Rules and Commitment in Communication: An Experimental Analysis

Guillaume R. Fréchette
New York University

Alessandro Lizzeri
New York University

Jacopo Perego
Columbia University

June 11, 2020

Abstract

We study the role of commitment in communication and its interactions with rules, which determine whether or not information is verifiable. Our framework nests models of cheap talk, information disclosure, and Bayesian persuasion. Our model predicts that commitment has opposite effects on information transmission under the two alternative rules. We leverage these contrasting forces to experimentally establish that subjects react to commitment in line with the main qualitative implications of the theory. Quantitatively, not all subjects behave as predicted. We show that a form of commitment blindness leads some senders to overcommunicate when information is verifiable and undercommunicate when it is not. This generates an unpredicted gap in information transmission across the two rules, suggesting a novel role for verifiable information in practice.

JEL Codes: C92, D83, D82, D91

We thank Andreas Blume, Elliot Lipnowski, Salvatore Nunnari, Santiago Oliveros, Sara Shahanaghi, and Emmanuel Vespa for useful comments. Fréchette and Lizzeri gratefully acknowledge financial support from the National Science Foundation via grant SES-1558857.

1 Introduction

The goal of this paper is to experimentally study the effects of *rules* and *commitment* in communication. We think of rules as restrictions on the language. They determine, for instance, whether an agent can freely misreport what she knows or whether she can only use verifiable information. Commitment captures the extent to which the agent can communicate according to predetermined protocols. In the literature, commitment has been used to model, for instance, how a school chooses a grading policy and a firm chooses an accounting standard.[1] Together, rules and commitment are defining features at the heart of any communication environment. For instance, models of cheap talk, information disclosure, and Bayesian persuasion differ among each other in ways that lead back to differences in rules and commitment. In many concrete applications, it is difficult to measure the exact degree of commitment available to an agent or the extent to which rules are enforced. Yet, rules and commitment do vary significantly in practice, depending on the context and observables such as the frequency of communication. Thus, studying their effects on communication is a natural question.

We present a simple model of communication under *partial* commitment and consider two opposing rules: *verifiable* and *unverifiable* information. The focus on partial commitment is a key feature of our analysis: it allows us to nest many existing communication models under the same umbrella and experimentally test key qualitative predictions about the role of commitment in communication. The contrast between verifiable and unverifiable information further enriches our analysis, as our main comparative static predictions have opposite signs under these two alternative rules. Our main results indicate clear treatment effects in line with the main qualitative predictions of the theory. We also uncover important quantitative deviations. Specifically, we find that rules matter in ways that are unpredicted by the theory; we propose a systematic rationalization for these departures.

We consider a sender-receiver model with binary states and actions. The sender wants the receiver to choose a high action, whereas the receiver wishes to match the state. There are three stages. In the commitment stage, the sender publicly commits to an information structure, which is a map between states and messages. Under unverifiable information, the sender can freely misreport her private information. Under verifiable information, she can only conceal it. In the revision stage, the sender learns the state and can privately revise the chosen information structure. In the last stage, the receiver observes a message and chooses an action. The message is generated with probability ρ from the commitment stage and with the remain-

[1] Bergemann and Morris (2019) and Kamenica (2019) survey this literature.

ing probability from the revision stage. We view the probability ρ as capturing the sender's commitment power: the higher ρ is, the higher the probability that the sender will *not* be able to revise her strategy after learning the state and thus, the higher the extent to which she is committed to her initial communication. Variations in commitment power generate predictions that are qualitatively different depending on the communication rule. For example, an increase in the sender's commitment power should *increase* the amount of information conveyed under unverifiable information, whereas it should *decrease* it under verifiable information. When the sender can fully commit, these two scenarios coincide and the equilibrium informativeness is independent of the communication rule. We exploit these predictions to experimentally test the role of commitment in communication.

With this framework, we nest models of cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007), disclosure (Grossman, 1981; Milgrom, 1981; Jovanovic, 1982; Okuno-Fujiwara et al., 1990), and Bayesian persuasion (Kamenica and Gentzkow, 2011). Thus, we span a considerable portion of the strategic information transmission models that have been discussed in the literature in the last few decades. This helps in organizing our analysis in two ways. First, the comparison *across* models generates asymmetric predictions that go to the heart of the strategic tension of communication under commitment. As we illustrate in the paper, these asymmetries discipline which explanations can be used to rationalize potential departures from the theory. Second, the framework itself informs a parsimonious experimental design. In our treatments, we change two parameters—the degree of commitment ρ and the verifiability of information—while leaving the underlying structure of the game unchanged.

We begin by establishing several patterns in the data that are consistent with the key qualitative predictions of the theory. More specifically, we present two main sets of findings. First, we show that both senders and receivers react to commitment. For senders, we exploit within-treatment variation to show that between the commitment and revision stage, their behavior changes in the directions predicted by the theory. When information is unverifiable, they reveal much more information in the commitment stage than in the revision stage. When information is verifiable, this ranking is reversed, as predicted by the theory. For receivers, we exploit across-treatment variation to show that, as commitment increases, they become more responsive to information from the commitment stage. That is, they understand that information conveyed in the commitment stage is more meaningful when the level of commitment is higher. In our second main finding, we test how increasing commitment power changes the overall informativeness of communication. In line with the theory, we find that informativeness increases with commitment in treatments with unverifiable information and decreases with commitment in treatments with verifiable information. Furthermore, we find that verifiability has the predicted effect of increasing the amount of information conveyed by senders. Overall, these strong treatment effects validate the qualitative implications of the theory.

We then analyze the main quantitative deviations from the theory that we observe in the data. In treatments with low commitment, we replicate existing findings in the literature by showing that relative to the predictions of the theory, senders undercommunicate when information is verifiable and overcommunicate when it is not.[2] However, we find that the opposite holds in treatments with high commitment: senders overcommunicate when information is verifiable and undercommunicate when it is not. These deviations create an *informativeness gap* between verifiable and unverifiable treatments, which is particularly apparent in the limiting case of full commitment: empirically, informativeness is higher when information is verifiable than when it is not, even though in theory, the informativeness should be the same. From a policy perspective, this excess informativeness presents a novel justification for making it more difficult for senders to misreport their information.

We discuss the extent to which a model with boundedly rational agents may help explain these deviations. We note that a number of plausible biases that have been explored in prior work—such as lying-averse senders or non-Bayesian receivers—cannot rationalize the observed deviations. Thus, we consider the possibility that a fraction of senders is *commitment blind*: they behave under commitment as if they had no commitment power whatsoever. These senders are incapable of exploiting commitment to their advantage. In both stages, they choose a strategy that is optimal under no commitment. This bias has different implications on informativeness depending on the communication rule and, in particular, could explain the observed informativeness gap. To find evidence for commitment blindness, we look at treatments with partial commitment, where we can observe the behavior of the same sender in scenarios with and without commitment power. Our analysis reveals that there is a group of senders who behave in ways that are compatible with commitment blindness. To evaluate whether this explanation is fully capable of accounting for the quantitative departures from theory, we estimate a structural model of Quantal Response Equilibrium (QRE). By clustering the observed senders' strategies in treatment-specific representative groups, we can capture the typical behavior of commitment-blind senders. For each treatment, we then simulate data from our estimated model and find that the model-implied equilibrium informativeness can explain a considerable part of the gap observed in the data.

[2] For cheap talk, see the survey by Blume et al. (2020). For information disclosure, see Jin et al. (2020) and references therein.

Related Literature. The role of commitment in communication is at the center of the recent literature on persuasion and information design (Kamenica, 2019; Bergemann and Morris, 2019). To study the effects of commitment, we innovate by considering a versatile framework in which commitment can be varied experimentally. Recent theoretical contributions by Lipnowski et al. (2018) and Min (2017) analyze in greater generality the implications of partial commitment under unverifiable information.[3] In a framework with no commitment, Kartik (2009) studies changes in lying costs, bridging models of cheap talk and information disclosure.

Our paper relates to a large body of experimental literature on cheap talk, which has been recently surveyed by Blume et al. (2020). Models of cheap talk feature no commitment and unverifiable information and have been used to study a variety of phenomena, including lobbying (Austen-Smith, 1993; Battaglini, 2002) and the interaction between legislative committees and a legislature (Gilligan and Krehbiel, 1987, 1989). Dickhaut et al. (1995) was the first experimental paper to test the central prediction of Crawford and Sobel (1982) 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. By contrast, in the experiment, 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 meanings emerge endogenously. Among other findings, they confirm that, as in Forsythe et al. (1999), receivers are gullible. Cai and Wang (2006) also vary preference alignment and find that senders overcommunicate relative to the predictions of the cheap-talk model and that receivers are overly trusting.[4]

Our paper also relates to the literature on information disclosure. Disclosure models feature no commitment but verifiable information and have been used to study quality disclosure by a privately informed seller (e.g., Verrecchia, 1983; Dye, 1985; Galor, 1985). Milgrom (2008) and Dranove and Jin (2010) survey this literature. In contrast to experiments on cheap talk, experiments on the disclosure of verifiable information typically find that sender undercommunicate relative to the theoretical predictions. For instance, Jin et al. (2020) find that receivers are insufficiently skeptical when senders do not provide any information, which in turn leads senders to under provide information.[5] Jin et al. (2019) and de Clippel and Rozen (2020) find

[3] Perez-Richet and Skreta (2018) study a model of interim information manipulation under full commitment.

[4] See also Sánchez-Pagés and Vorsatz (2007), Wang et al. (2010), and Wilson and Vespa (2020).

[5] See also Forsythe et al. (1989), King and Wallin (1991), Dickhaut et al. (2003), Forsythe et al. (1999),

evidence for strategic obfuscation of verifiable evidence in settings with no and full commitment, respectively. Information unraveling has also been studied in the field. For instance, Mathios (2000) and Jin and Leslie (2003) document the failures of information unraveling for food nutrition labels and hygiene grade cards in restaurants.

One of our treatments replicates the leading example in Kamenica and Gentzkow (2011) and is one of the first tests of Bayesian persuasion. This treatment features full commitment and unverifiable information. Other papers have studied a similar treatment with different designs and goals. Aristidou et al. (2019) compare the design of information and monetary incentives. Their remarkably simple implementation imposes some aspects of the equilibrium behavior onto subjects' tasks. In their findings, senders are able to extract a higher rent from receivers when using information rather than monetary incentives. On average, senders' strategies are close to equilibrium—a result that is in line with one of our findings. Au and Li (2018) augment Bayesian persuasion with reciprocity and test their model in the lab. In their implementation, senders directly choose posteriors instead of information structures. This simplifies senders' tasks and eliminates the need for receivers to do Bayesian updating. Their results highlight interesting inconsistencies relative to the standard theory. Finally, Nguyen (2017) uses an intuitive interface for senders and allows them to choose among a small set of precompiled communication strategies. Overall, given receivers' behavior, a large fraction of senders behave optimally and their behavior involves partial information transmission.

2 Theoretical Framework

In this section, we present our theoretical framework and discuss its main predictions. The model achieves two goals. First, it captures settings where the sender has only *partial* commitment power. Second, it highlights the contrast between *verifiable* and *unverifiable* information. These features generate a rich set of predictions that we then exploit in our experimental design.

2.1 Model

There are two players: a sender and a receiver. The sender has private information about the state, while the receiver can take an action that affects everyone's payoff. The sender communicates with the receiver by transmitting information, in an attempt to influence her action. More specifically, let $\Theta = \{\theta_L, \theta_H\}$ be the state space and $\mu_0 \in [0, 1]$ denote the common

Benndorf et al. (2015), Hagenbach et al. (2014), and Hagenback and Perez-Richet (2018).

prior probability that the state is θ_H. The receiver chooses an action in $A = \{a_L, a_H\}$, and her preferences are given by the following utility function:

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

Thus, the receiver wishes to match her actions to the state, and the relative cost of the mistakes in the two states is parametrized by q. A Bayesian receiver would choose action a_H whenever her posterior belief that the state is θ_H is larger than q. Thus, we call q the *persuasion threshold*. The sender's preferences are state-independent and given by $v(a) = \mathbb{1}(a = a_H)$. That is, the sender earns a positive payoff only if she successfully persuades the receiver to take action a_H. To make the problem interesting, we assume that $\mu_0 < q$. That is, absent further information, the receiver would choose a_L.

The sender communicates with the receiver by sending her information about the state. An information structure is a map $\pi : \Theta \to \Delta(M)$, with $M = \{\theta_L, \theta_H, n\}$ being the set of possible messages. Denote by Π^U the set of *all* such information structures and by Π the subset from which the sender can choose. The difference between Π and Π^U captures exogenous restrictions on the sender's strategies, or *communication rules*. We say that information is *unverifiable* if no restrictions are imposed on the sender, that is, $\Pi = \Pi^U$. We say that information is *verifiable* if, instead, $\Pi = \Pi^V := \{\pi \in \Pi^U : \pi(\theta_H|\theta_L) = \pi(\theta_L|\theta_H) = 0\}$. In other words, verifiability demands that message $m = \theta$ can only be sent by type θ. Therefore, we can interpret message $m = \theta$ as a certifiable statement asserting that the state is indeed θ. Conversely, message n is a statement that is neither true nor false and hence, cannot be verified.

The game unfolds in three consecutive stages. In the *commitment stage*, the sender publicly chooses a commitment strategy $\pi_C \in \Pi$ before learning the state θ. In the *revision stage*, at every history (π_C, θ), the sender privately observes θ and chooses $\pi_R \in \Pi$. Since π_R is chosen after observing θ, the sender has no commitment power in the revision stage. In the *guessing stage*, the final stage of the game, a message m realizes with probability $\rho \in [0, 1]$ from $\pi_C(\cdot|\theta)$ and $(1-\rho)$ from $\pi_R(\cdot|\theta)$. For every history $(\pi_C, \pi_R, \theta, m)$, the receiver observes (π_C, m) and takes an action $a(\pi_C, m) \in A$. The receiver updates her prior belief μ_0 according to some belief assessment $\mu(m, \pi_C, \pi_R)$ that assigns a posterior belief to each message m, possibly as a function of π_C and π_R. We use Perfect Bayesian Equilibrium (PBE) as a solution concept.

2.2 Discussion

We refer to ρ as the sender's *degree of commitment*. It measures the extent to which the sender is able to commit to her initial strategy π_C. For high values of ρ, the commitment strategy π_C is likely to be the one that determines the final message m. Conversely, for low values of ρ, the final message m is likely to be determined by the choice in the revision stage, after the sender has learned the state.[6]

Our framework is characterized by three main parameters: (i) the communication rule, Π^U versus Π^V, (ii) the degree of commitment ρ, (iii) the persuasion threshold q. This framework can nest several classic communication models as special cases. When $\rho = 0$ and information is unverifiable, our model captures cheap-talk communication. When $\rho = 0$ and information is verifiable, our model captures a disclosure game with verifiable communication. Finally, when $\rho = 1$ and information is unverifiable, our model becomes a Bayesian persuasion game.

An equilibrium outcome of particular interest for us is the informativeness of the equilibrium strategy, that is, the amount of information that the sender conveys to the receiver. We say that an equilibrium (π_C, π_R, a, μ) under parameters (Π, ρ, q) is more informative than equilibrium $(\pi'_C, \pi'_R, a', \mu')$ under parameters (Π', ρ', q') if the on-the-equilibrium-path information structure $\rho\pi_C + (1-\rho)\pi_R$ is more informative than $\rho'\pi'_C + (1-\rho')\pi'_R$. We measure the informativeness of an information structure $\pi \in \Pi$ as the correlation between the state and the action it induces.[7] We denote such correlation by $\phi^B(\pi)$. More formally, fix q and an arbitrary π. Define $a(m, \pi) \in A$ to be action that a Bayesian receiver would choose upon receiving message m from π. Then, $\phi^B(\pi) := \text{Corr}(\theta, a(m, \pi)) \in [0, 1]$. We say that an information structure is *uninformative* if $\phi^B(\pi) = 0$, and that it is *fully informative* if $\phi^B(\pi) = 1$. We say that π is *more informative* than π' if $\phi^B(\pi) \geq \phi^B(\pi')$.

As in many communication games, our framework allows for multiple PBEs. We provide a full equilibrium characterization in Appendix C. In the rest of the paper, we impose a simple tie-breaking rule on equilibrium behavior that is inspired by Hart et al. (2017). We say that a PBE is *truth leaning* if, whenever it is weakly optimal for type θ_H in the revision stage to tell the truth (i.e. to send message $m = \theta_H$), she does so. This tie-breaking rule is simple

[6] Alternative but equivalent interpretations are possible. One can think of the sender as having an *opportunity* to revise her commitment strategy after learning the state, which occurs only with probability $1 - \rho$. Another 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 ρ is then the probability that the sender is not this opportunistic type.

[7] There are other ways to measure informativeness. We offer a detailed discussion of this choice and its alternatives in Section 3.3.

but powerful. As we show in Proposition 1, it is sufficient to guarantee the uniqueness of equilibrium outcomes. Moreover, it is weaker than the refinement introduced in Hart et al. (2017). Indeed, it is not imposed on all types of senders in the revision stage, but only on type θ_H.[8] A fortiori, our tie-breaking rule is consistent with many of the equilibrium refinements that have been proposed in the literature.[9] Finally, this tie-breaking rule is consistent with our data. For example, when information is verifiable, the average $\pi_R(\theta_H|\theta_H)$ in our data is about 0.95. In the rest of the paper, we maintain the specialization to truth leaning PBEs and we refer to these as *equilibria*, without further qualification.

2.3 Main Predictions

We now describe the main comparative statics that we later bring to the lab. We begin with a characterization of equilibrium informativeness for a *fixed* level of commitment power ρ and contrast the equilibrium informativeness between the commitment and the revision stages.

Proposition 1. *Fix ρ. Let $\underline{\rho} := \frac{q-\mu_0}{q(1-\mu_0)}$ and $\bar{\rho} := \frac{q(1-\mu_0)}{q(1-\mu_0)+(1-q)\mu_0}$, and note that $\underline{\rho} \leq \bar{\rho}$:*

> [Unverifiable Information] *All equilibria at ρ have the same informativeness. In particular, these equilibria are* uninformative *if and only if $\rho < \underline{\rho}$. Moreover, when $\rho \geq \underline{\rho}$,* less *information is transmitted in the revision stage than in the commitment stage.*
>
> [Verifiable Information] *All equilibria at ρ have the same informativeness. In particular, these equilibria are* fully informative *if and only if $\rho < \bar{\rho}$. Moreover, when $\rho \geq \bar{\rho}$,* more *information is transmitted in the revision stage than in the commitment stage.*

This result establishes the uniqueness of the equilibrium *outcomes* for each rule and commitment level and it highlights the main tension between the commitment and revision stages. This tension manifests itself in opposite ways under the two alternative rules, thus providing useful and testable predictions that we will exploit in our experimental analysis. To understand this result, we first consider two extreme cases. When $\rho = 0$, the sender has no commitment power. Therefore, equilibria are fully informative when information is verifiable and uninformative otherwise. When $\rho = 1$, the sender has full commitment power. The equilibria feature partial information revelation in both of the verifiability scenarios that we consider. The

[8] More specifically, we do not require that, whenever type θ_L is indifferent in the revision stage she must send message θ_L. When information is unverifiable, this extra requirement is too strong and can lead to non existence of equilibria.

[9] See online Appendix C.5 of Hart et al. (2017)

intuition for Proposition 1 is then the following. Under both verifiable and unverifiable information, the sender would like to commit to persuading the receiver to choose the high action as often as possible, and this requires partial information revelation. However, in the revision stage, the sender is unable to resist the temptation to undo her commitments and manipulate information in her favor. Under verifiable information, this opportunity implies full information disclosure in the revision stage; under unverifiable information, it implies sending the message that induces the high action, regardless of the state (i.e. being uninformative). The presence of the revision stage changes the sender's problem in the commitment stage relative to the full commitment case: relative to the revision stage, the sender overcommunicates when information is unverifiable and undercommunicates when information is verifiable. These commitment strategies are an attempt to obtain final posteriors that are as close as possible to the full commitment scenario. When ρ is sufficiently high, partial information revelation occurs in both verifiability scenarios because the revision stage cannot completely undo the positive effect of the commitment stage. Overall, this result illustrates how changes in the rules can generate stark contrasts in the way senders react to commitment power.

Our next result describes how equilibrium informativeness changes with commitment power and how this depends on the communication rule.

Proposition 2. *Fix $q > \mu_0$. When information is unverifiable, equilibrium informativeness weakly increases in ρ. When information is verifiable, equilibrium informativeness weakly decreases in ρ. Moreover, when $\rho = 1$, equilibrium informativeness is independent of the communication rules.*

This result illustrates that changes in commitment affect equilibrium informativeness in starkly different ways depending on the communication rules.[10] The intuition for this result follows from the discussion above. As ρ increases, the revision stage becomes increasingly less likely, and the relevance of the commitment stage increases. This allows the sender to approach the optimal solution under full commitment, $\rho = 1$. In our game, the equilibrium outcome for $\rho = 1$ is independent of the rules of communication. To see this, note that when $\rho = 1$ and information is verifiable, the sender can replace the use of message θ_H with message n. By doing so, she can induce the same joint distribution over states and actions that is optimal under unverifiable information.

[10] These stark comparative statics hinge on the binary structure of our environments and the equilibrium refinement. They may fail in more general environments. See Appendix C.1 and Lipnowski et al. (2018).

3 Experimental Design

In this section, we describe the laboratory implementation of our model, the main treatments that we conducted, and the different ways with which we measure informativeness from the data. We view our experimental design as a particularly useful framework to organize our analysis of commitment and communication rules. As we illustrate in the next sections, subject behavior in any given treatment is heterogeneous and challenging to evaluate on its own. In contrast, the comparison across treatments, along with the asymmetric nature of our predictions, goes to the heart of the strategic tension in our model.

3.1 Lab Implementation and Treatments

We begin by describing the implementation of the base game. An urn contains three balls, two blue ($\theta_L = B$) and one red ($\theta_H = R$). A ball is drawn at random and $\mu_0(\theta = R) = 1/3$. The receiver takes a guess $a \in \{red, blue\}$ and earns \$2 if a matches the color of the ball θ. She earns nothing otherwise. The sender wins \$2 if $a = red$, irrespective of the state θ.

The game has three stages.[11] In the commitment stage, the sender chooses an information structure. She does so via a simple graphical interface. The sender selects $\pi_C(\cdot|\theta)$ by moving a slider, one for each state. The slider's bar is colored according to the conditional probabilities implied by the sender's choice. These probabilities are updated in real time in a table above the slider bar. In the revision stage, the sender learns the color of the ball θ. With the same interface as the one just described, she can revise the part of her strategy that concerns the *realized* state. We do not elicit the sender's choice for the state that did not realize. This design choice makes the revision stage simpler and highlights the stark contrast between the commitment and revision stage. Moreover, the revision stage is presented to the sender only when it matters, that is, when commitment is partial. To minimize confusion, we do not show the revision stage for treatments with full commitment. In the guessing stage, the receiver observes the information structure chosen by the sender in the commitment stage but not the one chosen in the revision stage. We use the strategy method to elicit the receiver's choice: She makes a guess for each possible message she could receive.

We have a 2×3 factorial between-subject design. Our experimental variables are the sender's commitment power ρ and the communication rules (verifiable versus unverifiable informa-

[11] In the lab, we referred to the these three stages with neutral labels: the *communication*, *update*, and *guessing* stage. In the remainder of the paper, we maintain instead the nomenclature introduced in the previous section.

Table 1: Treatments Denominations

	Sender's Commitment Power		
Information	$\rho = 0.20$	$\rho = 0.80$	$\rho = 1$
Verifiable	$V20$	$V80$	$V100$
Unverifiable	$U20$	$U80$	$U100$

tion). For each rule, we conducted three treatments with different degrees of commitment: $\rho \in \{0.20, 0.80, 1\}$. This gives us a total of six treatments, which constitute the bulk of our investigation. Treatments are denoted as illustrated in Table 1. In treatments with verifiable information, the interface prevents senders from assigning positive probability to a red message conditional on a blue ball or to a blue message conditional on a red ball. The interfaces are identical in all other respects.

Table 2 reports the equilibrium *strategy* predictions for each treatment. Figure 1 reports the predicted equilibrium *outcomes*. This set of treatments captures the key tensions of our model. First, treatments $V80$ and $U80$ reveal the tension between the commitment and the revision stage, as summarized by Proposition 1. This tension goes in opposite directions according to whether or not information is verifiable. Second, informativeness is increasing in ρ when information is unverifiable, while the opposite holds when information is verifiable. Third, treatments $U100$ and $V100$ are predicted to induce an identical outcome through senders' strategies that are substantially different. In the following sections, we will exploit these tensions to test the role of commitment and rules in communication.

For each treatment, we conducted four sessions, for a total of 24 sessions. Each session included 12 to 24 subjects (16 on average) for a total of 384 subjects recruited from the NYU undergraduate population using *hroot* (Bock et al., 2014). At the beginning of each session, instructions were read aloud, and subjects were randomly assigned into a fixed role: sender or receiver. In each session, subjects played 25 paid rounds of the game described above, with random rematching between rounds. At the end of every round, complete feedback was provided to both senders and receivers. Appendix E.2 contains the instructions for one of our treatments. In addition to their earnings from the experiment, subjects received a $10 show-up fee. Average earnings, including the show-up fee, were $36.55, ranging from $12 to $60. On average, sessions lasted 100 minutes. Our statistical analysis focuses on the last ten rounds to allow enough time for subjects to familiarize themselves with the interface and to learn the relevant strategic forces in the task they faced. As can be seen in Appendix D.3, some aspects of behavior change over the course of the experiments.

Table 2: Equilibrium Predictions

	Sender								Receiver		Correlation
	Commitment				*Revision*				*Guessing*		Coefficient
Treat.	State	Message			State	Message			Mes.	Guess	$\phi = \phi^B$
		r	b	n		r	b	n			
$V20$	R	1		0	R	1		0	r	*red*	1
	B		x	$1 - x$	B		x	$1 - x$	b	*blue*	
									n	*blue*	
$V80$	R	0		1	R	1		0	r	*red*	0.57
	B		3/4	1/4	B		0	1	b	*blue*	
									n	*red*	
$V100$	R	0		1					r	*red*	1/2
	B		1/2	1/2					b	*blue*	
									n	*red*	
$U20$	R	x	y	$1 - x - y$	R	1	0	0	r	*blue*	0
	B	x	y	$1 - x - y$	B	1	0	0	b	*blue*	
									n	*blue*	
$U80$	R	1	0	0	R	1	0	0	r	*red*	1/2
	B	3/8	5/8	0	B	1	0	0	b	*blue*	
									n	*blue*	
$U100$	R	1	0	0					r	*red*	1/2
	B	1/2	1/2	0					b	*blue*	
									n	*blue*	

x and y indicate any (feasible) probability.

3.2 Discussion of Design Choices

We briefly discuss our main design choices.

Treatments. It is instinctive to think of $\rho \in \{1/3, 2/3, 1\}$ as natural parametric choices. However, it is important to take into account the theoretical thresholds $\underline{\rho}$ and $\bar{\rho}$. We choose $\rho = 0.80$ to allow equal distance between the theoretical threshold $\bar{\rho}$—key for verifiable information—and the full-commitment benchmark. The choice of $\rho = 0.20$ ensures symmetry. In our treatments, we do not include the extreme case of $\rho = 0$ for two main reasons. First, this case is the only one for which there is already experimental evidence, both for verifiable and unverifiable information. Our main interest lies in treatments with partial and full commitment: these cases have not been tested in the lab and offer a unique opportunity to study the role of commitment in communication. Second, the equilibrium outcomes at $\rho = 0$ are identical to those at $\rho = 0.20$. In particular, results from the revision stage of treatments with $\rho = 0.20$ refer to a setting where senders have effectively no commitment power and thus, should be seen as proxies for $U0$ and $V0$.

Human Receivers. Senders' behavior is the central and more novel aspect of our experiment. Of course, senders' behavior depends on their expectation of how best to manipulate receivers, which in turn depends on the receivers' observed behavior. One may think that there could

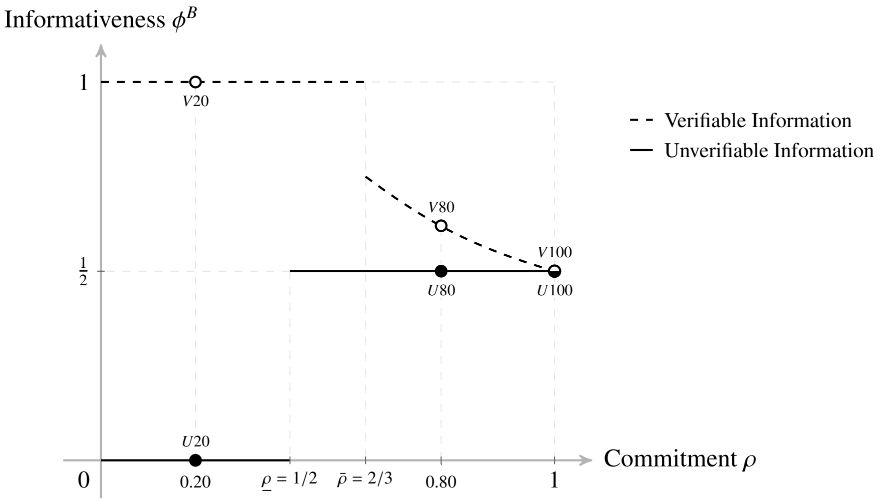


Figure 1: Predictions and Treatments.

be advantages to automating receivers' behavior to conform to the theory. We have three responses to this observation. First, we believe that senders' beliefs about how receivers interpret what message they see is central to understanding strategic communication. For instance, the main experimental finding in the literature on disclosure games, namely the failure of unraveling, would likely go undetected in a world with automated receivers. Second, the implementation of automated Bayesian receivers in the lab is far from trivial as it requires an explanation to senders of how the automated receivers behave. Failure to properly give this explanation defeats the potential purpose of introducing automated receivers. Third, as we show in Section 5.1 and Appendix A.1.2, many of our receivers are non-Bayesian, but their behavior is systematic and is monotone in information, a property that is sufficient for our comparative static exercise.

Message n. From a theoretical perspective, the inclusion of message n in treatments with unverifiable information may seem redundant. However, in the experiment, it allows us to switch from unverifiable to verifiable information with minimal changes to our design. This increases our ability to compare results between different communication rules. It is perhaps reassuring to note that the vast majority of senders in treatments with unverifiable information employs a "natural" language—that is, message n is only marginally used.[12]

Natural Language. Instead of using abstract labels for the messages, we label messages with colors that match the labels of the states—red and blue. In this way, messages can acquire a literal meaning. The focus of the paper is not on whether people understand how to coordinate on a language (Blume et al. (1998)). Thus, we wished to remove one potential obstacle to

[12] More specifically, the average total probability of message n, across all treatments with unverifiable information, is about 10%.

communication that would have complicated the subjects' task and our analysis. In the rest of the paper, as we explain our results, it is convenient to distinguish states and messages by denoting the former with upper case letters (R and B) and the latter with lower case letters (r, b, and n).

Fixed Roles. Before the beginning of the experiment, subjects played two unpaid practice rounds in which they played the game from both the sender's and the receiver's perspective. Then, subjects were assigned to a fixed role—sender or receiver—and played such role for the duration of the experiment. Because the tasks that subjects faced in our experiment were nontrivial, we thought it would be important for them to gain relevant experience in their role.

Additional Treatments. We conduct two robustness treatments, discussed in Appendix A. In our main treatments, payoffs are specified so that the persuasion threshold is $q = 1/2$. In an alternative payoff specification, we let $q = 3/4$. This allows us to test for changes in informativeness while keeping commitment and communication rules fixed. We also study a version of $U100$ with only two messages: r and b. This allows us to study the effects of the redundant message n in treatments with unverifiable information. We find that behavior in this robustness treatment is in line with $U100$, with slightly less noise.

Random Rematching. We chose to have random rematching of pairs of senders and receivers to simulate a one-shot interaction, while still allowing subjects to gain experience. Note, for instance, that experiments on duopoly games find that fixed pairing generates collusion, whereas random pairing does not Huck et al. (2001).

3.3 Measures of Informativeness

Empirically, we measure informativeness as the correlation between the color of the ball and the receiver's guess. This measure has been extensively used in the experimental literature on communication.[13] To fix ideas, suppose the sender truthfully discloses the color of the ball. Then, the receiver's final guess should be perfectly correlated with the state. Conversely, if the sender babbles, the receiver's final guess will be uncorrelated with the state.

To compute the correlation coefficient, we take advantage of our use of the strategy method in the communication and guessing stages to obtain significantly more precise measures of the correlation. However, in the revision stage, we only observe the sender's strategy conditional on the realized state. We circumvent this problem of missing data by imputing the session-

[13] See, for instance, Forsythe et al. (1999), Cai and Wang (2006), and Wang et al. (2010).

specific average behavior of the senders.[14] Thanks to this, we can analytically compute the Pearson correlation coefficients (specifically, the phi coefficients, since our variables are binary). Through simulations, we verified that the improvement in precision from using this method is significant.

We use two different versions of the correlation coefficient between the state and the guess. These two versions capture different aspects of equilibrium informativeness that are both useful in different ways. With our first measure, we compute the correlation by using the *observed* receiver's behavior. We denote such a measure by ϕ and refer to it as the *correlation coefficient*, without further qualifications. This way of measuring informativeness has the drawback of compounding the potential sender's inability to communicate with the potential receiver's unresponsiveness to information. Suppose, for instance, that the sender truthfully discloses the state, but the receiver does not listen. In this case, $\phi = 0$, although a great deal of information was offered to the receiver. To isolate the sender's behavior from the mistakes of the receivers, we use a second measure that we call *Bayesian correlation*, denoted by ϕ^B. This is the correlation coefficient implied by the sender's observed strategy combined with the guess of a hypothetical *Bayesian* receiver. Clearly, in a perfect Bayesian equilibrium, receivers are assumed to be Bayesian and thus $\phi = \phi^B$.

Informativeness of senders' strategies can also be measured by looking at moments of the distribution of induced (Bayesian) beliefs. This approach is akin to using ϕ^B, as it disregards the receiver's observed behavior. It differs from ϕ^B in the following sense. Consider a sender's strategy (π_C, π_R) inducing a distribution of posterior beliefs as follows: if $m = r$, the induced belief is just below $1/2$, if $m = b$, it is 0. Message n, instead, is sent with zero probability. This strategy does convey some information to the receiver. Yet, $\phi^B(\pi_C, \pi_R) = 0$ because a Bayesian receiver would guess *blue* regardless of the message. To circumvent this problem, we can measure informativeness by computing moments (e.g. variance) of the distribution of induced beliefs.

In summary, the correlation ϕ, the Bayesian correlation ϕ^B, and the moments of the distribution of induced beliefs present different advantages and are useful to highlight different aspects of our data. While choosing one or the other does not change the qualitative conclusion

[14] This choice seems natural and, due to the random rematching, receivers should hold comparable beliefs when facing a random sender in the last ten rounds of the experiment. Our results are robust to different imputation methods: For example, we can impute *subject*-specific averages and get essentially similar results. Also, it is important to note that the results for treatments with $\rho = 0.80$ (where we perform the imputation) are similar to those with $\rho = 1$ (where we do not need to use the imputation), suggesting the results are robust to our imputation method.

of our analysis nor the prediction of our theory,[15] these measures will be part of our toolbox in the next sections.

4 Treatment Effects

In this section, we present the treatment effects, which are in line with the main predictions of our theory. We will discuss two main sets of results. In Section 4.1, we test Proposition 1 by looking at how senders' behavior changes between the commitment and the revision stages as well as how receivers' responsiveness to information changes with commitment. In Section 4.2, we test Proposition 2 and analyze how informativeness changes as we vary the level of commitment. The predicted changes have opposite signs depending on the communication rules.

The results in this section also suggest that subjects' behavior is highly heterogeneous. The treatment effects that we document are the result of the aggregation of different communication "styles." While some subjects behave approximately as predicted by the theory, others under or over react to commitment and rules. In Section 5, we will focus exclusively on these deviations to better understand their sources and implications.

4.1 Commitment and Subjects' Behavior

4.1.1 Senders

We begin by focusing on senders' behavior. We explore the simplest and most direct evidence to test whether senders understand how to take advantage of commitment. By exploiting *within*-treatment variation, we observe how a sender's behavior changes between the commitment and the revision stages. Proposition 1 governs our predictions, which have opposite signs depending on whether the information is verifiable.

In Figure 2, we present the average difference in senders' strategies between the revision and the commitment stages in treatments $U80$ and $V80$. In the figure, a *positive* bar indicates a message that, conditional on the state, is sent more often in the revision stage. A *negative* bar indicates a message that is sent more often in the commitment stage.

Let us first consider treatment $U80$. From Table 2, the sender should be more informative in the commitment stage than in the revision stage. In particular, when in the revision stage she

[15] See Appendix D.1 for the case of posterior variance.

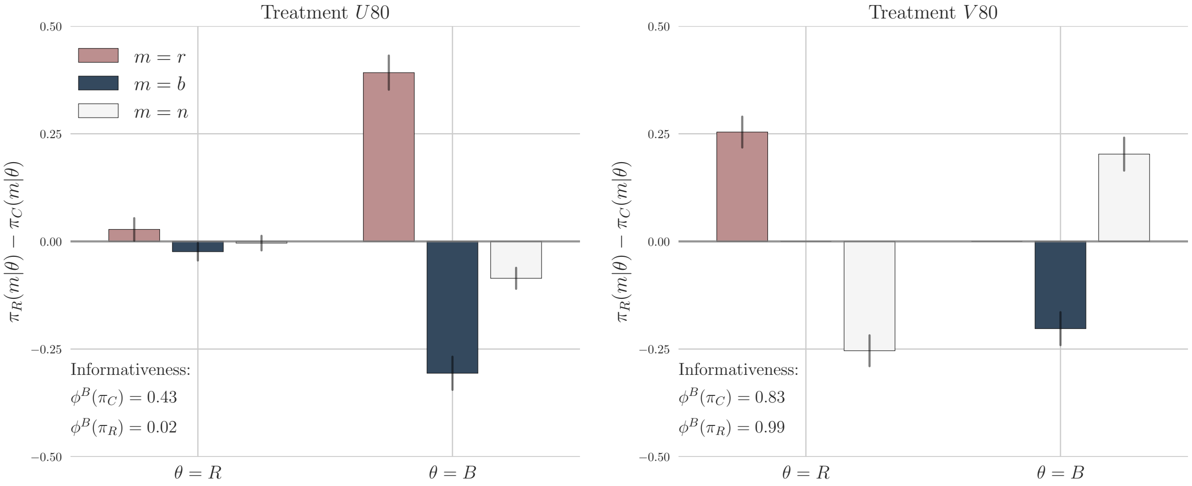


Figure 2: Sender's Strategy: Commitment vs. Revision, $\rho = 0.8$

learns that the state is B, she should replace message b with message r. That is, she should renege on her commitment to telling the truth. The results in the left panel of Figure 2 are very much in line with these predictions. More specifically, when the state is R, the equilibrium strategy is not predicted to change between the commitment and the revision stages. That is, all three bars should be of zero height. This is what we observe in the data. Although statistically significant changes occur for r and b, they are very small in magnitude.[16] Conversely, when the state is B, message r should replace b in the revision stage, whereas message n should not change. Again, qualitatively, this pattern is consistent with what we observe in the data. On average senders increase the frequency of message r at the expenses of b ($p < 0.01$). Overall, as predicted by Proposition 1, the average informativeness of senders' strategies is significantly higher ($p < 0.01$) in the commitment stage—$\phi^B(\pi_C) = 0.43$—than in the revision stage—$\phi^B(\pi_C) = 0.02$.

We now turn to treatment $V80$ (right panel of Figure 2). In contrast with the previous discussion, the sender should be less informative in the commitment stage than in the revision stage (Table 2). In particular, when learning that the state is R, she should replace message n with message r, thus revealing the state. Conversely, when learning that the state is B, she should replace message b with message n. These predicted changes are consistent with what we observe in the data. On average, when the ball is R, senders entering the revision stage increase the likelihood of message r at the expense of message n. Instead, when the ball is B, they increase

[16] Unless noted otherwise, all statistical results allow for random effects at the subject level and are clustered at the session level. We include random effects to account for persistent heterogeneity across subjects; clustering is motivated by potential session effects (see Fréchette, 2012). Results for alternative specifications are reported in the appendix. We note that the findings in the alternative specifications suggest that session-effects are not important in this setting.

the likelihood of message n at the expense of message b. Both changes are significant at the 1% level. Overall, we find that the directions of the predicted changes are matched by the data as shown. Moreover, as predicted by Proposition 1, the average informativeness of senders' strategies is significantly lower ($p < 0.01$) in the commitment stage—$\phi^B(\pi_C) = 0.83$—than in the revision stage—$\phi^B(\pi_C) = 0.99$.

From a quantitative point of view, it is not surprising to see that, on average, senders fall short of exactly matching the equilibrium predictions. It is perhaps more interesting to note that most of the quantitative deviations come from behavior in the commitment stage. In contrast, behavior in the revision stage is quite close to the theory. This distinction in the tendency of behavior to conform with theory in the two different stages has important consequences, as we discuss in Section 5.

In sum, the joint qualitative evidence arising from treatments $U80$ and $V80$ suggests that senders react to commitment and do so in ways that are consistent with the theory. One useful feature of considering different communication rules is that they generate opposing predictions within the same environment. On average, we see that senders exploit their commitment power to strategically hide good news (i.e., $m = n$ if $\theta = R$) when information is verifiable, and disclose bad news (i.e., $m = b$ if $\theta = B$) when information is unverifiable. Once in the revision stage, these commitments are no longer optimal, and indeed senders partially renege on them. We consistently observe the average informativeness of each stage changing as predicted.

4.1.2 Receivers

We now focus on receivers. Our goal is to evaluate the extent to which they understand the strategic implications of commitment and whether their reactions are consistent with the theory. To explicitly test for this hypothesis we exploit *across*-treatment variations. We first introduce the idea of *interim* vs *final* posteriors. Consider the posterior belief that a Bayesian receiver would hold upon observing message m given some (π_C, π_R), if she were to ignore the existence of the revision stage. We call such belief the *interim* posterior, which is formally equal to $\mu_0(R)\pi_C(m|R)/(\sum_\theta \mu_0(\theta)\pi_C(m|\theta))$. Clearly, interim and final posteriors—that is, those that do take into account the revision stage—coincide when $\rho = 1$. More generally, given π_C and π_R, the higher the degree of commitment ρ, the closer the interim posterior is to the final one. We use this simple observation to test whether receivers understand the strategic implications of different levels of commitment. We should observe *different* guessing behavior at *identical* interim beliefs for *different* degrees of commitment. In particular, at high levels of commitment, interim beliefs should be highly predictive of receivers' behavior; at low levels of commitment,

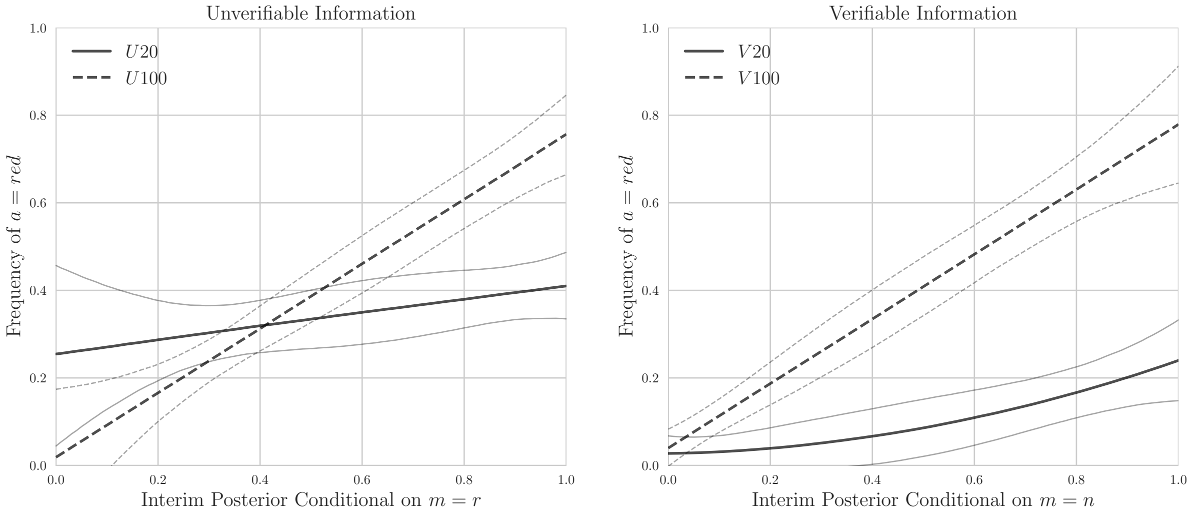


Figure 3: Receiver's Response to Persuasive Messages: $\rho = 0.2$ vs. $\rho = 1$

they should not.[17]

This analysis is carried out in Figure 3. We look at how receivers' responsiveness to interim posteriors changes in treatments with low ($\rho = 0.20$) versus high ($\rho = 1$) commitment.[18] We plot polynomial fits of the average receiver's guess as a function of the interim posterior induced by the observed sender's π_C, the strategy from the commitment stage, and message m.

We begin by comparing treatments $U20$ and $U100$. Our focus is on message $m = r$. If receivers understood the implications of commitment, this message should lead to a guess of *blue* in $U20$, irrespective of the interim posterior. In $U20$, the interim posterior should have little or no impact on the receiver's guess because it is likely that message r did not come from the observed π_C. Therefore, the interim posterior is likely to be far from the final posterior. By contrast, in $U100$, the interim posterior should have a substantial positive effect on the probability that the receiver guesses *red* (Table 2). Indeed, interim and final posteriors coincide in this case. We report our results in the left panel of Figure 3. Consistently with the predictions, the estimated receivers' response is mostly flat in $U20$ and unresponsive to interim beliefs, whereas it is strictly increasing in $U100$.[19]

Similar—if not stronger—evidence is found when comparing $V20$ and $V100$ (right panel of Figure 3). By the nature of verifiable information, messages r and b induce trivial interim beliefs of either 1 or 0. For this reason, we focus on message n, which is the one requiring

[17] We use Bayesian posteriors as a *benchmark* against which to compare actual receivers' behavior. The latter may of course be far from Bayesian, something that we will investigate in Section 5.

[18] In the online appendix, Figure D18 shows the comparison between $\rho = 0.20$ and $\rho = 0.80$.

[19] The linearity in posteriors may be suggestive of *probability matching*. In Appendix B, we show that it instead results from aggregating receivers who employ heterogeneous threshold strategies.

receivers to be sophisticated. We find that receivers' guessing behavior in $V20$ is quite flat in the interim posterior. In contrast, responsiveness is strong and positive for treatment $V100$.[20]

Overall, the joint evidence coming from Figure 3 suggests that, on average, receivers understand and react to commitment in ways that are consistent with the theory. They correctly anticipate senders' incentives to renege on their commitments. As a consequence, receivers understand that messages inducing identical interim beliefs should be treated differently for different degrees of commitment. While this shows that receivers react to commitment, their behavior could still be far from Bayesian. In line with a large body of experimental literature, Figure 3 suggests that this may be the case. We return to this point in Section 5 when we explore in detail the main quantitative deviations that we observe.

4.2 Commitment and Informativeness

The starkest prediction of our theory concerns how equilibrium informativeness changes with commitment under verifiable and unverifiable information. Proposition 2 predicts that equilibrium informativeness should increase with commitment under unverifiable information, whereas it should decrease under verifiable information. To test this prediction, we compute the average Bayesian correlation $\phi^B(\pi_C, \pi_R)$ for each sender and then plot its cumulative distribution function (CDF). We present the results in Figure 4. Each dot represents the average informativeness of a given sender in one of our treatments.

Two patterns emerge from this figure. First, when information is unverifiable (left panel), we observe a noticeable first-order stochastic *increase* in the informativeness of $U100$ and $U80$ relative to $U20$. That is, informativeness increases in commitment not just on average, but rather at all percentiles of the distribution. Moreover, $U80$ and $U100$ are unranked, as predicted by the theory (Table 2). Second, when information is verifiable (right panel), we observe a first-order stochastic *decrease* in informativeness of $V100$ relative to $V20$. This change is relatively less pronounced in $V80$ relative to $V20$. Nonetheless, informativeness appears to decrease in commitment not just on average, but at all (or most, for $V80$) percentiles of the distribution. Again, this is consistent with the theory (Table 2).

To provide further evidence on these comparative statics, we can also look at the posterior *distributions* that senders induce with their communication strategies. This is an alternative

[20] The probability that the receiver guesses *red* when the interim posterior is below $1/2$ does not differ statistically between $\rho = 0.2$ and $\rho = 1$, both for the case with unverifiable information (left panel) and verifiable information (right panel). Instead, for interim posteriors above $1/2$ we find a statistically significant difference in both cases ($p < 0.01$). Perhaps more importantly, the magnitude of the change—below and above $1/2$—is sizable: 56 versus 14 percentage points in the verifiable case, and 40 versus 6 percentage points in the unverifiable case.

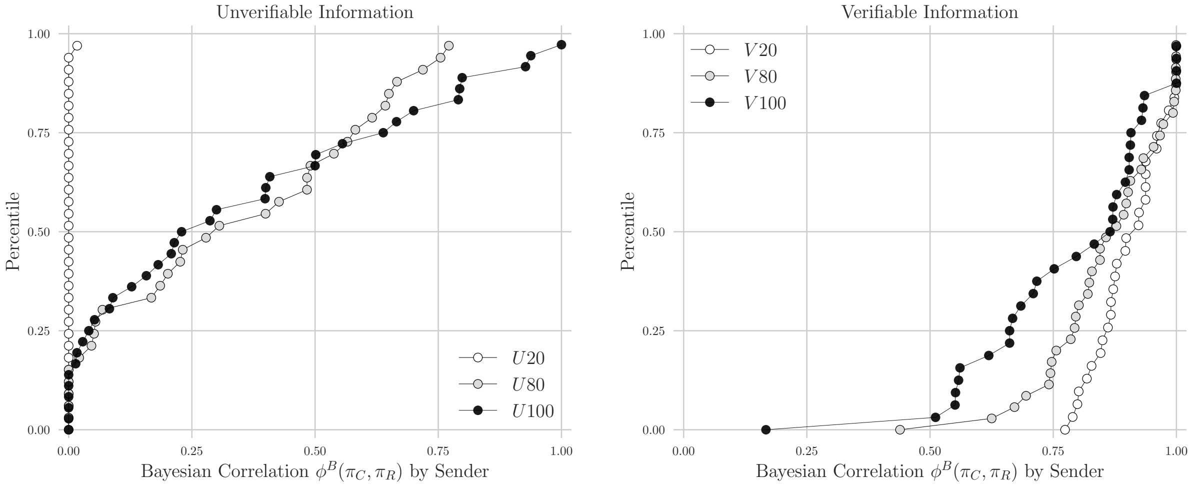
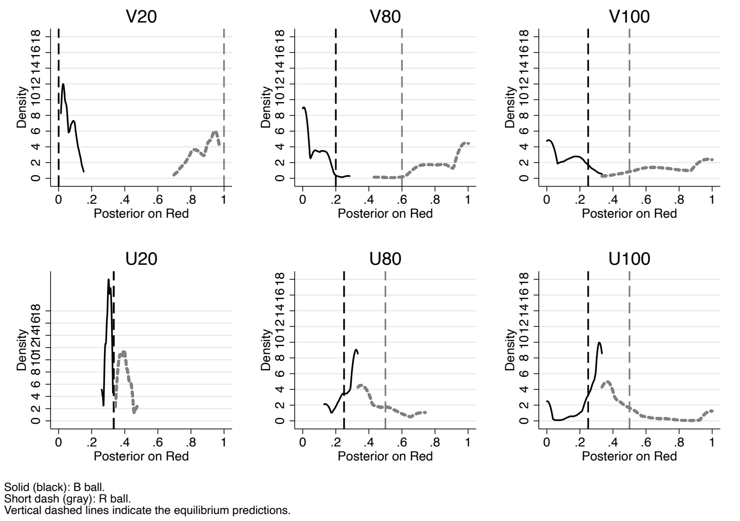


Figure 4: Cumulative Distribution of Sender-Average $\phi^B(\pi_C, \pi_R)$ by Treatment

measure of informativeness that, in particular, does not depend on payoffs. We focus attention on the expected posteriors conditional on the *state*. That is, given π_C, π_R and θ, we compute $\sum_m ((\rho\pi_C(m|\theta) + (1-\rho)\pi_R(m|\theta))\mu(m, \pi_C, \pi_R)$. The left panel of Figure 5 displays the kernel density estimates of the observed expected posteriors. The vertical dashed lines indicate the theoretical predictions. For instance, for $U100$, the expected posterior conditional on state R is at 1/2 because, in equilibrium, message r is sent with probability one and induces a posterior of 1/2. Conversely, the expected posterior conditional on state B is at $1/4$ because, in equilibrium, messages r and b are sent with 50% probability and induce posteriors of $1/2$ and 0, respectively.

In Figure 5, we see a sizable shift of the kernel distributions in the direction predicted by the theory, for both verifiable and unverifiable information. Moving from $U20$ to $U100$, the two distributions become more spread out, whereas moving from $V20$ to $V100$, the posteriors move closer, as predicted by theory. These shifts are quantified in the right panel of Figure 5, which reports the average difference between the expected posterior conditional on R (in solid black) and the one conditional on B (in dashed gray). The table shows that the data move in the right direction for both verifiable and unverifiable treatments, but that the mean difference is much closer to the theoretical predictions in the case of the unverifiable treatments than in the case of verifiable treatments.

Overall, the findings from Figure 4 and Figure 5 validate the asymmetric comparative statics of Proposition 2. The theory appears to explain the main qualitative features of how senders' behavior changes with commitment and rules. Under verifiable information, senders use commitment to decrease the total amount of information that they convey to receivers. Under unverifiable information, senders use commitment to increase the total amount of information



Solid (black): B ball.
Short dash (gray): R ball.
Vertical dashed lines indicate the equilibrium predictions.

	Commitment (ρ)					
	20%		**80%**		**100%**	
Verifiable						
Difference:	0.80 (1.00)		0.78 (0.40)		0.69 (0.25)	
	B	***R***	***B***	***R***	***B***	***R***
Mean:	0.07	0.87	0.07	0.86	0.10	0.79
Unverifiable						
Difference:	0.11 (0.00)		0.24 (0.25)		0.30 (0.25)	
	B	***R***	***B***	***R***	***B***	***R***
Mean:	0.30	0.41	0.25	0.49	0.23	0.53

Figure 5: On the left: Kernel Density of Expected Posterior Conditional on State. On the right: Average Differences in Expected Posteriors Conditional on State (theoretical values in parentheses)

that they convey to the receivers. This asymmetric use of commitment that we observe in the data shows that, on average, senders understand the strategic tension that underlies our model.

5 Understanding Departures from Theory

In the previous section, we showed evidence of treatment effects that match the main *qualitative* predictions of the model. Qualitatively, senders and receivers react to variations in commitment in the predicted ways. These treatment effects, however, hide substantial heterogeneity at the subject's level, which generates *quantitative* deviations from the theory. In this section, we document and explain these deviations.

We begin by looking at the average informativeness by treatment. Table 3 reports the predicted Bayesian correlation (left panel) and the observed one (right panel), averaged across sessions and subjects. As expected, we note again that informativeness moves in the right direction as commitment changes. Moreover, in treatments with partial commitment, we note that more information is conveyed by the senders under verifiable information than under unverifiable information. These changes are in line with Section 4.2 and with our theory. However, Table 3 also highlights important quantitative deviations.

For each communication rule, the observed changes are more muted relative to the theoretical predictions. In the case of unverifiable information for example, the observed increase in infor-

Table 3: Average Bayesian Correlations ϕ^B

ϕ^B – Theoretical Predictions

	Degree of Commitment ρ		
	$\rho = 0.2$	$\rho = 0.8$	$\rho = 1$
Verifiable	1	0.57	0.50
Unverifiable	0	0.50	0.50

ϕ^B – Observed

	Degree of Commitment (ρ)				
	$\rho = 0.2$		$\rho = 0.8$		$\rho = 1$
Verifiable	0.90	$\approx$	0.84	$>$	0.77
	$\vee$		$\vee$		$\vee$
Unverifiable	0.00	$<$	0.33	$\approx$	0.34

Notes: Symbol ">" indicates $p < 0.01$. Green symbol: as predicted. Red symbol: not as predicted.

mativeness from $U20$ to $U100$ is only 68% of the change predicted by the theory. In the case of verifiable information, the theory predicts that, moving from $V20$ to $V100$, we should observe a drop of 0.50 in the Bayesian correlation. Instead, in the data the corresponding reduction is only 0.13, or 26% of the predicted change. More specifically, in treatments with *low* commitment, we find evidence of undercommunication when information is verifiable and overcommunication when it is not. This is in line with the existing experimental literature on disclosure (e.g., Jin et al., 2020) and cheap talk (e.g., Cai and Wang, 2006), respectively.[21] Interestingly, these results do not extend to high-commitment environments. When the level of commitment is high, we find that the opposite holds: senders tend to overcommunicate in treatments with verifiable information and undercommunicate in treatments with unverifiable information.

As a consequence of this observation, communication rules affect informativeness even when this is not predicted by the theory. In particular, treatments $V100$ and $U100$ are predicted by Proposition 2 to be equally informative, but the observed Bayesian correlations are 0.78 and 0.34, respectively. This difference (significance at $p < 0.01$) represents a remarkable deviation from the theory. Furthermore, by comparing the black lines on the left and right panels of Figure 4, we can see that there is a gap at all percentiles of the distribution of ϕ^B.

We refer to these quantitative departures from the theory as the *informativeness gap*. In principle, this gap can be due to anomalous behavior on the part of receivers, senders, or an interaction between the two. In Section 5.1, we first explore receiver behavior and argue that despite their clear departures from the Bayesian benchmark, it is unlikely that receivers are primarily responsible for this gap. In Section 5.2, we turn our attention to sender behavior. We show evidence of a behavioral bias that could explain these deviations. We call this bias *commitment blindness* and show that indeed it generates opposing effects on informativeness depending on the communication rule and thus could generate a gap. Finally, in Section 5.3,

[21] In Appendix B, we show other ways in which our low-commitment treatments are in line with the existing experimental evidence that test no-commitment environments.

we estimate a structural model that accounts for such heterogeneity in senders' behavior and show that it can explain a large part of the observed deviations.

5.1 Can Receiver Behavior Explain the Informativeness Gap?

Although Section 4.1.2 illustrates that receivers do react to commitment, a large body of experimental literature suggests that their behavior is likely to be non-Bayesian.[22] In Appendix A.1.2 and B, we take a detailed look at receivers' behavior. Our analysis reveals that receiver behavior is indeed non-Bayesian. Yet, it is quite systematic. For example, most receiver behavior is consistent with threshold strategies: they guess *red* if the posterior is higher than some receiver-specific threshold. However, our analysis suggests that receiver behavior is unlikely to be the main explanation for the informativeness gap. We discuss three main reasons for this.

First, we note that this gap cannot be *directly* determined by receivers' non-Bayesian behavior. Indeed, we expressed these gaps in terms of ϕ^B, the Bayesian correlation coefficient. By construction, this measure is immune to receivers' mistakes, as explained in Section 3.[23]

Second, let us entertain the possibility that the informativeness gap could be *indirectly* generated by receivers' non-Bayesian behavior through its effects on sender behavior. Suppose that receivers are inherently skeptical of message n and senders know it (as in Jin et al., 2020). That is, suppose that receivers guess *blue* no matter how high the posterior induced by message n. In treatments with unverifiable information, such a receiver's bias would have negligible consequences on senders' behavior: message n can be avoided in equilibrium and indeed is not used often in the data. In contrast, in treatments $V80$ and $V100$, message n plays a key role in the equilibrium prediction. In the presence of such a bias against message n, senders' optimal strategy would then have to be fully informative, thus contributing to the informativeness gap. This rationalization is unsatisfactory for two reasons. The first one is theoretical, as this bias only explains the overcommunication in verifiable treatments and not the undercommunication in unverifiable ones. The second one is empirical: we do not see evidence of such a bias. Data show that receivers respond in similar ways to message n in treatments with verifiable information and r in treatments with unverifiable information. This can be seen in Figure 3. The dashed lines report receivers' responsiveness to message r in $U100$ (left panel) and n in

[22] See, e.g., Charness and Levin (2005) and (Holt, 2007, Chapter 30) for an overview of such literature.

[23] When we explicitly include receivers' behavior—that is, when we compute ϕ instead of ϕ^B—we find informativeness gaps of similar magnitudes. In particular, we find that ϕ is 0.22 and 0.68 for treatments $U100$ and $V100$, respectively. Similarly, we find that ϕ is 0.19 and 0.78 for treatments $U80$ and $V80$, respectively. Note that receivers' mistakes create a garbling in the mapping from states to guesses and therefore can only *decrease* the correlation ϕ relative to ϕ^B.

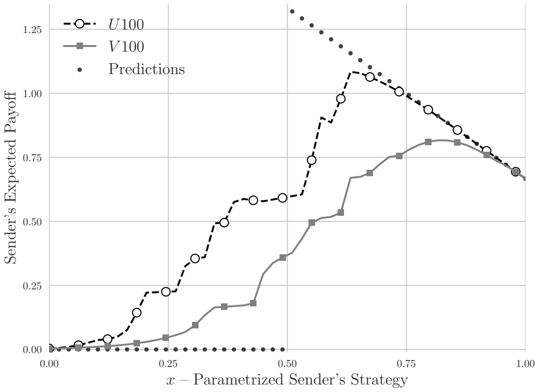


Figure 6: Sender's Empirical Expected Payoff

$V100$ (right panel), controlling for their induced posterior. Receivers' responsiveness does not appear to be significantly different in the two cases.

Third, let us again entertain the possibility that this gap could be *indirectly* generated by receivers' non-Bayesian behavior, but in ways that are more complicated than our previous argument. To address this point, we estimate a simple model of receivers' behavior and then compute the sender's empirical best response. To be concise, we focus attention on treatments with full commitment. In Figure 6, we report the expected payoff that a sender would earn by playing various strategies π_C when facing a *typical* receiver in our sample. For each treatment, we first fit a probit model to estimate the probability that $a = red$ given message m, its induced posterior, and subject fixed effects. Second, we use the estimated model to compute the expected payoff that a sender would earn when choosing various commitment strategies π_C. More specifically, we define a class of information structures parametrized by $x \in [0, 1]$. This class is rich enough to approximate most of the observed strategies, including the equilibrium strategies for these treatments. In particular, for $U100$, we consider strategies such that $\pi_C(r|R) = 1$ and $\pi_C(b|B) = 1 - \pi_C(r|B) = x$. For $V100$, we consider strategies such that $\pi_C(n|R) = 1$ and $\pi_C(b|B) = x$. In both $U100$ and $V100$, π_C is the equilibrium strategy when $x = 1/2$ (Table 2); it is uninformative when $x = 0$; and it is fully informative when $x = 1$. More generally, $\phi^B(\pi_C)$ is weakly increasing in x.

Figure 6 shows that receiver behavior leads to a payoff function for the sender that is flatter than it would be if all receivers were fully Bayesian. Moreover, for both treatments, the sender's best response to the receivers' behavior requires $x > 1/2$. This is intuitive: $x = 1/2$ is a knife-edge condition that leaves a Bayesian receiver just indifferent. Although receivers do not conform with the Bayesian paradigm, the vast majority of them are more likely to guess *red* following a message that carries more evidence in favor of the state being R. This monotone responsiveness in induced beliefs is a milder rationality requirement than Bayesianism, and it has been documented in other experiments (see Camerer (1998) for a discussion). Importantly, as shown by Figure 6, the extent of monotonicity displayed in our experiment is sufficient to confirm a key insight from models of communication under commitment, namely the fact that the best-response involves some degree of strategic obfuscation.[24] Therefore, an uninformative π_C is worse than a fully informative π_C, which is in turn worse than commitment to mixing. The finding that senders' empirical expected payoff is non monotone in the amount of information conveyed to the receiver is consistent with the theory.

More importantly, Figure 6 shows that receiver behavior alone appears insufficient to explain the large gaps in informativeness that we documented in Table 3. If senders were best-responding to the typical receivers' behavior, we would observe $\phi^B(\pi_C) = 0.60$ in treatment $U100$ and $\phi^B(\pi_C) = 0.75$ in treatment $V100$. This explanation is, therefore, unsatisfactory on two levels. First, it captures only a small fraction (35%) of the observed gap. Second, the empirical best response for $U100$ leads to an *increase* in informativeness—not a decrease, as it is observed.

Overall, the three points above suggest that receivers' non-equilibrium behavior is insufficient to explain the informativeness gap. As we show in the remainder of the section, senders are likely to be the primary drivers of these observed deviations.

5.2 Commitment Blindness

In this section, we introduce a simple bias in senders' behavior that can explain a large part of the informativeness gap. We begin by noting that senders employ very heterogeneous communication "styles" as illustrated in Figure 4. Understanding the sources of this heterogeneity is key to explaining the informativeness gap.

To this end, we introduce the notion of *commitment blindness*. A sender is commitment-blind if she behaves under commitment as if she had no commitment power at all. More specifically, her commitment strategy is the *equilibrium* strategy of a hypothetical game with $\rho = 0$. Commitment blindness has very different implications depending on the communication rule. Specifically, when information is unverifiable, $\rho = 0$ is equivalent to a cheap-talk game and the optimal strategy involves babbling. Such a strategy is *uninformative* ($\phi^B = 0$). If instead information is verifiable, $\rho = 0$ indicates an information-disclosure game and the optimal strategy

[24] Relatedly, de Clippel and Zhang (2020) explore the relative robustness of the Bayesian Persuasion model if the receiver is non-Bayesian.

involves unraveling; hence, it is *fully informative* ($\phi^B = 1$). These very different levels of informativeness suggest that commitment blindness could explain the unpredicted gap that we have documented: the same behavioral bias can inflate the average informativeness in treatments with verifiable information and deflate it in treatments with unverifiable information.[25]

Note that commitment blindness is different from lying aversion and has different implications. To see this, consider a sender who is fully averse to lying, regardless of her commitment power. To begin, when information is unverifiable, such a sender does not play the equilibrium strategy of a hypothetical game with $\rho = 0$. Moreover, she would play highly informative strategies *irrespective* of the communication rule. Thus, lying aversion cannot successfully generate the informativeness gap that we observe in the data.

We exploit our experimental design to test for the presence of senders who are compatible with commitment blindness. This can only be done in treatments with partial commitment. Indeed, one needs to observe how the *same* sender behaves in two opposing scenarios, with and without commitment power. Therefore, we focus our attention on treatments $U80$ and $V80$ and compare how sender behavior changes between the commitment and the revision stages. We seek to identify senders who (i) play the *same strategy* in both the commitment and revision stage and (ii) play the *equilibrium strategy* in the revision stage as defined in Table 2.

In contrast to our work from Section 4.1.1—where we focused on some features of average senders' strategies—we now describe their behavior at the observation level. Our goal is to identify the representative strategies (π_C, π_R) played in the treatments under consideration. Such an analysis presents a technical challenge, as senders' strategies are complex and high-dimensional objects. To organize the observed strategies, we use a standard machine-learning algorithm, the k-means, to cluster strategies into four representative groups (that is, $k = 4$).[26] We cluster the strategies by treatment and report the results in Figures 7 and 8 for treatments $U80$ and $V80$, respectively. To visualize all the data, we plot the clustered strategies onto two separate panels, one for π_C and one for π_R. The representative strategies are indicated with larger markers. Note that strategies that appear similar in the commitment (respectively revision) stage may belong to different clusters because they differ in the revision (respectively

[25] The experimental literature on Cournot competition with endogenous timing also studies commitment in the lab. A player can choose to publicly commit to a production quantity, thus emerging as a Stackelberg leader and increasing her payoff. See, for instance, Huck and Müller (2000), Huck et al. (2001), and Morgan and Várdy (2004, 2013)

[26] A commonly used method to group data is k-means clustering (see, MacQueen, 1967; Hastie et al., 2009; Murphy, 2012) The procedure selects points to be the centers of clusters: an observation is associated with the closest center, and the centers are iterated on to minimize the total within-cluster variance. We choose $k = 4$ and input entries: $\pi_C(m|\theta)$ and $\pi_R(m|\theta)$ for $m \in \{r, b\}$ and $\theta \in \{R, B\}$.

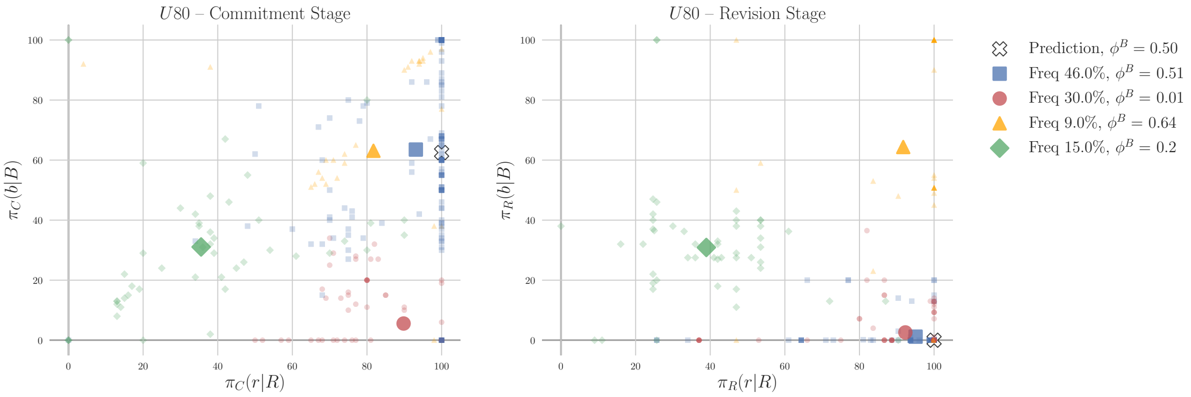


Figure 7: Treatment $U80$ – Clustering of Senders' Strategies

commitment) stage.[27]

We begin our analysis with treatment $U80$, that is, Figure 7. The strategies indicated by red circles are those compatible with commitment blindness. The representative strategy consists of sending message r regardless of the state, in both the commitment and the revision stage. This strategy coincides with equilibrium behavior in the revision stage (Table 2). As expected, this strategy is mostly uninformative, that is, $\phi^B = 0.01$. This strategy is also quite common: 30% of the observed strategies are of this kind. We now discuss the remaining clusters of Figure 7. The strategies indicated by blue squares are compatible with equilibrium behavior and are the most prevalent ones. These strategies drive most of the treatment effects documented in Section 4. Note that the induced informativeness $\phi^B = 0.51$ is remarkably close to the equilibrium prediction of 0.50. Strategies indicated by yellow circles are consistent with a weak form of lying aversion and are not prevalent in our data. Finally, strategies marked by green diamonds belong to a residual cluster that cannot be grouped in any of the categories above. We interpret these residual strategies as noise.

We now turn to the analysis of sender behavior in treatment $V80$ (Figure 8). Again, strategies indicated by red circles are those compatible with commitment blindness. The representative strategy consists of sending message r given R, and n given B, in both the commitment and the revision stages. This coincides with equilibrium behavior in the revision stage (Table 2). In contrast to $U80$, commitment-blind strategies are highly informative ($\phi^B = 0.94$). In terms of prevalence, 33% of the observed strategies are of this kind. We now discuss the remaining clusters of Figure 8. Strategies indicated by blue squares are consistent with equilibrium behavior. They react to commitment and induce an informativeness $\phi^B = 0.57$, which

[27] We present data at the observation level, but these clusters capture persistent senders' types, with a typical sender playing in the same cluster more than 80% of the times.

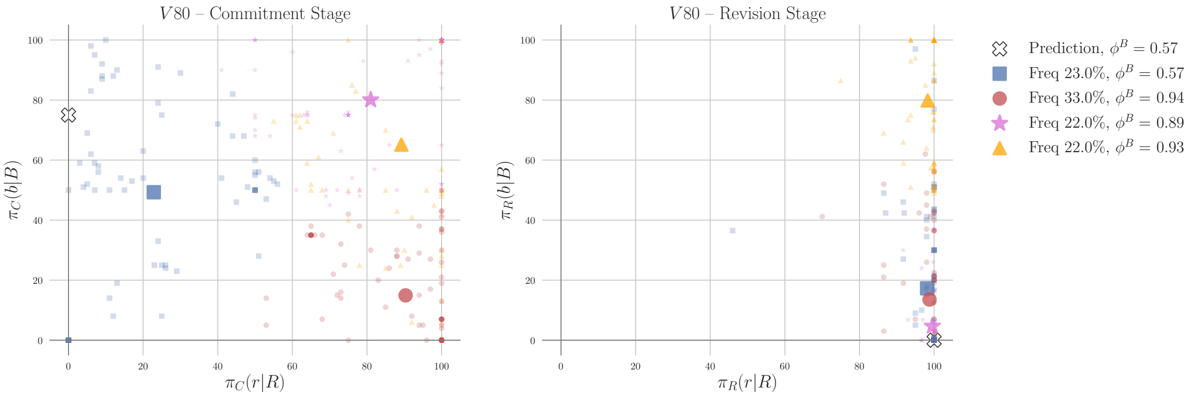


Figure 8: Treatment $V80$ – Clustering of Senders' Strategies

is at the equilibrium level. Strategies indicated by purple stars also react to commitment and play the equilibrium strategy in the revision stage, but fail to conceal information in the commitment stage. As a result, they induce higher-than-optimal levels of informativeness, that is, $\phi^B = 0.89$. Together, these last two clusters we discussed represent 45% of the data and drive the treatment effects documented in Section 4. Finally, strategies indicated by yellow triangles are consistent with lying aversion and induce high informativeness, that is, $\phi^B = 0.93$.

In sum, we have documented the existence of a behavioral type that is consistent with commitment blindness. Such behavior has opposite implications depending on the communication rule. Under unverifiable information, these senders tend to decrease the average informativeness. Under verifiable information, they tend to increase the average informativeness. Thus, the same behavioral bias could jointly explain the asymmetric departures documented in Table 3.

5.3 QRE: Quantifying Departures From Equilibrium

In this final part of the section, our goal is to quantitatively reproduce the gap in informativeness through a structural model. This model has two components. First, it accounts for the heterogeneity in senders' behavior that we documented. Second, it accounts for players' mistakes and noisy behavior through a quantal-response equilibrium (QRE).[28] We use the estimates from this structural model to compute the implied correlations—both ϕ^B and ϕ—and show that they reproduce between 70% and 80% of the observed informativeness gap.

In a QRE, players are assumed to respond with errors to their beliefs, which in turn correctly account for the errors that other players make. Two technical challenges make the estimation of our structural model nontrivial: first, senders choose among a continuum of high-dimensional

[28] See Goeree et al. (2016).

strategies, and second, our games have multiple stages and feature incomplete information. We address the first challenge by using the same k-means algorithm that we discussed in Section 5.2. We address the second challenge by using the methodology in Bajari and Hortacsu (2005). In the following paragraphs, we explain these two points in more detail. For simplicity, we will focus attention on treatments $U100$ and $V100$. Although a similar analysis could be performed under partial commitment, the focus on full commitment significantly simplifies our estimations. Moreover, the informativeness gap in these treatments is the farthest from the theoretical predictions and thus, the more interesting to explain.

Discretization and Senders' Heterogeneity. To estimate QRE, we first need to discretize senders' strategy space Π into a grid containing k representative strategies. This is usually achieved by manually gridding the strategy space (e.g., Camerer et al. (2016)). This approach is productive especially when the strategy space is sufficiently simple, e.g., one-dimensional. In our case, the high dimensionality of the strategy space renders this approach infeasible. A natural solution is to use a clustering algorithm. We use the k-means algorithm to identify the set of representative strategies Π_k.[29] As before, we set $k = 4$. Importantly, we compute Π_k separately for each treatment. This allows us to capture the very different ways in which senders play in treatments with verifiable and unverifiable information, as shown in Section 5.2. In particular, it allows us to capture the different implications of commitment blindness for these two treatments.

Multi-Stage QRE. For representative strategy $\pi_C \in \Pi_k$ and message $m \in M$, denote by $U(a, \pi_C, m)$ the receiver's expected payoff from choosing action $a \in \{a_L, a_H\}$. The Logit QRE model specifies that a receiver of type $\lambda_R \geq 0$ chooses action a_H with the following probability:

$$\mathbb{P}_R(a_H|\pi_C, m, \lambda_R) = \frac{e^{\lambda_R U(a_H,\pi_C,m)}}{e^{\lambda_R U(a_H,\pi_C,m)} + e^{\lambda_R U(a_L,\pi_C,m)}}.$$

That is, the probability of choosing the optimal action increases in the utility difference between the two possible actions. Given λ_R, the sender's expected utility from choosing π_C is given by $V(\pi_C|\lambda_R) := \sum_{\theta,m} \mu_0(\theta)\pi_C(m|\theta)\mathbb{P}_R(a_H|\pi_C, m, \lambda_R)$. That is, the sender takes the receiver's errors into account when computing her expected payoff from playing a certain strategy. As in the receiver's case, the probability that a sender of type $\lambda_S \geq 0$ chooses π_C is given by

$$\mathbb{P}_S(\pi_C|\lambda_S, \lambda_R) = \frac{e^{\lambda_S V(\pi_C|\lambda_R)}}{\sum_{\pi_C\in\Pi_k} e^{\lambda_S V(\pi_C|\lambda_R)}}.$$

[29] Figure D19 reports k-means clusters for treatments with full commitment.

Table 4: QRE-Implied Correlations

	Bayesian Correlation ϕ^B		Correlation ϕ	
Treatment	QRE-Implied	Observed	QRE-Implied	Observed
$V100$	0.72	0.77	0.64	0.68
$U100$	0.41	0.34	0.26	0.22

The parameters (λ_S, λ_R) capture the extent to which players best respond to their opponent's behavior. At one extreme, as $\lambda_i \to \infty$, the player in role i never makes a mistake. At the other extreme, when $\lambda_i = 0$, the player in role i randomizes uniformly across all available strategies. We allow $\lambda_S \neq \lambda_R$ since senders and receivers face substantially different tasks.

Estimation. We now describe how we estimate this model. Notice that our game has multiple stages. In treatments with full commitment, there are two stages and, importantly, the receiver perfectly observes the strategy chosen by the sender. Whether this strategy was chosen by mistake is irrelevant for the receiver, who simply responds as described above. Effectively, the receiver solves a single-agent decision problem. Thus, we can estimate $\hat{\lambda}_R$ independent of λ_S. Instead, the sender, moves before the receiver. Therefore, she must form expectations about the receiver's behavior. Since λ_R enters the payoff function $V(\pi_C|\lambda_R)$, the equilibrium λ_S depends on the true λ_R. We can consistently estimate $V(\pi_C|\lambda_R)$ for each strategy π_C by computing the average sender's expected payoffs of playing strategy π_C (Bajari and Hortacsu, 2005). Using maximum likelihood, it is then straightforward to estimate $(\hat{\lambda}_S, \hat{\lambda}_R)$.[30]

Simulation. Given these estimates, we simulate a dataset with 10^4 observations and compute counterfactual correlations ϕ and Bayesian correlations ϕ^B. The state θ is drawn at random from a Bernoulli distribution with parameter $1/3$, just as in the experiment. The sender chooses strategies in Π_k according to $\mathbb{P}_S(\pi_C|\hat{\lambda}_S, \hat{\lambda}_R)$. Message m is generated according to the chosen π_C and the realized state θ. The receiver chooses a_H with probability $\mathbb{P}_R(a_H|\pi_C, m, \hat{\lambda}_R)$.

In Table 4, we report both the QRE-implied correlations as well as the observed ones. The main conclusion from this table is that the combination of (i) treatment-specific clustering and (ii) noisy players' behavior as modeled by QRE can reproduce correlations that are remarkably similar to those we observed. In particular, the model explains between 70% and 80% of the observed gaps in informativeness. It is useful to point out that, in the procedure described above, we fit data in two separate steps. First, we use the data from each treatment to compute Π_k. That is, the representative strategies of treatment $U100$ are allowed to differ from those for treatment $V100$. The need for doing so follows from the discussion in Section 5.2, where

[30] For $U100$, we have $(\hat{\lambda}_S, \hat{\lambda}_R) = (0.41, 1.68)$. For $V100$, we have $(\hat{\lambda}_S, \hat{\lambda}_R) = (0.21, 1.28)$.

we argued that communication rules affect senders' play in a substantial and unpredicted way. Second, we use the data again to estimate treatment-specific $(\hat{\lambda}_S, \hat{\lambda}_R)$. By doing so, we account for inevitable noise in the behavior of senders and receivers. The combination of these methodologies generates correlations that closely fit the data.

5.4 Alternative Approaches

We now briefly discuss other theories that could in principle account for the informativeness gap: level-k, other-regarding preferences, and lying aversion. Although behaviors compatible with these theories may be present in our data to some extent, we argue that they are not the most natural avenues to explore, as they either fail to account for some of the key deviations or they would need to be enhanced relative to their standard specifications.

For instance, let us consider the simplest form of a level-k model.[31] A key component of a level-k analysis is the specification of level-0 players. First, in our full-commitment treatments, the strategy of the sender is fully observable by the receiver. Thus, there is little room for the different levels of strategic sophistication on the part of receivers to play a role. Second, in treatments with verifiable information, there is no leeway in specifying receivers' beliefs (or behavior) following the verifiable messages r or b. The only degree of freedom is in specifying non-equilibrium beliefs and play conditional on message n. The natural assumption is that a level-0 receiver naively updates in a passive manner, with a posterior of $1/3$, the same as the prior. However, in our setting, such belief leads the receiver to guess *blue*, the same guess she would take following message b. The fact that receivers' behavior is identical between level-0 and equilibrium play implies that this concept, taken as is, gives us little leverage to explain departures from equilibrium in our environment. This is of course not to say that a more elaborate version of level-k—possibly combined with other approaches—may not be a fruitful avenue to explore.

Other-regarding preferences have been successfully used to understand important patterns in a variety of experiments (see Cooper and Kagel, 2016). However, the informativeness gap entails departures that, in some cases, go in a direction that is opposite to the common prediction of such models—namely, away from equating players' payoffs. For instance, in $U100$, a commitment-blind sender plays an uninformative strategy and thus, earns the lowest possible payoff (see Figure 6), while the receiver can secure an expected payoff of \$1.33 (or \$2 times $2/3$) by guessing blue. By playing the empirical best response, the sender would instead in-

[31] Crawford et al. (2013) reviews this literature. In cheap-talk games, Cai and Wang (2006), Kawagoe and Takizawa (2009), and Wang et al. (2010) discuss level-k models.

crease her payoffs away from zero, while also increasing the payoff for the receiver. This suggests that commitment-blind senders do not behave in a way that is compatible with the spirit of many models of other-regarding preferences. Of course, this literature is incredibly rich, and there may be additional and more-complex types of behaviors that could be useful to explore in the future.

Finally, lying aversion has been studied in the context of cheap-talk experiments (e.g., Gneezy, 2005; Sánchez-Pagés and Vorsatz, 2007; Hurkens and Kartik, 2009). Lying aversion is consistent with the fraction of subjects who always tell the truth, as discussed in Section 5.2. However, such behavior is markedly different from the behavior of a commitment-blind sender, especially in treatments with unverifiable information. More importantly, it leads to implications that are, in principle, different from the observed departures: Lying aversion should indeed inflate the informativeness in treatments with unverifiable information, whereas the opposite happens in the data.

References

Aristidou, A., G. Coricelli, and A. Vostroknutov (2019): "Incentives or Persuasion? An Experimental Investigation," *Working Paper*.

Au, P. H. and K. K. Li (2018): "Bayesian Persuasion and Reciprocity Concern: Theory and Experiment," *Working Paper*.

Austen-Smith, D. (1993): "Information and Influence: Lobbying for Agendas and Votes," *American Journal of Political Science*, 37(3), 799–833.

Bajari, P. and A. Hortacsu (2005): "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data," *Journal of Political Economy*, Vol. 113, No. 4, pp. 703–741.

Battaglini, M. (2002): "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70(4), 1379–1401.

Battigalli, P. and M. Siniscalchi (2002): "Strong Belief and Forward-Induction Reasoning," *Journal of Economic Theory*.

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.

Bergemann, D. and S. Morris (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), 44–95.

Blume, A., D. De Jong, Y. Kim, and G. Sprinkle (1998): "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games," *American Economic Review*, 88, 1323–1340.

Blume, A., E. K. Lai, and W. Lim (2020): "Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations," in *Handbook of Experimental Game Theory*, ed. by C. M. Capra, R. Croson, M. Rigdon, and T. Rosenblat, Edward Elgar Publishing.

Bock, O., I. Baetge, and A. Nicklisch (2014): "hroot: Hamburg registration and organization online tool," *European Economic Review*, 71, 117–120.

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

Camerer, C. (1998): "Bounded Rationality in Individual Decision Making," *Experimental Economics*, 1, 163–183.

Camerer, C., S. Nunnari, and T. R. Palfrey (2016): "Quantal Response and Nonequilibrium Beliefs Explain Overbidding in Maximum-Value Auctions," *Games and Economic Behavior*, Vol 98, 243–263.

Cameron, A. C. and D. L. Miller (2015): "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50, 317–372.

Carter, A. V., K. T. Schnepel, and D. G. Steigerwald (2017): "Asymptotic Behavior of at Test Robust to Cluster Heterogeneity," *Review of Economics and Statistics*.

Charness, G. and D. Levin (2005): "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect," *American Economic Review*, 95, 1300–1309.

Cooper, D. J. and J. H. Kagel (2016): "Other-Regarding Preferences: A Selective Survey of Experimental Results," in *The Handbook of Experimental Economics, Volume 2*, ed. by J. H. Kagel and A. E. Roth, Princeton University Press.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51:1, 5–62.

Crawford, V. P. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

de Clippel, G. and K. Rozen (2020): "Communication, Perception, and Strategic Obfuscation," *Working Paper*.

de Clippel, G. and X. Zhang (2020): "Non-Bayesian Persuasion," *Working Paper*.

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.

Embrey, M., G. R. Fréchette, and S. Yuksel (2017): "Cooperation in the Finitely Repeated Prisoner's Dilemma," *Quarterly Journal of Economics*, 133, 509–551.

Forsythe, R., R. M. Isaac, and T. R. Palfrey (1989): "Theories and Tests of "Blind Bidding" in Sealed-bid Auctions," *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.

Fréchette, G. R. (2012): "Session-Effects in the Laboratory," *Experimental Economics*, 15, 485–498.

Galor, E. (1985): "Information Sharing in Oligopoly," *Econometrica*, 53, 329–343.

Gilligan, T. W. and K. Krehbiel (1987): "Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures," *Journal of Law, Economics*, 3, 287–335.

——— (1989): "Information and Legislative Rules with a Heterogeneous Committee," *American Journal of Political Science*, 33, 459–490.

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

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2016): *Quantal Response Equilibrium: A Stochastic Theory of Games*, Princeton University Press.

Green, J. R. and N. L. Stokey (2007): “A Two-person Game Of Information Transmission,” *Journal Of Economic Theory*, 135, 90–104.

Grossman, S. J. (1981): “The Informational Role of Warranties and Private Disclosure about Product Quality,” *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.

Hagenback, J. and E. Perez-Richet (2018): “Communication with Evidence in the Lab,” *Working Paper*.

Hart, S., I. Kremer, and M. Perry (2017): “Evidence Games: Truth and Commitment,” *American Economic Review*, 107(3), 690–713.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics New York, NY, USA:, second Edition.

Holt, C. A. (2007): *Markets, Games, and Strategic Behavior*, Pearson Addison Wesley Boston, MA.

Huck, S., W. Muller, and H.-T. Normann (2001): “Stackelberg Beats Cournot — On Collusion and Efficiency in Experimental Markets,” *Economic Journal*, 111, 749–765.

Huck, S. and W. Müller (2000): “Perfect versus Imperfect Observability—An Experimental Test of Bagwell’s Result,” *Games and Economic Behavior*, 31, 174 – 190.

Hurkens, S. and N. Kartik (2009): “Would I lie to you? On social preferences and lying aversion,” *Experimental Economics*, 12, 180–192.

Ibragimov, R. and U. K. Müller (2010): “t-Statistic Based Correlation and Heterogeneity Robust Inference,” *Journal of Business & Economic Statistics*, 28, 453–468.

Jin, G. and P. Leslie (2003): “The effect of information on product quality: Evidence from restaurant hygiene grade cards,” *Quarterly Journal of Economics*.

Jin, G., M. Luca, and D. Martin (2019): “Complex Disclosure,” *Working Paper*.

——— (2020): “Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure,” *American Economic Journal: Microeconomics*.

Jovanovic, B. (1982): “Truthful Disclosure of Information,” *Bell Journal of Economics*, 13, 36–44.

Kamenica, E. (2019): “Bayesian Persuasion and Information Design,” *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): “Bayesian persuasion,” *American Economic Review*, 101, 2590–2615.

Kartik, N. (2009): “Strategic communication with lying costs,” *The Review of Economic Studies*, 76, 1359–1395.

Kawagoe, T. and H. Takizawa (2009): “Equilibrium Refinement vs Level-k Analysis: An Experimental Study of Cheap-Talk Games with Private Information,” *Games and Economic Behavior*.

King, R. and D. Wallin (1991): “Market-induced information disclosures: An experimental markets investigation,” *Contemporary Accounting Research*, 8, 170–197.

Lipnowski, E., D. Ravid, and D. Shishkin (2018): “Persuasion via Weak Institutions,” *Working Paper*.

MacQueen, J. (1967): “Some Methods for Classification and Analysis of Multivariate Observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Oakland, CA, USA, vol. 1, 281–297.

Mathios, A. (2000): “The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market,” *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.

——— (2008): “What the Seller Won’t Tell You: Persuasion and Disclosure in Markets,” *Journal of Economic*

Perspectives, 22, 115–131.

Min, D. (2017): “Bayesian Persuasion under Partial Commitment,” *Working Paper*.

Morgan, J. and F. Várdy (2004): “An experimental study of commitment in Stackelberg games with observation costs,” *Games and Economic Behavior*, 49, 401 – 423.

——— (2013): “The Fragility of Commitment,” *Management Science*, Vol. 59, No. 6, 1344–1353.

Murphy, K. P. (2012): *Machine Learning: A Probabilistic Perspective*, MIT Press.

Nguyen, Q. (2017): “Bayesian Persuasion: Evidence from the Laboratory,” *Working Paper*.

Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura (1990): “Strategic Information Revelation,” *The Review of Economic Studies*, 57, 25–47.

Perez-Richet, E. and V. Skreta (2018): “Test Design under Falsification,” *Working Paper*.

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

Sánchez-Pagés, S. and M. Vorsatz (2007): “An Experimental Study of Truth-Telling in a Sender-Receiver Game,” *Games and Economic Behavior*, 61(1), 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,” *American Economic Review*, 100, 984–1007.

Wilson, A. and E. Vespa (2020): “Information Transmission Under the Shadow of the Future: An Experiment,” *American Economic Journal: Microeconomics*, Forthcoming.

A Additional Treatments

In this section, we explore two robustness treatments. In Section A.1, we simplify the message space by excluding message n. This message is redundant when information is unverifiable. We find that results *without* message n are comparable overall to those *with* message n, but feature less noise. We also take advantage of the simpler setup to take a deeper dive into receivers' behavior. We find that, although receivers do not conform to the Bayesian paradigm, their behavior is highly systematic and monotone to information. In Section A.2, we test a different comparative static result: instead of varying commitment or rules, we change the alignment between sender's and receivers' preferences.

A.1 Simplifying the Message Space

In our main treatments, senders can choose among three messages: r, b, and n. In theory, when information is unverifiable, one of these messages is redundant and its presence (or absence) does not change equilibrium outcomes. In practice, message n is convenient as it allows a clean comparison between treatments with and without verifiable information. In this section, we show that adding message n does not significantly alter agents' behavior. We report the results of a treatment with unverifiable information and full commitment where the message space includes only r and b. Every other aspect of this treatment, which we label $U100S$, is identical to $U100$.[32] Our main conclusion from the comparison of $U100$ and $U100S$ is that adding message n increases the noise, but does not significantly alter agents' behavior. We also take advantage of the simpler setting in $U100S$ to perform an analysis of receivers' behavior, which is representative of receivers' behavior in all other treatments (see Appendix B).

A.1.1 Comparison between $U100$ and $U100S$

We begin by comparing the senders' behavior in treatments $U100$ and $U100S$. The left panel of Figure A9 reports the main clusters for these treatments computed through a k-means algorithm, as described in Section 5.2. Solid markers indicate the representative strategies for $U100S$. Hollow markers indicate those for $U100$. A9 shows that the strategies that senders play in these two treatments are highly comparable, despite the difference in the message space.

[32] We conducted four sessions of $U100S$, each with 14-20 subjects (17.5 on average per session) for a total of 70 subjects. In addition to their earnings from the experiment, subjects received a \$10 show-up fee. Average earnings, including the show-up fee, were \$34 (ranging from \$14 to \$52) per session.

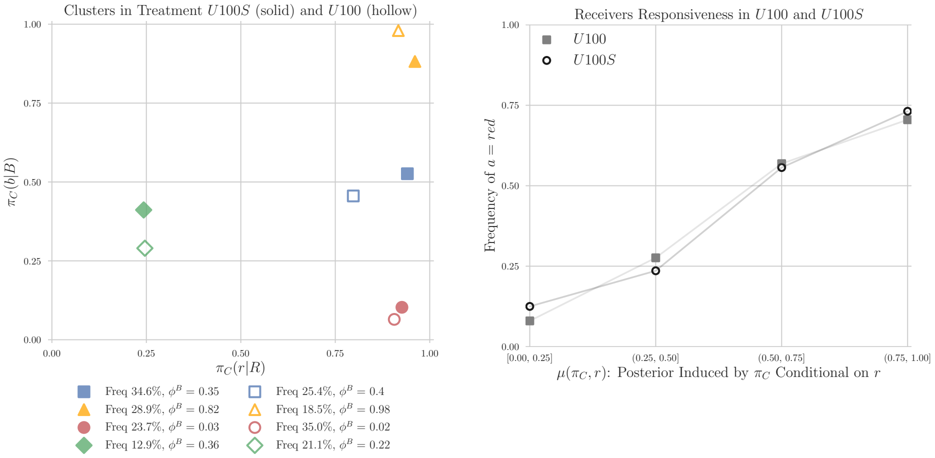


Figure A9: Senders' (left panel) and Receivers' (right panel) Behavior in $U100$ and $U100S$

We note that the behavior in $U100S$ is less noisy than in $U100$. This can be deduced from the fact that the residual cluster, indicated by green diamonds, has a lower frequency in $U100S$ (12.9%) relative to $U100$ (21.1%). There is a higher frequency of senders who approximately best respond to receiver $U100S$ relative to $U100$. From Figure 6 and A10, we can deduce that, in these treatments, the best response involves a combination of blue squares and yellow triangles. These represent 63.5% and 44% of the data in $U100S$ and $U100$, respectively. This last observation is also reflected in the *average* Bayesian correlation that is induced by senders in these two treatments. We find that $\phi^B(\pi_C) = 0.41$ in $U100S$. This is significantly lower ($p < 0.01$) than the equilibrium prediction of 0.5, but the sign is significantly higher ($p < 0.05$)) than in $U100$. We conclude that senders' behavior in $U100S$ is qualitatively comparable to $U100$, but cleaner and less noisy than in $U100$.

We now compare receivers' behavior in treatments $U100$ and $U100S$. The right panel of Figure A9 reports the average receivers' responsiveness to Bayesian posteriors belonging to four key intervals (horizontal axis). We focus attention on the posteriors induced by message $m = r$, the potentially persuasive message. In none of the intervals is the receivers' behavior significantly different in the two treatments considered. We conclude that receivers do not seem to react in unexpected ways to the presence of the redundant message n.

A.1.2 A Closer Look at Receivers' Behavior

We take advantage of the relative simplicity of treatment $U100S$ to take a closer look at receivers' behavior. A similar analysis for all the other treatments can be found in Appendix B. We begin by describing some aggregate features of the data. First, receivers' responsiveness is monotonic in the induced posterior. That is, on average, receivers are more persuaded to guess *red* by messages that carry more evidence in favor of the state being R. As highlighted in Section 5.1, this is a robust feature of receivers' behavior that holds across all our treatments, including $U100S$. For $U100S$, this is illustrated graphically in Figure A9 when $m = r$. When pooling message r and b, we find that, for posteriors above $\frac{1}{2}$, receivers guess *red* 57% of the time, whereas they guess *red* only 11% of the time for posteriors below $\frac{1}{2}$ ($p \leq 0.01$).

The extent of monotonicity that we observe in receivers' behavior is sufficient to confirm one of the main insights from models of communication under commitment, namely that the best response involves some degree of strategic obfuscation: an uninformative π_C is worse than a fully informative π_C, which is worse than using commitment to mix. In Figure A10, we replicate the same exercise performed in Figure 6 for $U100S$. As was the case for $U100$ and $V100$, we find that senders' empirical expected payoff is non monotone in the amount of information conveyed to the receiver, in line with the theory.

Monotonicity is, of course, a mild requirement for receivers' rationality. A Bayesian receiver should choose $a = red$ for any posterior $\mu(m, \pi_C) \geq \frac{1}{2}$ and $a = blue$ otherwise. The aggregate evidence presented in Figure A9 fails to satisfy this stronger requirement of rationality. Furthermore, receivers respond to the color of the message independently of the posterior this color conveys. When $\mu(m, \pi_C) \geq \frac{1}{2}$, receivers guess $a = red$ 62% of the time if $m = r$ and 38% of the time if $m = b$. In contrast, when $\mu(m, \pi_C) < \frac{1}{2}$, receivers guess $a = red$ 21% if $m = r$ and 5% of the time if $m = b$. These differences, which are significant at the 1% level, are inconsistent with the behavior of a Bayesian receiver. Even when provided with conclusive evidence that the state is R, that is, even when $\mu(m, \pi_C) \approx 1$, some receivers nonetheless guess *blue* at least some of the time. To summarize, at the aggregate level, receivers are non-Bayesian, an observation that is in line with a large body of experimental literature (e.g., Charness and Levin, 2005; Holt, 2007, Ch. 30).

To understand better whether the deviations are driven by a few subjects or shared by most, we turn to individual behavior. We demonstrate that, despite not being Bayesian, receivers react to information as summarized by the posterior belief in systematic ways. In particular, we consider the possibility that subjects follow (potentially different) *threshold strategies*. A $\bar{\mu}$-threshold strategy, for $\bar{\mu} \in [0, 1]$, consists of guessing $a = red$ if and only if $\mu(m, \pi_C) \geq \bar{\mu}$. When $\bar{\mu} = \frac{1}{2}$, the receiver is Bayesian. When $\bar{\mu} \neq \frac{1}{2}$ the receiver is non-Bayesian, but behaves systematically: she requires stronger or weaker than needed evidence to choose $a = red$. Given our data, we can estimate a receiver-specific threshold that rationalizes the greatest fraction of her guesses.[33] We find that the behavior of many subjects is consistent with a

[33] See Appendix B for more details. In $U100S$, when focusing on the last ten rounds of the game, we observe a

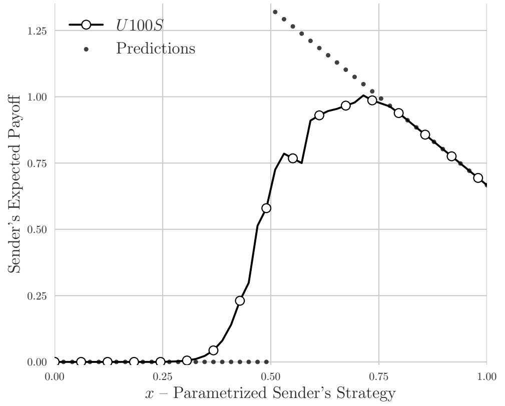
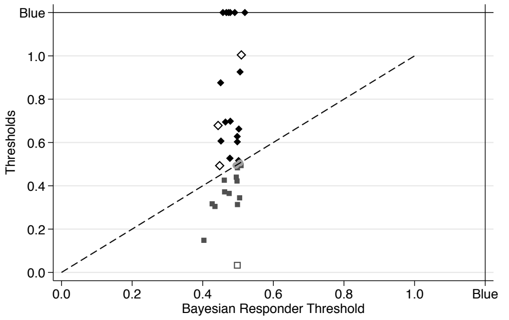


Figure A10: Probability of Guessing Red by Posterior and Message



Diamonds for subjects harder to convince than a Bayesian, square for subjects easier to convince than a Bayesian, circle otherwis
Hollow marker for subjects whose threshold is accurate in less than 80% of cases.
Blue indicates a subject who's choices are best explained by the rule always guess blue.
Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Figure A11: Estimated Thresholds: Actual Receivers vs Bayesians

threshold strategy. Almost half the receivers (46%) display behavior that is always consistent with a threshold strategy, and almost nine out of ten receivers (89%) behave consistently with a threshold strategy for more than 80% of their guesses. Figure A11 plots the estimated threshold for each receiver (vertical axis). We compare these thresholds with the thresholds that we would have estimated from the same data if receivers were Bayesian (horizontal axis).[34] The figure shows substantial heterogeneity in receivers' behavior. Dots lying above the 45-degree line indicate receivers who are reluctant to guess *red*, even when a Bayesian would conclude that there is enough evidence. By contrast, the points below the 45-degree line indicate subjects who are too eager to guess *red*, despite insufficient evidence from the perspective of a Bayesian agent. The aggregation of this heterogeneous behavior is partly responsible for the smoothness of aggregate responses to the posterior that is displayed in Figure A9 (right panel). Also note that Figure A11 shows a sizable fraction of receivers who exhibit behavior consistent with the Bayesian benchmark: one-quarter of the receivers have thresholds within 5 percentage points of being consistent with a Bayesian receiver; the number increases to one third if we are more permissive and allow for a band of 10 percentage points around the Bayesian receiver.

Overall, this threshold analysis reveals three important aspects of receivers' behavior. First, the vast majority of receivers appear to behave in systematic ways, as summarized by threshold strategies. Second, there is a substantial heterogeneity in the thresholds: some receivers are

receiver's guess on 20 occasions following r and b messages. We look for the threshold that best describes these 20 observations. This procedure typically results in a *range* of best-fitting thresholds. We report the average of these thresholds. This method akin to *perceptrons* in machine learning; see, for instance, Murphy (2012).

[34] Given the finite sample, even a Bayesian receiver can have an estimated average threshold that is different from 1/2. As an example, imagine a receiver who is perfectly Bayesian, but for whom the closest posteriors to 0.5 that we observe are 0.45 and 0.65. Her estimated threshold would then be 0.55. Figure B16 in the appendix presents the estimated threshold and their respective precision.

skeptical, some are approximately Bayesian, some others are gullible. Third, virtually all receivers respond to information in monotonic ways. It is thanks to this that the senders' empirical best responses (Figure A10) are qualitatively in line with the theory.

A.2 Changing Receiver's Incentives

Propositions 1 and 2 in Section 2 constitute the bulk of our experimental strategy, which revolves around the idea of partial commitment. A different kind of comparative static exercise that we considered does not vary the degree of commitment ρ nor the communication rule. Instead, it shows how equilibrium informativeness changes with the persuasion threshold q. As we explain below, this can be done experimentally by changing the preferences of the receiver. Formally, the prediction that we test is the following.

Proposition 3. *Fix $q' > q > \mu_0$ and consider any $\rho \geq \frac{q'-\mu_0}{q'(1-\mu_0)}$. Equilibrium informativeness under q' is strictly higher than under q, irrespective of the rules of communication.*

This result shows that when ρ is sufficiently high, an increase in q increases equilibrium informativeness irrespective of the communication rules. In particular, when $\rho = 1$, raising q strictly increases the equilibrium informativeness for both verifiability scenarios.

Based on this idea, we designed an additional treatment with full commitment ($\rho = 1$) and unverifiable information. We label this treatment $U100H$ and compare it directly to $U100$. Payoffs are as follows. As in all other treatments, the receiver earns nothing if she guesses incorrectly. In contrast to our main treatments however, the receiver earns \$2 if she correctly guesses that $\theta = B$, but only 67¢ if she correctly guesses that $\theta = R$. This payoff structure increases the persuasion threshold from $q = 1/2$ to $q = 3/4$. Since the receiver is harder to persuade, the sender is automatically worse off relative to $U100$. Therefore, to guarantee the comparability between treatments, we also modify the sender's payoff in $U100H$. In particular, she earns \$3 (instead of \$2) whenever $a = red$. In this way, her expected equilibrium payoff is the same for $U100$ and $U100S$. In equilibrium, the sender is to choose $\pi_C(r|R) = 1$ and $\pi_C(b|B) = 5/6$ and the predicted informativeness is $\phi^B(\pi_C) = 5/\sqrt{40} \approx 0.79$. We conducted four sessions of $U100H$, each with 16-20 subjects (72 in total).[35]

The left panel of Figure A12 reports the main clusters of senders' behavior in treatment $U100H$. These are computed through a k-means algorithm, as described in Section 5.2. A large fraction of senders, indicated by a blue square, choose strategies that are close to equilibrium

[35] The sessions lasted approximately 100 minutes. Subjects earned on average \$32, including a show-up fee of \$10. On average, senders and receivers made \$23 and \$40, respectively.

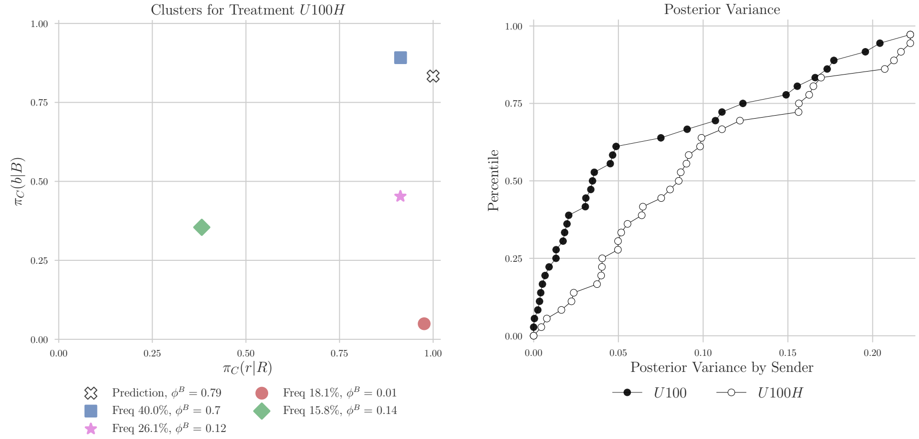
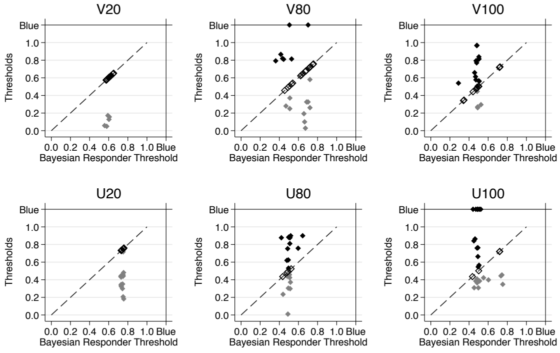


Figure A12: Strategy Clusters (left) and Posterior Variance (right) in Treatment $U100H$

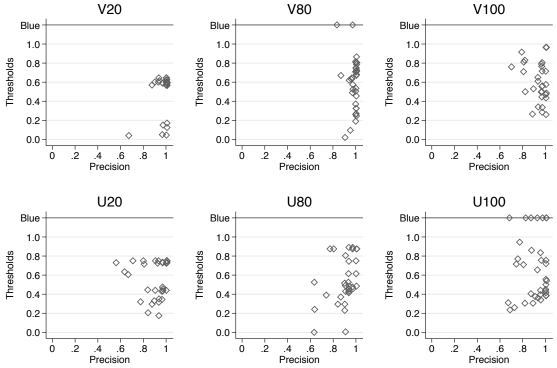
behavior. A smaller but significant fraction of senders, indicated by a purple star, choose a strategy that would be close to equilibrium behavior in $U100$ but are not informative enough to persuade a Bayesian receiver in $U100H$. The strategies summarized by the red circle capture commitment blindness, while those summarized by the green diamond capture a cluster of residual strategies that should be interpreted as noise. When comparing these clusters with those computed for treatment $U100$ (Figure D19, right panel) or $U80$ (Figure 7), we observe an overall shift toward more-informative strategies, as predicted by the theory (upper-right corner).

Quantifying this shift is complicated by the fact that receivers' preferences between $U100$ and $U100H$ have changed. Therefore, Bayesian correlations ϕ^B have to be computed using different Bayesian receivers in the two treatments. For example, a posterior of 0.74 induces $a = red$ for a Bayesian receiver in $U100$, but $a = blue$ in $U100H$. When using Bayesian correlation ϕ^B to measure informativeness, we do not find a significant change between $U100$ and $U100H$. However, this is caused by the fact that ϕ^B is a conservative measure of informativeness: when a sender induces a posterior distribution with support $\{0, q-\epsilon\}$, the Bayesian correlation is zero, even if a great deal of information was conveyed by the sender. To avoid this problem, we can compute the *variance* of the posterior distribution induced by π_C. As described in Section 3.3 and D.1, the variance of induced posteriors is an alternative measure of informativeness which, unlike ϕ^B, does not depend on q, and thus, may be more appropriate when comparing data from treatments that feature different q's. The posterior variance in $U100$ is 0.067 (predicted 0.055); in $U100H$, it is 0.094 (predicted 0.14). The increase from $U100$ to $U100H$ is significant ($p < 0.01$), in line with Proposition 3. Moreover, the sender-by-sender CDF of the posterior variance increases from $U100$ to $U100H$ in a first-order stochastic sense, as reported in the right panel of Figure A12.



Black for subjects harder to convince than a Bayesian, gray for subjects easier to convince than a Bayesian.
Blue indicates a subject who's choices are best explained by the rule always guess blue.
Values are jittered slightly to make multiple overlapping thresholds distinguishable.
Only includes responders for whom the best fitting threshold accounts for at least 80% of choices.

Figure B13: Estimated Threshold: Actual Receivers Against Bayesian



Blue indicates a subject who's choices are best explained by the rule always guess blue.
Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Figure B14: Estimated Threshold and Precision

B Threshold Strategies in Main Treatments

The relevant data for the estimation of threshold strategies comprises pairs of induced posteriors μ and guesses a for each receiver and message. We look for a threshold $\bar{\mu} \in [0, 1]$ that minimizes $\#\{a \neq \mathbb{1}\{\mu \geq \bar{\mu}\}\}$ where a takes a value of 1 for *red* and 0 for *blue*. In other words, we find the threshold $\bar{\mu}$ that rationalizes the greatest number of choices a receiver has made. We refer to the fraction of choices properly accounted for by the threshold as the *precision* of $\bar{\mu}$. Given that the sample is finite and thresholds exist on the unit interval, there will be an infinite number of thresholds with the same precision. For instance, imagine a hypothetical sample comprising only two observations: a receiver that guessed *red* given a posterior of 0.7 and guessed *blue* when the posterior was 0.4. In this case, any threshold $\bar{\mu} \in [0.4, 0.7]$ would have the same precision, namely 1. We report the midpoints of the estimated ranges.

The theory assumes receivers are Bayesian. However, notice that even a Bayesian receiver is unlikely to yield a threshold of 0.5. This is because the sample is finite. For instance, in the two-observation example proposed above, the estimated threshold is 0.55, even if the agent behaves as a Bayesian. To account for this, we compare thresholds for the receivers in our experiment with the hypothetical thresholds that we would estimate given the observed sample if the receivers were Bayesian.

Figure B13 and B14 illustrate the best-fitting thresholds and their precisions for the main treatments. Unlike for the $U100S$ treatment, these thresholds are computed with 30 choices per subject, thus achieving high precision is more difficult. Nonetheless, precision is still high: the treatment with the lowest precision still has 81% of subjects with 80% precision; across all treatments, 90% of subjects meet that criteria.

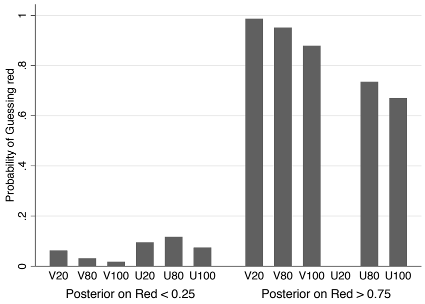
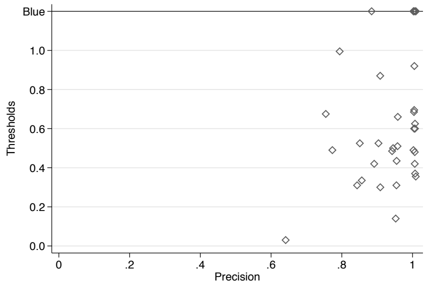


Figure B15: Frequency of $a = red$ for All Messages Given Posterior



Blue indicates a subject who's choices are best explained by the rule always guess blue.
Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Figure B16: Estimated Threshold and Precision for Treatment U100S

Figure B14 also shows that precision is particularly high when information is verifiable: 55% of receivers always choose in a way that is consistent with a threshold. That number is 24% for the treatments with unverifiable messages. From Figure B13, we deduce that receiver behavior is highly heterogeneous. A nontrivial fraction of subjects are close to the behavior Bayesian receivers would exhibit. There is also a substantial fraction of subjects who are skeptical, that is, they require higher-than-needed evidence to guess *red*, and there is a fraction of subjects who are instead, gullible. Finally, note that in the treatment that comes closest to the setup of a cheap-talk experiment, namely $U20$, all receivers that are not compatible with the Bayesian benchmark are classified as gullible. This is in line with one of the main findings in Cai and Wang (2006). Overall, the aggregation of this heterogeneous behavior is partly responsible for the linearity of aggregate responses to the posterior that is displayed in Figure 3.[36]

Finally, in all treatments, receivers' responsiveness is monotone increasing in information. However, there are some expected differences between communication rules. As Figure B15 illustrates, in treatments with verifiable information, receivers are more likely to guess $a = red$ conditional on any message m that leads to a posterior above $3/4$. This is in part due to the fact that, in these treatments, the frequency of extreme posteriors, that is $\mu = 1$, is higher, since information is verifiable. Conversely, the frequency of $a = red$ conditional on any message m that leads to a posterior below $1/4$ is lower in the verifiable treatments (it is already very low in the unverifiable treatments). Again, this is in part because the frequency of extreme posteriors, in this case $\mu = 0$, is higher in treatments with verifiable information.

[36] This linearity may appear consistent with *probability matching*. That is, subjects guess *red* with a probability equal to the posterior belief. To test for this, we compute for each subject the mean-squared error (MSE) of the predicted guess using the estimated threshold strategies and compare it with the MSE of the probability-matching model. Across all treatments, we find that for about 80% of the receivers, threshold strategies have lower MSE than probability matching.

Online Appendix for

Rules and Commitment in Communication: An Experimental Analysis

Guillaume Fréchette
New York University

Alessandro Lizzeri
New York University

Jacopo Perego
Columbia University

Contents

C Equilibrium, Refinement and Proofs

C.1 Equilibrium Characterization

In this section, we characterize the set of Perfect Bayesian Equilibria (PBE) for the framework introduced in Section 2. In a PBE, the sender chooses an information structure $\pi_C \in \Pi$ in the commitment stage. Then, at every history π'_C, the sender chooses $\pi'_R \in \Pi$, possibly as a function of π'_C. Finally, the receiver observes history (m, π_C) and responds with an action in $\{a_H, a_L\}$. We call such an action $a(m, \pi_C)$. Finally, a belief assessment μ assigns a belief to every triple (m, π'_C, π'_R). For notational simplicity, we omit the dependence of π'_R on π'_C.

Definition 1. *Fix* (Π, ρ, q)*. The tuple* (π_C, π_R, a, μ) *is a Perfect Bayesian Equilibrium if:*

(1) π_C *maximizes* $\sum_{\theta,m} \mu_0(\theta)(\rho\pi_C(m|\theta) + (1-\rho)\pi_R(m|\theta))v(a(m,\pi_C))$*;*

(2) For all (π'_C, θ)*,* π'_R *maximizes* $\sum_m \pi'_R(m|\theta)v(a(m,\pi'_C))$*;*

(3) For all (m, π'_C)*,* $a(m, \pi'_C) = a_H$ *iff* $\mu(m, \pi'_C, \pi'_R) \geq q$*;*

(4) For all (m, π'_C, π'_R)*, posterior belief* $\mu(m, \pi'_C, \pi'_R)$ *is computed from* $\pi := \rho\pi'_C + (1-\rho)\pi'_R$ *using Bayes' rule whenever possible.* [37]

Next, we provide a characterization of the equilibrium set, before imposing a refinement on our equilibrium notion. We say that an equilibrium (π_C, π_R, a, μ) under (Π, ρ, q) achieves *full-commitment informativeness* (FCI) if $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$. This is the equilibrium informativeness that can be achieved under full commitment and unverifiable information. It is also useful to define two thresholds for ρ: $\underline{\rho} := \frac{q-\mu_0}{q(1-\mu_0)}$ and $\bar{\rho} = \frac{q(1-\mu_0)}{q(1-\mu_0)+(1-q)\mu_0}$. Note that $\underline{\rho} \leq \bar{\rho}$. We begin from the case of unverifiable information.

Lemma 1. *Fix* $q > \mu_0$ *and let information be unverifiable.*

(a) If $\rho < \underline{\rho}$*, all equilibria are uninformative.*

(b) If $\rho \in [\underline{\rho}, \bar{\rho})$*, uninformative equilibria exist. Equilibria that are as informative or more informative than FCI exist.*

(c) If $\rho \geq \bar{\rho}$*, there is no uninformative equilibrium. Equilibria that are as informative or more informative than FCI exist.*

[37] When information is verifiable, we naturally assume that an off-path message θ_H (resp. θ_L), i.e. the verifiable message, leads to an off-path belief of 1 (resp. 0). This is in the spirit of Battigalli and Siniscalchi (2002).

The proof for this and the next result are in Section C.3. When commitment power is low, all equilibria are uninformative. When commitment power is sufficiently high, instead, all equilibria are strictly informative. Furthermore, in this case, FCI can be achieved in equilibrium, even if the sender may lack full commitment power. She does so by appropriately overcommunicating in the commitment stage, anticipating that her behavior in the revision stage will reduce the credibility of her communication.

Next, we present the equilibrium characterization for the case of verifiable information.

Lemma 2. *Fix $q > \mu_0$ and let that information be verifiable.*

(a) If $\rho < \underline{\rho}$, all equilibria are fully informative.

(b) If $\rho \in [\underline{\rho}, \bar{\rho})$, the least informative equilibrium is FCI; fully informative equilibria exist.

(c) If $\rho \geq \bar{\rho}$, there are no fully informative equilibria; the least informative equilibrium is FCI.

From this result, we appreciate the contrast between verifiable and unverifiable information. When information is verifiable, all equilibria are fully informative when the commitment power is low. This is in stark contrast with the unverifiable case. Moreover, when the commitment is sufficiently high, there are no fully informative equilibria and, thus, the sender can avoid the unattractive scenario where she fully disclose all her private information.

Overall, these Lemmas provide a complete characterization of the equilibrium set. Modulo the equilibrium multiplicity, which we will address in the next subsection, these results partially replicate the comparative statics that we presented in Section 2.

C.2 Truth-Leaning Equilibria

In this section, we provide two examples—one for unverifiable and one for verifiable information—of PBEs that do not satisfy the truth-leaning tie-breaking rule that we introduced in Section 2. We use these examples to argue that equilibria that are not truth-leaning feature behavior in the revision stage that is somewhat unreasonable.

Example 1: *Unverifiable Information.*

Assume that information is unverifiable and let $\rho = \frac{3}{5}$, the persuasion threshold to $q = \frac{1}{2}$, and the prior to $\mu_0 = \frac{1}{3}$. Consider the pair (π_C, π_R) that is reported in Table C5. First note that $\mu(\theta_H, \pi_C, \pi_R) < q$ and $\mu(\theta_L, \pi_C, \pi_R) < q$. That is, despite the fact that π_C is fully informative, the sender's behavior in the revision stage entirely garbles the information from the commitment stage.

Table C5

π_C	θ_H	θ_L	n
θ_H	1	0	0
θ_L	0	1	0

π_R	θ_H	θ_L	n
θ_H	0	1	0
θ_L	1	0	0

We show that the pair (π_C, π_R) can be used to construct an equilibrium. Suppose that for any deviation at the commitment stage π'_C, the sender chooses an appropriate π'_R at the revision stage so as to make the pair (π'_C, π'_R) uninformative. The Proof of Lemma 1.(b) establishes that, for ρ sufficiently low, such a π'_R exists. Given the receiver's beliefs about the revision stage strategy, the receiver would choose action a_L for both messages. Thus, the sender is indifferent among all her strategies in the revision stage and is willing to choose π'_R. Furthermore, given the receiver's expectation about π'_R, in the commitment stage the sender is also indifferent among all his strategies: all of them lead to a payoff of zero. This particularly strange behavior of the sender in the revision stage is ruled out by the truth-leaning refinement. In this equilibrium, in the revision stage the sender of type θ_H is indifferent between sending message θ_L and being truthful. Truth-leaning requires that such a sender choose $\pi_R(\theta_H|\theta_L) = 1$ instead.

Example 2: *Verifiable Information.*

Now assume that information is verifiable. As above, let $\rho = \frac{3}{5}$, $q = \frac{1}{2}$ and $\mu_0 = \frac{1}{3}$. We consider the pair (π_C, π_R) that is described in Table C6.

Table C6

π_C	θ_H	θ_L	n
θ_H	0	0	1
θ_L	0	$\frac{5}{6}$	$\frac{1}{6}$

π_R	θ_H	θ_L	n
θ_H	0	0	1
θ_L	0	0	1

Given such a π_C, in the revision stage the sender of type θ_L strictly prefers message n to message θ_L, whereas the sender of type θ_H is indifferent among the two feasible messages. Furthermore, the pair (π_C, π_R) described in the table is FCI, i.e., it leads to the maximal achievable equilibrium payoff for the sender. Therefore, the sender has no incentive to deviate at the commitment stage. This equilibrium relies on unrealistic behavior at the revision stage. To see this, consider the on-path decision of the sender of type θ_H in the revision stage. She can choose between sending message n, inducing an on-path belief of $\frac{1}{2}$, or sending an off-path message θ_H, inducing an off-path belief of 1 (See Footnote 37). Both messages trigger action a_H by the receiver. Therefore, the sender is indifferent and, yet, not truthful. Hence, while consistent with the requirement of PBE, this equilibrium is not truth-leaning.

The truth-leaning refinement is a simple tie-breaking rule, but it is powerful enough to select a unique equilibrium outcome for each ρ. This is formalized in the next result.

Lemma 3. *Fix $\rho \in [0, 1]$ and $q > \mu_0$.*

> *(Unverifiable) If $\rho < \underline{\rho}$, truth-leaning equilibria are uninformative. If $\rho \geq \underline{\rho}$, truth-leaning equilibria are FCI.*
>
> *(Verifiable) If $\rho < \bar{\rho}$, truth-leaning equilibria are fully informative. If $\rho \geq \bar{\rho}$, all truth-leaning equilibria are equally informative.*

The proof for this result is relegated to Online Appendix C.3.

C.3 Proofs

C.3.1 Proof of Proposition 1

In Lemma 3, we have established that, for any given ρ and $q > \mu_0$ and verifiability scenario, all truth-leaning equilibria are equally informative. Assume that information is unverifiable. Lemma 3 also establishes that truth-leaning equilibria are uninformative if $\rho < \underline{\rho}$ and FCI otherwise. Moreover, $\phi^B(\rho\pi_C+(1-\rho)\pi_R) = (\frac{q-\mu_0)}{(1-\mu_0)})^{\frac{1}{2}} > 0$, since $q > \mu_0$. Finally, we want to show that, when $\rho \geq \underline{\rho}$, any truth-leaning equilibrium (π_C, π_R, μ, a) satisfies $\phi^B(\pi_C) > \phi^B(\pi_R)$. Since the equilibrium is strictly informative, there exists a message m' inducing action a_H. Then, $\pi_R(m'|\theta) = 1$, for all θ. Therefore, $\phi^B(\pi_R) = 0$. However, $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}} > 0$, implying that $\phi^B(\pi_C) > 0$. We conclude that π_C is more informative than π_R. Now assume that information is verifiable. In Lemma 3, we established that truth-leaning equilibria are fully informative if $\rho < \bar{\rho}$. Moreover, we also established that, if $\rho \geq \bar{\rho}$, any truth-leaning equilibrium (π_C, π_R, μ, a) has $\phi^B(\rho\pi_C+(1-\rho)\pi_R) = (\frac{q-\mu_0(\rho+q(1-\rho))}{(1-\mu_0)(\rho+q(1-\rho))})^{\frac{1}{2}} < 1$. In this case, we argued that the fact that truth-leaning equilibria are not fully informative pins down the on-path sender behavior (π_C, π_R). In particular, we showed that $\pi_C(n|\theta_H) = 1$, $\pi_C(n|\theta_L) = (1-\underline{\rho}) - \frac{1-\rho}{\rho} \in [0, 1]$ and that $\pi_R(\theta_H|\theta_H) = \pi_R(n|\theta_L) = 1$. Given this, it is straightforward to conclude that $\phi^B(\pi_C) < \phi^B(\pi_R)$. □

C.3.2 Proof of Proposition 2

When information is unverifiable, Lemma 3 established that, any truth-leaning equilibrium (π_C, π_R, μ, a) satisfies

$$\phi^B(\rho\pi_C + (1-\rho)\pi_R) = \begin{cases} 0 & \text{if } \rho < \underline{\rho} \\ (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}} & \text{if } \rho \geq \underline{\rho} \end{cases}$$

Therefore, when information is unverifiable, equilibrium informativeness is weakly increasing in ρ. Assume now that information is verifiable. In the proof of Lemma 3, we established that any truth-leaning equilibrium (π_C, π_R, μ, a) satisfies

$$\phi^B(\rho\pi_C + (1-\rho)\pi_R) = \begin{cases} 1 & \text{if } \rho < \bar{\rho} \\ (\frac{q-\mu_0(\rho+q(1-\rho))}{(1-\mu_0)(\rho+q(1-\rho))})^{\frac{1}{2}} & \text{if } \rho \geq \bar{\rho} \end{cases}$$

It is easy to verify that $(\frac{q-\mu_0(\rho+q(1-\rho))}{(1-\mu_0)(\rho+q(1-\rho))})^{\frac{1}{2}}$ is decreasing (strictly) in ρ. Therefore, we conclude that equilibrium informativeness when information is verifiable is weakly decreasing in ρ. Finally, consider the extreme case, $\rho = 1$. It's immediate to check that in this case, irrespective of weather information is verifiable or not, equilibrium informativeness coincides and it is equal to $(\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$. □

C.3.3 Proof of Proposition 3

Assume that information is unverifiable. Fix $q' > q > \mu_0$ and consider $\rho \geq \frac{q'-\mu_0}{q'(1-\mu_0)}$. We want to show that the informativeness of truth-leaning equilibria under q' is higher than under q. To see this, note that ρ is large enough that equilibria are strictly informative, for both q' and q. In particular, due to Lemma 3 and Proposition 2, we know that under q equilibrium informativeness is equal to $(\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$ and, since $\frac{q-\mu_0}{1-\mu_0} < \frac{q-\mu_0}{1-\mu_0}$, we conclude that the informativeness of truth-leaning equilibria under q' is higher than under q. Now assume that information is verifiable. Then, both $\bar{\rho}$ and $(\frac{q-\mu_0(\rho+q(1-\rho))}{(1-\mu_0)(\rho+q(1-\rho))})^{\frac{1}{2}}$ are increasing in q. Therefore, for any value of ρ, equilibrium informativeness under q' is higher than under q. □

C.3.4 Proof of Lemma 1

Proof of Lemma 1.(a). Let information be unverifiable and $\rho < \underline{\rho}$. Suppose by way of contradiction that there is an equilibrium (π_C, π_R, a, μ) such that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) > 0$. This implies that there are positive probability messages that lead to action a_H. There are two cases to consider.

Case 1. There exists exactly one positive probability message m' such that $a(m, \pi_C, \pi_R) = a_H$. In this case, the equilibrium conditions imply that $\pi_R(m'|\theta) = 1$ for all θ. However, given this we have that

$$\begin{aligned} q \le \mu(m') &= \frac{\mu_0(\rho\pi_C m'|\theta_H) + (1-\rho))}{\mu_0(\rho\pi_C(m'|\theta_H) + (1-\rho)) + (1-\mu_0)(\rho\pi_C(m'|\theta_L) + (1-\rho))} \\ &\le \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\rho)} \\ &< \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\underline{\rho})} = q. \end{aligned}$$

The first inequality holds because m' leads to action a_H. The first equality follows from Bayes' rule. The second inequality holds because $\mu(m')$ is maximized when we set $\pi_C(m'|\theta_H) = 1 - \pi_C(m'|\theta_L) = 1$. The third inequality holds because $\rho < \underline{\rho}$. This leads to a contradiction, and therefore we can rule out Case 1.

Case 2. There are exactly two positive probability messages $m', m'' \in M$ such that $a(m, \pi_C, \pi_R) = a_H$, for $m \in \{m', m''\}$. Define $\pi_i(m', m''|\theta) := \pi_i(m'|\theta) + \pi_i(m''|\theta)$, for all θ and $i \in \{C, R\}$. Because both m' and m'' lead to a_H, equilibrium conditions imply that $\pi_R(m', m''|\theta) = 1$ for all θ. Denote by $\mu(m', m'')$ the posterior belief conditional on observing m' or m''. That is,

$$\begin{aligned} \mu(m', m'') &= \frac{\mu_0(\rho(\pi_C(m', m''|\theta_H) + (1-\rho))}{\mu_0\rho\pi_C(m', m''|\theta_H) + (1-\mu_0)\rho\pi_C(m', m''|\theta_L) + (1-\rho)} \\ &\le \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\rho)} \\ &< \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\underline{\rho})} = q. \end{aligned}$$

The first inequality holds because $\mu(m', m'')$ is maximized when $\pi_C(m', m''|\theta_H) = 1 - \pi_C(m', m''|\theta_L) = 1$. This shows that $\mu(m', m'') < q$. However, Bayes' rule also implies that, for appropriately chosen weight β,[38]

$$\mu(m', m'') = \beta\mu(m') + (1-\beta)\mu(m'') \ge q.$$

Therefore, we have $q \le \mu(m', m'') < q$, a contradiction. We conclude that the equilibrium cannot be informative. □

Proof of Lemma 1.(b).

Existence of FCI equilibria. Fix $\rho \ge \underline{\rho}$. We first show that FCI equilibria exist. We do so by constructing such an equilibrium. We start by defining strategies on the equilibrium path.

[38] More specifically, $\beta := \frac{\sum_\theta \mu_0(\theta)(\rho\pi_C(m'|\theta) + (1-\rho)\pi_R(m'|\theta))}{\sum_\theta \mu_0(\theta)(\rho\pi_C(m', m''|\theta) + (1-\rho)\pi_R(m', m''|\theta))}$

For the commitment stage, let $\pi_C(m'|\theta_H) = 1$, $\pi_C(m'|\theta_L) = x$ and $\pi_C(m''|\theta_L) = 1 - x$, where $x = \frac{1}{\rho}(\frac{\mu_0(1-q)}{q(1-\mu_0)} - (1-\rho))$. Note that π_C is well-defined. On the one hand, $x \geq 0$ if $\frac{\mu_0(1-q)}{q(1-\mu_0)} \geq 1 - \underline{\rho} \geq 1 - \rho$, which is true since $1 - \underline{\rho} = \frac{\mu_0(1-q)}{q(1-\mu_0)}$. On the other hand, $x \leq 1$ follows directly from our maintained assumption $q > \mu_0$. For the revision stage, let $\pi_R(m'|\theta) = 1$, for all θ. Given this choice of π_C and π_R, we have that $\mu(m', \pi_C, \pi_R) = q$ and $\mu(m'', \pi_C, \pi_R) = 0$, hence let $a(m', \pi_C) = a_H$ and $a(m'', \pi_C) = a_L$. It is straightforward to check that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$, i.e. it is consistent with FCI. We now define strategies off the equilibrium path. For any π'_C, let $\mu(m, \pi'_C) = \frac{\mu_0\pi'_C(m|\theta_H)}{\mu_0\pi'_C(m|\theta_H)+(1-\mu_0)\pi'_C(m|\theta_L)}$. Let $\bar{m}$ be such that $\mu(\bar{m}, \pi'_C) \geq \mu(m, \pi'_C)$, for all $m \in M$. Let $\pi'_R(\bar{m}|\theta) = 1$ for all θ. For such pairs (π'_C, π'_R), let $a(m, \pi'_C) = a_H$ if and only if $\mu(m, \pi'_C, \pi'_R) \geq q$. Whenever a message m has zero probability let $\mu(m, \pi'_C, \pi'_R) = 0$. It is straightforward to check that this strategy is indeed an equilibrium and, as noted above, FCI.

Existence of uninformative equilibria.
Next, we show that when $\rho \in [\underline{\rho}, \bar{\rho})$, an uninformative equilibrium exists. The proof is by construction and consists in finding, for each possible history π_C, a revision strategy π_R such that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = 0$. The existence of such π_R for each history π_C guarantees the existence of an uninformative equilibrium. To this end, consider an arbitrary π_C. If $\mu(m, \pi_C) < q$, for all $m \in M$, then let $\pi_R = \pi_C$, which gives $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = 0$. Conversely, suppose that there exists a message m such that $\mu(m, \pi_C) \geq q$. For arbitrary π_C, Bayes plausibility requires that there exists at least one message, call it m''', such that $\mu(m''', \pi_C) \leq \mu_0$. To simplify notation, let $\pi_C(m'|\theta_H) = a'$, $\pi_C(m''|\theta_H) = a''$, $\pi_C(m'''|\theta_H) = a'''$, $\pi_C(m'|\theta_L) = b'$, $\pi_C(m''|\theta_L) = b''$, $\pi_C(m'''|\theta_L) = b'''$. Define the revision strategy as follows: $\pi_R(m'''|\theta_H) = 1$, and let $\pi_R(m'|\theta_L) = x'$, $\pi_R(m''|\theta_L) = x''$ and $\pi_R(m'''|\theta_L) = x'''$. We want to show that there exists (x', x'', x''') such that $x' + x'' + x''' = 1$ and $\pi(m, \pi_C, \pi_R) < q$, for all $m \in M$. We have that $\mu(m', \pi_C, \pi_R) < q$ is equivalent to:

$$x' > \Phi' := \frac{\rho}{1-\rho}\Big((1-\underline{\rho})a' - b'\Big).$$

Similarly, $\mu(m'', \pi_C, \pi_R) < q$ is equivalent to:

$$x'' > \Phi'' := \frac{\rho}{1-\rho}\Big((1-\underline{\rho})a'' - b''\Big).$$

Finally, the last condition $\mu(m''', \pi_C, \pi_R) < q$ is equivalent to:

$$x' + x'' < \bar{\Phi} := \underline{\rho} + \frac{\rho}{1-\rho}\Big(b''' - a''' + \underline{\rho}a'''\Big).$$

It is straightforward to check that $\Phi' + \Phi'' < \bar{\Phi}$ and also that $\Phi' + \Phi'' < 1$ if and only if $\rho < \bar{\rho}$. Therefore, x' and x'' can be found so that the thus defined π_R is an information structure and $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = 0$.

It is straightforward to complete the construction of the uninformative equilibrium. Note that the sender has no profitable deviation in the commitment stage. In fact, all possible deviations π'_C lead to a π'_R that, by construction, only induces beliefs strictly below q, hence a guess a_L. Similarly, the sender has no profitable deviation in the revision stage, for reasons that are similar to the existence of a babbling equilibrium in a cheap talk game.

Existence of equilibria that are more informative than FCI. The construction of these equilibria is tightly related to the construction of uninformative equilibria above. Fix $\rho \geq \underline{\rho}$. We start by constructing the sender's strategies on the equilibrium path. Let $\pi_C(m'|\theta_H) = \pi_C(m''|\theta_L) = 1$, that is, π_C is fully informative. Let $\pi_R(m'|\theta) = 1$ for all θ. Following these choices, the receiver's guesses and beliefs are naturally pinned down. For all "off-path" $\pi'_C \neq \pi_C$, we associate a π'_R that is constructed as in the case of an uninformative equilibrium, as explained above. This means that for all $\pi'_C \neq \pi_C$, $\pi^B(\rho\pi'_C + (1-\rho)\pi'_R) = 0$, the receiver always guesses a_L and the sender's expected utility is 0. Clearly, in light of this construction, the sender in the commitment stage has no incentive to deviate from π_C. Thus, this defines an equilibrium. Moreover, it is easy to verify that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = (\frac{\mu_0\rho}{1-\rho(1-\mu_0)})^{\frac{1}{2}}$, which is higher than FCI for all $\rho \geq \underline{\rho}$. □

Proof of Lemma 1.(c).

The existence of FCI equilibria as well as the existence of equilibria that are more informative than FCI follows directly from the Proof of Lemma 1.(b).

Non-existence of uninformative equilibria. We now prove that when $\rho \geq \bar{\rho}$ all equilibria are strictly informative. Suppose not. That is let $\rho \geq \bar{\rho}$ and let (π_C, π_R, μ, a) be an uninformative equilibrium. Thus, the sender earns a payoff of zero. We construct a profitable deviation π'_C under which there exists a message m' that induces action a_H with strictly positive probability. We construct this deviation to be fully informative, namely, $\pi'_C(m'|\theta_H) = 1$ and $\pi'_C(m''|\theta_L) = 1$, for $m'' \neq m'$. Call π'_R the continuation strategy of the sender in the revision stage. We have that,

$$\mu(m', \pi'_C, \pi'_R) = \frac{\mu_0(\rho + \pi'_R(m'|\theta_H))}{\mu_0(\rho + \pi'_R(m'|\theta_H)) + (1-\mu_0)(1-\rho)\pi'_R(m'|\theta_L)}$$

$$\geq \frac{\mu_0\rho}{\mu_0\rho + (1-\mu_0)(1-\rho)} \geq \frac{\mu_0\bar{\rho}}{\mu_0\bar{\rho} + (1-\mu_0)(1-\bar{\rho})} = q$$

The first inequality holds because setting $\pi'_R(m'|\theta_H) = 0$ and $\pi'_R(m'|\theta_L) = 1$ induces a lower bound for $\mu(m', \pi'_C, \pi'_R)$. The second inequality holds because $\rho \geq \bar{\rho}$, by assumption. Therefore, in the continuation game following deviation π'_C, the receiver plays a_H with positive probability so that the deviation is strictly profitable. □

C.3.5 Proof of Lemma 2

Proof of Lemma 2.(a). Assume now that information is verifiable and $\rho < \underline{\rho}$. We want to show that all equilibria are fully informative. Suppose not. That is, (π_C, π_R, a, μ) is an equilibrium with $\phi^B(\rho\pi_C + (1-\rho)\pi_R) < 1$. Because information is verifiable, such a situation implies that $a(n, \pi_C) = a_H$. Therefore, $\mu(n, \pi_C, \pi_R) \geq q$. However, equilibrium conditions also imply that $\pi_R(n|\theta_L) = 1$, i.e., in the revision stage, the sender of type θ_L always sends message n. Therefore, we have that:

$$q \leq \mu(n, \pi_C, \pi_R) \leq \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\rho)} < \frac{\mu_0\underline{\rho}}{\mu_0\underline{\rho} + (1-\mu_0)(1-\underline{\rho})} = q.$$

The first inequality comes from $a(n, \pi_C) = a_H$. The second inequality holds because setting $\pi_C(n|\theta_H) = 1$, $\pi_C(n|\theta_L) = 0$ and $\pi_R(n|\theta_R) = 1$ generates an upper bound for the value of $\mu(n, \pi_C, \pi_R)$. The last inequality holds because $\rho < \underline{\rho}$, by assumption. Therefore, $q \leq \mu(n, \pi_C, \pi_R) < q$, a contradiction. □

Proof of Lemma 2.(b).

Existence of FCI equilibria. We prove this by construction. Fix $\rho \geq \underline{\rho}$. Let π_C and π_R be such that $\pi_C(n|\theta_H) = \pi_R(n|\theta_H) = \pi_R(n|\theta_L) = 1$ and $\pi_C(n|\theta_L) = x$. Let $x := \frac{1}{\rho}(\rho - \underline{\rho})$ and note that, by assumption, $\rho \geq \underline{\rho}$, hence $x \in [0, 1]$. Moreover, it is easy to verify that $\mu(n, \pi_C, \pi_R) = q$. Let $a(n, \pi_C) = a_H$, therefore π_R is a best response to π_C given the receiver's behavior. It is also easy to verify that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$. Therefore, (π_C, π_R) is FCI. As a consequence, no profitable deviation away from π_C exists. Thus, we have have constructed an equilibrium that is FCI. Moreover, this is the least informative equilibrium that exists in this case. To see this, note that, because of the nature of verifiable information, $\phi^B(\rho\pi_C + (1-\rho)\pi_R) < 1$ requires that $\mu(n, \pi_C, \pi_R) \in [q, 1)$. Moreover, $\phi^B(\rho\pi_C + (1-\rho)\pi_R)$ is increasing in $\mu(n, \pi_C, \pi_R)$. The equilibrium that we constructed above has $\mu(n, \pi_C, \pi_R) = q$ and it is therefore minimally informative.

Existence of fully informative equilibria. We prove this by construction. Consider a revision strategy π_R defined as $\pi_R(\theta_H|\theta_H) = \pi_R(n|\theta_L) = 1$. Moreover, suppose that π_R is played for all histories π'_C. Consider an arbitrary history π'_C. Note that, for all $\rho < \bar{\rho}$,

$$\mu(n, \pi'_C, \pi_R) = \frac{\mu_0 \rho \pi'_C(n|\theta_H)}{\mu_0 \rho \pi'_C(n|\theta_H) + (1-\mu_0)(\rho \pi'_C(n|\theta_L) + (1-\rho))} < q.$$

Moreover, note that π_R is a best-response to this arbitrary π'_C. Finally, note that, in the subgame indexed by π'_C, the sender expects to receive a payoff of $\mu_0(\rho\pi'_C(\theta_H|\theta_H) + (1-\rho)\pi_R(\theta_H|\theta_H) \leq \mu_0$. Now consider the strategy $\pi_C = \pi_R$. This strategy gives a payoff of μ_0 and, due to the argument above, no profitable deviation from this strategy exists. Moreover, $\phi^B(\rho\pi'_C + (1-\rho)\pi'_R) = 1$. □

Proof of Lemma 2.(c). The existence of FCI equilibria as well as the fact that these are the least informative equilibria follows directly from the Proof of Lemma 2.(b).

Non-existence of fully informative equilibria. We first show that when $\rho \geq \bar{\rho}$, there exist no fully informative equilibrium. When $\rho = 1$ the result is a straightforward consequence of full commitment, so let us focus on the case $\rho \in [\bar{\rho}, 1)$. Suppose that there exists an equilibrium (π_C, π_R, a, μ) such that $\phi^B(\rho\pi_C + (1-\rho)\pi_R) = 1$. In this equilibrium, the sender expects to earn μ_0. Consider a deviation π'_C such that $\pi'_C(n|\theta_H) = 1$ and $\pi'_C(\theta_L|\theta_L) = 1$. We argue that this deviation leads to a subgame in which the sender earns strictly more than μ_0. First, note that for all π'_R,

$$\mu(n, \pi'_C, \pi'_R) = \frac{\mu_0(\rho + (1-\rho)\pi'_R(n|\theta_H))}{\mu_0(\rho + (1-\rho)\pi'_R(n|\theta_H)) + (1-\mu_0)(1-\rho)\pi'_R(n|\theta_L)} \geq$$

$$\geq \frac{\mu_0 \rho}{\mu_0 \rho + (1-\mu_0)(1-\rho)} \geq \frac{\mu_0 \bar{\rho}}{\mu_0 \bar{\rho} + (1-\mu_0)(1-\bar{\rho})} = q.$$

Therefore, $a(\pi'_C, n) = a_H$. This implies that $\pi'_R(n|\theta_L) = 1$. Hence, the expected payoff for the sender in the commitment stage is bounded below by $\mu_0(\rho\pi'_C(n|\theta_H) + (1-\rho)\pi'_R(\theta_H|\theta_H) + (1 - \mu_0)(1-\rho)\pi'_R(n|\theta_L) = \mu_0 + (1-\rho)(1-\mu_0) > \mu_0$. Therefore, π'_C is a profitable deviation. Moreover, irrespective of what $\pi'_R(n|\theta_H)$ is, the fact that n is sent with strictly positive probability in both states implies that, as long as $\rho < 1$, $\mu(n, \pi'_C, \pi'_R) < 1$; hence, $\phi^B(\rho\pi'_C + (1-\rho)\pi'_R) < 1$. □

C.3.6 Proof of Lemma 3

Unverifiable Information. If $\rho < \underline{\rho}$, Lemma 1.(a), all PBEs are uninformative. A fortiori, under this assumption, all truth-leaning are uninformative. Note that, truth-leaning equilibria exist in this case. For example, let π_C and π_R be defined as $\pi_C(\theta|\theta) = 1$ for all θ and $\pi_R(\theta_H|\theta) = 1$ for all θ, $\mu(m, \pi_C, \pi_R) = \mu_0$, and $a(m, \pi_C) = a_L$. Therefore, consider instead the case $\rho \geq \underline{\rho}$. We want to argue that all truth-leaning equilibria are FCI. In order to do so, we argue that there exists a pair (π_C, π_R) such that (1) π_R is a best-response to π_C, (2) π_R is uniquely pinned down by the truth-leaning refinement and, moreover, (3) $\phi^B(\rho\pi_C+(1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$. To this end, let $\pi_C(\theta_H|\theta_H) = 1$, $\pi_C(\theta_H|\theta_L) = x$ and $\pi_C(\theta_L|\theta_L) = 1 - x$, where $x := \frac{1}{\rho}(\rho - \underline{\rho})$. Note that $x \in [0, 1]$, hence π_C is well-defined. Conversely, let π_R be such that $\pi_R(\theta_H|\theta) = 1$ for all θ. First, let us establish that π_R best-responds to π_C. To see this note that, by construction, $\mu(\theta_L, \pi_C, \pi_R) = 0$ and $\mu(\theta_H, \pi_C, \pi_R) = q$. Therefore, $a(\theta_H, \pi_C) = a_H$ and $a(\theta_L, \pi_C) = a_L$. Consistently, π_R gives positive probability to $m = \theta_H$ only. Hence π_R best-responds to π_C. Second, let us argue that π_R is indeed truth-leaning. To see this, just notice that, in the revision stage, the sender of type θ_H is being truthful, hence π_R is truth-leaning. Type θ_L is also truth-leaning since she is not indifferent between $m = \theta_H$ and $m = \theta_L$. Finally, it is straightforward to verify that, given this choice of (π_C, π_R), we have that $\phi^B(\rho\pi_C+(1-\rho)\pi_R) = (\frac{q-\mu_0}{1-\mu_0})^{\frac{1}{2}}$, i.e. it is FCI. This implies that, if the pair (π_C, π_R) is played on the equilibrium path, it leads to the first-best payoff, namely $\frac{\mu}{q}$. This proves that all truth-leaning equilibria of the grand-game are FCI. To see this, suppose that this is not the case, i.e. there exists a truth-leaning equilibrium $(\pi'_C, \pi'_R, \mu', a')$ that is not FCI, so that the sender's expected payoff in this equilibrium is strictly smaller than $\frac{\mu}{q}$. However, a deviation at the commitment stage exists, namely strategy π_C, that leads to a unique best-response in the revision stage, namely π_R, that is consistent with truth-leaning and that achieves the first-best payoff, namely $\frac{\mu}{q}$. Therefore, such deviation is strictly profitable and $(\pi'_C, \pi'_R, \mu', a')$ is not an equilibrium.

Verifiable Information. If $\rho < \underline{\rho}$, Lemma 2.(a) shows that all PBE are fully informative. A fortiori, all truth-leaning equilibria are fully informative. Trivially, a truth-leaning equilibrium exists. For example, $\pi_i(\theta|\theta) = 1$ for all θ and $i \in \{C, R\}$; $\mu(m, \pi_C, \pi_R) = 1$ if $m = \theta_H$ and 0 otherwise; $a(m, \pi_C) = a_H$ iff $m = \theta_H$ and a_L otherwise. Therefore, consider instead the case $\rho \in [\underline{\rho}, \bar{\rho})$. We want to show that all truth-leaning equilibria are fully informative. Suppose not, namely let (π_C, π_R, μ, a) be a truth-leaning equilibrium such that $\phi^B(\rho\pi_C + (1 - \rho)\pi_R) < 1$. Since equilibrium informativeness is strictly less than one, there must exist a message m such that $\mu(m, \pi_C, \pi_R) \in (0, 1)$. When information is verifiable, it is necessarily the case that $m = n$. Moreover, $a(n, \pi_C) = a_H$. If this were not the case, $\pi_i(n|\theta_H) = 0$, for $i \in \{C, R\}$, hence $\mu(n, \pi_C, \pi_R) = 0$, a contradiction. Therefore, let $\mu(n, \pi_C, \pi_R) \in [q, 1)$. On the one hand, equilibrium requires that $\pi_R(n|\theta_L) = 1$. (Note that this is consistent with truth-leaning since the two messages lead to different payoffs). On the other hand, type θ_H in the revision stage is indifferent between θ_H and n, as they both lead to action a_H. The truth-leaning refinement requires that $\pi_R(\theta_H|\theta_H) = 1$. Therefore, the fact that the equilibrium is not fully informative uniquely pins down π_R. Given this, we note that:

$$\mu(n, \pi_C, \pi_R) \le \frac{\mu\rho}{\mu\rho + (1 - \mu)(1 - \rho)} < \frac{\mu\bar{\rho}}{\mu\bar{\rho} + (1 - \mu)(1 - \bar{\rho})} = q.$$

Table D7: Predicted and Observed Posterior Variances by Treatment

ψ^B – Theoretical Predictions

	Commitment (ρ)		
	20%	**80%**	**100%**
Verifiable	0.22	0.08	0.05
Unverifiable	0.00	0.05	0.05

ψ^B – Empirical Posterior Variance

	Commitment (ρ)		
	20%	**80%**	**100%**
Verifiable	0.18	0.17	0.15
Unverifiable	0.02	0.05	0.06

Hence, $\mu(m, \pi_C, \pi_R) < q$, a contradiction. Finally, let us consider the case $\rho \geq \bar{\rho}$. We want to show that all truth-leaning equilibria are equally informative. Let (π_C, π_R, μ, a) be a truth-leaning equilibrium. By Lemma 2.(c), no equilibrium is fully informative. Therefore, by the argument made above, $\mu(n, \pi_C, \pi_R) \in [q, 1)$ and π_R is uniquely pinned down. Moreover, π_R is independent of π_C. Therefore, there exists a unique best-response π_C to such a revision strategy π_R. Such π_C is given by $\pi_C(n|\theta_H) = 1$ and $\pi_C(n|\theta_L) = x$, where $x := (1 - \underline{\rho}) - \frac{1-\rho}{\rho} \in [0, 1]$. This strategy π_C satisfies $\mu(n, \pi_C, \pi_R) = q$, while maximizing the ex-ante probability of sending message n. By construction, all truth-leaning equilibria share the same on-path sender behavior (π_C, π_R). Therefore, all truth-leaning equilibria have to be equally informative. Moreover, it is easy to verify that $\phi^B(\rho\pi_C + (1 - \rho)\pi_R) = (\frac{q-\mu_0(\rho+q(1-\rho))}{(1-\mu_0)(\rho+q(1-\rho))})^{\frac{1}{2}}$. □

D Additional Material

D.1 Alternative Measures of Informativeness: Posteriors Variance

In the paper, Bayesian correlation ϕ^B has been our principal way to measure the informativeness of a sender's strategy. In Section 3.3, we discussed the merits of this measure and how it relates to the existing literature. In this section, we re-evaluate our main comparative static exercise from Section 4 using an alternative measure of informativeness, the variance of induced posteriors. More formally, a strategy (π_C, π_R) induces a distribution $\tau \in \Delta(\Delta(\Theta))$ over posterior beliefs $\mu(m, \pi_C, \pi_R)$. The variance of τ, denoted $\psi^B := \mathbb{E}_\tau((\mu - \mu_0)^2)$, is what we call the *variance of induced posteriors*. Clearly, ϕ^B and ψ^B have much in common. First, they are highly correlated: Across all our treatments, the correlation between ϕ^B and ψ^B is above 0.95. Second, they are both immune to receivers' mistakes. The main difference between the two is that ψ^B does not require the specification of a payoff function for the (Bayesian) receiver. Thus, it is perhaps better suited to compare treatments with different q, like $U100$ and $U100H$.

In Table D7, we report the average posterior variance across treatments together with the

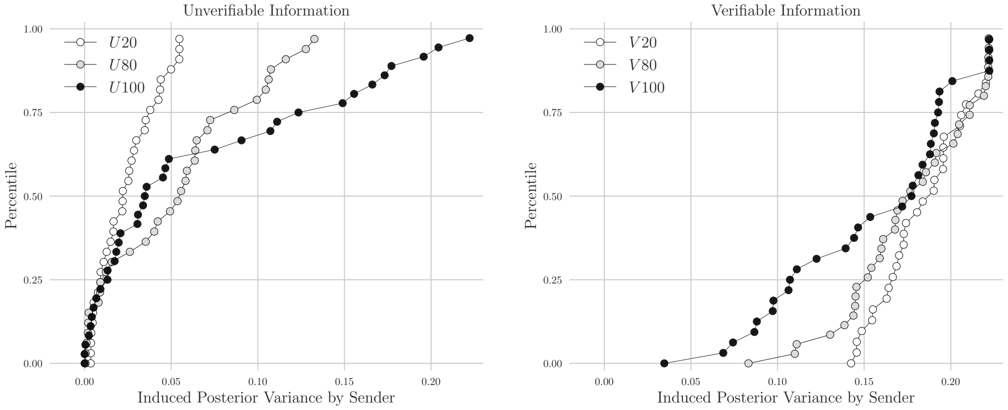


Figure D17: CDF of Sender-Average Variance of Induced Posteriors (ψ)

theoretical predictions. As for Table 3, this measure of informativeness moves in the direction predicted by the theory. Namely, it increases in treatments with unverifiable information and it decreases in treatments with verifiable information. Yet, as for ϕ^B, the point-predictions are far from the empirical averages. In particular, senders in $V100$ appear to be overly informative relative to the prediction and there is a large gap between $V100$ and $U100$. This is all in line with the evidence reported in Section 5. Similarly, Figure D17 reports the CDF of sender-average ψ^B, which replicates Figure 4 in the main text from the perspective of posterior variance.

D.2 Statistical Tests

The p-values reported in the main text are obtained by regressing the variable of interest on the relevant regressor (sometimes an indicator variable) with subject-level random effects and clustering of the variance-covariance matrix at the session-level. This specification has the advantage of being uniform (the same throughout the paper), it directly accounts for heterogeneity across subjects via the random effects (as the paper documents, there is clear evidence of heterogeneity between subjects), and it permits unmodeled dependencies between observations from the same session (see Fréchette, 2012, where such possibilities are discussed). However, it does not directly account for the fact that we are often dealing with a limited dependent variable. Also, clustering with a small number of clusters can lead to insufficient corrections (see Cameron and Miller, 2015, for a survey). But this relies mostly on simulations that do not necessarily mirror the situation of most laboratory experiments. In particular, the extent of the problem is found to depend on the size of the within session correlation (see, for example, Carter et al., 2017). For many experiment, such correlation can be expected to be low (once the appropriate factors are controlled for). Hence, we are more concerned with controlling for the source of dependencies across the observations of a given subject than for the within-session

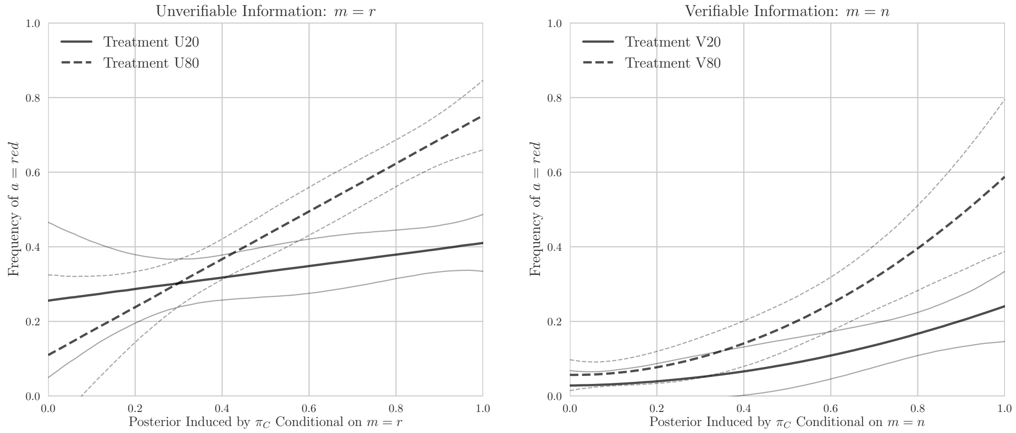


Figure D18: Receiver's Response to Persuasive Messages: ρ = 0.2 vs. ρ = 1

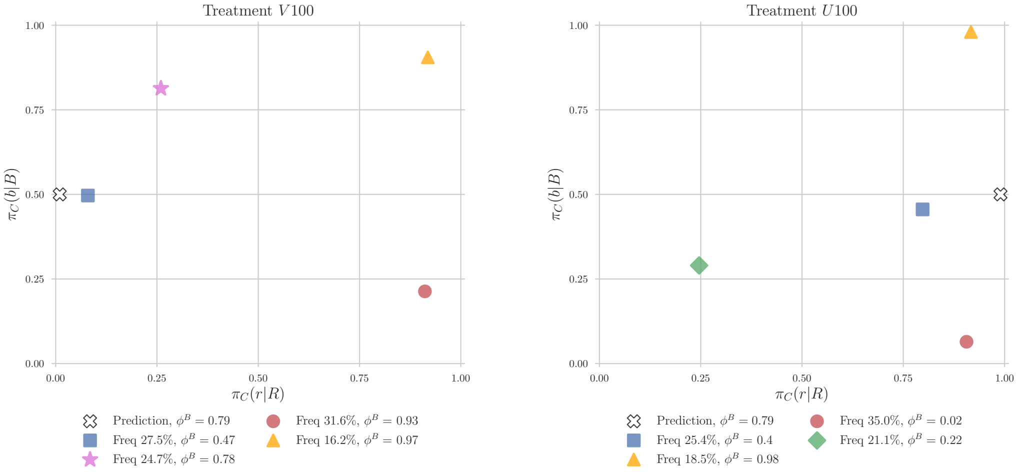


Figure D19: k-Means – Representative Strategies in Treatments with Full Commitment

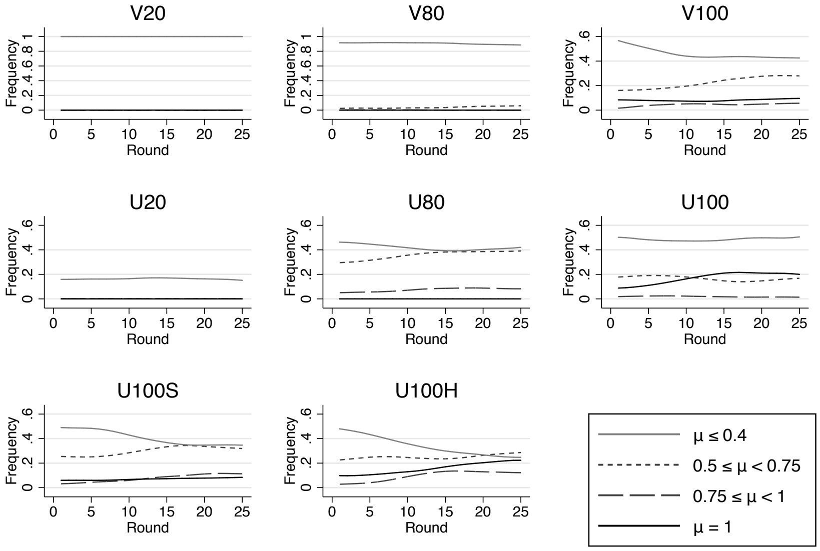
Table D8: P-Values of Statistical Tests

Model	Linear	Linear	Pr(T)obit	Pr(T)obit	Linear	Linear
Subject	RE	RE	RE	RE	FE	FE
Session	Cluster	RE	Cluster	RE	Cluster	Cluster
Bootstrap					CATs	
Test						
$\Pr\left(red \mid \mu < \frac{1}{2}\right) = \Pr\left(red \mid \mu \geq \frac{1}{2}\right)$	0.000	0.000	0.000	0.000	0.011	0.012
$\Pr\left(red \mid m = r, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = r, \mu \geq \frac{1}{2}\right)$	0.000	0.000	0.000	0.000	0.010	0.014
$\Pr\left(red \mid m = b, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = b, \mu \geq \frac{1}{2}\right)$	0.010	0.000	0.003	0.000	0.047	0.078
Left panel Figure 2, all bars = 0 when ball is R	0.000	0.000				
Left panel Figure 2, all bars = 0 when ball is B	0.000	0.000				
Right panel Figure 2, r message bar = 0 when ball is R	0.000	0.000				
$\phi_C^B = \phi_R^B$ in U80	0.000	0.000	0.000	0.996		
$\phi_C^B = \phi_R^B$ in V80	0.000	0.000	0.006	0.000		
$\Pr\left(red \mid m = r, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = r, \mu \geq \frac{1}{2}\right)$ in U20	0.053	0.002	0.083	0.004	0.150	0.126
$\Pr\left(red \mid m = r, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = r, \mu \geq \frac{1}{2}\right)$ in U100	0.000	0.000	0.024	0.000	0.040	0.021
$\Pr\left(red \mid m = r, \mu < \frac{1}{2}, U20\right) = \Pr\left(red \mid m = r, \mu < \frac{1}{2}, U100\right)$	0.627	0.535	0.718	0.610		
$\Pr\left(red \mid m = r, \mu \geq \frac{1}{2}, U20\right) = \Pr\left(red \mid m = r, \mu \geq \frac{1}{2}, U100\right)$	0.000	0.001	0.002	0.003		
$\Pr\left(red \mid m = n, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = n, \mu \geq \frac{1}{2}\right)$ in V20	0.038	0.002	0.133	0.006	0.257	0.163
$\Pr\left(red \mid m = n, \mu < \frac{1}{2}\right) = \Pr\left(red \mid m = n, \mu \geq \frac{1}{2}\right)$ in V100	0.000	0.000	0.000	0.000	0.022	0.014
$\Pr\left(red \mid m = r, \mu < \frac{1}{2}, V20\right) = \Pr\left(red \mid m = r, \mu < \frac{1}{2}, V100\right)$	0.566	0.674	0.536	0.452		
$\Pr\left(red \mid m = r, \mu \geq \frac{1}{2}, V20\right) = \Pr\left(red \mid m = r, \mu \geq \frac{1}{2}, V100\right)$	0.000	0.000	0.000	0.000		
$\phi(V20) = \phi(V80)$	0.217	0.215				
$\phi(V80) = \phi(V100)$	0.001	0.020	0.258	0.451		
$\phi(U20) = \phi(U80)$	0.002	0.001				
$\phi(U80) = \phi(U100)$	0.696	0.676	0.486	0.441		
$\phi(V20) = \phi(U20)$	0.000	0.000				
$\phi(V80) = \phi(U80)$	0.000	0.000				
$\phi(V100) = \phi(U100)$	0.000	0.000	0.000	0.000		
$\phi^B(V20) = \phi^B(V80)$	0.156	0.130				
$\phi^B(V80) = \phi^B(V100)$	0.032	0.052	0.608	0.648		
$\phi^B(U20) = \phi^B(U80)$	0.000	0.000				
$\phi^B(U80) = \phi^B(U100)$	0.957	0.925	0.711	0.661		
$\phi^B(V20) = \phi^B(U20)$	0.000	0.000				
$\phi^B(V80) = \phi^B(U80)$	0.000	0.000				
$\phi^B(V100) = \phi^B(U100)$	0.000	0.000	0.000	0.000		
$\phi^B(U100) = \phi^B(U100H)$	0.144	0.116	0.205	0.180		
$\phi^B(U100) = \phi^B(U100H)$ in last 3 matches	0.052	0.038	0.061	0.056		
$\Pr\left(red \mid \mu < \frac{1}{2}, U100\right) = \Pr\left(red \mid \mu < \frac{1}{2}, U100H\right)$	0.069	0.053	0.026	0.026		
$\Pr\left(red \mid \frac{1}{2} \leq \mu < \frac{3}{4}, U100\right) = \Pr\left(red \mid \frac{1}{2} \leq \mu < \frac{3}{4}, U100H\right)$	0.008	0.110	0.011	0.125		
$\Pr\left(red \mid \mu \geq \frac{3}{4}, U100\right) = \Pr\left(red \mid \mu \geq \frac{3}{4}, U100H\right)$	0.001	0.014	0.008	0.046		

correlations (see also Appendix A.4 of Embrey et al. (2017) for a discussion of these issues).

In Table D8 we document the robustness of the tests reported in the text by exploring alternative specifications. These include directly accounting for the limited nature of the dependent variable by using a probit or Tobit when appropriate. When possible we also report bootstrapped estimates that have been shown to perform better when the number of clusters is small (cluster-adjusted t-statistics or CAT) and allow for subject-specific fixed-effects (Ibragimov and Müller, 2010). When we report those we also include results from a standard subject specific fixed-effects estimation with session clustering to provide a benchmark. As can be seen, p-values are not systematically larger for CATs than with the "standard" clustering, nor are they very different when estimating a probit or tobit.[39] As a whole, results are fairly robust:

[39] Note that if a tobit could have been estimated but is not reported, it means that the dependant variable was



Posterior following a critical message: no message for V treatments and red message for U treatments
V20 and V80 are drawn with a different y-axis.

Figure D20: Senders' Frequency of Inducing $\mu(m, \pi_C, \pi_R)$

out of the 35 hypotheses tested, for only six of them are results not the same for all tests reported (in the sense of being consistently significant–or not–at the 10% level). The few cases where there are differences are for the most part not difficult to make sense of. Two of them involve comparing $V80$ and $V100$, where the difference is small in magnitude. Hence, whether or not the difference is statistically significant is not clear, but either way it is not large. In most other cases, the p-values are either under the 0.1 cutoff or just slightly above.

D.3 Subjects' Behavior Over Time

Figures D20 and D21 illustrate changes in behavior over the course of the experiment.

Senders. Figure D20 studies senders by coarsely separating their strategies by the posterior they induce conditional on the *persuasive* message; that is, message n under verifiable information and r otherwise. Four posterior intervals are considered: low ($\mu < 0.4$), close to full-commitment equilibrium ($0.5 \geq \mu < 0.75$), high ($0.75 \leq \mu < 1$), and maximal ($\mu = 1$). We excluded posteriors in the interval $0.4 < \mu < 0.5$. As the figure shows, overall there are very few changes over time (at least, no change across these groups of posteriors). Notable exceptions are treatment $U100H$ and, to a lesser extent $U100S$ and $U100$, where senders seem to learn to provide more information over time.

Receivers. Figure D21 studies receivers and displays changes in terms of the likelihood a

not actually censored.

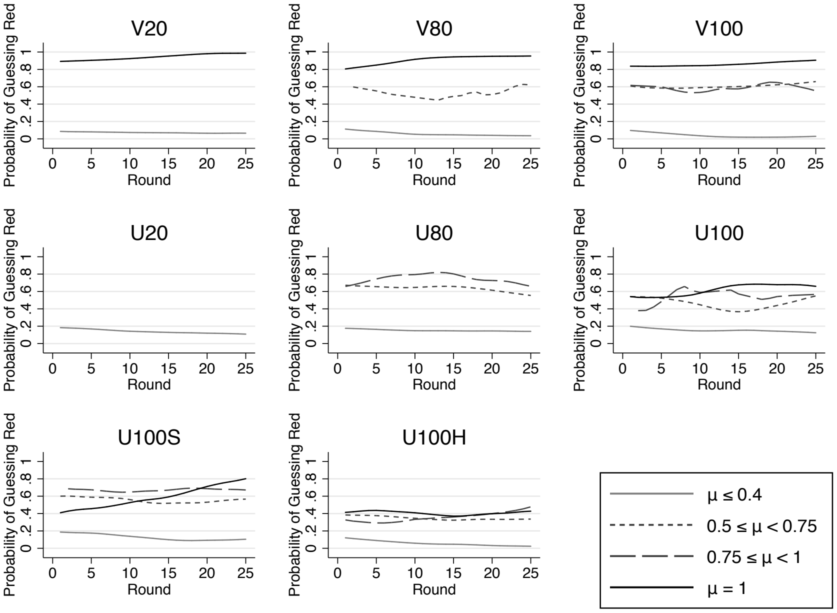


Figure D21: Receivers's Frequency of a = *red* Grouped by Posterior (μ)

given posterior leads to a $a = red$. Overall, time effects are limited. There appears to be an increase in the frequency of $a = red$ conditional on higher posteriors ($U80$ is one exception) and a decrease of such frequency conditional on lower posteriors.

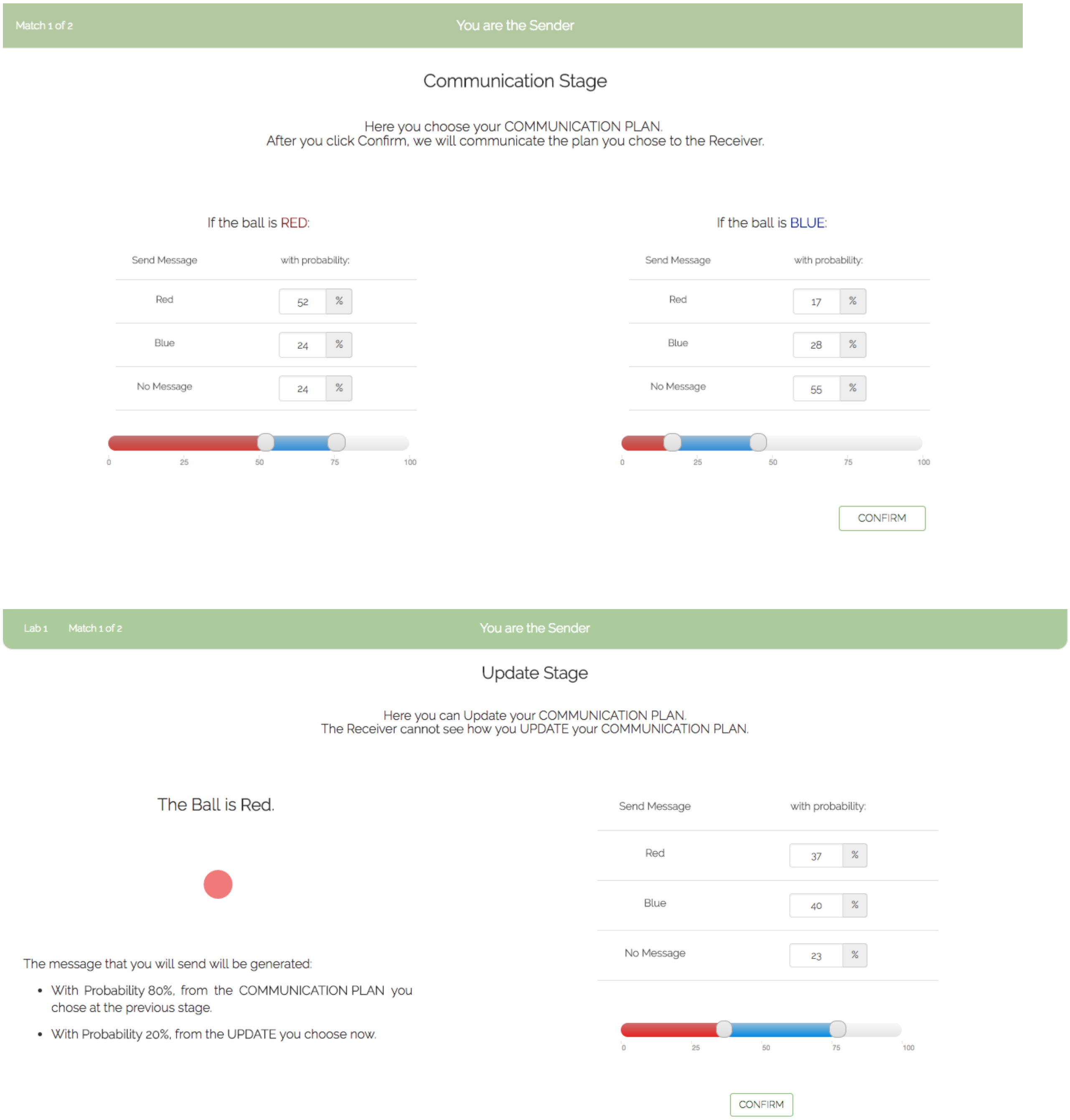
E Design

E.1 Graphical Interface

Figures D22 and D23 show the software interface of our experiment. More specifically, Figures D22 show the commitment and revision stages. To avoid framing, the experiment referred to these stages as “communication” and “update.” Figure D23 show the guessing stage and the feedback screen. In the feedback screen, all relevant information are reported to both players, with the exception of the sender’s choices in the Revision stage.

E.2 Sample Instructions

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 these instructions, screenshots similar to those in Figures D22 and D23 were shown with a projector,



Match 1 of 2

You are the Sender

Communication Stage

Here you choose your COMMUNICATION PLAN.
After you click Confirm, we will communicate the plan you chose to the Receiver.

If the ball is RED:

Send Message	with probability:
Red	52 %
Blue	24 %
No Message	24 %

If the ball is BLUE:

Send Message	with probability:
Red	17 %
Blue	28 %
No Message	55 %

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.

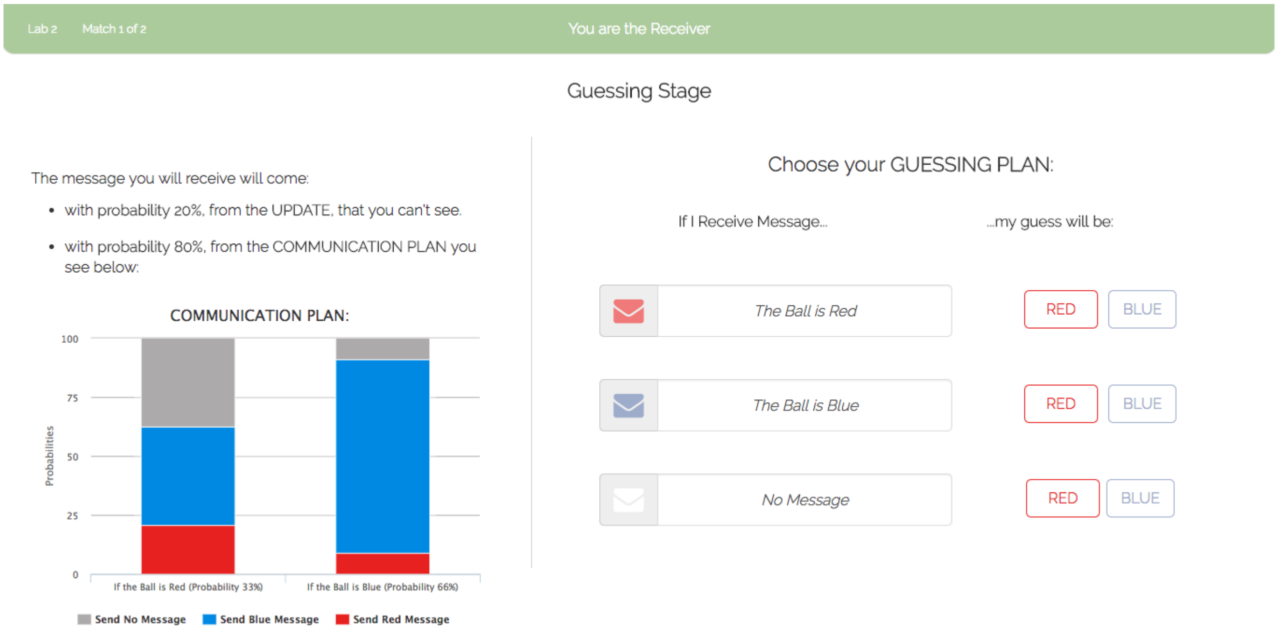
The Ball is Red.

Send Message	with probability:
Red	37 %
Blue	40 %
No Message	23 %

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.

Figure D22: Samples Screenshots U80, Commitment and Revision Stages



Lab 2 Match 1 of 2 You are the Receiver

Guessing Stage

The message you will receive will come:

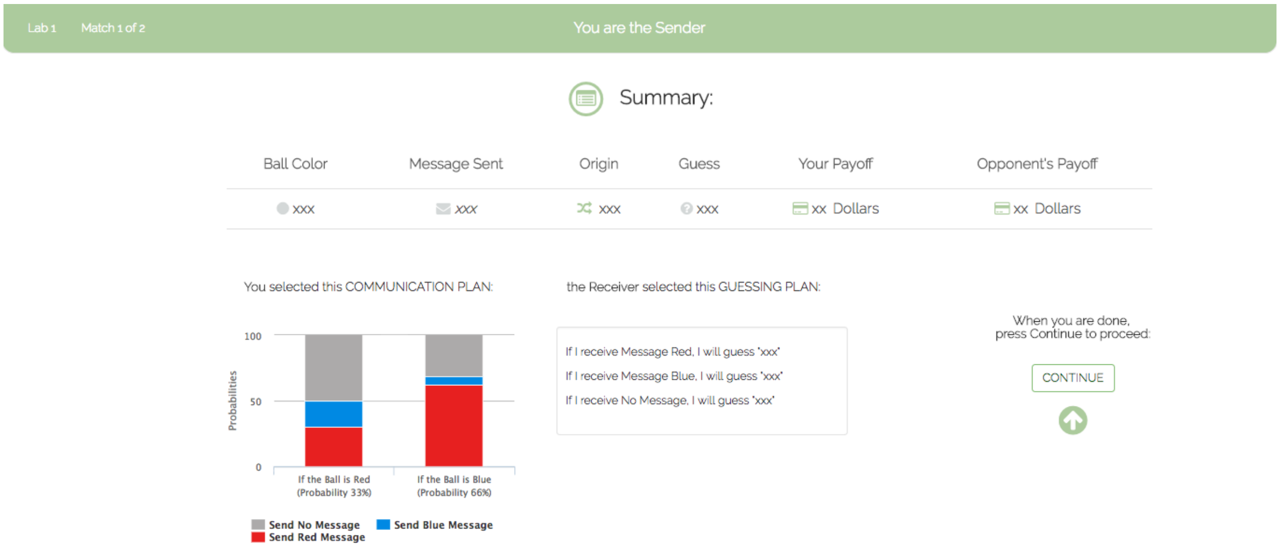
- with probability 20%, from the UPDATE, that you can't see.
- with probability 80%, from the COMMUNICATION PLAN you see below:

COMMUNICATION PLAN:

Probabilities
100
75
50
25
0
If the Ball is Red (Probability 33%)
If the Ball is Blue (Probability 66%)
Send No Message
Send Blue Message
Send Red Message

Choose your GUESSING PLAN:

If I Receive Message...	...my guess will be:
The Ball is Red	RED BLUE
The Ball is Blue	RED BLUE
No Message	RED 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:

Probabilities
100
50
0
If the Ball is Red (Probability 33%)
If the Ball is Blue (Probability 66%)
Send No Message
Send Blue Message
Send Red Message

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 D23: Samples Screenshots U80, Guessing Stage and Feedback

to ease the exposition and the understanding of the tasks.

E.2.1 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 the main features of the session. If you have any questions during this period, raise your hand and your question will be answered privately.

E.2.2 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 E9.

	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 E9: 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 E10.

If Ball is Red:

- Message: "*The Ball is Red.*"
- No Message.

If Ball is Blue:

- Message: "*The Ball is Blue.*"
- No Message.

Table E10: 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.

E.2.3 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.

E.2.4 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.

E.2.5 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.

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)

E.2.6 How is a message generated?

E.2.7 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.

E.2.8 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.