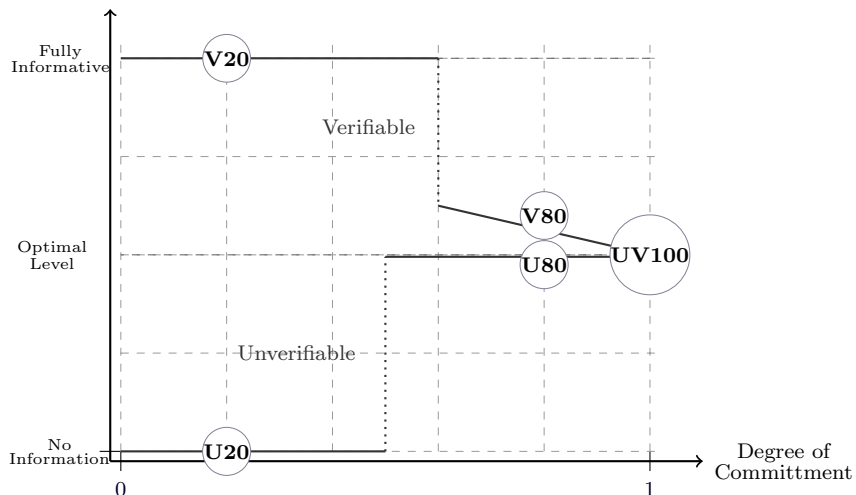


Figure 2: Predictions and Treatments.

predictions for our specific design.

We also run a treatment exploring a different dimension of comparative statics. This treatment involves full commitment ( $\rho = 1$ ) and unverifiable information. We call this treatment  $U100H$ . In this treatment we only change payoffs so that receivers require more persuasion in order to choose the senders' favorite action. Payoffs are as follows. As in all other treatments the receiver obtains zero payoff if he makes the wrong guess. In contrast with the treatments above, the receiver wins different amounts if he correctly guesses the color of the ball: 2 if ball is Blue,  $\frac{2}{3}$  if ball is Red. The sender wins 3 if receiver guesses Red. With this new treatment, the persuasion threshold changes from 0.5 to 0.75. Thus, in equilibrium, the sender should provide more information. The strategy of the sender involves sending  $r$  with probability one if the ball is Red, sending  $r$  with probability  $1/6$  and  $b$  with probability  $5/6$  if the ball is blue.

### 3.2 Design implementation

The experiment involved 25 rounds of the game described above, played for money in fixed roles with random re-matching between rounds, preceded by two practice rounds. Before the practice rounds, instructions were read aloud (see the Appendix D for an example).

One advantage of our design is that predictions are independent of risk preferences because outcomes are binary.

We conducted four sessions per treatment for a total of 24 session. Each session included

12 to 24 subjects (16 on average per session) for a total of 384 subjects. In addition to their earnings from the experiment, subjects received a \$10 show-up fee. Average earnings, excluding the show-up fee, were \$16.55 (ranging from \$11.28 to \$21.66) per session. On average, sessions last approximately 1 hour and 40 minutes.

### 3.3 Equilibrium behavior

Table 2 summarizes the theoretical equilibrium predictions for all treatments in our specific design.

Table 2: Equilibrium Predictions

Treat.	Sender						Receiver			
	Ball	<i>Commitment</i>			Ball	<i>Revision</i>			<i>Guessing</i>	
		red	blue	no		red	blue	no	Mes.	Guess
V20	R	1		0	R	1		0	red	<i>red</i>
	B		$x$	$1 - x$	B		$x$	$1 - x$	blue no	<i>blue</i> <i>blue</i>
V80	R	0		1	R	1		0	red	<i>red</i>
	B		$\frac{3}{4}$	$\frac{1}{4}$	B		0	1	blue no	<i>blue</i> <i>red</i>
V100	R	0		1					red	<i>red</i>
	B		$\frac{1}{2}$	$\frac{1}{2}$					blue no	<i>blue</i> <i>red</i>
U20	R	$x$	$y$	$1 - x - y$	R	1	0	0	red	<i>blue</i>
	B	$x$	$y$	$1 - x - y$	B	1	0	0	blue no	<i>blue</i> <i>blue</i>
U80	R	1	0	0	R	1	0	0	red	<i>red</i>
	B	$\frac{3}{8}$	$\frac{5}{8}$	0	B	1	0	0	blue no	<i>blue</i> <i>blue</i>
U100	R	1	0	0					red	<i>red</i>
	B	$\frac{1}{2}$	$\frac{1}{2}$	0					blue no	<i>blue</i> <i>blue</i>

It is worth discussing two main features of the equilibria in the different scenarios.

- (i) Consider first the extreme cases of  $U100$  and  $V100$ . The  $U100$  scenario is the classic example of Bayesian persuasion: given the prior of  $\frac{1}{3}$  on the red ball, an optimal strategy for the sender is to commit to always send message  $r$  when the state is  $R$  (i.e.,

to always disclose the “good news”), and to randomize 50-50 between  $r$  and  $b$  when the state is  $B$ . This strategy induces a posterior of zero on  $R$  when the message is  $b$ , and of 0.5 on  $R$  when the message is  $r$ . This is the lowest possible value of the posterior on the red state that still induces the receiver to choose the red action. Therefore, this information structure maximizes the ex-ante probability that the receiver chooses the red action. Now consider the case of V100. Because messages are verifiable, the sender can no longer send the red message if the state is blue. Therefore, the equilibrium strategy described for the case of U100 is no longer feasible. However, the sender can use “no message” to achieve the same outcome: she should always send  $n$  when the state is  $R$  and randomize 50-50 between  $b$  and  $n$  if the state is  $B$ . This strategy effectively replaces the red message with “no message” and induces the same posteriors for the receiver as in the case of U100, and therefore is optimal.

- (ii) Consider now cases in which the probability of commitment is lower, so that there is a non trivial revision stage. It is easy to see that the equilibrium strategy in the revision stage is the same as if the sender had no commitment at all. Thus, in the verifiable treatments the sender always reveals all the information, while in the non verifiable treatments the sender never reveals any information. Anticipating their own behavior in the revision stage, at the commitment stage the sender attempts to compensate by committing to reveal more than under full commitment in the unverifiable treatment, and less than under full commitment in the verifiable treatment. In the case of  $\rho = 0.8$ , for instance, this turns out to balance perfectly in the unverifiable treatment, thereby generating the same outcome as under full commitment, but to balance only partially in the verifiable treatment, thereby generating more information than under full commitment.<sup>9</sup> However, when  $\rho = 0.2$ , in neither treatment is the sender able to use the commitment stage to undo the behavior in the revision stage, and the outcome is the same as if the sender had no commitment at all. Because of this, in the  $\rho = 0.2$  case, there is a degree of indeterminacy in the commitment stage in both verifiable and unverifiable scenarios. However, this indeterminacy is immaterial for the equilibrium outcome. Note that the case of  $\rho = 0.8$  displays the behavior described in Proposition 1: Under V80 more information is transmitted in the revision stage than

---

<sup>9</sup>In the V80 treatment there is another equilibrium in which the Sender’s payoffs are the same as in V100. However, this requires the sender to always send no message in the revision stage, which is weakly dominated. Indeed, if the ball is red, sending the red message dominates sending the empty message; the latter behavior is also consistent with what we see in the data in the revision stage.

in the commitment stage, while the opposite is true under U80.

### 3.4 Data Analysis Preliminaries

The behavior by senders and receivers jointly determines the degree to which decisions match the state. For instance, even if a sender is perfectly truthful so that the receiver is in principle able to perfectly infer the state from the sender’s messages, the correlation between the state and the decision would be zero if the receiver did not act upon this information. To measure informativeness, we will report the correlation between the state and the decision; we denote this by  $\phi$ .

However, in much of the analysis, it will be more informative to construct an alternative measure that isolates the informativeness of the sender’s strategies. To do this, we compute the correlation  $\phi^B$  that would result under the assumption that all receivers are Bayesian, in the sense that they form the “correct” posteriors given the information conveyed by the senders, and then act optimally given this information.<sup>10</sup>

The average correlation provides a useful summary of the patterns in the data but hides important details. For instance, if senders manage to induce posteriors that are “almost” correct, but come up just short of what is needed for a receiver to choose the action desired by the sender, then the correlation (whether or not we use actual choices or Bayesian receivers) will be 0. Hence, we also provide a different way to capture the amount of informativeness conveyed by senders that does not suffer from this shortcoming. Consider the posterior that is generated for each state. The dispersion of these posteriors is a measure of the informativeness of the sender’s strategies. The further apart are the posteriors in the two states, the more information is conveyed by the sender. Of course, the posteriors must be linked to the prior consistently with Bayes’ rule. The theory predicts that, as commitment increases, the posterior dispersion should *decrease* under verifiable information and *increase* under unverifiable information.

Recall that a state induces a message, via  $\hat{\pi}$ , and a message induces a posterior, via Bayes’ Rule, that is:

$$\theta \rightarrow m \rightarrow \mu(R|m) := \frac{\mu_0(R)\hat{\pi}(m|R)}{\mu_0(R)\hat{\pi}(m|R) + \mu_0(B)\hat{\pi}(m|B)} \in [0, 1],$$

where  $\hat{\pi}$  can be computed from the data as follows:

$$\hat{\pi}(m|\theta) := \rho\pi_C(m|\theta) + (1 - \rho)\hat{\pi}_U(m|\theta).$$

---

<sup>10</sup>We discuss below how the “correct” posteriors are obtained.



The priors  $\mu_0(R)$  and  $\mu_0(B)$  as well as the parameter  $\rho$  are treatment parameters. The likelihoods from the commitment stage  $\pi_C(m|\theta)$  are observable directly from the data. However, when  $\rho < 1$ , in order to obtain the overall posterior, we need to incorporate the role of the information conveyed in the revision stage. Since the revision strategy is only specified conditional on the realized state: the likelihoods at the revision stage,  $\hat{\pi}_U(m|\theta)$  are not directly observable. To circumvent this, for treatments with commitment 20% and 80%, we use the average behavior of senders in the session, therefore, we approximate  $\hat{\pi}_U(m|\theta)$  with the average population behavior in the revision stage of the relevant session. This seems to be a natural approach, and, on average this is what receivers ought to learn to expect. We note, also, that the results for treatments with  $\rho = 0.8$  (where we perform such an approximation) are similar to those with  $\rho = 1$  (where we do not need to use this approximation); this suggests that the approximation does not make a big difference.

## 4 Results

We now proceed to an analysis of the data. The first subsection explores the simplest and most direct evidence to determine whether senders and receivers react to commitment. In line with Proposition 1, we compare the behavior of the senders in the commitment stage with their behavior in the revision stage. We then evaluate how the receivers' reaction to the information contained in the commitment stage depends on the degree of commitment, i.e., we compare the receivers' behavior across different commitment levels. The subsequent subsection uses the predictions from Proposition 3 to understand how senders react to the persuasion threshold  $q$ : this compares treatments U100H and U100. This is followed by an overall comparison of the impact of  $\rho$  under each set of rules. Finally V100 and U100 are compared more specifically: i.e. comparing the impact of rules holding commitment constant. The last results subsection documents aspects of behavior such as heterogeneity and the extent to which subjects best respond to each other.

### 4.1 Response to Commitment

The power of commitment is that it allows senders to persuade receivers to choose the action favored by the sender. In the standard Bayesian persuasion model, this commitment power is expressed as a comparison with the outcome of a cheap talk model: e.g., as in comparing a treatment of  $U0$  with a treatment  $U100$ . Some of our later discussion will perform similar

comparisons. However, our framework allows us to evaluate the power of commitment within a particular treatment, by comparing senders' behavior in the commitment stage with their behavior in the revision stage. The clearest manifestation of this for senders can be seen when  $\rho = 0.8$ , in which case the commitment strategies and revision strategies display particularly stark differences. As shown in Proposition 1, when information is unverifiable, senders should reveal more information in the commitment stage than in the revision stage; while when information is verifiable senders should reveal more in the revision stage. In other words, commitment allows senders to tell the truth only some of the time when they have bad news if information is unverifiable. On the other hand, it allows them to hide the truth when they have good news and information is verifiable.

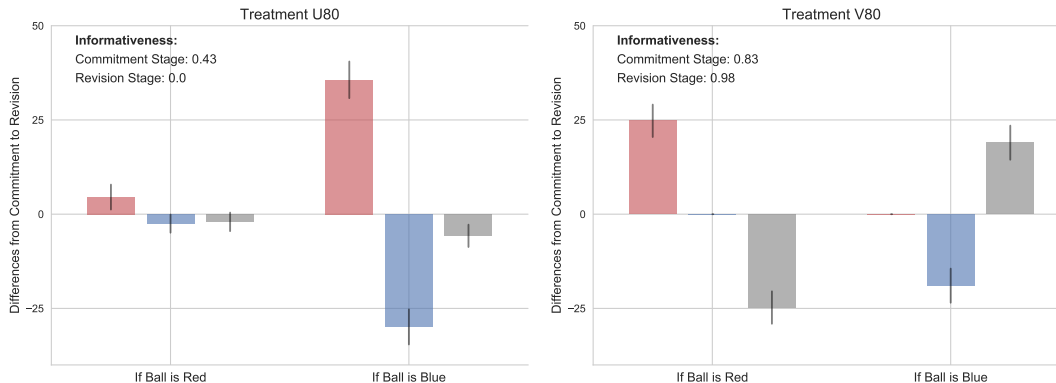


Figure 3: Sender's Strategy: Commitment vs. Revision,  $\rho = 0.8$

These changes in strategies are presented in Figure 3. Looking at the U80 treatment, it is clear that there are important differences in the strategies when the state is B.<sup>11</sup> Indeed, when the state is B, the revision strategy specifies a red message much more often than the commitment strategy.<sup>12</sup> These differences in strategies between the two stages have a significant ( $p < 0.01$ ) and quantitatively important impact on the (hypothetical) induced correlation between state and decision: in the commitment stage the correlation is  $\phi_C^B = 0.43$ , while the correlation drops to  $\phi_R^B = 0.00$  in the revision stage.

As predicted by the theory, in the case of the V80 treatment, the changes are in the opposite direction, with  $\phi^B = 0.83$  in the commitment stage, increasing to  $\phi^B = 0.98$  in the

<sup>11</sup>The changes in strategy are jointly statistically different from zero both when the state is R ( $p < 0.1$ ) and when it is B ( $p < 0.05$ ).

<sup>12</sup>Although there are statistically significant changes in the same direction when the state is R, those are small in magnitude.

revision stage.<sup>13</sup> The strategies underlying this change imply replacing “no message” with red message when going from commitment to revision if the state is R. In addition, but to a slightly lesser extent, senders replace blue with “no message” when the state is B.

In the case of the V80 treatment, the changes are in the opposite direction, with  $\phi^B = 0.83$  in the commitment stage, increasing to  $\phi^B = 0.98$  in the revision stage. The increase in  $\phi^B$  is statistically significant ( $p < 0.01$ ) and the changes in strategies are statistically significant as well ( $p < 0.01$  for both state R and state B). The strategies underlying this change imply replacing “no message” with red message when going from commitment to revision if the state is R. In addition, but to a slightly lesser extent, senders replace blue with “no message” when the state is B.

Hence, in both U80 and V80, there is clear evidence that senders adjust their behavior between the commitment and revision stages. Those adjustments are in line with the forces highlighted in Proposition 1, namely commitment allows senders to hide good information under verifiable information and to misrepresent good information under unverifiable information.

We now wish to evaluate whether the receivers understand the difference in the information content of the commitment and the revision stage. This is a way to assess whether the receivers react to commitment in a way that is consistent with the theory. In order to do so, we must compare receiver behavior across treatments. First, note that under unverifiable information, the key message, which we will refer to as the *persuasive message*, is the red message. Red is the message the sender uses to attempt to convince the receiver to choose red even when the state is  $B$ . Under verifiable information however, the critical message, or *persuasive message* is “no message.” That is the message the sender uses to hide good information and convince the receiver to guess red even if sometimes the state is  $B$ . We now compare how receivers respond to these persuasive messages depending on the level of commitment available to the sender.

Figure 4 displays the probability that receivers guess red as a function of the posterior induced by the strategy of the sender in the commitment stage and the message they received, in this case the persuasive message. The data is separated by rules, but also depending on whether the sender had low commitment or full commitment. Since the message is unlikely to come from the commitment stage when  $\rho = 0.2$ , the posterior induced at the commitment stage should have little impact on receivers. On the other hand, when  $\rho = 1$ , receivers should

---

<sup>13</sup>The increase in  $\phi^B$  is statistically significant ( $p < 0.01$ ) and the differences in strategies are statistically significant as well ( $p < 0.01$ ) for both state R and state B).

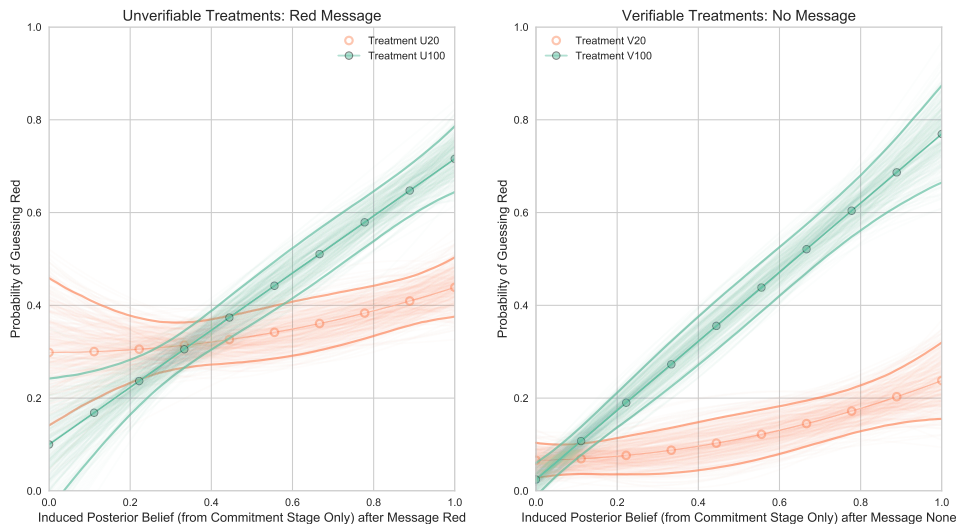


Figure 4: Receiver's Response to Persuasive Messages:  $\rho = 0.2$  vs.  $\rho = 1$

guess red for high enough posteriors. Indeed, the figures show this to be the case for both unverifiable and verifiable implementations, although the difference is more pronounced in the case of verifiable information.<sup>14</sup>

## 4.2 Does Commitment Have the Predicted Effect?

The previous section establishes that both sender's and receivers react to the power of commitment. This section further explores the extent to which the reaction they display is in the direction suggested by equilibrium forces. One of the more direct implication of commitment, which is stated in Proposition 3, is that increasing  $q$  should lead to more informative communication from senders. Consequently, in treatment U100H where it is more valuable for the receiver to match the state when the state is B, senders should convey more information.

Figure 5 shows that the distribution of  $\phi^B$  for the U100H treatment is to the right of that for the U100 treatment. The median subject goes from inducing a correlation of 0.25

<sup>14</sup>The probability that the Receiver guesses Red when the induced posterior from the commitment stage is below 0.5 does not differ statistically as we vary the commitment level (i.e. when going from  $\rho = 0.2$  to  $\rho = 0.8$ ). This is true both for the case with verifiable information and the case with unverifiable information. For posteriors of 0.5 and above, there is a statistically significant difference in both cases ( $p < 0.05$  when messages are unverifiable and  $p < 0.01$  in the verifiable case).

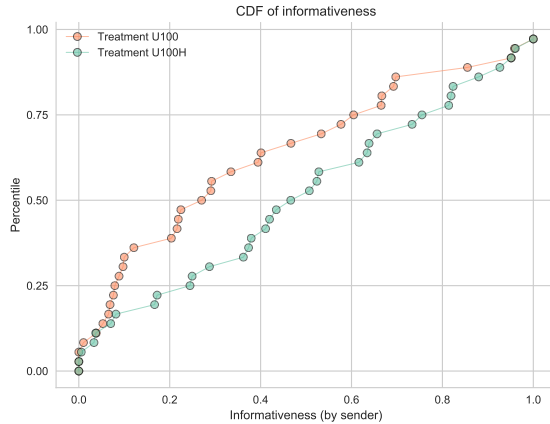


Figure 5: CDF of  $\phi^B$  for treatments U100 and U100H

in U100 to one of 0.52 in U100H. This shift, however, is not statistically significant; which is not overly surprising considering that we have only two sessions of U100H treatment so far.

We now discuss the other key prediction of the theory, stated in Proposition 2, concerning what happens to information transmission as  $\rho$  increases. Up to this point, we have reported only  $\phi^B$ , but now we will also consider  $\phi$ , which is jointly determined by the behavior of senders and receivers. This approach allows us to partly disentangle where potential losses in information come from.

Table 3 reports aggregate correlations between state and decision. The top left panel displays the theoretical predictions; the top right panel indicates the correlation in the data; the bottom left panel reports the correlation that would result if one replaced the receivers in the data with Bayesian receivers. The symbols between the numbers indicate their “statistical relations” at the 10% level, i.e.  $\approx$  means that the p-value of the equality of the two is larger than 0.1.

Looking at the top-right panel, we see that subjects react to commitment in the expected direction; both for verifiable and unverifiable cases. However, the effects are more muted than in the predictions of the theory. In the case of verifiable treatments, the change from the treatment with  $\rho = 0.2$  to the treatment with  $\rho = 1$  is, in theory, predicted to reduce the correlation from 1 to 0.5; but the difference in the data is 0.13; or 26% of the predicted change. For the unverifiable cases, the changes are also in the expected directions, and of the same magnitude. Comparing this to the bottom left panel suggests that some, but not all, of the “missing effect” comes from receiver behavior. The correlations assuming a

Table 3: Correlation Between Ball and Guess

Theory				Data							
	Commitment ( $\rho$ )				Commitment ( $\rho$ )						
	20%	80%	100%		20%	80%	100%		20%	80%	100%
<b>Verifiable</b>	1	0.57	0.50	<b>Verifiable</b>	0.80	$\approx$	0.78	$>$	0.67		
					$\vee$		$\vee$		$\vee$		
<b>Unverifiable</b>	0	0.50	0.50	<b>Unverifiable</b>	0.09	$<$	0.21	$\approx$	0.21		

Sender Data with Bayesian Receiver					
	Commitment ( $\rho$ )				
	20%	80%	100%		
<b>Verifiable</b>	0.89	$\approx$	0.85	$>$	0.78
	$\vee$		$\vee$		$\vee$
<b>Unverifiable</b>	0.00	$<$	0.33	$\approx$	0.34

Bayesian receiver reveals larger effects of commitment for the unverifiable treatments and smaller effects for the verifiable treatments: 68% in the case of the unverifiable treatments and 22% of the predicted change for verifiable treatments.

These correlations can be investigated further by looking at the CDF of  $\phi^B$  for each treatment. Figure 6 shows two interesting patterns. First, the distribution for U80 and U100 are very similar, as predicted, and to the right of U20. Second, the distributions for the verifiable treatments are, loosely speaking, ordered as predicted by theory. However, it is also clear, especially for the unverifiable treatments, that many sender's have correlations that are positive, but low enough to suggest that some of their strategies must not move the posteriors enough. To understand this better we now turn to posteriors.

Figure 7 displays the kernel density estimates of the posteriors conditional on the state. The vertical dashed lines indicate the theoretical predictions; the solid lines present the data under the different treatments. The solid red line is at .5, because in equilibrium, the posterior following the red state is .5. The solid blue line is at .25 because in equilibrium, in the blue state the posterior is .5 with 50% probability (when the sender sends the red

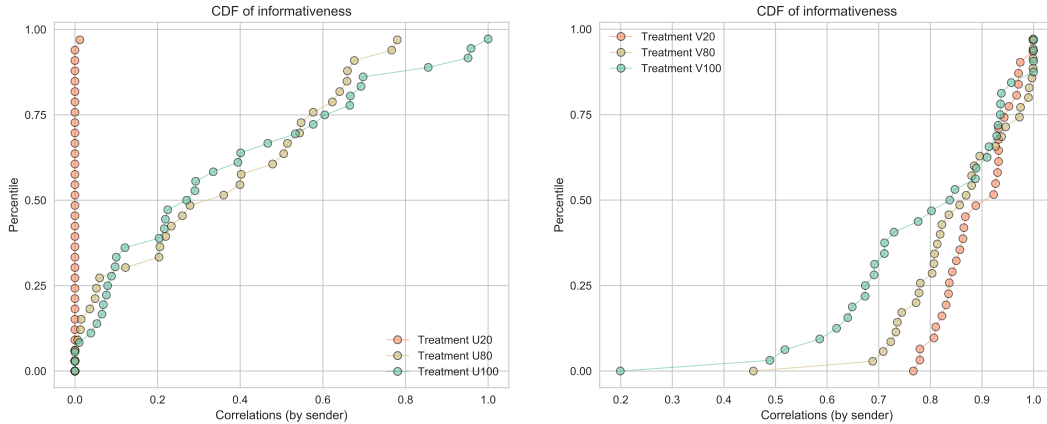


Figure 6: CDF of  $\phi^B$  for treatments V20, V80, V100, U20, U80, and U100

message) and 0 with 50% probability (when the sender sends the blue message).

In all cases there is a sizable response to the treatment in the direction predicted by the theory, more so in the unverifiable than in the verifiable treatments. Moving from U20 to U100, the posteriors become more spread out, while moving from V20 to V100, the posteriors move closer, as predicted by theory. However, there are some important discrepancies: While in the case of V20 the posteriors are inside the lines describing the theoretical predictions, in the other two verifiability treatments, most of the mass of the posteriors lies outside of the relevant (theory-consistent) lines. In other words, senders are not informative enough under V20, and too informative in the other cases.

Table 4 reports the difference between the mean posteriors when the ball is red relative to the case in which the ball is blue. The table makes it clear that the data moves in the right direction for both verifiable and unverifiable treatments, but that average behavior is closer to the theoretical predictions in the case of the unverifiable treatments than in the case of verifiable treatments.

According to the data presented in Table 4, average behavior is most consistent with the theory in U100 and least consistent in V100. In order to better understand behavior, we now focus on the differences between the U100 and V100 treatments.

Table 5 highlights an important feature of the V100 treatment. Senders ought to strategically hide “good news” in order to make it possible to sometimes hide the bad news. Instead, in the aggregate data, subjects commit to approximately 50-50 randomization with the empty message following **both** states. This undermines the logic of using the empty

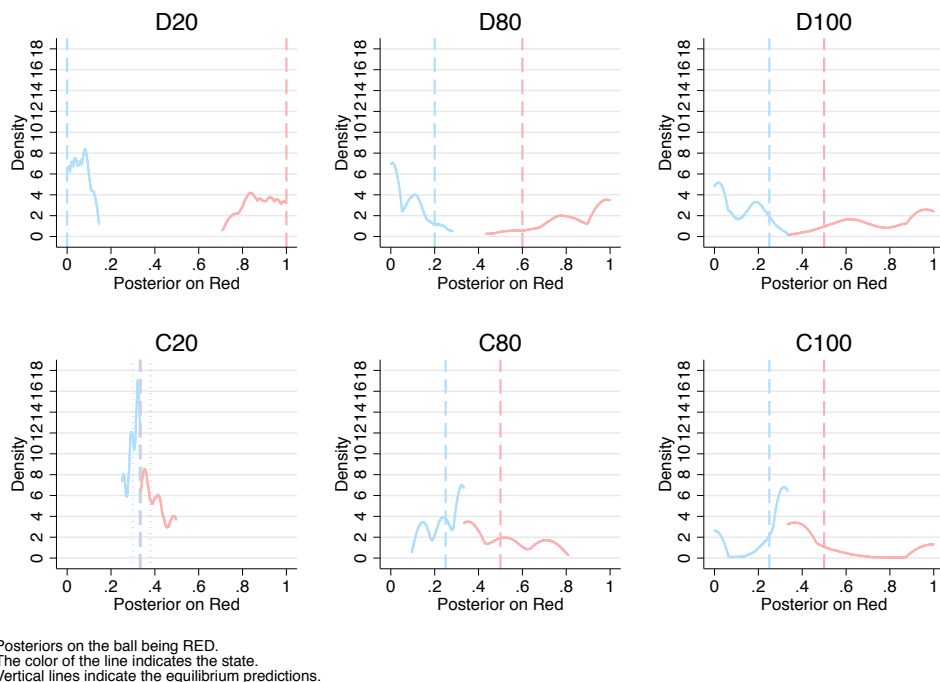


Figure 7: Posterior on R as a Function of the Ball

Table 4: Difference Between the Mean Conditional Posteriors (theoretical values in parentheses)

	Commitment ( $\rho$ )					
	20%		80%		100%	
<b>Verifiable</b>	0.86	(1.00)	0.78	(0.40)	0.69	(0.25)
	<b>B</b>	<b>R</b>	<b>B</b>	<b>R</b>	<b>B</b>	<b>R</b>
	0.05	0.91	0.07	0.85	0.10	0.80
<b>Unverifiable</b>	0.11	(0.00)	0.23	(0.25)	0.30	(0.25)
	<b>B</b>	<b>R</b>	<b>B</b>	<b>R</b>	<b>B</b>	<b>R</b>
	0.30	0.40	0.26	0.49	0.23	0.53



Table 5: Theoretical Predictions and Data: V100 and U100

U100		Messages			V100		Messages		
		$r$	$b$	$n$			$r$	$b$	$n$
States	R	100%	0	0	States	R	0	0	100%
	B	50%	50%	0		B	0	50%	50%

		Messages					Messages		
		$r$	$b$	$n$			$r$	$b$	$n$
States	R	74%	12%	14%	States	R	<b>51%</b>	0	<b>49%</b>
	B	44%	39%	17%		B	0	58%	42%

message to substitute the red one in constructing the optimal Bayesian persuasion strategy. In contrast, in U100, senders have to strategically lie about “bad news.” The aggregate data suggests that, on balance subjects are less successful in implementing a persuasion strategy in the verifiable treatment. However, as we will see below, the aggregate data hides rich heterogeneity in behavior, and in fact, a substantial fraction of subjects do attempt to persuade in the V100 treatment.

### 4.3 Behavior in More Detail

[This section is incomplete. For the moment it reports on various aspects of heterogeneity and other aspects of behavior that we will develop in the final version of the paper. We also plan on adding a QRE analysis that we have completed but that needs to be written up.]

#### 4.3.1 sender behavior

We now take a more detailed look at individual senders’ behavior. Figure 8 depicts one key feature of this behavior. On the horizontal axis, we report for each subject the mean correlation (over the last ten rounds) between the state and the guess, assuming Bayesian receivers. On the vertical axis, we report the standard error of this mean. So, for instance, a subject with a mean of 0.5 and a standard error of 0 would correspond to equilibrium behavior in the U100 treatment. However, a subject with a mean of 0.5, but a high standard error, is sometimes producing low correlations and others high correlations. In particular, variation

on the horizontal dimension represents cross-subject heterogeneity, whereas variation on the vertical dimension represents within-subject heterogeneity.

Some features of this Figure stand out. First, there is a marked difference across treatments; this difference is qualitatively in line with the theory. Second, there is a lot of heterogeneity among subjects, except in the U20 and V20 treatments where the correlation is very close to equilibrium for most subjects. For instance, in U80 and U100 treatments, roughly a third of the subjects generate a very low correlation between state and decision (below .25), while some subjects generate much higher correlations. Third, there seems to be more “noise,” both at the individual, and at the aggregate level, in the unverifiable treatments. Subjects are more clustered toward high correlations and with relatively little individual level noise in the treatments with verifiable information.

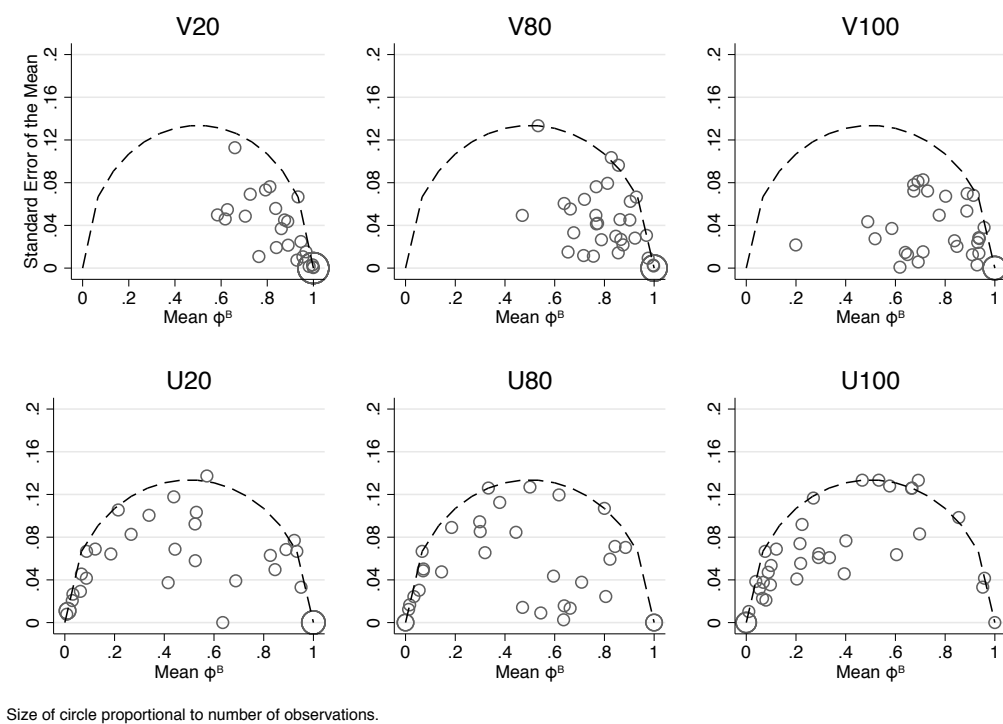


Figure 8: Senders’ heterogeneity.

Figure 9 presents a sender level analysis of the difference across states in the posteriors induced by senders and how this varies across treatments. The horizontal axis is the mean difference in posteriors, while the vertical axis is the standard deviation of this mean. As in the case of the sender level correlations discussed above, these figures show a remarkable

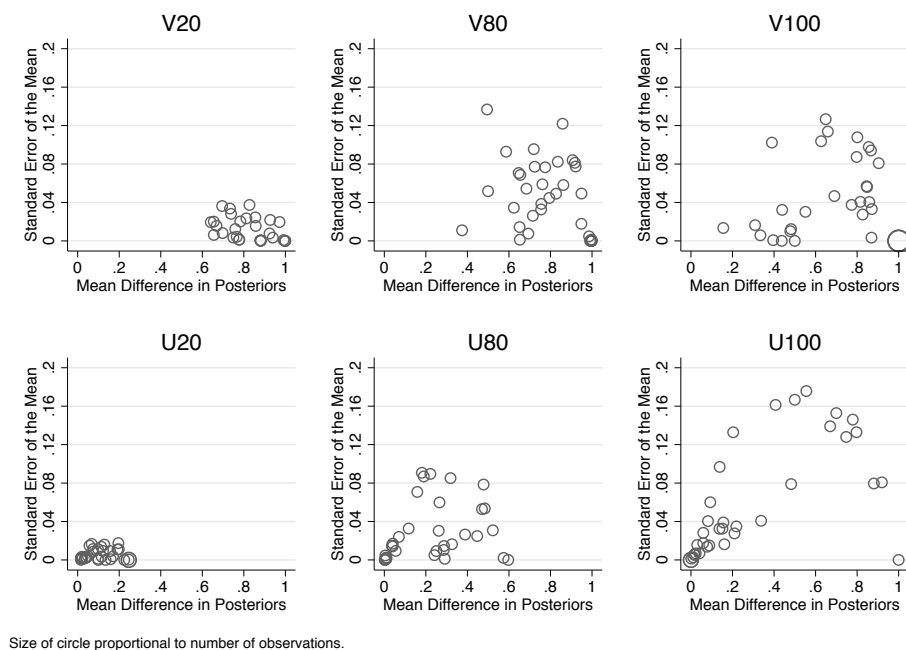


Figure 9: Mean Difference across States in the Posteriors Induced by Senders.

amount of sender heterogeneity, and some amount of within sender variation, except in the U20 and V20 treatments where values are clustered right around equilibrium predictions. In the V80 and V100 treatments, most senders display too large a difference in means; i.e., they convey too much information. In contrast, in the U80 and U100 treatments, senders are widely dispersed, with many conveying little to no information, and a significant number conveying too much information relative to equilibrium predictions.

The clustering of behavior in U20 and V20 treatments around the equilibrium predictions is of interest because it suggests that it is not likely to be the complicated structure of the game that generates departures from equilibrium in the other treatments: the structure with two rounds of communication, including a commitment round is the same in U20, V20, U80 and V80 treatments. However, the senders seem to behave in a manner that is consistent with the equilibrium in U20 and V20. Notably, these cases have *opposite* predictions regarding the amount of information that is transmitted, so this conformity with equilibrium predictions is not a matter of following the same default convention: in U20 subjects are uninformative, whereas in V20 subjects are very informative.

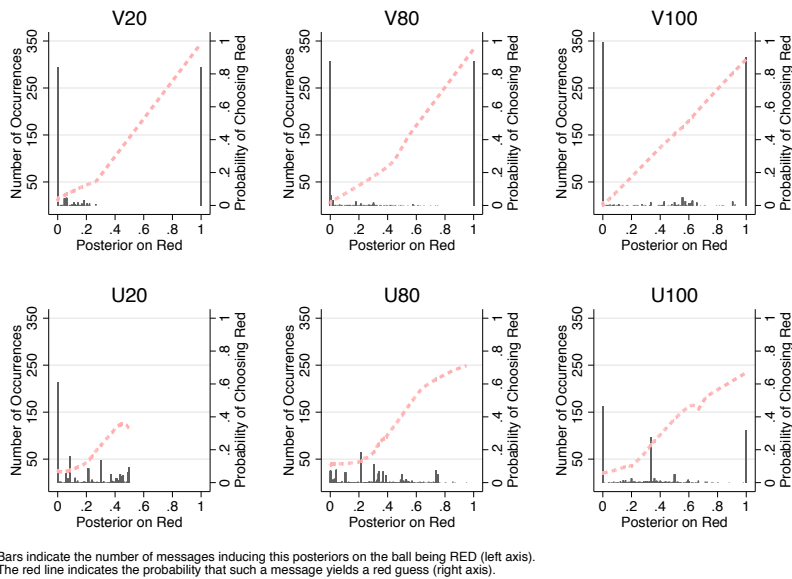


Figure 10: Guess of *red* as a Function of the Posterior on R and Frequency of Posterior

### 4.3.2 receiver behavior

We now discuss receivers' behavior. We first consider how receivers react to the posteriors induced by the senders. This is directly of interest and will allow some interesting comparison with the literature on cheap talk experiments. It is also important, because characterizing messages in terms of their posterior (as we have done above for the senders) is a useful way to describe the data only if receivers react to posteriors in a sensible (and predictable) manner. We first discuss aggregate data and we then look at individual subject behavior. Figure 10 shows, for each treatment, the probability that receivers guessed red as a function of the posterior induced by the message of the sender. It also shows via grey bars the frequency of each posterior.

The first finding is that the probability with which receivers guess *red* increases with the posterior.<sup>15</sup> This indicates that at least they react in a reasonable way to theoretically more convincing messages. The figure also reveals that messages that generate the same posterior generate different behavior by the receiver depending on whether we are in a verifiable or unverifiable treatments. That is, conditional on a high posterior, receivers are more likely to choose a red action following a red message if senders are not allowed to lie about the state, then when they are.

<sup>15</sup>Note that points for which there are no posteriors are interpolated.

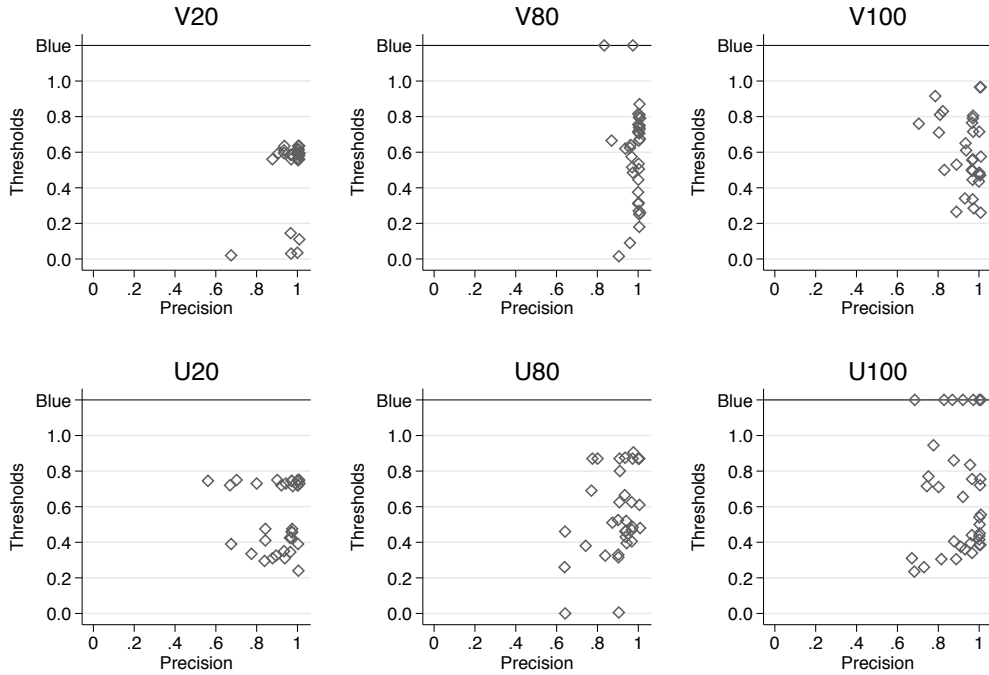
However, the probability of guessing *red* is not a step function, which is what we would observe for a population of Bayesian receivers. This can be due to two reasons. The first possibility is that, at the individual level, reaction functions are step functions, but because different subjects use different cutoffs, aggregation produces a smooth increasing line. The second possibility is that individual subjects react probabilistically (or inconsistently) to a given posterior. This could in turn be due to the fact that they have an imperfect (or approximate) understanding of posteriors (or how to compute them) or to the fact that our computation of the posterior does not appropriately account for how subjects consider the revision stage. Note, however, that this last possibility cannot account for the pattern in V100 and U100 since in those two treatments there is no revision stage.

We now consider whether subjects actually do behave according to threshold strategies. Figure 11 shows the best fitting threshold for each subject. Under the assumption that the subject uses a threshold strategy, the best fitting threshold is defined to be one that is consistent with the highest fraction of choices by the receiver. The *precision* of the threshold indicates the fraction of observations that are consistent with the threshold. A precision of one, means all choices can be rationalized by a threshold strategy on the posteriors, and that this is true for all messages (remember that we elicit the receiver’s choice for all possible message). For many subjects, multiple thresholds have the same precision, in those cases the figure reports the mid-point. Figure B13 in the Appendix displays the entire threshold range. A point on the top horizontal line labeled *Blue* means that, assuming that these subjects always chose blue regardless of the posterior fits the data better than any threshold  $\leq 1$ .

The figure reveals that a relatively high fraction of subjects displays behavior mostly consistent with the use of a threshold strategy. Specifically, at least 51% and up to 91% of subjects per treatment make choices consistent with a threshold strategy 90% or more of the time. If instead we consider a precision of 80% or more, threshold strategies can account for 93% or more of subjects in each of the verifiable treatments; and 80% or more of the subjects in the unverifiable treatments.

In addition to suggesting that most subject’s behavior is in line with the use of a threshold strategy, Figure 11 also clearly suggests a fair degree of heterogeneity in behavior. However, this is only suggestive as the figure does not easily allow to separate heterogeneity in receiver behavior from variability in the posteriors that the receivers were exposed too.

We now turn to a comparison between the receivers behavior in the data with the behavior that would be displayed by Bayesian receivers. Specifically, we wish to understand whether



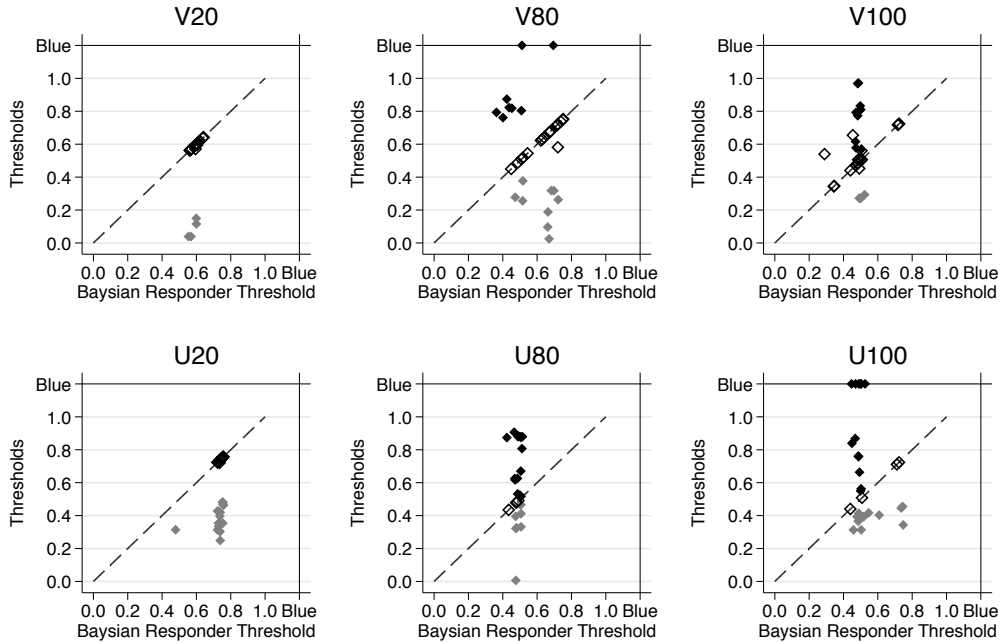
Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Figure 11: Estimated Thresholds and Fraction of Observations They Account For

our subjects are more or less skeptical than a Bayesian receiver. The answer to this question can be evaluated in Figure 12. In this figure we restricting attention to subjects whose behavior is relatively well described by a threshold strategy (i.e. their precision is at least 0.8).<sup>16</sup> The figure compares the estimated thresholds to those of a Bayesian receiver who faces the same data, i.e., we assume that the data is generated by Bayesian receivers but we use the same procedure for calculating a best-fitting threshold as we did for the receivers in our data. Hence, we recover our estimated threshold for hypothetical Bayesian receivers who would have been presented with the data generated by the senders in our experiment. In order to understand why Bayesian receivers are represented by a distribution of thresholds, note that there are posteriors that the receivers are never exposed to, therefore the estimated thresholds may be different from 0.5 (the theoretical value in equilibrium).<sup>17</sup> The figure plots the estimated thresholds (the mid-points) of our actual subjects against those that we would recover from simulated Bayesian receivers. If the point lies on the 45 degree line, then we

<sup>16</sup>For other subjects, thresholds are not a meaningful way to describe their strategies, hence a comparison with Bayesian threshold is not informative.

<sup>17</sup>Namely, there are multiple best fitting thresholds and the midpoint may well be different from .5.



Bayesian Responder Threshold are the thresholds that would be estimated if the responders were Bayesian given the posteriors in the data. Black for subjects harder to convince than a Bayesian, gray for subjects easier to convince than a Bayesian. Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Figure 12: Thresholds vs Bayesian Thresholds

cannot distinguish the behavior of such a receiver from that displayed by a Bayesian receiver. The markers are color-coded so that, considering all thresholds consistent with a subject's choice (from the lowest to the highest), if they are higher than those of the Bayesian receiver, they are marked with black symbols; in contrast, if they are lower than those of a Bayesian receiver they are marked with grey symbols. The hollow markers indicate a subject whose behavior cannot be unambiguously ranked relative to a Bayesian receiver because her lowest possible threshold and highest possible threshold are not both higher (or lower) than those of the Bayesian receiver.

The figure shows that many subjects are consistent with the behavior of a Bayesian receiver (they lie on the 45 degrees line). In some treatments, subjects whose behavior is not in line with Bayesian behavior need less convincing evidence to guess red, such as in the V20 and U20 treatments. In line with prior literature, we call these subjects credulous. In other treatments we have coexistence of subjects who are easier to convince and others who are more difficult to convince. For instance, in the V100 and U100 more subjects need stronger evidence to guess red than the Bayesian benchmark (we call these subjects skeptical), than

there are credulous subjects. Furthermore, the skeptical subjects are further away on average from the Bayesian benchmark than those who are easier to convince.

To summarize, receivers react to posteriors in the expected direction. Many receivers display behavior that is highly consistent with the use of a threshold strategy. The thresholds are heterogeneous. Some use thresholds consistent with those of a Bayesian receiver, but others do not. In low commitment treatments receivers tend to be more credulous than Bayesians, while in high commitment treatments, receivers appear to be more skeptical than Bayesian. This last finding is particularly interesting in light of prior experimental literature on communication that studies games without commitment and commonly finds that subjects are relatively gullible or trusting.<sup>18</sup> This prior literature only studies treatments with no commitment.

Table 6: Random Effects Probit of Determinants of Guess in U Treatments

	U20		U80		U100	
	Est.	Marg.	Est.	Marg.	Est.	Marg.
main						
Posterior	2.055** (0.855)	0.484** (0.213)	1.623*** (0.412)	0.411*** (0.104)	2.297*** (0.379)	0.532*** (0.084)
Blue Message	-0.541** (0.275)	-0.127** (0.061)	-0.718*** (0.210)	-0.182*** (0.042)	-0.495* (0.292)	-0.115 (0.070)
No Message	-0.148 (0.202)	-0.035 (0.047)	-0.619*** (0.184)	-0.157*** (0.035)	-0.068 (0.147)	-0.016 (0.033)
Observations	963	963	990	990	828	828
p-value						
Blue = No M.	0.03		0.22		0.08	

Notes: Standard errors clustered (at the session level) in parentheses. \*\*\* 1%, \*\* 5%, \* 10% significance.

A natural question arises next: to what extent do deviations from theory result from the specific message (ball color) that the sender’s strategy produces? Table 6 reports random effects (subject level) probit estimates of the impact of the posterior and the color of the message on guesses for all three unverifiable treatments.<sup>19</sup> In theory, only the posterior should matter for how receivers guess, so, for a Bayesian receiver the effect of the posterior

<sup>18</sup>As mentioned in the introduction these are findings are reported for instance in [Blume et al. \(1998\)](#), [Forsythe et al. \(1999\)](#), and [Cai and Wang \(2006\)](#).

<sup>19</sup>The V treatments are not included as interpreting the estimates is less clear for those. For instance, in the V20 treatment, the color of the message is in theory enough to determine what to guess.



going from zero to one should be 1, and the effect of the colors of the ball should be zero. In contrast, while our subjects do react strongly to the posterior, the reaction is about half what it would be for Bayesian receivers. Furthermore, in contrast with Bayesian receivers, there is also a significant effect on the choice made by the receiver of the color of the ball chosen by the sender.

## **5 Conclusion**

[To Be Done.]

# Appendix

## A Proofs (Incomplete)

### A.1 Proof of Proposition 1

**Part (i)** First consider the unverifiable scenario.

(1) We first show that, if the receiver chooses the action  $a_H$  with positive probability, then information transmission can only take place at the commitment stage. For the receiver to choose  $a_H$  with positive probability, because  $\mu < q$ , there must be a message  $m_1$  that induces a higher posterior for the receiver than some other message  $m_2$ , which in turn must induce a posterior below  $\mu$ . Message  $m_1$  induces the receiver to choose  $a_H$  with positive probability while  $m_2$  induces the receiver to choose  $a_L$  with probability one. But then the sender always chooses  $m_1$  in the revision stage.

(2) We now establish that there is a threshold  $0 < \hat{\rho} < 1$  such that, for  $\rho > \hat{\rho}$  the receiver chooses  $a_H$  with positive probability in equilibrium. First note that, because  $\mu < q$ , the receiver must receive some information in order to choose  $a_H$ . Thus, the sender would like to use commitment to transmit information. If some information is transmitted in the revision stage, then it is clear that positive information is transmitted. Assume then that no information is transmitted in the revision stage. Consider a strategy by the sender that chooses full revelation in the commitment stage, for instance, by choosing an information structure of  $\theta_H$  with probability one if the state is  $\theta_H$  and  $\theta_L$  with probability one if the state is  $\theta_L$ . If  $\rho$  is sufficiently high, then the posterior induced by this strategy following a message of  $\theta_H$  is larger than  $\mu$ , so the receiver is happy to choose  $a_H$  following such a strategy. Therefore, for sufficiently high  $\rho$  the sender will choose to transmit a positive amount of information. By (1), if the receiver chooses  $a_H$ , the revision stage must be uninformative. But for low  $\rho$ , there is now way to induce a posterior above  $\mu$ . However, as shown in (1), for sufficiently high  $\rho$  this is possible, and profitable. Thus, for some  $\hat{\rho}$  the sender will induce the receiver to choose  $a_H$ .

Consider now the verifiable scenario.

(3) First note that any undominated strategy in the revision stage involves disclosing the high state. All these strategies in the revision stage imply full information revelation at this stage.

(4) Now, recall that, if the state is  $\theta_H$  the sender can either send  $\theta_H$  or  $n$  but cannot

send  $\theta_L$ . Analogously, if the state is  $\theta_L$ , the sender cannot send  $\theta_H$ . Thus, in order to attempt to maximally persuade the receiver, the commitment stage must involve using the  $n$  message when the state is  $\theta_H$  and randomizing between  $\theta_L$  and  $n$  when the state is  $\theta_L$ . However, if observing  $n$  induces the receiver to choose  $a_H$ , then, at the revision stage, the sender optimally chooses  $n$  with probability 1 if the state is  $\theta_L$ . If  $\rho$  is low, then this means that there is no way that a choice of  $n$  in the commitment stage that leads the receiver to choose  $a_H$ . Thus, this leads the receiver to choose  $a_L$  following both messages,  $n$  and  $\theta_L$ . As a result, this strategy leads to lower payoffs for the sender than the outcome of full information revelation that involves sending  $\theta_H$  with probability one in the commitment stage. This means that, for low values of  $\rho$ , the equilibrium involves full information revelation.

Now, consider a  $\rho$  close to 1. Assume that the sender at the commitment stage chooses  $n$  with probability one if the state is  $\theta_H$ , and  $\theta_L$  with probability one when the state is  $\theta_L$ , then even if the sender chooses  $n$  in the revision stage when the state is  $\theta_L$ , this is not enough to overwhelm the information contained in  $n$  from the commitment stage. Thus, following message  $n$ , the receiver chooses  $a_H$ . In fact, because the receiver now chooses  $a_H$  with positive probability even when the state is  $\theta_L$  (following the choice of  $n$  in the revision stage), the overall payoff of this specific commitment strategy for the sender is higher than the one obtained under full disclosure, proving that, for high enough  $\rho$ , it is beneficial for the sender to choose a strategy that leads to lower information transmission. In fact, for high enough  $\rho$ , at the commitment stage the sender can choose to randomize between  $n$  and  $\theta_L$  when the state is  $\theta_L$  to induce and even higher probability of choice of  $a_H$ .

**Part (ii)** This is an immediate consequences of the comparison between steps 1 and 2 for the unverifiable case, and steps 3 and 4 for the verifiable case.

□

## A.2 Proof of Proposition 2

We first argue that, when  $\rho = 1$ , the equilibrium outcome is independent of verifiability. In both scenarios, the sender wishes to maximize the probability that the receiver chooses  $a_H$ , subject to the Bayes' constraint that the average posterior must equal the prior. The optimal strategy therefore must involve the selection of a persuasive message that leads the receiver to choose  $a_H$  while being indifferent between  $a_H$  and  $a_L$ . Thus, the posterior of the receiver upon receiving the persuasive message must be equal to the prior, whereas the posterior following any other message must be equal to zero. The following strategy implements this

optimum in the unverifiable scenario: if the state is  $\theta_H$  send message  $\theta_H$  with probability one; if state is  $\theta_L$  randomize between  $\theta_H$  and  $\theta_L$  in such a way that the posterior following a message of  $\theta_H$  is equal to  $q$ . In the verifiable case, this strategy is not feasible, but the following alternative strategy by the sender implements the same beliefs and actions for the receiver: if state is  $\theta_H$  send message  $n$  with probability one; if state is  $\theta_L$  randomize between  $n$  and  $\theta_L$  in such a way that the posterior following a message of  $\theta_H$  is equal to  $q$ .

When  $\rho$  decreases, the revision stage imposes constraints on what the sender can achieve with the commitment strategy. As shown in the proof of Proposition 1, the strategy in the revision stage is independent of  $\rho$  for both verifiable and unverifiable scenarios, with no information transmitted in the unverifiable scenario and full information transmission in the verifiable scenario. Thus, if the strategy at the commitment stage is left unchanged as  $\rho$  decreases, the revision stage leads to overall less information communicated for lower  $\rho$  with unverifiable information and more information with verifiable information. The sender will therefore change the commitment strategy to attempt to minimize the effect of the revision stage. As  $\rho$  decreases, the sender will reduce the amount of information transmission in the verifiable case and increase it in the unverifiable case. However, the smaller  $\rho$ , the less the sender is able to do this. Since this constraint is monotonically becoming tighter as  $\rho$  decreases, the amount of information that is communicated overall between the commitment and revision stages also changes monotonically.  $\square$

### A.3 Proof of Proposition 3

Note first that changing  $q$  has no effect on the equilibrium strategies in the revision stage: as shown in Proposition 1, there is full revelation in the verifiable case and no revelation in the unverifiable case. Now recall that  $q$  equals the minimal posterior for the receiver to choose  $a_H$ . As  $q$  increases, the sender has less room for randomizing between messages in the low state while still maintaining incentives for the receiver. For  $\rho = 1$ , the result is immediate since the equilibrium posterior conditional on the persuasive message must equal  $q$ . Thus, as  $q$  increases, the sender reduces the probability of sending the persuasive message when the state is  $\theta_L$ , and therefore reveals more information. For  $\rho < 1$ , the same logic is similar: in both verifiable and unverifiable cases the sender uses the commitment stage to undo the outcome in the revision stage as much as possible to maintain the incentives of the receiver. As  $q$  increases, this involves revealing more information.

## B Additional Material

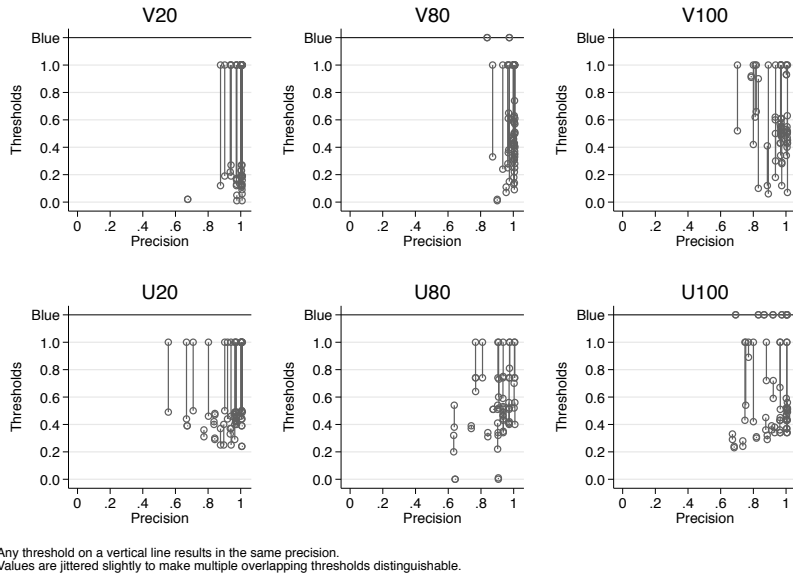


Figure B13: Ranges for estimated threshold strategies.

Figure B14 presents an analysis of individual sender's mean payoffs across treatments, and the standard deviation of this mean. This payoffs are computed by paring the actual behavior of a sender with the optimal behavior of a Bayesian receiver.

## C Design.

Figure C15 complements Figure 1 showing the other three relevant screenshots from our experiment. The top panel of Figure C15 shows what sender could see during Stage 2, conditional on the ball begin Red. A similar screen (omitted) was shown in case of a Blue ball. The middle panel shows the tasks of the receiver in Stage 3. On the left, we reported to the receiver the Communication Plan chosen by the sender and reminded the receiver of the basic rule of the game. On the right, the receiver was supposed to express her best guess for every possible message she could receive. Finally, the bottom panel of Figure C15 shows the Feedback screen. All relevant information were reported to both players, with the exception of the sender's decision in the Revision stage.

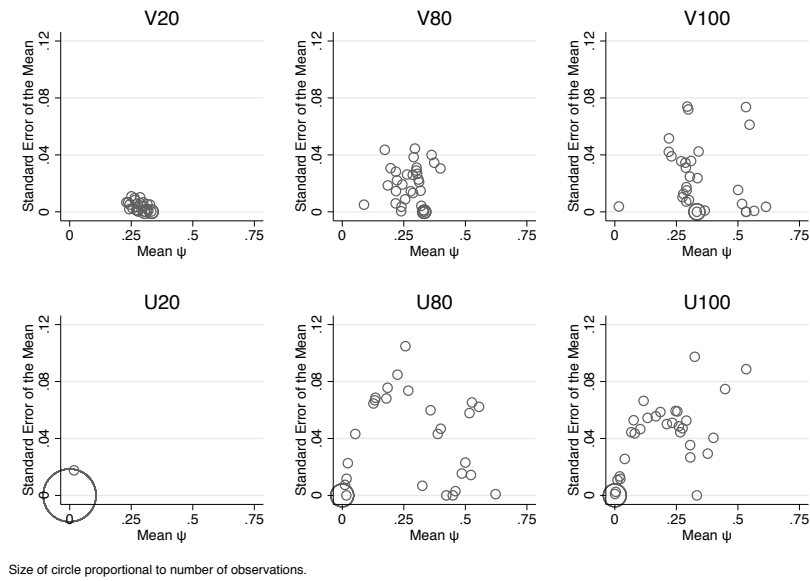


Figure B14: Senders' Payoffs Distribution when paired with Bayesian Receiver.

## D Instructions for V80.

In this section, we reproduce instruction for one of our treatment, V80. These instruction were read out aloud so that everybody could hear. A copy of these instructions was handout to the subject and available at any point during the experiment. Finally, while reading Section D.2.1, screenshots similar to those in Appendix C, were shown to subjects, to ease the exposition and the understanding of the tasks.

### D.1 Welcome:

You are about to participate in a session on decision-making, and you will be paid for your participation with cash vouchers (privately) at the end of the session. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance. On top of what you will earn during the session, you will receive an additional \$10 as show-up fee.


Please turn off phones and tablets now. The entire session will take place through computers. All interaction among you will take place through computers. Please do not talk or in any way try to communicate with other participants during the session. We will start with a brief instruction period. During the instruction period you will be given a description of

Lab 1 Match 1 of 2 You are the Sender

### Update Stage

Here you can Update your COMMUNICATION PLAN.  
The Receiver cannot see how you UPDATE your COMMUNICATION PLAN.

The Ball is Red.




The message that you will send will be generated:

- With Probability 80%, from the COMMUNICATION PLAN you chose at the previous stage.
- With Probability 20%, from the UPDATE you choose now.

Send Message with probability:

Red	37 %
Blue	40 %
No Message	23 %



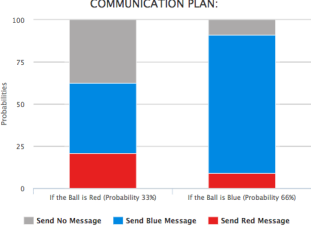
Lab 2 Match 1 of 2 You are the Receiver

### Guessing Stage

The message you will receive will come:

- with probability 20%, from the UPDATE, that you can't see.
- with probability 80%, from the COMMUNICATION PLAN you see below.

**COMMUNICATION PLAN:**



Choose your GUESSING PLAN:

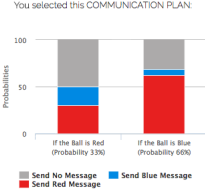
If I Receive Message...	...my guess will be:
The Ball is Red	<input type="button" value="RED"/> <input type="button" value="BLUE"/>
The Ball is Blue	<input type="button" value="RED"/> <input type="button" value="BLUE"/>
No Message	<input type="button" value="RED"/> <input type="button" value="BLUE"/>

Lab 1 Match 1 of 2 You are the Sender

### Summary:

Ball Color	Message Sent	Origin	Guess	Your Payoff	Opponent's Payoff
● xxx	✉ xxx	✉ xxx	⊙ xxx	xx Dollars	xx Dollars

You selected this COMMUNICATION PLAN:



the Receiver selected this GUESSING PLAN:

If I receive Message Red, I will guess "xxx"  
If I receive Message Blue, I will guess "xxx"  
If I receive No Message, I will guess "xxx"

When you are done, press Continue to proceed




Figure C15: Screenshots from the experiment, Treatment U80.

the main features of the session. If you have any questions during this period, raise your hand and your question will be answered privately.

## D.2 Instructions:

You will play for 25 matches in either of two roles: **sender** or **receiver**. At the beginning of every Match one ball is drawn at random from an urn with three balls. Two balls are BLUE and one is RED. The receiver earns \$2 if she guesses the right color of the ball. The sender's payoff only depends on the receiver's guess. She earns \$2 only if the receiver guesses RED. Specifically, payoffs are determined illustrated in Table D7.

	If Ball is Red		If Ball is Blue	
If Receiver guesses Red	Receiver \$2	Sender \$2	Receiver \$0	Sender \$2
If Receiver guesses Blue	Receiver \$0	Sender \$0	Receiver \$2	Sender \$0

Table D7: Payoffs

The sender learns the color of the ball. The receiver does not. The sender can send a message to the receiver. The messages that the sender can choose among are reported in Table D8.

If Ball is Red:

- Message: “*The Ball is Red.*”
- No Message.

If Ball is Blue:

- Message: “*The Ball is Blue.*”
- No Message.

Table D8: Messages

Each Match is divided in three stages: Communication, Update and Guessing.



1. Communication Stage: before knowing the true color of the ball, the sender chooses a COMMUNICATION PLAN to send a message to the receiver.
2. Update Stage: A ball is drawn from the urn. The computer reveals its color to the sender. The sender can now UPDATE the plan she previously chose.
3. Guessing Stage: The actual message received by the receiver may come from the Communication stage or the Update stage. Specifically, with probability 80% the message comes from the Communication Stage and with probability 20% it comes from the Update Stage. The receiver will not be informed what stage the message comes from. The receiver can see the COMMUNICATION PLAN, but she cannot see the UPDATE. Given this information, the receiver has to guess the color of the ball.

At the end of a Match, subjects are randomly matched into new pairs. We now describe what happens in each one of these stages and what each screen looks like:

### **D.2.1 Communication Stage: (Only the sender plays)**

In this stage, the sender doesn't yet know the true color of the ball. However, she instructs the computer on what message to send once the ball is drawn. In the left panel, the sender decides what message to send if the Ball is Red. In the right panel, she decides what message to send if the Ball is Blue. We call this a COMMUNICATION PLAN.

Every time you see this screen, pointers in each slider will appear in a different random initial position. The position you see now is completely random. If I had to reproduce the screen once again I would get a different initial position. By sliding these pointers, the sender can color the bar in different ways and change the probabilities with which each message will be sent. The implied probabilities of your current choice can be read in the table above the sliders.

When clicking Confirm, the COMMUNICATION PLAN is submitted and immediately reported to the receiver.

### **D.2.2 Update Stage: (Only the sender plays)**

In this Stage, the sender learns the true color of the ball. She can now update the COMMUNICATION PLAN she selected at the previous stage. We call this decision UPDATE. The

receiver will not be informed whether at this stage the sender updated her COMMUNICATION PLAN.

### D.2.3 Guessing Stage. (Only the receiver plays)

While the sender is in Update Stage, the receiver will have to guess the color of the ball. On the left, she can see the COMMUNICATION PLAN that the sender selected in the Communication Stage. By hovering on the bars, she can read the probabilities the sender chose in the Communication Stage. Notice that the receiver cannot see whether and how the sender updated her COMMUNICATION PLAN in the Update Stage. On the right, the receiver needs to express her best guess for each possible message she could receive. We call this A GUESSING PLAN. Notice that once you click on these buttons, you won't be able to change your choice. Every click is final.

### D.2.4 How is a message generated?

With 80% probability	With 20% probability
The message is sent according to COMMUNICATION PLAN	The message is sent according to UPDATE
(Remember: COMMUNICATION PLAN is always seen by the Receiver)	(Remember: UPDATE is never seen by the Receiver)

## D.3 Practice Rounds:

Before the beginning of the experiment, you will play 2 Practice rounds. These rounds are meant for you to familiarize yourselves with the screens and tasks of both roles. You will be both the sender and the receiver at the same time. All the choices that you make in the Practice Rounds are unpaid. They do not affect the actual experiment.

## D.4 Final Summary:

Before we start, let me remind you that.

- The receiver wins \$2 if she guesses the right color of the ball.

- The sender wins \$2 if the receiver says the ball is Red, regardless of its true color.
- There are three balls in the urn: two are Blue (66.6% probability), one is Red (33.3% probability). After the Practice rounds, you will play in a given role for the rest of the experiment.
- The message the receiver sees is sent with probability 80% using COMMUNICATION PLAN and with probability 20% using UPDATE.
- The choice in the Communication Stage is communicated to the receiver. The choice in the Update stage is not.
- At the end of each Match you are randomly paired with a new player.

## References

- ALONSO, R. AND O. CAMARA (2016): “Persuading Voters,” *The American Economic Review*, 106, 3590–3605(16).
- AUSTEN-SMITH, D. (1993): “Information and Influence: Lobbying for Agendas and Votes,” *American Journal of Political Science*, 37(3), 799–833.
- BATTAGLINI, M. (2002): “Multiple Referrals and Multidimensional Cheap Talk,” *Econometrica*, 70(4), 1379–1401.
- BENNDORF, V., D. KÜBLER, AND H.-T. NORMANN (2015): “Privacy concerns, voluntary disclosure of information, and unraveling: An experiment,” *European Economic Review*, 75, 43–59.
- BLUME, A., D. DE JONG, Y. KIM, AND G. SPRINKLE (1998): “Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games,” *The American Economic Review*, 88, 1323–1340.
- BLUME, A., E. K. LAI, AND W. LIM (2017): “Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations,” *Working Paper*.
- CAI, H. AND J. T. Y. WANG (2006): “Overcommunication in strategic information transmission games,” *Games and Economic Behavior*, 56, 7–36.
- CRAWFORD, V. P. AND J. SOBEL (1982): “Strategic Information Transmission,” *Econometrica*, 50, 1431–1451.
- DICKHAUT, J., M. LEDYARD, A. MUKHERJI, AND H. SAPRA (2003): “Information management and valuation: an experimental investigation,” *Games and Economic Behavior*, 44, 26–53.
- DICKHAUT, J., K. MCCABE, AND A. MUKHERJI (1995): “An Experimental Study of Strategic Information Transmission,” *Economic Theory*, 6, 389–403.
- DRANOVE, D. AND G. JIN (2010): “Quality Disclosure and Certification: Theory and Practice,” *Journal of Economic Literature*, 48, 935–963.
- DYE, R. (1985): “Disclosure of Nonproprietary Information,” *Journal of Accounting Research*, 23(1), 123–145.

- FORSYTHE, R., R. M. ISAAC, AND T. R. PALFREY (1989): “Theories and Tests of ”Blind Bidding” in Sealed-bid Auctions,” *The RAND Journal of Economics*, 20, 214–238.
- FORSYTHE, R., R. LUNDHOLM, AND T. RIETZ (1999): “Cheap Talk, Fraud, and Adverse Selection in Financial Markets: Some Experimental Evidence,” *The Review of Financial Studies*, 12, 481–518.
- GALOR, E. (1985): “Information Sharing in Oligopoly,” *Econometrica*, 53, 329–343.
- GENTZKOW, M. AND E. KAMENICA (2014): “Costly Persuasion,” *American Economic Review*, 104, 457–462.
- (2016): “Competition in Persuasion,” *The Review of Economic Studies*, 84(1), 300–322.
- GILLIGAN, T. W. AND K. KREHBIEL (1987): “Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures,” *Journal of Law, Economics, and Organization*, 3, 287–335.
- (1989): “Information and Legislative Rules with a Heterogeneous Committee,” *American Journal of Political Science*, 33, 459–490.
- GROSSMAN, S. J. (1981): “The Informational Role of Warranties and Private Disclosure about Product Quality,” *The Journal of Law and Economics*, 24, 461.
- HAGENBACH, J., F. KOESSLER, AND E. PEREZ-RICHET (2014): “Certifiable Pre-Play Communication: Full Disclosure,” *Econometrica*, 82(3), 1093–1131.
- JIN, G. AND P. LESLIE (2003): “The effect of information on product quality: Evidence from restaurant hygiene grade cards,” *The Quarterly Journal of Economics*.
- JIN, G., M. LUCA, AND D. MARTIN (2016): “Is no news (perceived as) bad news? An experimental investigation of information disclosure,” *NBER Working Paper*.
- JOVANOVIC, B. (1982): “Truthful Disclosure of Information,” *Bell Journal of Economics*, 13, 36–44.
- KAMENICA, E. AND M. GENTZKOW (2011): “Bayesian persuasion,” *American Economic Review*, 101, 2590–2615.

- KING, R. AND D. WALLIN (1991): “Market-induced information disclosures: An experimental markets investigation ,” *Contemporary Accounting Research*, 8, 170197.
- MATHIOS, A. (2000): “The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market,” *The Journal of Law and Economics*.
- MILGROM, P. (1981): “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 12, 380–391.
- MIN, D. (2017): “Bayesian Persuasion under Partial Commitment,” *Working Paper*.
- OKUNO-FUJIWARA, M., A. POSTLEWAITE, AND K. SUZUMURA (1990): “Strategic Information Revelation,” *The Review of Economic Studies*, 57, 25–47.
- SÁNCHEZ-PAGÉS, S. AND M. VORSATZ (2007): “An experimental study of truth-telling in a senderreceiver game,” *Games and Economic Behavior*, 61, 86–112.
- VERRECCHIA, R. E. (1983): “Discretionary Disclosure.” *Journal of Accounting and Economics*, 5, 179–194.
- WANG, J., M. SPEZIO, AND C. CAMERER (2010): “Pinocchio’s Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games,” *The American Economic Review*, 100, 984–1007.