

**CREDIBILITY IN PERSUASION:
A LABORATORY EXPERIMENT**

Yuki Shiomi
Nobuyuki Hanaki

October 2026

The Institute of Social and Economic Research
The University of Osaka
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Credibility in Persuasion: A Laboratory Experiment^{*}

Yuki Shiomi[†]

Nobuyuki Hanaki[‡]

September 29, 2026

Abstract

We experimentally examine whether partial commitment to the message distribution can substitute for full commitment to the information structure in Bayesian persuasion. Theory predicts that when the players' marginal incentives are aligned, an equilibrium with the same level of information transmission as under full commitment can be sustained even under partial commitment, whereas when marginal incentives are misaligned, no information can be credibly transmitted. Our central finding is that, even when marginal incentives are aligned, information transmission is substantially lower under partial commitment than under full commitment: partial commitment does not fully substitute for full commitment in practice. When marginal incentives are misaligned, information transmission declines further under partial commitment, as theory predicts. We also find that misaligned marginal incentives reduce information transmission even under full commitment, where the standard theory predicts no such effect. Our analysis of sender behavior shows that, in all of these cases, senders transmit less information than the theoretical benchmark implies, suggesting that sender undercommunication plays an important role in the lower information transmission we observe.

JEL Classification: C92, D82, D83

Keywords: Bayesian persuasion, commitment, laboratory experiment

^{*}We thank participants at the 2nd East Asia Experimental and Behavioral Economics Workshop, the 19th Annual Meeting of the Association of Behavioral Economics and Finance, the Asia-Pacific Meeting of the Economic Science Association 2026, and the Summer Workshop on Economic Theory 2026 for their helpful comments and discussions. The experiment reported in this paper was approved by the IRB of the Institute of Social and Economic Research, the University of Osaka (No. 20250701), and pre-registered at AsPredicted (#257058). We are grateful for the financial support from the Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (No. 23H00055 and 25H00388). The authors used LLMs (OpenAI ChatGPT and Anthropic Claude) to proofread and polish the manuscript.

[†]Graduate School of Economics, the University of Osaka. E-mail: u282915h@ecs.osaka-u.ac.jp

[‡]Institute of Social and Economic Research, the University of Osaka, and University of Limassol. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

1 Introduction

The persuasion problem, in which an information sender provides information to a receiver in order to induce the receiver to take an action desirable to the sender, is commonly observed in many real-world situations. For example, schools disclose information about their students' abilities through grade transcripts to potential employers in order to encourage them to hire their students. Similarly, used-car sellers disclose information about the condition of their vehicles through inspection reports to potential buyers in order to promote sales.

Bayesian persuasion, formalized by [Kamenica and Gentzkow \(2011\)](#), is now a standard framework for analyzing such persuasion problems. The central feature of this model is that the sender can design an information structure and fully commit to it. In the school example, the information structure corresponds to the school's grading policy, specifying, for example, what fraction of high-ability students receive high grades and what fraction of others receive low grades. The assumption that the sender can commit to the information structure implies that the school can explicitly announce and adhere to this grading policy. Under this assumption, the Bayesian persuasion literature has shown that the sender can optimally design the information structure to influence the receiver's beliefs and thereby induce behavior that would not arise without such information (see, e.g., [Bergemann and Morris, 2019](#); [Kamenica, 2019](#)).

However, in practice, the assumption of full commitment is often violated due to limitations in the *observability* and *verifiability* of the information structure. For example, a school's grading policy is typically not directly observable (observability constraint), and even when the grading policy is publicly announced, employers cannot verify whether students' grades are actually assigned according to the announced policy (verifiability constraint).

In contrast, as pointed out by [Lin and Liu \(2024\)](#), what is often announced in practice is not the grading policy itself but the grade distribution that the policy generates. Indeed, many universities make the distribution of students' grades publicly available. This

observation has important implications for both observability and verifiability. First, even if the grading policy itself is unobservable, the resulting grade distribution may still be observable to employers. Second, even if the grading policy is observable but not verifiable, the school cannot secretly deviate to another policy that changes the grade distribution, because such a deviation would be detected through changes in the observed grade distribution. Taken together, these considerations suggest that, in many real-world situations, the sender’s commitment is effectively limited not to the information structure itself but to the message distribution induced by that structure.

A natural question that arises is whether partial commitment to the induced message distribution can substitute for full commitment to the information structure. [Lin and Liu \(2024\)](#) provide a theoretical answer to this question. They study Bayesian persuasion under such partial commitment, and define an information structure as credible if no profitable deviation exists to another information structure that induces the same message distribution. They then show that the alignment of marginal incentives between the sender and the receiver is key. When marginal incentives are aligned, the optimal information structure under full commitment remains credible under some additional conditions, and partial commitment can therefore substitute for full commitment. By contrast, when marginal incentives are misaligned, no meaningful information can be credibly transmitted, and no such substitution is possible.

This result can serve as a guide for applying the Bayesian persuasion framework to real-world environments.¹ Whether the substitution works in practice, however, is an empirical question. We therefore conducted a laboratory experiment examining how limited commitment and the alignment of marginal incentives affect information transmission in Bayesian persuasion. Specifically, we exogenously varied three commitment structures (Full commitment, Partial commitment, and Announced Partial (A-Partial) commitment) and two incentive structures (marginally aligned and marginally misaligned), yielding six

¹[Xiang \(2026\)](#) structurally estimates a Bayesian persuasion model of information transmission in the physician-patient relationship. In this setting, where a physician’s full commitment to a treatment recommendation policy is not necessarily guaranteed, [Xiang \(2026\)](#), citing [Lin and Liu \(2024\)](#), points to the alignment of marginal incentives between physicians and patients and patients’ ability to observe the distribution of recommendations as grounds for applying the framework.

experimental treatments.

In our experiment, Full commitment corresponds to the standard Bayesian persuasion setting in which the sender can fully commit to the information structure. By contrast, under partial commitment, the sender’s commitment is limited not to the information structure itself but only to the message distribution induced by that structure. Accordingly, under our Partial commitment, the sender cannot directly disclose the information structure, and the receiver observes only the resulting message distribution. This setting reflects an observability constraint on information structures (i.e., the information structure itself cannot be observed). Furthermore, we implemented a distinct version of partial commitment in the laboratory, referred to as A-Partial commitment. Under A-Partial commitment, the sender publicly announces the designed information structure but may deviate to another structure as long as the induced message distribution remains unchanged. The receiver understands this possibility but cannot observe the post-deviation structure. This setting reflects a verifiability constraint (i.e., although the information structure can be observed, the receiver cannot verify whether it has actually been implemented).

In our simple experimental environment, the predictions differ sharply across treatments. Under Full commitment, a partially informative equilibrium exemplified by the prosecutor-judge example in [Kamenica and Gentzkow \(2011\)](#) (hereafter, the KG equilibrium) can arise regardless of the incentive structure. Under Partial and A-Partial commitment, by contrast, whether the KG equilibrium can be sustained depends on the alignment of marginal incentives ([Lin and Liu, 2024](#)). When marginal incentives are aligned, the KG equilibrium remains implementable; when they are misaligned, a babbling equilibrium arises in which no information is transmitted.

Our central finding is that partial commitment is not a full substitute for full commitment. Even when marginal incentives are aligned, information transmission between players is significantly lower under Partial and A-Partial commitment than under Full commitment. That is, although theory predicts that the same level of information transmission is attainable regardless of the commitment structure when marginal incentives are

aligned, that level is not attained in practice. When marginal incentives are misaligned, by contrast, we observe a pattern consistent with the babbling equilibrium that theory predicts: information transmission under Partial and A-Partial commitment is lower than under Full commitment, and it also falls below the level observed when marginal incentives are aligned. We further find that information transmission declines when marginal incentives change from aligned to misaligned even under Full commitment—an environment in which the standard framework of [Kamenica and Gentzkow \(2011\)](#) predicts the same level of transmission regardless of the incentive structure.

To better understand subjects’ behavior, we analyze the decisions of senders and receivers separately. Our analysis of sender behavior reveals that, among the treatments in which the KG equilibrium is theoretically implementable, senders transmit on average less information than the equilibrium implies in all but Full commitment with aligned marginal incentives. We refer to this phenomenon as undercommunication. A closer look at senders’ strategies shows that strategies are highly dispersed in these treatments: while many highly informative information structures are observed, so too are strategies with low informativeness and strategies that correlate messages negatively with the state. This heterogeneity in strategies contributes to the undercommunication we observe.

Our analysis of receiver behavior shows that receivers’ decisions are broadly consistent with Bayesian decision-making under Full commitment, but not under Partial and A-Partial commitment. Receivers in these environments cannot observe the information structure the sender actually used and must infer it from the message distribution alone. To examine what structure receivers infer, we elicited their beliefs about the sender’s information structure in an incentivized manner. These beliefs reveal that receivers in these environments respond not to the information structure the sender actually used, but to the one they believe the sender used.

Finally, we explore several factors that may account for senders’ undercommunication. We first show that it cannot be explained as an empirical best response to the actual behavior of receivers in the experiment. We then discuss the possibility that the undercommunication observed under full commitment and that observed under partial

commitment arise from different sources. We also discuss why lying aversion and social preferences are unlikely to account for it.

Taken together, our results show that limited commitment can reduce information transmission below the theoretical benchmark even when marginal incentives are aligned. Our findings suggest that raising the observability or verifiability of information structures, and thereby strengthening the sender’s ability to commit, can increase information transmission even in such environments.

1.1 Related Literature

Our study is related, first, to the theoretical literature on Bayesian persuasion that relaxes the assumption of full commitment. Since [Kamenica and Gentzkow \(2011\)](#), the framework of Bayesian persuasion, which assumes that the sender can fully commit to an information structure, has been widely adopted and has continued to develop theoretically ([Bergemann and Morris, 2019](#); [Kamenica, 2019](#)). At the same time, a growing theoretical literature has relaxed this assumption by introducing various forms of partial commitment. For example, [Min \(2021\)](#) and [Lipnowski et al. \(2022\)](#) analyze settings in which the sender, after publicly announcing an information structure, may deviate to another structure with some exogenously given probability. By contrast, [Lin and Liu \(2024\)](#) formalize a different form of partial commitment in which the sender cannot commit to the information structure itself but can commit to the message distribution it induces, and characterize the conditions under which partial commitment can substitute for full commitment.²

Second, our study contributes to the rapidly growing experimental literature on Bayesian persuasion ([Nguyen, 2017](#); [Aristidou et al., 2019](#); [Fr  chette et al., 2022](#); [Wu and Ye, 2023](#); [Zhou, 2023](#); [Rentschler and Samad, 2025](#)). Among these studies, [Fr  chette et al. \(2022\)](#) and [Zhou \(2023\)](#) explicitly examine the role of the sender’s commitment. [Fr  chette et al.](#)

²Other approaches to relaxing the commitment assumption include introducing costly communication ([Nguyen and Tan, 2021](#)) and studying informed senders without ex ante commitment ([Koessler and Skreta, 2023](#)). Other studies provide microfoundations for commitment through repeated interactions and reputation ([Best and Quigley, 2024](#)) or organizational contracts ([Deb et al., 2026](#)).

(2022) conduct experiments that provide a unified comparison of disclosure, cheap talk, and Bayesian persuasion by varying the strength of commitment and whether information is hard or soft. Zhou (2023) experimentally analyzes the interaction between the sender’s partial commitment and the receiver’s verification costs. In both studies, however, partial commitment is modeled in a way close to Min (2021) and Lipnowski et al. (2022), whereby the sender can deviate from a publicly announced information structure with some exogenously given probability. By contrast, following Lin and Liu (2024), we experimentally study an environment in which the sender’s commitment is limited to the induced message distribution and examine the extent to which such partial commitment can substitute for full commitment in actual behavior.

Third, our study is related to the broader experimental literature on communication games that examines the relationship between incentive alignment and information transmission. Cai and Wang (2006) experimentally test the cheap talk model of Crawford and Sobel (1982) by manipulating the degree of preference alignment between the sender and the receiver and comparing observed information transmission with theoretical predictions. We adopt a similar experimental approach by exogenously varying the alignment of marginal incentives between the sender and the receiver and comparing information transmission with theoretical predictions. Our object of analysis, however, is Bayesian persuasion.

Finally, our study is related to the experimental literature documenting systematic deviations of information transmission from theoretical predictions in communication games. For example, in cheap talk experiments, senders tend to exhibit overcommunication, transmitting more information than predicted by theory (e.g., Cai and Wang, 2006), whereas in information disclosure experiments, senders tend to exhibit undercommunication, disclosing less than the full revelation predicted by theory (e.g., Jin et al., 2021). Recent Bayesian persuasion experiments, such as Fréchette et al. (2022) and Rentschler and Samad (2025), also document undercommunication under full commitment. We likewise observe systematic undercommunication in Bayesian persuasion, but unlike these previous studies, we study an environment in which the sender’s payoff is state-dependent. We

further show that the amount of information transmitted by senders depends on both the commitment structure and the alignment of marginal incentives.

2 Theoretical Framework

2.1 Setup

There are two players, a sender (S ; she) and a receiver (R ; he), and two possible states of the world, Red and Blue, denoted by $\theta \in \Theta = \{\text{Red}, \text{Blue}\}$. Neither player directly observes the realized state $\theta \in \Theta$, but they share a common prior $\mu_0 = \Pr(\text{Red}) < 1/2$.

The receiver chooses an action $a \in A = \{\text{red}, \text{blue}\}$. The payoffs of both players depend on the realized state $\theta \in \Theta$ and the receiver's action $a \in A$. Specifically, the receiver's payoff is given by

$$u_R(\theta, a) = \begin{cases} \bar{v} & \text{if } (\theta, a) \in \{(\text{Red}, \text{red}), (\text{Blue}, \text{blue})\}, \\ \underline{v} & \text{otherwise,} \end{cases}$$

where $0 < \underline{v} < \bar{v}$. Hence, the receiver prefers to match his action with the true state. The sender's payoff is defined as

$$u_S(\theta, a) = \begin{cases} \bar{v} + \varepsilon & \text{if } (\theta, a) = (\text{Red}, \text{red}), \\ \bar{v} - \varepsilon & \text{if } (\theta, a) = (\text{Blue}, \text{red}), \\ \underline{v} & \text{if } a = \text{blue}, \end{cases}$$

where $\varepsilon \in (\underline{v} - \bar{v}, \bar{v} - \underline{v})$. As discussed below, the parameter ε determines the sender's marginal incentives across states. This payoff structure is also illustrated in Table 1.

In this setting, the receiver's optimal behavior is characterized by a threshold strategy:

$$a(\mu) = \begin{cases} \text{red} & \text{if } \mu \geq \frac{1}{2}, \\ \text{blue} & \text{if } \mu < \frac{1}{2}, \end{cases}$$

Table 1: Payoff table

$\Theta \backslash A$	<i>red</i>	<i>blue</i>
Red	$\bar{v} + \varepsilon, \bar{v}$	$\underline{v}, \underline{v}$
Blue	$\bar{v} - \varepsilon, \underline{v}$	$\underline{v}, \bar{v}$

Notes: In each cell, the left number represents the sender's payoff and the right number represents the receiver's payoff.

where μ denotes the receiver's belief that the state is Red. Given that $\mu_0 < 1/2$, the receiver would, under the prior, always choose the action *blue*. Hence, the sender must employ some form of information transmission to persuade the receiver.

We assume that the sender transmits information to the receiver by designing an information structure $\pi : \Theta \rightarrow \Delta\mathcal{M}$. Here, $\mathcal{M} = \{r, b\}$ denotes the set of messages available to the sender, where r and b denote messages labeled Red and Blue, respectively.

2.2 Theoretical Predictions

Full commitment. As a benchmark, consider the situation in which, as in the standard Bayesian persuasion framework, the sender can fully commit to the information structure she designs. That is, in this case, the receiver observes both the information structure designed by the sender and the realized message, namely (π, m) .

In this setting, our framework essentially reproduces the prosecutor-judge example in [Kamenica and Gentzkow \(2011\)](#). Accordingly, for any $\varepsilon \in (\underline{v} - \bar{v}, \bar{v} - \underline{v})$, the sender's optimal information structure in equilibrium is given by

$$\pi(r|\text{Red}) = 1, \quad \pi(b|\text{Blue}) = 1 - \frac{\mu_0}{1 - \mu_0}.$$

The receiver's optimal strategy and posterior beliefs are as follows:

$$a(m) = \begin{cases} \text{red} & \text{if } m = r, \\ \text{blue} & \text{if } m = b, \end{cases} \quad \mu(m) = \begin{cases} \frac{1}{2} & \text{if } m = r, \\ 0 & \text{if } m = b. \end{cases}$$

That is, the sender commits to truthfully sending message r when the state is Red.

When the state is Blue, she commits to mixing between messages r and b . In this case, even though the receiver recognizes that the sender’s message r may be false with positive probability, it is still optimal for him to follow the sender’s message. In what follows, we refer to this equilibrium as the *KG equilibrium*.

Partial commitment. Following [Lin and Liu \(2024\)](#), we define partial commitment as a situation in which the sender cannot commit to the information structure π itself, but only to the message distribution it induces. Let $\mathbb{P}_\pi(m) := \sum_\theta \pi(m|\theta)\mu_0(\theta)$ denote the message distribution induced by π . Under partial commitment, the receiver observes $(\mathbb{P}_\pi, m)$ rather than (π, m) .

Under this form of partial commitment, [Lin and Liu \(2024\)](#) define *credibility* as follows: an information structure is *credible* if there is no profitable deviation to another information structure that induces the same message distribution. They show that whether the optimal information structure derived from the KG equilibrium survives this credibility requirement crucially depends on the alignment of marginal incentives between the sender and the receiver.³

In our setting, this alignment is captured by the parameter ε . Specifically, ε determines how the sender’s marginal payoff from inducing the receiver to switch from *blue* to *red* varies across states. When $\varepsilon \geq 0$, this marginal payoff is (weakly) higher in state Red than in state Blue.⁴ This is in the same direction as the receiver’s incentive, for whom *red* becomes relatively more attractive as the state changes from Blue to Red, and hence the players’ marginal incentives are aligned. [Lin and Liu \(2024\)](#) show that under such alignment, the optimal information structure in the KG equilibrium remains credible under partial commitment and can therefore be sustained as an equilibrium.⁵

In contrast, when $\varepsilon < 0$, the sender’s marginal payoff from inducing the receiver to

³The marginal incentives characterize the modularity of each player’s payoff function. When the marginal incentives are aligned, the sender’s and the receiver’s payoff functions exhibit the same sign of modularity.

⁴Strictly speaking, when $\varepsilon = 0$ the sender has state-independent preferences. However, in our simple environment, the equilibrium implications under $\varepsilon = 0$ coincide with those under $\varepsilon > 0$. For expositional convenience, we therefore treat the two cases jointly as $\varepsilon \geq 0$. For a detailed analysis of the case $\varepsilon = 0$, see Observation 1 in [Lin and Liu \(2024\)](#).

⁵This result holds in our simple environment, where $|A| = 2$ and u_S is strictly supermodular when $\varepsilon > 0$. See Proposition 4 in [Lin and Liu \(2024\)](#) and its proof for details.

switch from *blue* to *red* is higher in state Blue than in state Red. This runs in the opposite direction to the receiver’s incentive, so the players’ marginal incentives are misaligned. Under such misalignment, no informative outcome is credible.⁶ As a result, the only credible persuasion outcome is the babbling equilibrium, in which the receiver ignores the sender’s message and simply chooses blue based on the prior.

Intuitive explanation. The intuition under partial commitment is as follows. Suppose that the receiver always follows the sender’s message as in the KG equilibrium. Consider first the case in which $\varepsilon < 0$. Since $u_S(\text{Blue}, \text{red}) > u_S(\text{Red}, \text{red})$, the sender has an incentive to deviate by increasing the probability of sending message r when the state is Blue. Importantly, such a deviation can be made without altering the induced message distribution. Therefore, any information structure that conveys meaningful information about the state is not credible, and the equilibrium collapses into a babbling equilibrium.

In contrast, when $\varepsilon \geq 0$, making the same type of deviation does not yield a profitable gain for the sender. Hence, the optimal information structure under full commitment remains credible and can be sustained as an equilibrium under partial commitment.

3 Experimental Design

3.1 Basic Environment and Incentive Structure

We first describe the basic environment of the game implemented in the experiment. A box contains 10 balls in total: 3 Red balls and 7 Blue balls. The computer randomly draws one ball from the box. While neither player observes the realized color of the drawn ball, both players know the composition of the box. Thus, the realized color of the ball represents the state of the world ($\theta \in \{\text{Red}, \text{Blue}\}$) with a common prior $\mu_0 = 3/10$.

The receiver chooses an action $a \in \{\text{red}, \text{blue}\}$, which corresponds to guessing the color of the ball. If the guess is correct, the receiver earns 150 points; if incorrect, the receiver earns 80 points. The sender’s payoff is determined as follows. If the receiver chooses *red*,

⁶See Proposition 2 in [Lin and Liu \(2024\)](#) and its proof for details.

Table 2: Experimental payoff tables

(a) ALG			(b) MIS		
$\Theta \backslash A$	<i>red</i>	<i>blue</i>	$\Theta \backslash A$	<i>red</i>	<i>blue</i>
Red	200, 150	80, 80	Red	100, 150	80, 80
Blue	100, 80	80, 150	Blue	200, 80	80, 150

the sender receives $150 + \varepsilon$ points when the state is Red, and $150 - \varepsilon$ points when the state is Blue. If the receiver chooses *blue*, the sender receives 80 points regardless of the state (i.e., $\bar{v} = 150, v = 80$).

In the experiment, the parameter ε takes two values, $\varepsilon \in \{50, -50\}$. When $\varepsilon = 50$, the sender’s and receiver’s marginal incentives are aligned (ALG); when $\varepsilon = -50$, they are misaligned (MIS). The corresponding payoff tables are presented in Table 2.

3.2 Commitment Structure

In addition to varying the incentive structure, we manipulate the sender’s ability to commit to an information structure. We consider three types of commitment, which differ in whether and how the receiver can observe the sender’s information structure.

3.2.1 Full Commitment

We first describe Full commitment. Under Full commitment, the game proceeds in two stages:

Stage 1 (Design of the Information Structure). The sender designs an information structure π without observing the realized state. The information structure $\pi(m|\theta)$ specifies, for each state $\theta \in \{\text{Red}, \text{Blue}\}$, the probability of sending each message $m \in \{r, b\}$. Following Fr chet te et al. (2022), we implement the design of π using a slider-based interface. Figure 1 shows the decision screen used by the sender. The two horizontal sliders at the top of the screen correspond to the conditional probabilities for the two states. By adjusting the positions of these sliders, the sender can freely specify the information structure in increments of one percent.

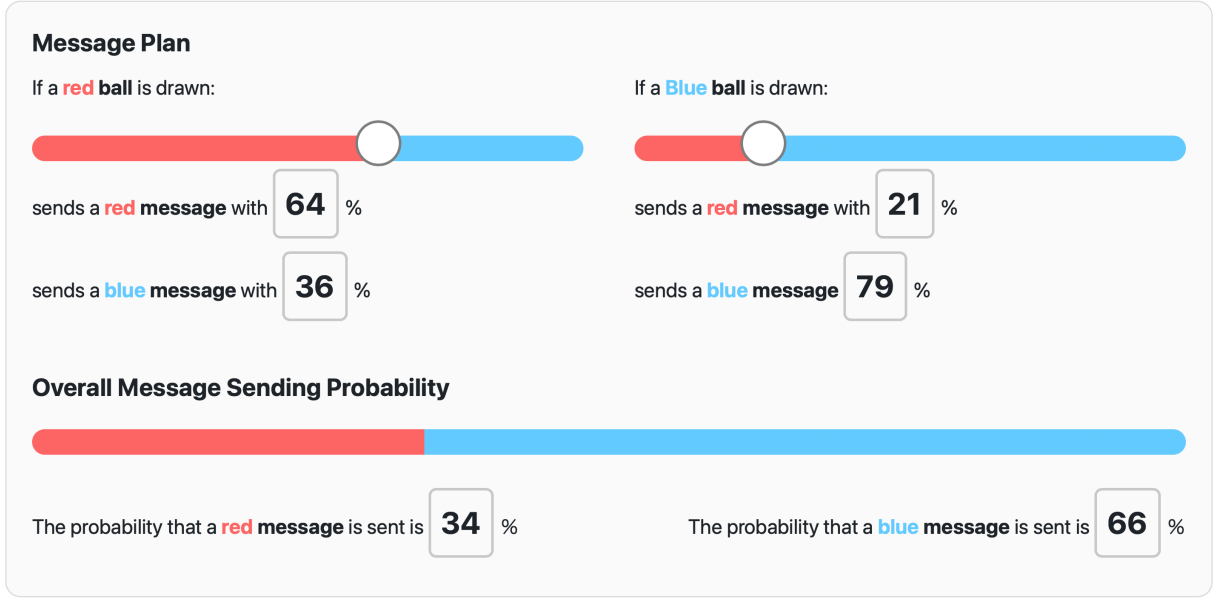


Figure 1: Sender screen: Design of the information structure

The bottom slider displays the message distribution implied by the information structure chosen through the top sliders.⁷ This message distribution is computed automatically by the computer, and the sender cannot directly manipulate it. Below each slider, the corresponding probability values are displayed and updated in real time as the sender moves the sliders. This interface allows the sender to design π directly while simultaneously observing the resulting distribution of messages.

Stage 2 (Receiver’s Guess). The receiver observes the information structure π designed in Stage 1 and the induced message distribution, and then forms a guess about the ball color. The decision is elicited using the strategy method: the receiver states his action for each possible message, i.e., which color he would guess if the message were r , and which color he would guess if the message were b (see Figure 2).

3.2.2 Partial Commitment

We now describe Partial commitment. As in Full commitment, the sender designs an information structure π in Stage 1. However, under Partial commitment, the receiver

⁷During the experiment, we referred to the information structure as the “message plan,” and to the message distribution as the “overall message sending probability.”

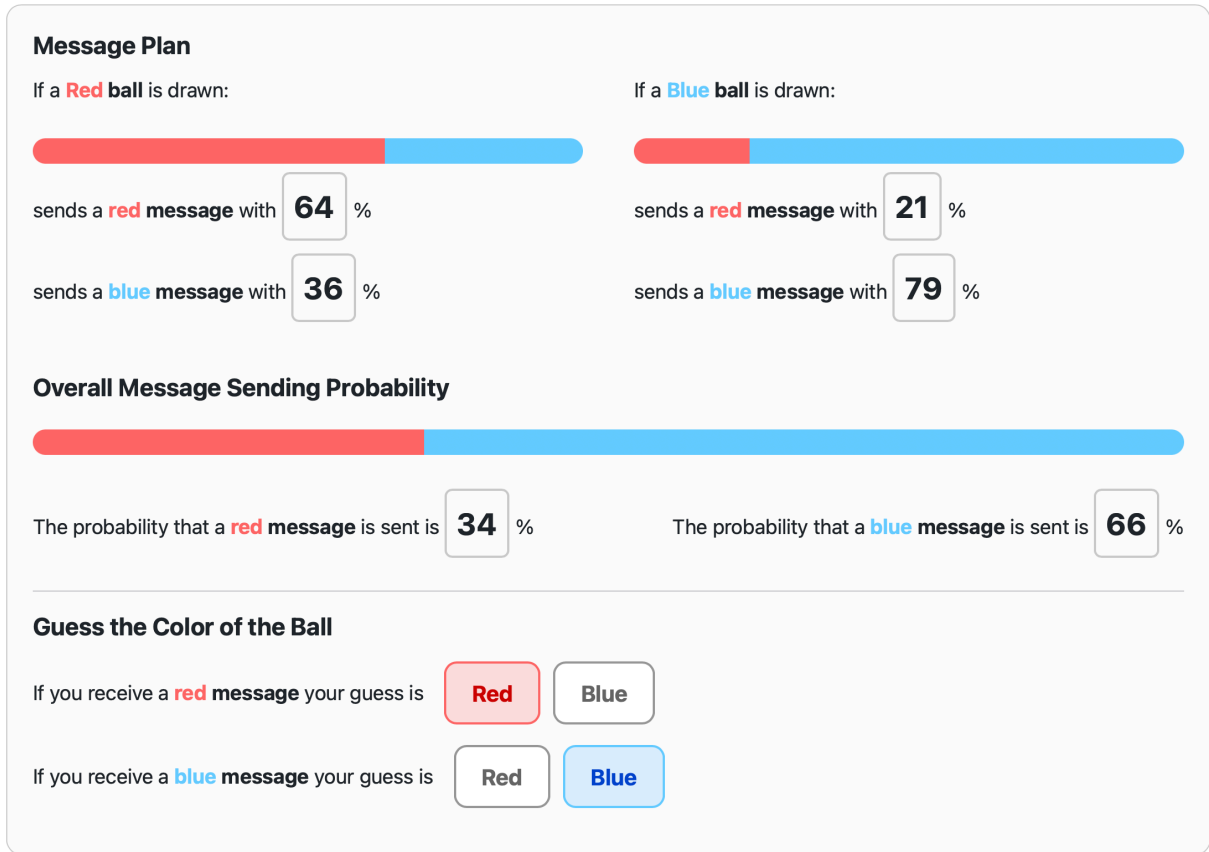


Figure 2: Receiver screen: Guess under Full and A-Partial commitment

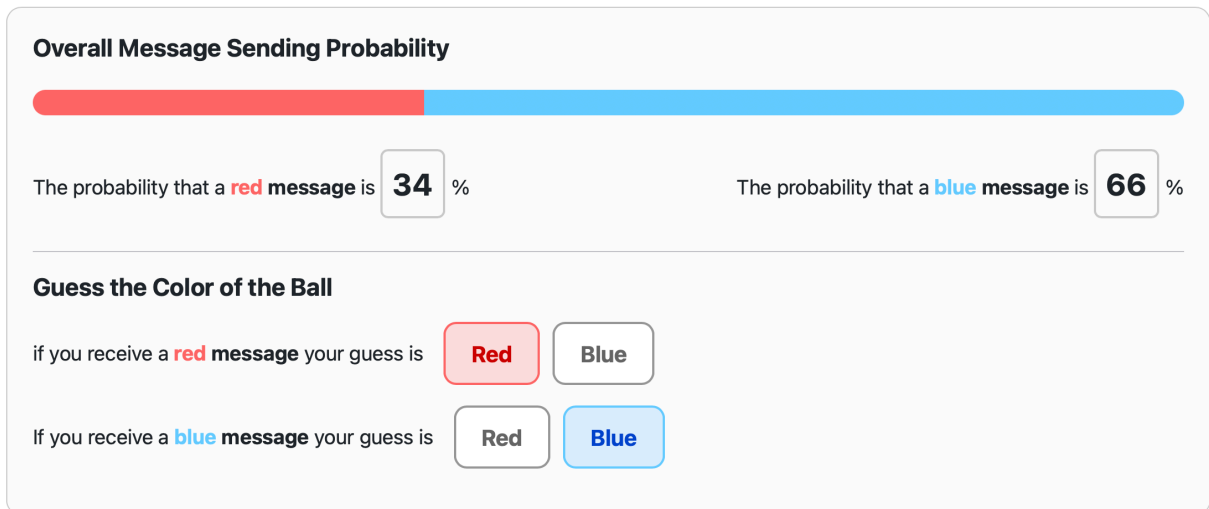


Figure 3: Receiver screen: Guess under Partial commitment

cannot observe the information structure directly. Instead, the receiver only observes the message distribution induced by the information structure. Therefore, the only difference from Full commitment is that in Stage 2 the receiver is not shown the information structure itself (i.e., the top two sliders). The decision screen used by the receiver under Partial

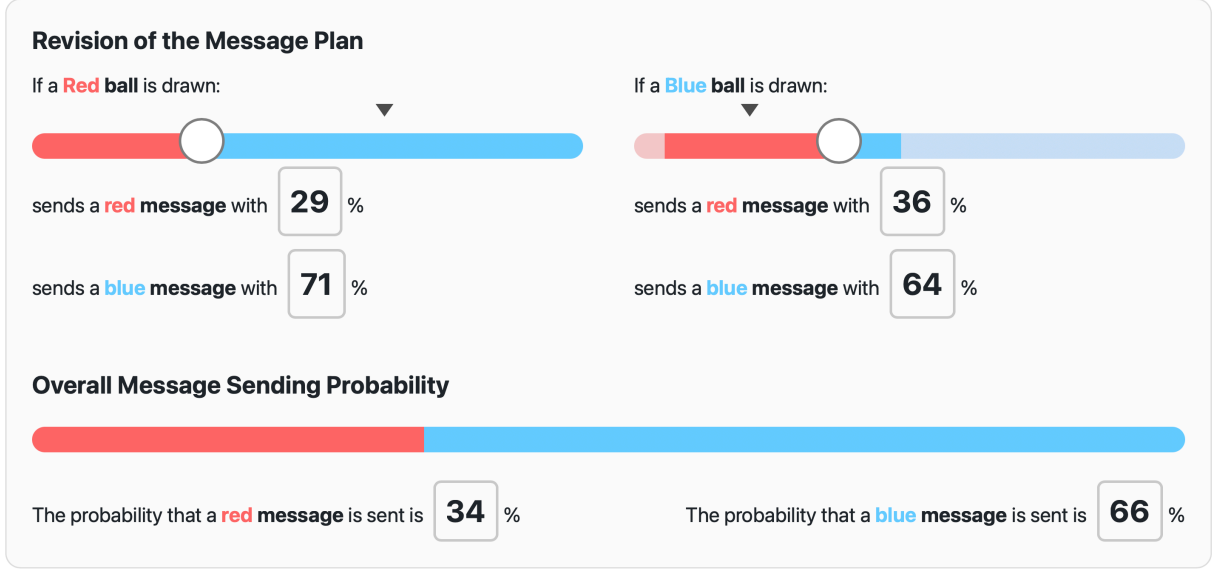


Figure 4: Sender screen: Revision of the information structure

commitment is shown in Figure 3. All other procedures are identical to those under Full commitment.

3.2.3 Announced Partial Commitment

In addition, we introduce another version of Partial commitment, which we call Announced Partial (A-Partial) commitment. In this condition, the game proceeds in three stages:

Stage 1 (Announcement of Information Structure). The sender designs an information structure $\tilde{\pi}$, which is publicly disclosed to the receiver. The interface for this stage is identical to that shown in Figure 1.

Stage 2 (Revision of the Information Structure). The sender can revise $\tilde{\pi}$ to another information structure π , provided that π induces the same message distribution as $\tilde{\pi}$. The revised information structure π is used to generate messages.

Figure 4 presents the decision screen used by the sender in this stage. The initial positions of the sliders correspond to the information structure $\tilde{\pi}$ designed in Stage 1 and are indicated by a marker on each slider. The sender may adjust the sliders to revise $\tilde{\pi}$, but only to the extent that the message distribution shown at the bottom of the screen remains

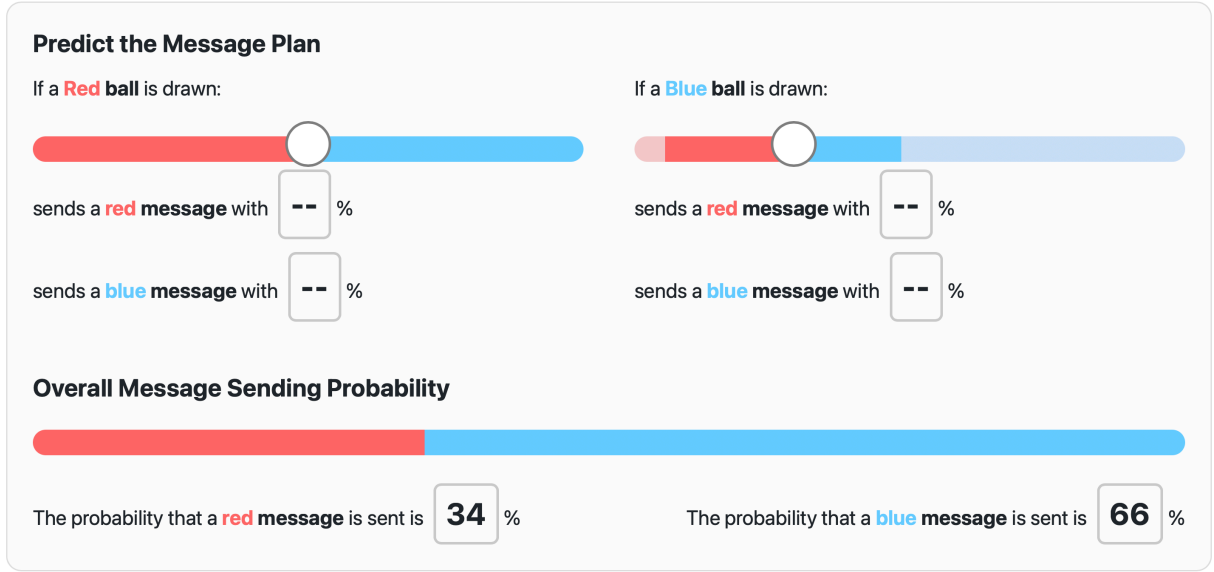


Figure 5: Receiver screen: Belief elicitation

unchanged. Accordingly, when the sender moves one of the two sliders, the other slider is automatically adjusted by the computer to preserve the message distribution. In addition, each slider is restricted to a feasible range, and regions outside this range are displayed in a lighter color, indicating that the sender cannot move the slider into those areas. This implementation allows subjects to intuitively understand the relationship between the information structure and the induced message distribution, thereby reducing the cognitive burden associated with the task.

Stage 3 (Receiver’s Guess). The receiver observes the information structure $\tilde{\pi}$ announced in Stage 1, along with the message distribution. The receiver understands that the sender may have revised $\tilde{\pi}$ in Stage 2, but the message distribution remains unchanged. The receiver then reports his guesses for both possible messages r and b . The interface used in this stage is the same as that shown in Figure 2, with the only difference being that the information structure $\tilde{\pi}$ is displayed above the message distribution.

3.3 Belief Elicitation

Under Partial and A-Partial commitment, the receiver cannot directly observe the information structure π actually used to generate the messages. Receivers in these treatments

must therefore infer the sender’s information structure from the observable message distribution, and this inference may play an important role in their decision making. We therefore elicit receivers’ beliefs about the information structure in an incentivized manner.

The belief elicitation takes place immediately after the receiver reports his guesses about the color of the ball. Figure 5 displays the decision screen used in this stage. The interface is nearly identical to the sender’s revision screen in A-Partial commitment. The receiver is presented with the message distribution and selects, from among the information structures consistent with this distribution, the one he believes the sender actually used; we denote this reported information structure by $\hat{\pi}$.

The receiver earns a higher payoff the closer $\hat{\pi}$ is to the true information structure employed by the sender.⁸ However, to avoid hedging by the receiver, only one of the two rewards—either the reward based on the receiver’s action (color guess) or the reward based on the reported belief—is randomly selected for the final payment.

3.4 Experimental Procedures

We implemented a total of six treatments in a between-subjects design, combining two incentive structures (ALG and MIS) with three commitment structures (Full, Partial, and A-Partial). A total of 276 subjects participated in the experiment, with 42 to 48 subjects in each treatment. Each treatment was conducted in two sessions. The experiment was carried out at the Institute of Social and Economic Research, The University of Osaka, between November 2025 and March 2026.⁹

Each session consisted of the sender-receiver game described above, followed by a questionnaire on preferences and cognitive ability. Before the start of the game, par-

⁸Formally, the receiver’s payoff from the belief-elicitation task is given by

$$\max \left\{ 80, 150 - 1120 \left(\frac{\pi(r|\text{Red}) - \hat{\pi}(r|\text{Red})}{\Delta} \right)^2 \right\},$$

where Δ is the maximal admissible range over which $\hat{\pi}(r|\text{Red})$ can vary.

⁹The experiment was programmed using oTree (Chen et al., 2016), and subjects were recruited from ISER’s subject pool through ORSEE (Greiner, 2015).

ticipants received written instructions, the contents of which were read aloud through a computer-generated voice.¹⁰ Participants then completed a comprehension quiz, followed by two unpaid practice rounds, during which all subjects experienced both the sender and receiver roles.

After the practice rounds, participants were randomly assigned to either the sender or the receiver role, and this assignment remained fixed throughout the experiment. The main game consisted of 20 rounds. In each round, participants were randomly rematched into sender-receiver pairs. At the end of each round, participants received feedback showing the realized state (ball color), the sender’s message, the receiver’s action, and the resulting payoffs for both players.¹¹

After completing the game, participants answered a questionnaire measuring social preferences, risk preferences, and cognitive ability. Social preferences were measured using the task of [Kerschbamer \(2015\)](#), risk preferences using the task of [Holt and Laury \(2002\)](#), and cognitive ability using the Berlin Numeracy Test ([Cokely et al., 2012](#)).

Participants received a participation fee of 500 Japanese yen (JPY). After the completion of the experiment, one of the 20 rounds of the sender-receiver game was randomly selected, and the earnings from that round were converted at a rate of 1 point = 10 JPY. In addition, participants received payments based on randomly selected questions from the questionnaire. The average payment was 2547 JPY, about 16 USD at the time.

3.5 Equilibrium Predictions

Table 3 summarizes the labels of the six treatments and their corresponding equilibrium predictions. As shown in Section 2.2, under Full commitment the KG equilibrium is predicted to arise regardless of whether the players’ marginal incentives are aligned or misaligned (AF and MF).

Under Partial and A-Partial commitment, by contrast, multiple equilibria may gen-

¹⁰English translations of the experimental materials are provided in Online Appendix E.

¹¹Before being assigned to the sender or receiver role, all participants received the same instructions, completed the comprehension quiz, and were required to experience both roles in the practice rounds. These procedures were intended to ensure that all aspects of the game structure described above were common knowledge among participants.

Table 3: Treatment labels and equilibrium predictions

		Commitment		
		Full	Partial	A-Partial
ALG	Labels	AF	AP	AP-A
	Predictions	KG	KG	KG
MIS	Labels	MF	MP	MP-A
	Predictions	KG	Babbling	Babbling

erally exist. Since our primary interest is whether these forms of partial commitment can substitute for full commitment, we focus on whether the KG equilibrium can still be sustained in these environments. The key determinant is whether the players' marginal incentives are aligned. When incentives are aligned, the KG equilibrium remains sustainable (AP and AP-A). When incentives are misaligned, however, the KG equilibrium is no longer sustainable, and communication collapses to a babbling equilibrium (MP and MP-A).

Finally, note that A-Partial commitment differs from Partial commitment in that the sender publicly announces an information structure $\tilde{\pi}$ before communication takes place. This announcement is non-binding, as the sender can subsequently revise $\tilde{\pi}$ while preserving the induced message distribution. Thus, in both environments, the sender can effectively commit only to the induced message distribution, and the two commitment structures generate identical equilibrium predictions.¹²

4 Results

In this section, we examine the main predictions of the theoretical framework. We begin in Section 4.1 by quantifying the extent of information transmission achieved through com-

¹²Although Partial and A-Partial commitment yield identical equilibrium predictions, we implement both treatments for two reasons. First, they capture distinct constraints that arise in real-world information environments: Partial commitment captures an observability constraint, whereas A-Partial commitment captures a verifiability constraint. Second, and more importantly, comparing the two allows us to examine how the presence of a non-binding publicly announced information structure $\tilde{\pi}$ affects behavior. See Online Appendix A for related results.

munication between senders and receivers and comparing the observed outcomes with the equilibrium predictions. We then examine subjects’ behavior in detail to understand what underlies these results, analyzing sender behavior in Section 4.2 and receiver behavior in Section 4.3.

4.1 Overall Informativeness

In this section, we directly test the equilibrium predictions reported in Table 3. To do so, we quantify the extent to which information about the state is transmitted through communication between senders and receivers. Specifically, we use a measure of *overall informativeness* defined as $\phi(\theta, a) := \text{Corr}(\theta, a; \pi)$. This measure captures the correlation between the state and the receiver’s action. It takes the value $\phi(\theta, a) = 1$ when the sender fully reveals the state and the receiver perfectly follows the sender’s message. By contrast, $\phi(\theta, a) = 0$ under a babbling equilibrium in which the receiver completely ignores the message and chooses an action solely based on the prior. In our experimental environment, the KG equilibrium predicts $\phi(\theta, a) = 0.54$.¹³ Measuring informativeness using the correlation between the state and the receiver’s action is standard in the experimental literature on communication games (Cai and Wang, 2006; Fr chet te et al., 2022).

Table 4 reports the average level of overall informativeness in each treatment together with the corresponding theoretical values. Rather than assessing whether the observed values exactly match the theoretical values, we examine whether the relative differences across treatments predicted by the theory are supported by the data.

Our first prediction concerns the role of commitment when marginal incentives are aligned. In this case, the KG equilibrium can be sustained without full commitment, yielding the same benchmark level of overall informativeness in AF, AP, and AP-A. However, the data do not support this prediction. Overall informativeness falls sharply from 0.43 in AF to 0.13 in AP and 0.21 in AP-A, with both differences being statistically sig-

¹³Because the sliders are in increments of one percent, the largest value of $\pi(r|\text{Blue})$ for which $\mu(r) \geq 1/2$ is 0.42 rather than $3/7$. The theoretical value of 0.54 is computed under this implementable information structure; with $\pi(r|\text{Blue}) = 3/7$ it would be 0.53.

Table 4: Overall informativeness by treatment

		Commitment		
		Full	Partial	A-Partial
ALG	Data	0.43 [0.34, 0.53]	0.13 [0.07, 0.20]	0.21 [0.10, 0.31]
	Theory	0.54	0.54	0.54
MIS	Data	0.24 [0.17, 0.31]	0.04 [0.00, 0.07]	−0.02 [−0.06, 0.02]
	Theory	0.54	0	0

Notes: The Data rows report the mean of $\phi(\theta, a)$ and, in brackets, its 95% confidence interval, based on sender-level aggregation. The treatment comparisons in the text use Mann–Whitney U tests on the same aggregation, together with session-level clustered inference on subject-round observations.

nificant (Mann–Whitney U test, both p -value < 0.01 ; clustered, both p -value < 0.05).¹⁴ This shows that partial commitment is not a full substitute for full commitment, even when marginal incentives are aligned.¹⁵

Our second prediction concerns the role of commitment when marginal incentives are misaligned. The theory predicts that, once full commitment is lost, the KG equilibrium can no longer be sustained and communication collapses to a babbling equilibrium. The data are broadly consistent with this prediction. Overall informativeness declines significantly from 0.24 in MF to 0.04 in MP and to -0.02 in MP-A (M–W U tests, both p -value < 0.01 ; clustered, p -value = 0.051 and p -value < 0.05 , respectively). Moreover, informativeness in MP-A is significantly lower than in AP-A (M–W U test, p -value < 0.01 ; clustered, p -value < 0.05), and informativeness in MP falls below that in AP (M–W U test, p -value = 0.052; clustered, p -value = 0.029). These results are consistent with the

¹⁴Throughout, we report two p -values for each treatment comparison. The first is from the Mann–Whitney U test on subject-level means, as specified in our pre-registration (AsPredicted #257058). Since subjects are randomly rematched within a session, the strictly independent unit is the session rather than the subject. We therefore also report p -values from a regression on subject-round observations with subject random effects, clustered at the session level (hereafter, *clustered*). We also report the results based only on the Round 1 data. See Online Appendix C for details.

¹⁵AF serves as a useful benchmark in our experiment. It corresponds to the standard Bayesian persuasion setting with full commitment and aligned incentives. Indeed, AF exhibits the highest level of overall informativeness among all treatments and is relatively close to the theoretical benchmark. Although the observed level of informativeness in AF (0.43) falls somewhat short of the theoretical prediction, similar deviations have been reported in Bayesian persuasion experiments conducted in comparable environments (Fr chet te et al., 2022; Rentschler and Samad, 2025).

prediction that a babbling equilibrium arises once full commitment is lost when marginal incentives are misaligned.

Our third prediction concerns the role of the incentive structure under full commitment. According to the theory, misaligned incentives should not affect overall informativeness, implying identical levels of informativeness in AF and MF. The data, however, contradict this prediction. Overall informativeness declines from 0.43 in AF to 0.24 in MF (M–W U test, p -value < 0.01 ; clustered, p -value $= 0.070$), suggesting that misaligned incentives hinder information transmission even when the sender can fully commit to the information structure.

In the following sections, we examine the behavior of senders and receivers separately to better understand the behavior underlying these patterns.

4.2 Senders’ Behavior

4.2.1 Senders’ Informativeness

In this section, we examine sender behavior. To isolate the sender’s contribution to information transmission, we focus on how much information senders’ messages convey about the state, abstracting from receivers’ actions. To this end, we introduce a measure of *senders’ informativeness*, defined as $\phi(\theta, m) := \text{Corr}(\theta, m; \pi)$, that is, the correlation between the state and the message generated by the sender’s information structure π .¹⁶

Table 5 reports the average values of senders’ informativeness in each treatment together with the corresponding theoretical values. The treatment comparisons closely mirror those observed for overall informativeness in Table 4. Senders’ informativeness is

¹⁶This measure is directional and is not invariant to relabeling the messages. It therefore distinguishes an information structure that is consistent with the meaning of the message labels from one in which the labels are swapped. By contrast, the measure often used in Bayesian persuasion experiments (e.g., Fr chet te et al., 2022), the correlation between the state and the action of a Bayesian receiver under π , is invariant to such relabeling. We adopt the former for two reasons. First, the latter presumes that the receiver correctly infers the sender’s information structure and interprets the message labels accordingly; under Partial and A-Partial commitment, where the receiver cannot observe π , this presumption is generally not satisfied, as we show below. We therefore regard relabeling in these treatments as an attempt to mislead the receiver rather than as an alternative information structure that conveys the same information. Second, message labels are explicit in our experiment, so fixing their meaning is consistent with the implementation. Our definition follows cheap talk experiments such as Cai and Wang (2006).

Table 5: Senders’ informativeness by treatment

		Commitment		
		Full	Partial	A-Partial
ALG	Data	0.60 [0.51, 0.68]	0.28 [0.15, 0.40]	0.27 [0.12, 0.43]
	Theory	0.54	0.54	0.54
MIS	Data	0.23 [0.09, 0.36]	0.04 [−0.10, 0.17]	−0.15 [−0.28, −0.02]
	Theory	0.54	0	0

Notes: The Data rows report the mean of $\phi(\theta, m)$ and, in brackets, its 95% confidence interval, based on sender-level aggregation. The treatment comparisons in the text use Mann–Whitney U tests on the same aggregation, together with session-level clustered inference on subject-round observations.

significantly lower in AP and AP-A than in AF (M–W U tests, both p -value < 0.01 ; clustered, both p -value < 0.05), even though the theory allows the same level to be attained across these treatments. It also declines from MF to both MP and MP-A, from AP to MP, and from AP-A to MP-A (M–W U tests, all p -value < 0.05 ; clustered, p -value = 0.249, p -value = 0.092, p -value < 0.01 , and p -value < 0.05 , respectively). These declines are in the direction predicted by the babbling equilibrium. Finally, senders’ informativeness is lower in MF than in AF (M–W U test, p -value < 0.01 ; clustered, p -value = 0.094), even though the theory predicts no difference under full commitment.

We now turn to a quantitative comparison with the theoretical values. First, senders’ informativeness in AF (0.60) is slightly higher than the theoretical value of 0.54. However, this difference is not statistically significant (Wilcoxon signed-rank test, p -value = 0.16). By contrast, senders’ informativeness in AP (0.28), AP-A (0.27), and MF (0.23) falls significantly below the theoretical value of 0.54 (all p -value < 0.01).¹⁷ In other words, senders communicate substantially less information than predicted by the KG equilibrium, even in treatments where informative communication is theoretically sustainable. We refer to this systematic downward deviation from the theoretical value observed in AP, AP-A,

¹⁷The alternative measure, the correlation between the state and the action of a Bayesian receiver, is meaningful under Full commitment, where the receiver observes π . Our conclusions for AF and MF are unchanged under this measure: it is 0.54 in AF, not significantly different from the theoretical value of 0.54 (p -value = 0.85), and 0.31 in MF, significantly below the theoretical value (p -value < 0.01). The difference between AF and MF remains significant (M–W U test, p -value < 0.01 ; clustered, p -value = 0.012).

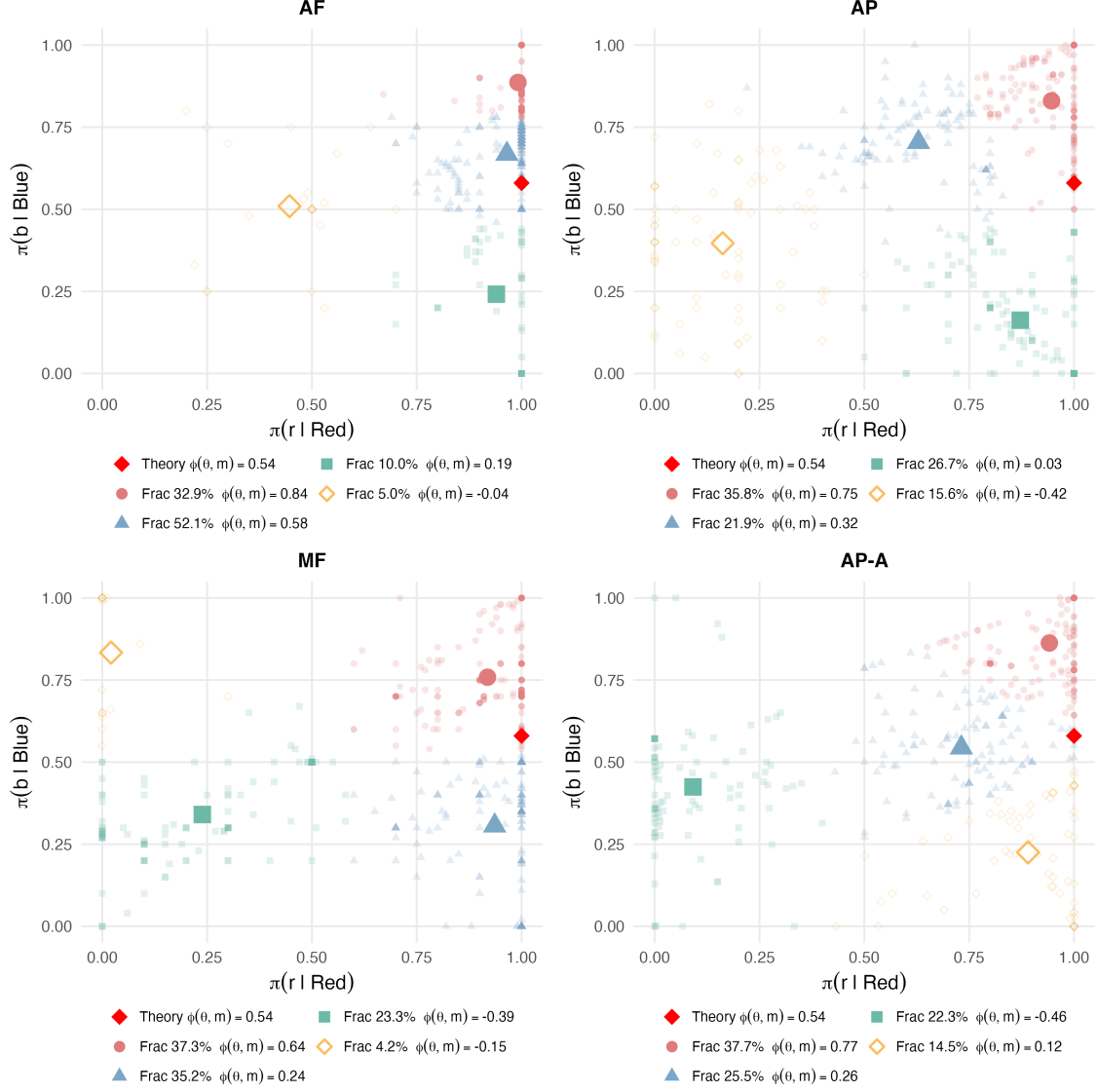


Figure 6: Clustering senders' information structures by treatment

and MF as *undercommunication*.

4.2.2 Heterogeneity in Senders' Strategies

In this section, we analyze heterogeneity in sender behavior in order to better understand the sender undercommunication documented in the previous section. To this end, we cluster the information structures observed in AF, MF, AP, and AP-A and identify representative sender strategies. Following Fr chet te et al. (2022), we apply the standard k -means clustering algorithm and classify them into four clusters within each treatment.

Figure 6 reports the results.¹⁸ Each panel corresponds to one treatment, and the senders' informativeness induced by each cluster together with its relative frequency is reported below the panel.

We begin with AF. In this treatment, the high-informativeness cluster represented by red circles accounts for 32.9% of senders' strategies, while the medium-informativeness cluster represented by blue triangles accounts for 52.1%. Both clusters induce levels of senders' informativeness that are close to or above the theoretical value, and together they account for approximately 85% of all strategies. That is, strategies in AF are relatively homogeneous, concentrated in clusters that are close to or above the theoretical value.

Turning to MF, AP, and AP-A, a markedly different pattern emerges. Relative to AF, senders' strategies are substantially more heterogeneous. Nevertheless, the most frequently observed cluster remains highly informative. In each treatment, the largest cluster (red circles) accounts for approximately 36–38% of senders' strategies and induces levels of senders' informativeness ranging from 0.64 to 0.77, all of which exceed the theoretical value. In addition, the cluster represented by blue triangles induces levels of senders' informativeness between 0.24 and 0.32, close to the average level observed in each treatment.

At the same time, a sizable share of strategies in these treatments are substantially less informative. In AP, 26.7% of strategies belong to the cluster represented by green squares, which induces a level of senders' informativeness of only $\phi(\theta, m) = 0.03$ and is therefore close to a babbling strategy. In MF and AP-A, 23.3% and 22.3% of strategies, respectively, fall into green-square clusters with negative senders' informativeness ($\phi(\theta, m) = -0.39$ and $\phi(\theta, m) = -0.46$). These information structures generate messages that are negatively correlated with the state. The remaining cluster, represented by yellow diamonds, is either small in size or low in informativeness, and does not exhibit a consistent pattern across treatments.

Taken together, these results suggest that sender undercommunication is not driven

¹⁸The corresponding clustering results for MP and MP-A are reported in Figure D.3 of Online Appendix D.2.

by the disappearance of highly informative strategies. Indeed, the high-informativeness cluster remains the most frequently observed one in MF, AP, and AP-A. Rather, senders' strategies are substantially more heterogeneous in these treatments than in AF, including a sizable share of babbling-like or negatively correlated information structures. The prevalence of these strategies is reflected in the lower average level of senders' informativeness documented in Table 5.

4.3 Receivers' Behavior

4.3.1 Do Receivers Behave as Bayesian?

In this section, we examine whether receivers' behavior is consistent with Bayesian decision-making. We focus on receivers' responses to message r . Under our parameterization, a Bayesian receiver follows a threshold strategy. Specifically, the receiver chooses action *red* whenever the posterior belief that the state is Red satisfies $\mu(r) \geq 1/2$, and chooses *blue* otherwise.

Figure 7 plots the relationship between the posterior belief $\mu(r)$ implied by the sender's information structure π and the frequency with which receivers choose action *red*. The dashed line represents the theoretically predicted threshold strategy. As shown in the figure, in the treatments with full commitment (AF and MF), the proportion of receivers choosing *red* increases with the posterior belief, broadly consistent with Bayesian behavior.¹⁹ In contrast, under Partial and A-Partial commitment, the probability of choosing *red* shows little responsiveness to the posterior belief, suggesting that receiver behavior may be inconsistent with Bayesian decision-making.

However, recall that under Partial and A-Partial commitment, receivers cannot directly observe the information structure π actually employed by the sender; they observe only the induced message distribution $\mathbb{P}_\pi(m)$. Consequently, receivers in these treatments must engage in a two-step inference process: they must first infer the sender's

¹⁹This does not imply that receivers perfectly follow the theoretically predicted Bayesian threshold strategy. As shown in Figure 7, their responses to posterior beliefs are more gradual than the theoretical threshold strategy.

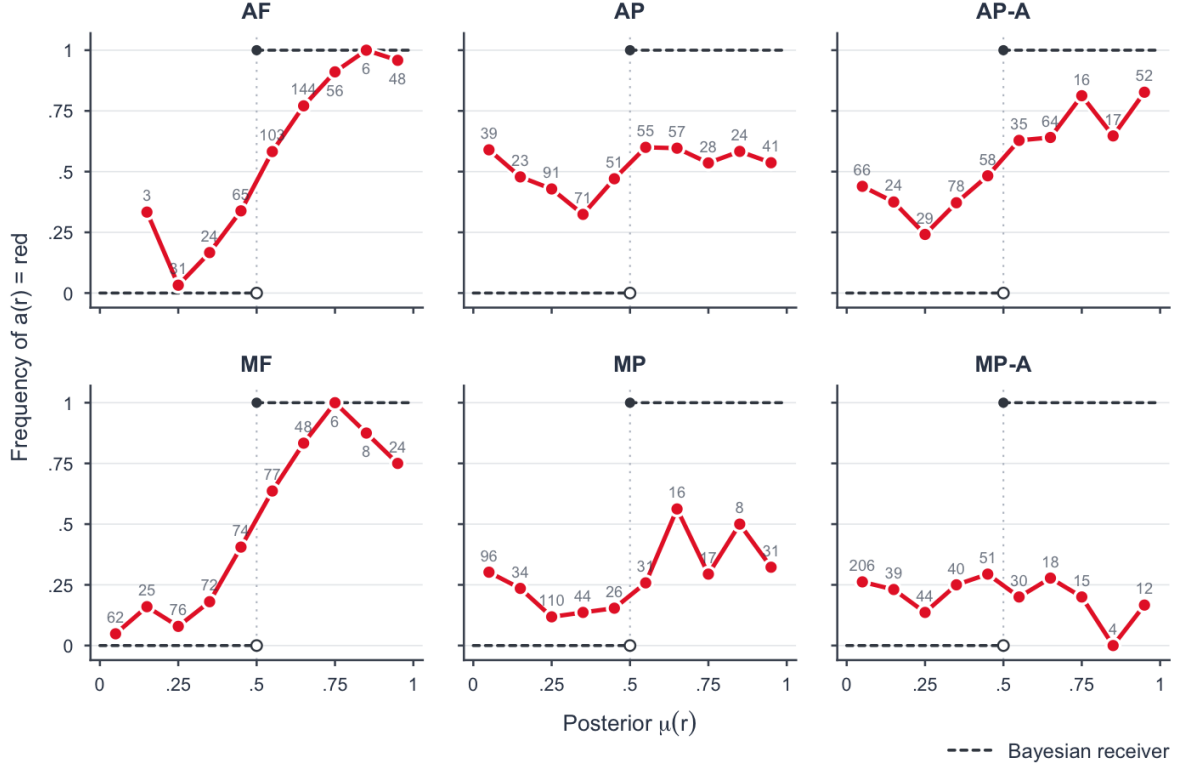


Figure 7: Receiver responses to objective posterior beliefs

information structure from the observed message distribution and then perform Bayesian updating based on the inferred structure. Therefore, Figure 7 alone cannot distinguish whether the observed pattern is driven by non-Bayesian decision-making by receivers, inaccurate beliefs about the sender’s information structure, or both. To disentangle these possibilities, as discussed in Section 3.3, we elicited receivers’ beliefs about the sender’s information structure, denoted by $\hat{\pi}$.

Figure 8 replots Figure 7 using the subjective posterior belief $\hat{\mu}(r)$ implied by the receiver’s reported belief $\hat{\pi}$. Compared with Figure 7, a clearer positive relationship emerges between the subjective posterior belief and the frequency of choosing *red*. That is, conditional on their beliefs about the sender’s information structure, receivers respond to the implied subjective posterior beliefs in a manner broadly consistent with Bayesian decision-making.²⁰ This result also suggests that receivers do not necessarily form accurate beliefs about the sender’s information structure.²¹

²⁰The corresponding regression results are reported in Table D.2 of Online Appendix D.3.

²¹Figure D.4 in Online Appendix D.3 plots the relationship between the objective posterior belief $\mu(r)$

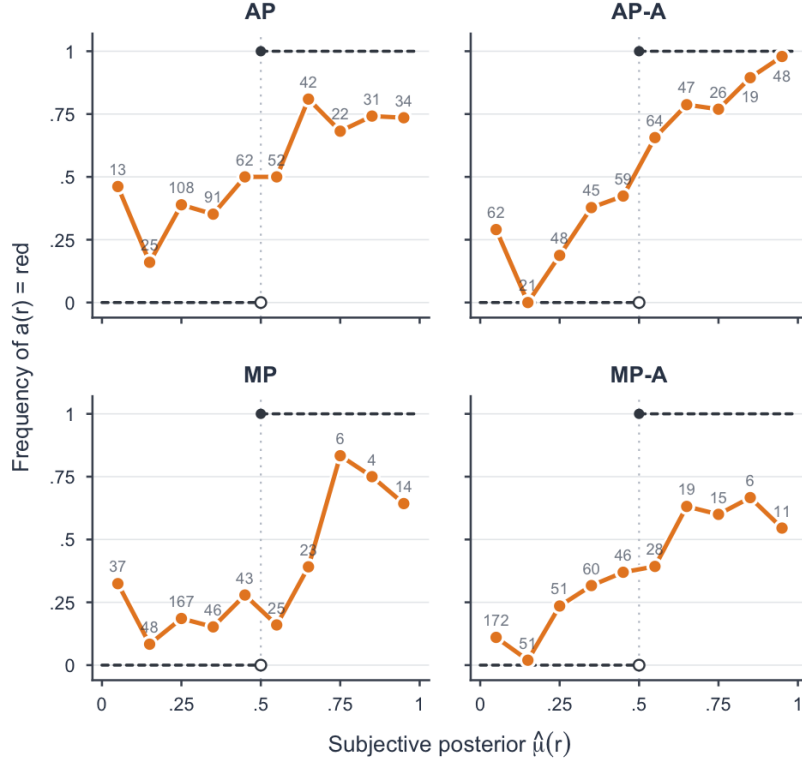


Figure 8: Receiver responses to subjective posterior beliefs

Taken together, under Full commitment, receivers respond to the posterior beliefs implied by the information structure designed by the sender in a manner broadly consistent with Bayesian decision-making. Under Partial and A-Partial commitment, by contrast, receivers show little responsiveness to the posterior beliefs implied by the information structure actually used by the sender. Instead, they respond to the subjective posterior beliefs implied by their own beliefs about the sender's information structure.

4.3.2 Receivers' Skepticism

In the previous section, we noted that, under Partial and A-Partial commitment, receivers do not necessarily form accurate beliefs about the information structure actually employed by the sender. In these treatments, however, a more important question is not whether receivers can accurately infer the sender's information structure, but whether they understand how the sender's incentives affect the messages that are sent. Specifically, do

and the subjective posterior belief $\hat{\mu}(r)$, showing that the two do not generally coincide.

Table 6: Receivers’ perceived informativeness by treatment

	Commitment	
	Partial	A-Partial
ALG	0.28 [0.19, 0.38]	0.28 [0.14, 0.41]
MIS	0.05 [−0.03, 0.13]	−0.12 [−0.22, −0.02]

Notes: Each cell reports the mean of $\hat{\phi}(\theta, m)$ and, in brackets, its 95% confidence interval, based on receiver-level aggregation. The treatment comparisons in the text use Mann–Whitney U tests on the same aggregation, together with session-level clustered inference on subject-round observations.

receivers recognize that senders with misaligned marginal incentives have an incentive to distort information, and therefore become more skeptical of communication? This section examines this question.

To examine this, we consider receivers’ *perceived informativeness*, defined as $\hat{\phi}(\theta, m) := \text{Corr}(\theta, m; \hat{\pi})$, that is, the correlation between the state and the message under the receiver’s belief $\hat{\pi}$ about the sender’s information structure. This measure captures the extent to which receivers believe that the sender’s messages convey information about the underlying state. If receivers are indeed more skeptical under MIS, we expect $\hat{\phi}(\theta, m)$ to be lower in MP than in AP, and lower in MP-A than in AP-A.

Table 6 reports average perceived informativeness across treatments. A clear pattern emerges. Under Partial commitment, perceived informativeness equals 0.28 in AP but only 0.05 in MP. Similarly, under A-Partial commitment, perceived informativeness equals 0.28 in AP-A but falls to −0.12 in MP-A. In both cases, the differences are statistically significant under the Mann–Whitney test (both p -value < 0.01). Under clustered inference, the difference between AP-A and MP-A remains significant (p -value < 0.01), whereas that between AP and MP does not (p -value = 0.175).

These results suggest that, although receivers do not necessarily form accurate beliefs about the sender’s information structure, their beliefs systematically vary with the sender’s incentives. Specifically, under Partial and A-Partial commitment, receivers substantially discount messages from senders with misaligned marginal incentives relative to

those from senders with aligned incentives.

5 Understanding Deviations from Theory

In the previous section, we documented several systematic deviations from the theoretical benchmark. First, overall informativeness is significantly lower in AP, AP-A, and MF than in AF, even though the KG equilibrium can be sustained in all four treatments. Correspondingly, senders' informativeness is close to the theoretical benchmark in AF, whereas senders undercommunicate in AP, AP-A, and MF. By contrast, although receivers' behavior does not perfectly conform to the theoretical predictions, we do not observe a common pattern of deviations from theory across these three treatments. Taken together, these results suggest that sender undercommunication plays an important role in the lower overall informativeness observed in AP, AP-A, and MF.

Moreover, this undercommunication does not appear to dissipate with experience. Table D.1 in Online Appendix D.1 reports regressions of senders' informativeness on the round number. In MF, AP, and AP-A, the coefficient on the round number is not significantly different from zero, and the point estimates are small. We therefore focus in this section on examining what factors may account for sender undercommunication and, more broadly, the observed deviations from the theoretical predictions.

First, Section 5.1 examines whether undercommunication can be explained as senders' empirical best response to the actual behavior of receivers in the experiment and shows that it cannot. Next, Section 5.2 discusses the possibility that the undercommunication observed in MF and that observed in AP and AP-A arise from different sources. Finally, Section 5.3 considers other potential explanations for undercommunication and shows that they are unlikely to account for the observed pattern.

5.1 Do Senders Best Respond to Empirical Receivers?

As shown in Section 4.3.1, under Partial and A-Partial commitment, receivers' behavior is largely unresponsive to the posterior beliefs induced by the information structure π

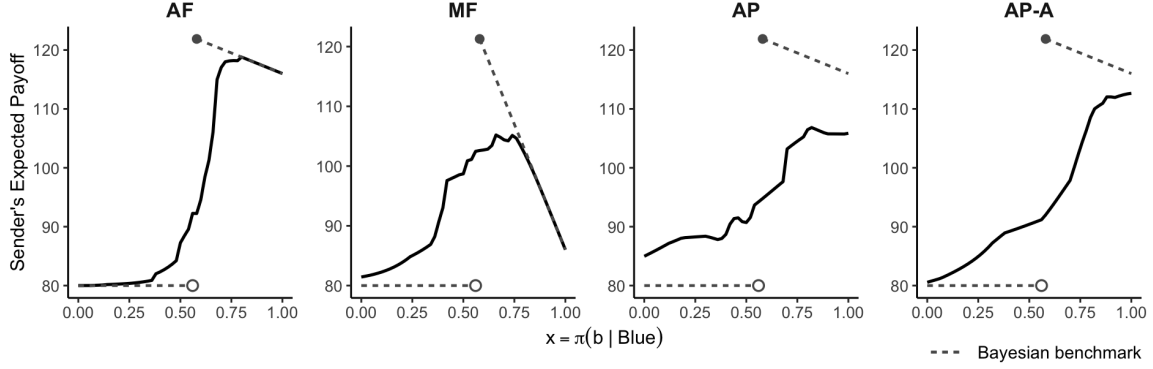


Figure 9: Sender expected payoff under empirical receiver response functions

actually used by the sender. Moreover, even in MF, receivers respond more gradually than predicted by the theoretical threshold strategy. Such receiver behavior may affect senders' choice of information structure and thereby indirectly give rise to undercommunication.

We first estimate receivers' response functions in order to describe their behavior in the experiment, using a probit model estimated separately for each receiver within each treatment, with an indicator for choosing action *red* as the dependent variable. The explanatory variable differs across treatments, reflecting what receivers can observe. Under Full commitment (AF and MF), where receivers observe the sender's information structure, we use the posterior belief $\mu(m)$ induced by that structure. Under Partial and A-Partial commitment (AP and AP-A), where receivers cannot observe the information structure, we use the message probability $\mathbb{P}_\pi(m)$. In both cases, a message dummy is also included as an explanatory variable. Using the estimated response functions, we compute senders' expected payoffs. Following Fr chet te et al. (2022), we evaluate the expected payoff against each receiver and take the median across receivers. For simplicity, we parameterize the sender's strategy in one dimension by fixing $\pi(r|\text{Red}) = 1$ and varying $x := \pi(b|\text{Blue})$.²²

Figure 9 plots the sender's expected payoff as a function of x , with the dashed line showing, as a benchmark, the expected payoff against a Bayesian receiver. In AF, the expected payoff corresponds relatively closely to the benchmark, whereas in the other

²²We restrict the sender's strategy to one dimension so that the expected payoff can be displayed as a function of a single variable, as in Figure 9. Extending the grid search to two dimensions over information structures with $\pi(r|\text{Red}) \geq \pi(r|\text{Blue})$ yields the same best responses in all four treatments.

Table 7: Observed and empirical best response outcomes

Treatment	$\phi(\theta, m)$		Sender's expected payoff	
	Observed	Empirical BR	Observed	Empirical BR
AF	0.60	0.74	107	119
MF	0.23	0.61	93	105
AP	0.28	0.76	96	107
AP-A	0.27	1.00	97	113

Notes: Under Sender's expected payoff, the Observed column does not report the average payoff senders actually earned in the experiment. It reports expected payoffs computed from the information structures senders actually used and the estimated receiver response functions, so as to be comparable with the expected payoffs under Empirical BR.

treatments it deviates substantially. This indicates that, given the actual behavior of receivers, it was difficult for senders to attain the theoretical maximum payoff even when behaving optimally. In AP and AP-A in particular, the expected payoff remains well below the theoretical benchmark even at or close to full revelation. This is consistent with the fact that, under Partial and A-Partial commitment, receivers respond to their subjective posterior beliefs rather than to the posterior induced by the actual information structure.

Table 7 reports the senders' informativeness $\phi(\theta, m)$ and the sender's expected payoff under the empirical best response, alongside the observed values. Importantly, in every treatment the empirical best response yields a level of $\phi(\theta, m)$ between 0.61 and 1.00, exceeding the KG equilibrium prediction of 0.54. This indicates that, if anything, best responding to the receivers that senders actually face would call for greater informativeness than the equilibrium prediction, not less. The sender's expected payoff under the observed strategies, moreover, falls below the payoff under the empirical best response in every treatment.

Taken together, the observed undercommunication cannot be rationalized as a best response to the actual behavior of receivers.

5.2 Different Sources of Undercommunication

In the previous section, we examined whether the undercommunication observed in MF, AP, and AP-A can be explained in a unified way as senders' empirical best response, and showed that it cannot. In this section, we discuss the possibility that the undercommunication observed in MF and that observed in AP and AP-A arise from different sources.

The two environments differ substantially. In MF, the sender can fully commit to the information structure while the players' marginal incentives are misaligned. In AP and AP-A, by contrast, the sender can commit only to the induced message distribution while marginal incentives are aligned. Receivers' behavior also differs between the two, as we saw in Section 4.3.1: in MF receivers respond to the posterior belief implied by the sender's information structure, whereas in AP and AP-A they respond not to the structure the sender designed but to the one they themselves infer.

We begin with MF. Because the sender can fully commit to the information structure in MF, the receiver observes it directly. As Figure 7 shows, receivers indeed respond to the posterior belief that the structure implies. Senders could therefore have induced receivers' actions relatively easily by choosing a more informative structure, and, as Table 7 indicates, could have earned a higher expected payoff by doing so. Yet senders in fact undercommunicate: relative to AF, the probability of sending message r in state Blue is higher. This is the direction that corresponds to the sender's payoff structure under MIS, where the sender earns the highest payoff when the receiver chooses action *red* in state Blue. Since full commitment is available, however, such a structure is observed by the receiver, and this strategy is therefore not optimal. Senders appear not to take this fully into account and instead respond naively to the payoff structure, and such behavior may be what gives rise to undercommunication in MF.²³ However, even if this interpretation is correct, why senders exhibit such a naive response remains a separate question. Understanding the source of this behavior is left for future research.

²³As a related concept, Fr chet te et al. (2022) refer to senders who, in a game with commitment power, play the strategy that would be optimal absent commitment as exhibiting commitment blindness. Our claim about MF is weaker: senders distort their strategies in a direction consistent with MIS.

We turn next to AP and AP-A. Because the incentive structure is ALG in these treatments, senders do not have an incentive to distort information in the same way as in MF. It is therefore unlikely that undercommunication in these treatments arises, as in MF, from senders responding to the payoff structure. Instead, we document two notable behavioral patterns in these treatments. First, as documented in Section 4.2.2, senders' strategies are highly heterogeneous. Second, as documented in Section 4.3.1, receivers respond to their subjective posterior beliefs rather than to the posterior beliefs induced by the information structure actually used by the sender. These patterns suggest that both senders and receivers face difficulties in these treatments. From the sender's perspective, because receivers do not respond to the posterior beliefs induced by the actual information structure, the sender's ability to influence receivers' actions through the choice of information structure is limited. From the receiver's perspective, in the presence of substantial heterogeneity in senders' strategies, it is difficult to form accurate beliefs about the information structure actually used by the sender based on the observable message distribution. These difficulties may be one reason why information transmission in AP and AP-A does not reach the level observed in AF.

These two difficulties are closely related, but it is unclear which gives rise to which. Heterogeneity in senders' strategies may make it difficult for receivers to infer the information structure; alternatively, the limited influence that senders can exert on receivers may generate heterogeneity in senders' strategies.²⁴ Our data, however, cannot distinguish between these possibilities, and additional experiments would be required.²⁵

5.3 Other Possibilities

In this section, we discuss other possibilities that we have not considered so far.

²⁴On the latter possibility, Section 5.1 shows that undercommunication is not an empirical best response to the actual behavior of receivers. This does not rule out the possibility that receivers' behavior affected senders' behavior. Senders who recognized that their influence on receivers was limited may have responded in some way other than best responding, and this may in turn have produced undercommunication.

²⁵To test whether heterogeneity in senders' strategies makes receivers' inference difficult, for instance, one could computerize a subset of senders so that they always use the optimal information structure under full commitment, thereby exogenously reducing the heterogeneity of strategies.

The first is senders’ lying aversion, which has been discussed in the literature on communication games, and on cheap talk games in particular. In our setting, lying aversion on the part of senders works in the direction of raising senders’ informativeness in every treatment. Indeed, among the clusters we identify, strategies that induce higher senders’ informativeness than the equilibrium predicts are frequently observed, and these are consistent with lying aversion. What we record overall, however, is undercommunication, which lying aversion cannot account for.

The second possibility is senders’ social preferences. In every treatment, the average payoff of senders is lower than that of receivers. Had senders played the empirical best response derived in Section 5.1, both their own expected payoff and that of receivers would have risen (in AP, for instance, from 96 to 107 for senders and from 128 to 133 for receivers).²⁶ Moreover, except in AP-A, this improvement would have narrowed the gap between senders’ and receivers’ expected payoffs. Senders in our experiment nonetheless did not do so. It therefore seems difficult to account for undercommunication by standard forms of altruistic preferences or inequality aversion.

In addition, in Online Appendix B, we estimate a quantal response equilibrium (QRE) that incorporates heterogeneity in senders’ strategies and stochastic deviations from best responses by both players. The QRE-implied levels of senders’ informativeness and overall informativeness are close to the observed levels in AF, MF, AP, and AP-A, indicating that the observed undercommunication is consistent with stochastic best responding by both players.

6 Conclusion

We conducted an experiment that varied the commitment structure and the alignment of marginal incentives to examine whether partial commitment to the induced message distribution can substitute for full commitment to the information structure. Our main finding is that partial commitment does not fully substitute for full commitment. Even

²⁶Receivers’ expected payoffs under the empirical best response are computed using the response functions estimated in Section 5.1.

when marginal incentives are aligned and the KG equilibrium can theoretically be sustained under partial commitment, information transmission is substantially lower under partial commitment than under full commitment.

In addition to this main finding, we obtain two further results. First, even when full commitment is available, information transmission decreases when marginal incentives are misaligned. This result shows that senders' marginal incentives affect actual information transmission even in the presence of full commitment. Second, under partial commitment, information transmission is lower when marginal incentives are misaligned than when they are aligned. This result is consistent with the theoretical prediction that the KG equilibrium can no longer be sustained under partial commitment when marginal incentives are misaligned. Furthermore, our analysis of sender behavior shows that senders undercommunicate in MF, AP, and AP-A, suggesting that this plays an important role in the lower information transmission observed in these treatments.

Our main finding also has implications for institutional design. Even in environments where marginal incentives are aligned and information transmission can theoretically be sustained under partial commitment, our results suggest that strengthening senders' commitment ability by increasing the observability or verifiability of the information structure can promote information transmission in practice.

An important task for future research is to identify the mechanisms that generate deviations from the theoretical benchmark, including sender undercommunication. In [Section 5.2](#), we discussed hypotheses consistent with our data: in MF, senders may respond naively to their marginal incentives, whereas in AP and AP-A, difficulties in strategic reasoning about the information structure may impede information transmission. Future research could test these hypotheses and, for example, directly elicit senders' beliefs about how receivers will behave under the information structures they choose, thereby further investigating the mechanisms underlying sender undercommunication.

References

- Aristidou, A., G. Coricelli, and A. Vostroknutov (2019): “Incentives or persuasion? an experimental investigation,” Working paper, Maastricht University.
- Bell, R. M. and D. F. McCaffrey (2002): “Bias reduction in standard errors for linear regression with multi-stage samples,” *Survey Methodology*, 28, 169–181.
- Bergemann, D. and S. Morris (2019): “Information design: A unified perspective,” *Journal of Economic Literature*, 57, 44–95.
- Best, J. and D. Quigley (2024): “Persuasion for the long run,” *Journal of Political Economy*, 132, 1740–1791.
- Cai, H. and J. T.-Y. Wang (2006): “Overcommunication in strategic information transmission games,” *Games and Economic Behavior*, 56, 7–36.
- Chen, D. L., M. Schonger, and C. Wickens (2016): “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero (2012): “Measuring risk literacy: The Berlin numeracy test,” *Judgment and Decision Making*, 7, 25–47.
- Crawford, V. P. and J. Sobel (1982): “Strategic information transmission,” *Econometrica*, 50, 1431–1451.
- Deb, R., M. M. Pai, and M. Said (2026): “Indirect persuasion,” *Journal of Political Economy*, 134, 1210–1244.
- Fréchette, G. R., A. Lizzeri, and J. Perego (2022): “Rules and commitment in communication: An experimental analysis,” *Econometrica*, 90, 2283–2318.
- Greiner, B. (2015): “Subject pool recruitment procedures: organizing experiments with ORSEE,” *Journal of the Economic Science Association*, 1, 114–125.

- Holt, C. A. and S. K. Laury (2002): “Risk aversion and incentive effects,” *American Economic Review*, 92, 1644–1655.
- Jin, G. Z., M. Luca, and D. Martin (2021): “Is no news (perceived as) bad news? An experimental investigation of information disclosure,” *American Economic Journal: Microeconomics*, 13, 141–173.
- 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.
- Kerschbamer, R. (2015): “The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test,” *European Economic Review*, 76, 85–103.
- Koessler, F. and V. Skreta (2023): “Informed information design,” *Journal of Political Economy*, 131, 3186–3232.
- Lin, X. and C. Liu (2024): “Credible persuasion,” *Journal of Political Economy*, 132, 2228–2273.
- Lipnowski, E., D. Ravid, and D. Shishkin (2022): “Persuasion via weak institutions,” *Journal of Political Economy*, 130, 2705–2730.
- Min, D. (2021): “Bayesian persuasion under partial commitment,” *Economic Theory*, 72, 743–764.
- Nguyen, A. and T. Y. Tan (2021): “Bayesian persuasion with costly messages,” *Journal of Economic Theory*, 193, 105212.
- Nguyen, Q. (2017): “Bayesian persuasion: evidence from the laboratory,” Working paper.

- Pustejovsky, J. E. and E. Tipton (2018): “Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models,” *Journal of Business & Economic Statistics*, 36, 672–683.
- Rentschler, L. and Z. Samad (2025): “Bayesian persuasion with fact-checking: An experimental investigation,” *European Economic Review*, 179, 105114.
- Wu, W. and B. Ye (2023): “Competition in persuasion: An experiment,” *Games and Economic Behavior*, 138, 72–89.
- Xiang, J. (2026): “Physicians as persuaders: Evidence from hospitals in China,” *The RAND Journal of Economics*, 57, 451–468.
- Zhou, J. (2023): “Costly verification and commitment in persuasion,” *Journal of Economic Behavior & Organization*, 212, 1100–1142.

Online Appendix for

Credibility in Persuasion: A Laboratory Experiment

A The Role of Announced Information Structure

In this section, we examine how the existence of an announced information structure affects information transmission and the behavior of senders and receivers. Recall that the only difference between Partial and A-Partial commitment is whether an announced information structure $\tilde{\pi}$ is present. Under A-Partial commitment, even after announcing $\tilde{\pi}$, the sender can deviate to any π that leaves the message distribution unchanged. The sender is thus effectively bound only by the message distribution, just as under Partial commitment. The existence of $\tilde{\pi}$ therefore does not affect the equilibrium predictions, and the predictions are identical across the two treatments. At the behavioral level, however, the announcement of $\tilde{\pi}$ may influence players' behavior.

A.1 Treatment Effects

Table A.1: The effect of the announced information structure

		Commitment	
		Partial	A-Partial
ALG	$\phi(\theta, a)$	0.13 [0.07, 0.20]	0.21 [0.10, 0.31]
	$\phi(\theta, m)$	0.28 [0.15, 0.40]	0.27 [0.12, 0.43]
MIS	$\phi(\theta, a)$	0.04 [0.00, 0.07]	-0.02 [-0.06, 0.02]
	$\phi(\theta, m)$	0.04 [-0.10, 0.17]	-0.15 [-0.28, -0.02]

Notes: Each cell reports the mean and, in brackets, its 95% confidence interval, based on sender-level aggregation. The treatment comparisons in the text use Mann–Whitney U tests on the same aggregation, together with session-level clustered inference on subject-round observations.

Table A.1 compares Partial and A-Partial commitment in terms of overall informative-

ness $\phi(\theta, a)$ and senders' informativeness $\phi(\theta, m)$. Consider first overall informativeness. Under ALG, the announcement of $\tilde{\pi}$ tends to improve information transmission (AP: 0.13 vs. AP-A: 0.21), although the difference is not statistically significant (Mann–Whitney U test, p -value = 0.32; clustered, p -value = 0.144). Under MIS, by contrast, $\phi(\theta, a)$ is significantly lower in MP-A (MP: 0.04 vs. MP-A: -0.02 , M–W U test, p -value < 0.05; clustered, p -value < 0.05).

Consider next senders' informativeness. Under ALG, the means in the two treatments are very close and the difference is not significant (AP: 0.28 vs. AP-A: 0.27, M–W U test, p -value = 0.94; clustered, p -value = 0.94). Under MIS, by contrast, senders' informativeness is significantly lower in MP-A and takes a negative value (MP: 0.04 vs. MP-A: -0.15 , M–W U test, p -value < 0.05; clustered, p -value < 0.05). That is, under MIS, senders correlate their messages more negatively with the state in the presence of an announced $\tilde{\pi}$.

To summarize, when marginal incentives are aligned, the announcement of $\tilde{\pi}$ tends to improve overall information transmission, though not significantly, while having little effect on senders' informativeness. When marginal incentives are misaligned, by contrast, the announcement of $\tilde{\pi}$ leads to a small but statistically significant reduction in overall informativeness and induces senders to correlate their messages more negatively with the state.

A.2 How Senders Use the Announced Information Structure

In this section, we show how senders use $\tilde{\pi}$ as part of their strategy under A-Partial commitment. In the k -means analysis presented in Figure 6 of Section 4.2.2 and Figure D.3 of Online Appendix D.2, the sender's strategy was represented solely by the information structure π actually used. Under A-Partial commitment, however, the sender announces $\tilde{\pi}$ in Stage 1 and then chooses π in Stage 2, so that the sender may use both $\tilde{\pi}$ and π strategically. We therefore represent each sender's strategy in AP-A and MP-A as a four-dimensional vector $(\tilde{\pi}, \pi)$ and classify senders into four clusters within each treatment.

Figures A.1 and A.2 present the results, separately for the plane of $\tilde{\pi}$ (Stage 1) and the plane of π (Stage 2). Throughout this section, we denote by $\tilde{\phi}(\theta, m) := \text{Corr}(\theta, m; \tilde{\pi})$ the senders' informativeness based on the announced information structure $\tilde{\pi}$, in parallel with $\phi(\theta, m)$ based on the actual information structure π .

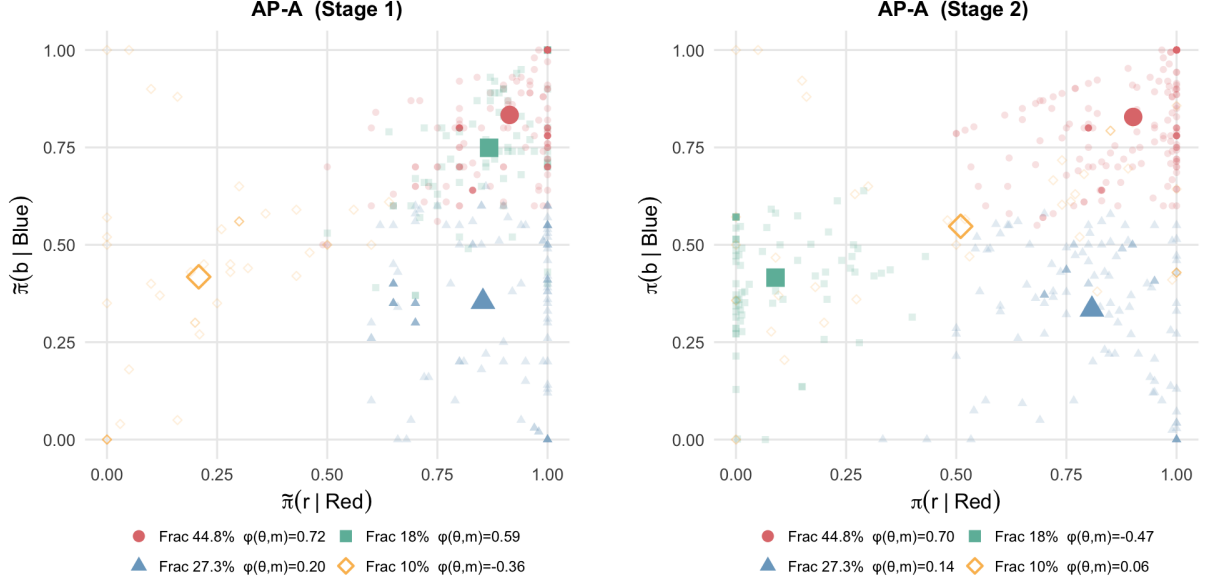


Figure A.1: Clustering senders' strategy $(\tilde{\pi}, \pi)$ in AP-A

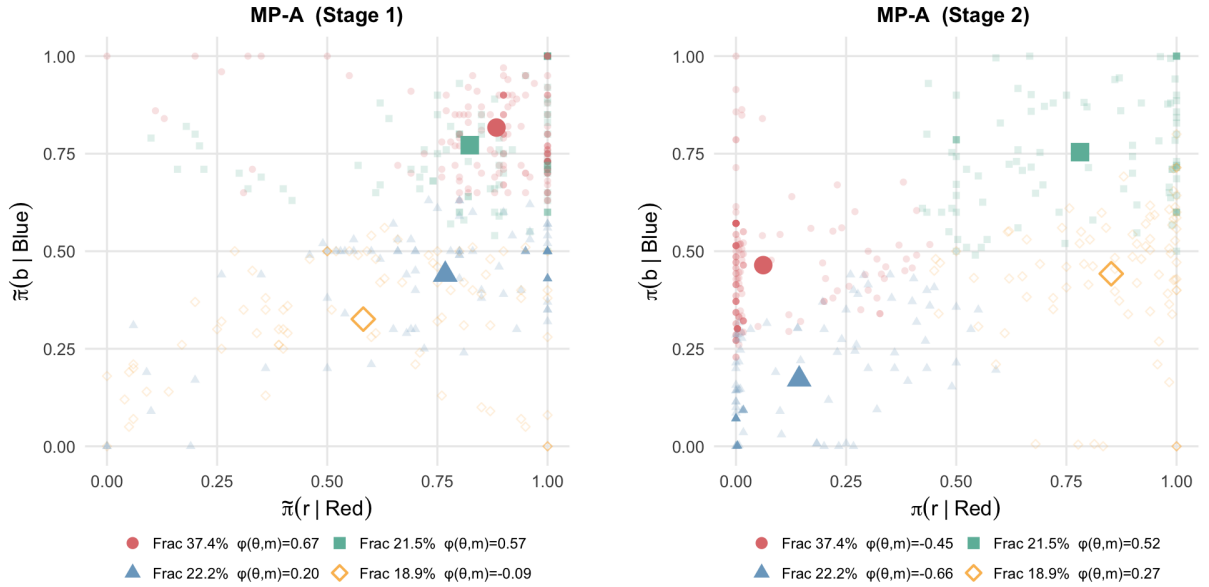


Figure A.2: Clustering senders' strategy $(\tilde{\pi}, \pi)$ in MP-A

Under AP-A, the largest cluster, represented by red circles (44.8%), is highly informative in both $\tilde{\pi}$ and π ($\tilde{\phi}(\theta, m) = 0.72$ and $\phi(\theta, m) = 0.70$, respectively), so that these

senders execute essentially the information structure they announced. This can also be interpreted as senders using an informative $\tilde{\pi}$ to indicate the information structure they subsequently implement. However, not all senders implement the information structure they announced. Most striking is the cluster represented by green squares (18.0%), which is relatively informative in $\tilde{\pi}$ ($\tilde{\phi}(\theta, m) = 0.59$) yet turns to negative informativeness in the actual π by swapping messages ($\phi(\theta, m) = -0.47$). Such deviations from the announced information structure also appear in the aggregate measures. In AP-A, the average senders' informativeness at the announcement stage, $\tilde{\phi}(\theta, m)$, is relatively high at 0.45. However, the average senders' informativeness based on the actual information structure, $\phi(\theta, m)$, falls to 0.27, which is almost identical to the level observed in AP (0.28), where no announcement is made.

Under MP-A, by contrast, clusters in which the announced $\tilde{\pi}$ and the actual π diverge substantially are dominant. The largest cluster, represented by red circles (37.4%), announces an informative $\tilde{\pi}$ ($\tilde{\phi}(\theta, m) = 0.67$) while correlating messages negatively with the state in the actual π ($\phi(\theta, m) = -0.45$). The second cluster, represented by blue triangles (22.2%), similarly deviates in the negative direction from $\tilde{\pi}$ to π (from $\tilde{\phi}(\theta, m) = 0.20$ to $\phi(\theta, m) = -0.66$). These senders appear to announce that they will design an informative π by displaying an informative $\tilde{\pi}$ to the receiver, while in fact swapping messages so as to induce the receiver to choose red in state Blue. This pattern is consistent with the deviation incentive that gives rise to the babbling equilibrium under misaligned marginal incentives.

A.3 How Receivers Respond to the Announced Information Structure

In the previous section, we examined how senders use the announced information structure $\tilde{\pi}$. In this section, we examine how receivers respond to $\tilde{\pi}$. As we saw in the previous section, under AP-A, 44.8% of sender strategies implemented a π that was almost identical to the highly informative $\tilde{\pi}$ they had announced. If the announced $\tilde{\pi}$ is informative about

Table A.2: Receivers' response to the announced posterior

	(1) AP-A	(2) MP-A	(3) Pooled
$\tilde{\mu}(r)$	0.710*** (0.110)	0.325** (0.115)	0.325*** (0.114)
$\tilde{\mu}(r) \times \text{ALG}$			0.385** (0.157)
Receiver fixed effects	Yes	Yes	Yes
Observations	439	459	898
R^2	0.525	0.209	0.442

Notes: The dependent variable is $a(r)$, an indicator equal to one if the receiver chooses the action *red* after observing message r . $\tilde{\mu}(r)$ denotes the posterior probability of state Red implied by the announced information structure $\tilde{\pi}$. Columns (1) and (2) report separate regressions for AP-A and MP-A, respectively. Column (3) pools the two treatments and interacts $\tilde{\mu}(r)$ with an indicator for aligned marginal incentives (ALG). The main effect of the ALG indicator is absorbed by the receiver fixed effects. All regressions are linear probability models with receiver fixed effects. Observations in which the sender's information structure assigns zero probability to message r in both states (so that $\tilde{\mu}(r)$ is undefined) are excluded: 1 in AP-A and 1 in MP-A. Standard errors clustered at the receiver level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

the π that is actually implemented in this way, receivers may use the information contained in $\tilde{\pi}$ when choosing their actions. Under MP-A, by contrast, senders have an incentive to deviate from an informative $\tilde{\pi}$, and, as we saw in the previous section, clusters in which the announced $\tilde{\pi}$ and the actual π diverge substantially are indeed frequently observed. We therefore expect receivers to respond less strongly to $\tilde{\pi}$ under MP-A than under AP-A.

Table A.2 reports regressions of the receiver's action $a(r)$ on the posterior belief implied by the announced information structure $\tilde{\pi}$. Under AP-A, receivers respond strongly to the posterior implied by the announced information structure (coefficient 0.710, p -value < 0.01). Although the coefficient is also positive and significant under MP-A (coefficient 0.325, p -value < 0.05), the response is significantly stronger under AP-A than under MP-A. Indeed, the coefficient on the interaction term in Column (3) is positive and significant (coefficient 0.385, p -value < 0.05). This result suggests that, under AP-A, the announced $\tilde{\pi}$ may provide receivers with information about the information structure π that is actually implemented.

This interpretation is also consistent with other findings. Panel A of Table D.2 in Online Appendix D.3 reports regressions of the receiver's action on the true posterior

belief $\mu(r)$. Apart from Full commitment treatments, AP-A is the only treatment in which the coefficient is significant (coefficient 0.426, p -value < 0.01). Moreover, as shown in Section A.1, although the difference is not statistically significant, overall informativeness $\phi(\theta, a)$ is higher in AP-A (0.21) than in AP (0.13). These findings are consistent with the possibility that the announced $\tilde{\pi}$ facilitates information transmission under AP-A by providing receivers with information about the actual π .

At the same time, it should be noted that the coefficient on the posterior implied by $\tilde{\pi}$ is also positive and significant under MP-A. Despite senders having an incentive to deviate from an informative $\tilde{\pi}$ under MP-A, receivers still respond to the announced $\tilde{\pi}$. This suggests that some receivers may respond to the information contained in the announced $\tilde{\pi}$ without fully taking into account senders' incentives to deviate.

B QRE

In this section, we estimate a quantal response equilibrium (QRE) that allows the sender and the receiver to deviate stochastically from best responses. The goal of this section is to examine whether the senders' undercommunication is quantitatively consistent with a QRE. In AF and MF, the sender has full commitment to the information structure. We can therefore apply the QRE approach of [Fr chet te et al. \(2022\)](#) directly. In AP and AP-A, by contrast, the receiver does not observe π but only the induced message distribution $\mathbb{P}_\pi(m)$. We address this issue by using the subjective posteriors that receivers report in the belief-elicitation task.

We assume that the sender and the receiver each have a treatment-specific type $\lambda_S \geq 0$ and $\lambda_R \geq 0$, and that these are common knowledge between the players. Each type captures how closely the corresponding player adheres to best responding: when $\lambda_i = 0$, player i randomizes uniformly over the available choices, and as $\lambda_i \rightarrow \infty$, the player's choice converges to a best response.

We first model the receiver's behavior as a logit quantal response. Let μ denote the receiver's posterior belief that the state is Red. Under the experimental payoff structure,

the receiver’s expected payoffs from the two actions are $U(\text{red}|\mu) = 150\mu + 80(1 - \mu)$ and $U(\text{blue}|\mu) = 80\mu + 150(1 - \mu)$. A receiver of type λ_R then chooses action *red* with probability

$$\mathbb{P}(\text{red}|\mu, \lambda_R) = \frac{e^{\lambda_R U(\text{red}|\mu)}}{e^{\lambda_R U(\text{red}|\mu)} + e^{\lambda_R U(\text{blue}|\mu)}}.$$

We now turn to the sender. Given the receiver behavior described above, the sender’s expected payoff from choosing an information structure π is $V(\pi|\lambda_R) = \sum_{\theta} \sum_m \mu_0(\theta) \pi(m|\theta) \sum_a \mathbb{P}(a|\mu(m), \lambda_R) u_S(\theta, a)$. Because the sender’s strategy space is continuous, it must be discretized to fit the discrete-choice framework of QRE. Following [Fr chet te et al. \(2022\)](#), we use the same k -means procedure employed in the analysis of senders’ heterogeneity in Section 4.2.2 to discretize the observed sender strategies into $k = 4$ representative strategies. Each representative strategy is defined as the empirical mean of the observations belonging to the corresponding cluster. We denote this set of representative strategies by Π_k . A sender of type λ_S chooses a representative strategy $\pi \in \Pi_k$ with probability

$$\mathbb{Q}(\pi|\lambda_S, \lambda_R) = \frac{e^{\lambda_S V(\pi|\lambda_R)}}{\sum_{\pi' \in \Pi_k} e^{\lambda_S V(\pi'|\lambda_R)}}.$$

We estimate λ_S and λ_R by maximum likelihood. Under full commitment, the receiver directly observes the information structure π actually used to generate the message. As a result, any error in the sender’s choice of strategy is irrelevant to the receiver, who can be regarded as facing a single-agent decision problem. We can therefore estimate λ_R independently of λ_S ([Fr chet te et al., 2022](#)). In AP and AP-A, by contrast, the receiver does not observe π , so this single-agent interpretation does not hold directly. In this paper, we take as given, for each representative strategy, the average subjective posterior reported by receivers who faced strategies in that cluster, treat the receiver’s action choice as a single-agent decision problem under that belief, and estimate λ_R accordingly.

In all cases, the sender moves before the receiver, so the estimate of λ_S depends on λ_R . To evaluate the expected payoff $V(\pi|\lambda_R)$ of each representative strategy, we follow [Fr chet te et al. \(2022\)](#) and compute the empirical average of the sender’s expected payoff

Table B.1: Observed and QRE-implied informativeness

Treatment	$\phi(\theta, m)$		$\phi(\theta, a)$	
	Observed	QRE-Implied	Observed	QRE-Implied
AF	0.60	0.56	0.43	0.48
MF	0.23	0.26	0.24	0.25
AP	0.28	0.27	0.13	0.14
AP-A	0.27	0.28	0.21	0.19

across the strategies belonging to the same cluster, using this average to approximate $V(\pi|\lambda_R)$ for each representative strategy. That is, we evaluate the receiver’s response from the behavior observed in the data rather than from a theoretical choice probability. Taking this empirical average as given, we estimate λ_S by maximum likelihood. Because both the set of representative strategies and the expected payoffs used in the sender’s choice probabilities are constructed from the observed data, this exercise should be read as a consistency check rather than as an out-of-sample prediction.

Using the estimated (λ_S, λ_R) , we generate a dataset of 10^4 simulated sender-receiver interactions.²⁷ Each interaction consists of a state θ drawn from the prior, an information structure drawn according to the sender’s choice probability $\mathbb{Q}$, a message m generated from that information structure, and an action a chosen according to the receiver’s choice probability $\mathbb{P}$. Using this simulated dataset, we compute two QRE-implied measures of information transmission for each treatment: $\phi(\theta, m)$ and $\phi(\theta, a)$.

Table B.1 reports the QRE-implied and observed levels of senders’ informativeness $\phi(\theta, m)$ and overall informativeness $\phi(\theta, a)$ for AF, MF, AP, and AP-A. Across all four treatments, the QRE-implied values are close to the observed ones. In particular, in MF, AP, and AP-A the QRE-implied value of $\phi(\theta, m)$ falls below the equilibrium prediction of 0.54 as well as the value in AF, in line with the observed undercommunication.

²⁷The estimated precision parameters are $(\lambda_S, \lambda_R) = (0.047, 0.057)$ for AF, $(0.169, 0.054)$ for MF, $(0.081, 0.047)$ for AP and $(0.036, 0.060)$ for AP-A.

Table C.1: Means by treatment and session

Treatment	Session	$\phi(\theta, a)$	$\phi(\theta, m)$	$\hat{\phi}(\theta, m)$	Subjects
AF	1	0.41	0.56	—	26
	2	0.46	0.64	—	22
MF	1	0.19	0.11	—	24
	2	0.29	0.34	—	24
AP	1	0.15	0.30	0.23	26
	2	0.12	0.26	0.35	22
MP	1	0.04	0.04	−0.03	26
	2	0.03	0.03	0.17	16
AP-A	1	0.19	0.23	0.26	26
	2	0.24	0.33	0.30	18
MP-A	1	−0.03	−0.19	−0.11	22
	2	−0.02	−0.12	−0.13	24

C Independent Units and Statistical Inference

Each session constitutes a single matching group, and subjects are randomly rematched within a session in each round. Each treatment was run in two sessions. The strictly independent unit is therefore the session rather than the subject, and the number of independent units is two per treatment. The Mann–Whitney U tests reported in the main text are conducted on subject-level means, following our pre-registration, and thus do not account for within-session dependence. We address this in two complementary ways. Section C.1 clusters inference at the session level using all rounds; Section C.2 restricts attention to Round 1, in which no interaction has yet taken place and subject-level observations can be treated as independent.

C.1 Session-Level Clustered Inference

We estimate a linear mixed model on subject-round observations with a treatment dummy, round fixed effects, and subject random effects. For inference, we use the bias-reduced cluster-robust variance estimator (CR2) of [Bell and McCaffrey \(2002\)](#), clustered at the

Table C.2: Treatment comparisons: Mann–Whitney U tests, session-level clustered inference, and Round 1 only

Comparison	All rounds			Round 1 only	
	Difference	M–W U	Clustered	Difference	M–W U
<i>Panel A. Overall informativeness $\phi(\theta, a)$</i>					
AF vs. AP	0.298	0.000	0.011	0.185	0.086
AF vs. AP-A	0.223	0.003	0.029	0.149	0.309
AF vs. MF	0.193	0.002	0.070	0.004	0.948
MF vs. MP	0.201	0.000	0.051	0.280	0.002
MF vs. MP-A	0.260	0.000	0.031	0.359	0.002
AP vs. MP	0.096	0.052	0.029	0.099	0.345
AP-A vs. MP-A	0.230	0.000	0.015	0.215	0.133
AP vs. AP-A	−0.075	0.317	0.144	−0.036	0.569
MP vs. MP-A	0.060	0.010	0.013	0.080	0.939
<i>Panel B. Senders’ informativeness $\phi(\theta, m)$</i>					
AF vs. AP	0.320	0.000	0.019	0.305	0.009
AF vs. AP-A	0.324	0.001	0.034	0.222	0.104
AF vs. MF	0.373	0.000	0.094	0.130	0.163
MF vs. MP	0.189	0.021	0.249	0.287	0.018
MF vs. MP-A	0.378	0.000	0.092	0.436	0.005
AP vs. MP	0.242	0.010	0.008	0.112	0.356
AP-A vs. MP-A	0.426	0.000	0.020	0.344	0.037
AP vs. AP-A	0.005	0.939	0.937	−0.084	0.328
MP vs. MP-A	0.188	0.039	0.041	0.149	0.372
<i>Panel C. Receivers’ perceived informativeness $\hat{\phi}(\theta, m)$</i>					
AP vs. MP	0.236	0.001	0.175	0.201	0.048
AP-A vs. MP-A	0.396	0.000	0.002	0.100	0.555

Notes: Difference is the mean of the first treatment minus the mean of the second. Under All rounds, means are subject-level averages over 20 rounds; M–W U reports p -values from Mann–Whitney U tests on these averages, and Clustered reports p -values from the session-level clustered inference described above. Under Round 1 only, means and Mann–Whitney U tests are based on Round 1 observations, one per subject (24, 24, 24, 21, 22, and 23 subjects per role in AF, MF, AP, MP, AP-A, and MP-A, respectively).

session level, with Satterthwaite degrees of freedom ([Pustejovsky and Tipton, 2018](#)).

Table C.1 reports means by treatment and session, and Table C.2 reports, in its first three columns, p -values from both the Mann–Whitney U test and the clustered test for the 20 treatment comparisons discussed in the paper. In 18 of the 20 comparisons, the two approaches yield the same sign and the same verdict at the 10% level. The exceptions

Table C.3: Overall informativeness by treatment: Round 1 only

		Commitment		
		Full	Partial	A-Partial
ALG	Data	0.32 [0.17, 0.47]	0.14 [−0.02, 0.29]	0.17 [−0.02, 0.36]
	Theory	0.54	0.54	0.54
MIS	Data	0.32 [0.19, 0.44]	0.04 [−0.09, 0.17]	−0.04 [−0.21, 0.12]
	Theory	0.54	0	0

Notes: The Data rows report the mean of $\phi(\theta, a)$ in Round 1 and, in brackets, its 95% confidence interval.

are the comparison between MF and MP for senders' informativeness, $\phi(\theta, m)$, and that between AP and MP for receivers' perceived informativeness, $\hat{\phi}(\theta, m)$. These likely reflect dispersion across sessions. Specifically, $\phi(\theta, m)$ is 0.11 and 0.34 in the two MF sessions, respectively, while $\hat{\phi}(\theta, m)$ is −0.03 and 0.17 in the two MP sessions, respectively.

C.2 Round 1 Only

Although subjects are randomly rematched within a session, no interaction has taken place before Round 1, so subject-level observations in Round 1 can be treated as independent. Tables C.3–C.5 report the Round 1 counterparts of Tables 4–6, and the last two columns of Table C.2 report Mann–Whitney U tests on Round 1 observations for the same 20 comparisons. Because each subject contributes a single observation, these tests have considerably lower power than those in the main text.

For overall and senders' informativeness, three patterns emerge that are consistent with the all-rounds analysis. First, the sign of every treatment difference in the main text is preserved in Round 1, with the single exception of the comparison between AP and AP-A in senders' informativeness, which is not significant in either sample. Second, our main finding—that information transmission is lower under partial commitment even when marginal incentives are aligned—also holds in Round 1. Both overall and senders' informativeness are lower in AP than in AF (p -value = 0.086 and p -value = 0.009, re-

Table C.4: Senders' informativeness by treatment: Round 1 only

		Commitment		
		Full	Partial	A-Partial
ALG	Data	0.49 [0.36, 0.61]	0.18 [0.00, 0.36]	0.26 [0.05, 0.48]
	Theory	0.54	0.54	0.54
MIS	Data	0.36 [0.20, 0.51]	0.07 [-0.14, 0.27]	-0.08 [-0.30, 0.14]
	Theory	0.54	0	0

Notes: The Data rows report the mean of $\phi(\theta, m)$ in Round 1 and, in brackets, its 95% confidence interval.

spectively), and both are lower in AP-A than in AF, although these differences are not significant (p -value = 0.309 and p -value = 0.104). Third, the collapse of communication under partial commitment when marginal incentives are misaligned also holds in Round 1. Both overall and senders' informativeness are significantly lower in MP and MP-A than in MF (p -value < 0.02 in all four comparisons), and senders' informativeness is significantly lower in MP-A than in AP-A (p -value = 0.037). Both measures are lower in MP than in AP, and overall informativeness is lower in MP-A than in AP-A, although these differences are not significant (p -value = 0.345, 0.356, and 0.133).

The comparison between AF and MF—the effect of misaligned incentives under full commitment—is the one case in which Round 1 differs qualitatively from the all-rounds sample. Senders' informativeness is lower in MF than in AF already in Round 1 (0.36 versus 0.49), consistent with the all-rounds results, although the difference is not significant (p -value = 0.163). Overall informativeness, by contrast, is 0.32 in both treatments (p -value = 0.948), unlike in the all-rounds results. The gap in overall informativeness between AF and MF thus appears to emerge over the course of the experiment, consistent with the dynamics in Figure D.2, in which receivers increasingly follow senders' messages in AF but not in MF.

Finally, we turn to receivers' perceived informativeness (Table C.5). Under Partial commitment, receivers already perceive significantly less information in MP than in AP

Table C.5: Receivers' perceived informativeness by treatment: Round 1 only

	Commitment	
	Partial	A-Partial
ALG	0.27 [0.14, 0.40]	0.08 [−0.16, 0.32]
MIS	0.07 [−0.10, 0.24]	−0.02 [−0.20, 0.16]

Notes: Each cell reports the mean of $\hat{\phi}(\theta, m)$ in Round 1 and, in brackets, its 95% confidence interval.

in Round 1 (0.07 versus 0.27, p -value = 0.048), consistent with Section 4.3.2. Under A-Partial commitment, perceived informativeness in Round 1 is 0.08 in AP-A and −0.02 in MP-A, and the difference is not significant (p -value = 0.555). The Round 1 value in AP-A is well below its average over all rounds (0.28), suggesting that perceived informativeness in AP-A increases over the course of the experiment.

D Additional Figures and Tables

D.1 The Dynamics of Players' Behavior

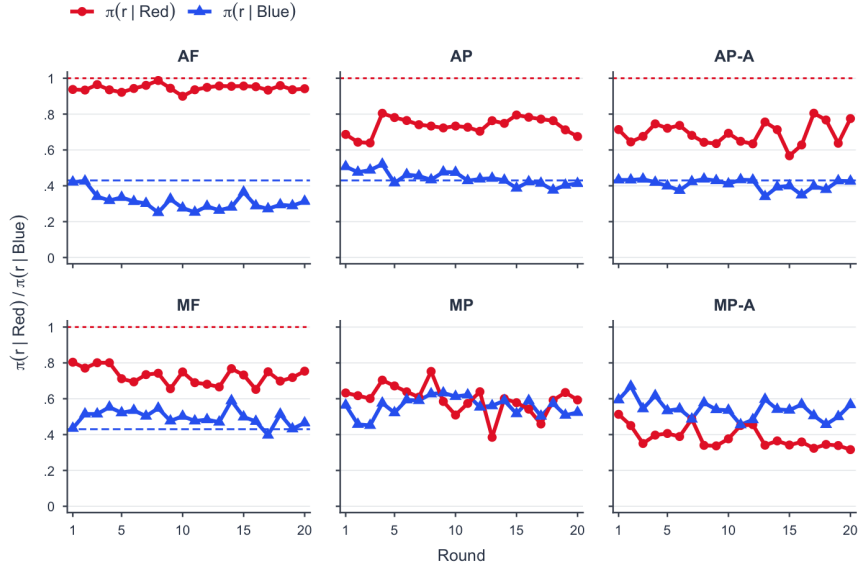


Figure D.1: Senders' strategy over rounds

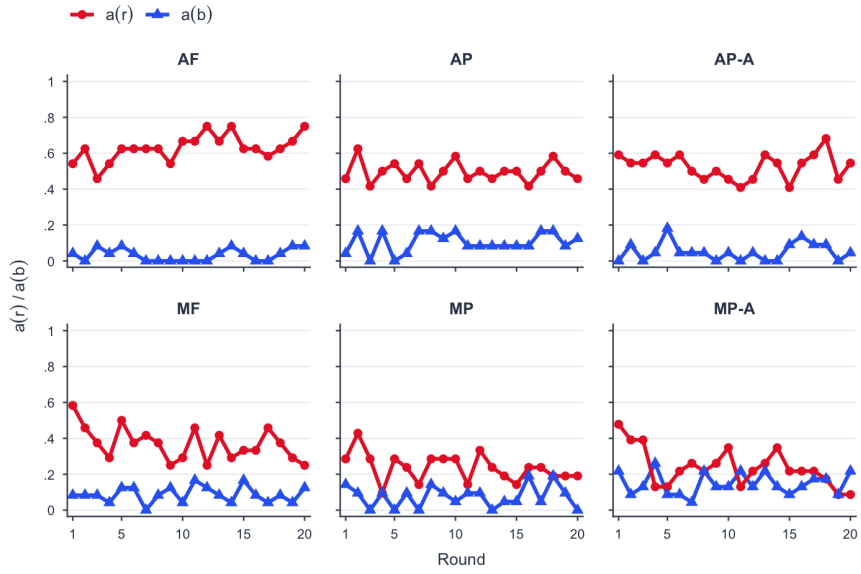


Figure D.2: Receivers' strategy over rounds

Table D.1: Senders' informativeness over rounds

	(1) AF	(2) MF	(3) AP	(4) AP-A	(5) MP	(6) MP-A
Round	0.0044* (0.0023)	-0.0009 (0.0053)	0.0073 (0.0058)	0.0032 (0.0045)	-0.0051 (0.0075)	-0.0010 (0.0058)
Sender fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	480	480	480	440	420	460
R^2	0.529	0.477	0.432	0.457	0.321	0.322

Notes: The dependent variable is senders' informativeness $\phi(\theta, m)$, measured at the sender-round level. Round is the round number, from 1 to 20. All regressions are estimated by OLS with sender fixed effects. Standard errors clustered at the sender level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D.2 Additional Results on Senders' Behavior

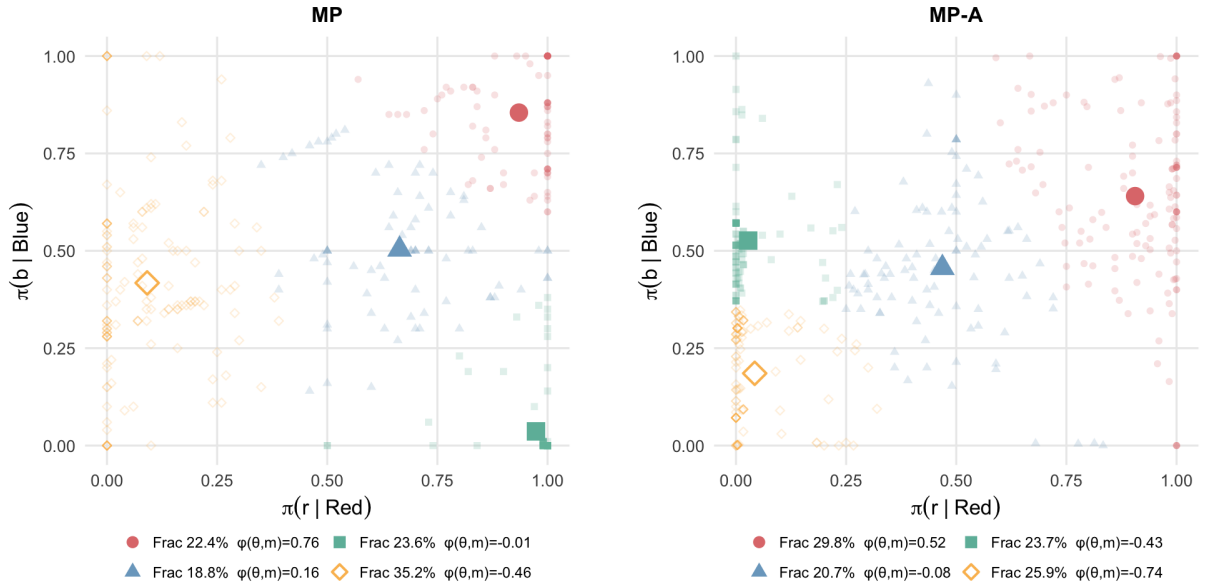


Figure D.3: Clustering senders' information structure in MP and MP-A

D.3 Additional Results on Receivers' Behavior

Table D.2: Receivers' responses to objective and subjective posteriors

	(1) AF	(2) MF	(3) AP	(4) AP-A	(5) MP	(6) MP-A
Panel A. Objective Posterior						
$\mu(r)$	1.435*** (0.128)	1.065*** (0.126)	0.144 (0.102)	0.426*** (0.086)	0.116 (0.100)	-0.045 (0.073)
Observations	480	472	480	439	413	459
R^2	0.493	0.465	0.384	0.454	0.232	0.179
Panel B. Subjective Posterior						
$\hat{\mu}(r)$			0.665*** (0.157)	0.806*** (0.104)	0.568** (0.203)	0.618*** (0.126)
Observations			480	439	413	459
R^2			0.451	0.555	0.286	0.289
Panel C. Objective and Subjective Posterior						
$\mu(r)$			-0.081 (0.087)	0.184** (0.077)	0.009 (0.098)	-0.058 (0.061)
$\hat{\mu}(r)$			0.706*** (0.149)	0.721*** (0.102)	0.564** (0.215)	0.619*** (0.125)
Observations			480	439	413	459
R^2			0.453	0.565	0.286	0.290
Receiver fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is $a(r)$, an indicator equal to one if the receiver chooses the action *red* after observing message r . $\mu(r)$ denotes the objective posterior probability of state Red implied by the sender's actual information structure, while $\hat{\mu}(r)$ denotes the subjective posterior probability implied by the receiver's elicited belief about the sender's information structure. Panels B and C are estimated only under Partial and A-Partial commitment (AP, AP-A, MP, MP-A), in which receivers' beliefs about the sender's information structure are elicited. Observations in which the sender's information structure assigns zero probability to message r in both states (so that $\mu(r)$ and $\hat{\mu}(r)$ are undefined) are excluded: 8 in MF, 1 in AP-A, 7 in MP, and 1 in MP-A. All regressions are linear probability models with receiver fixed effects. Standard errors clustered at the receiver level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D.3: Correlation between messages and actions by treatment

	Commitment		
	Full	Partial	A-Partial
ALG	0.60 [0.49, 0.71]	0.38 [0.25, 0.51]	0.48 [0.35, 0.61]
MIS	0.28 [0.18, 0.38]	0.14 [0.06, 0.23]	0.09 [−0.02, 0.20]

Notes: Each cell reports the mean of $\phi(m, a) := \text{Corr}(m, a; \pi)$, the correlation between the sender’s message and the receiver’s action, and, in brackets, its 95% confidence interval, based on receiver-level aggregation. This measure quantifies the extent to which receivers follow the sender’s messages.

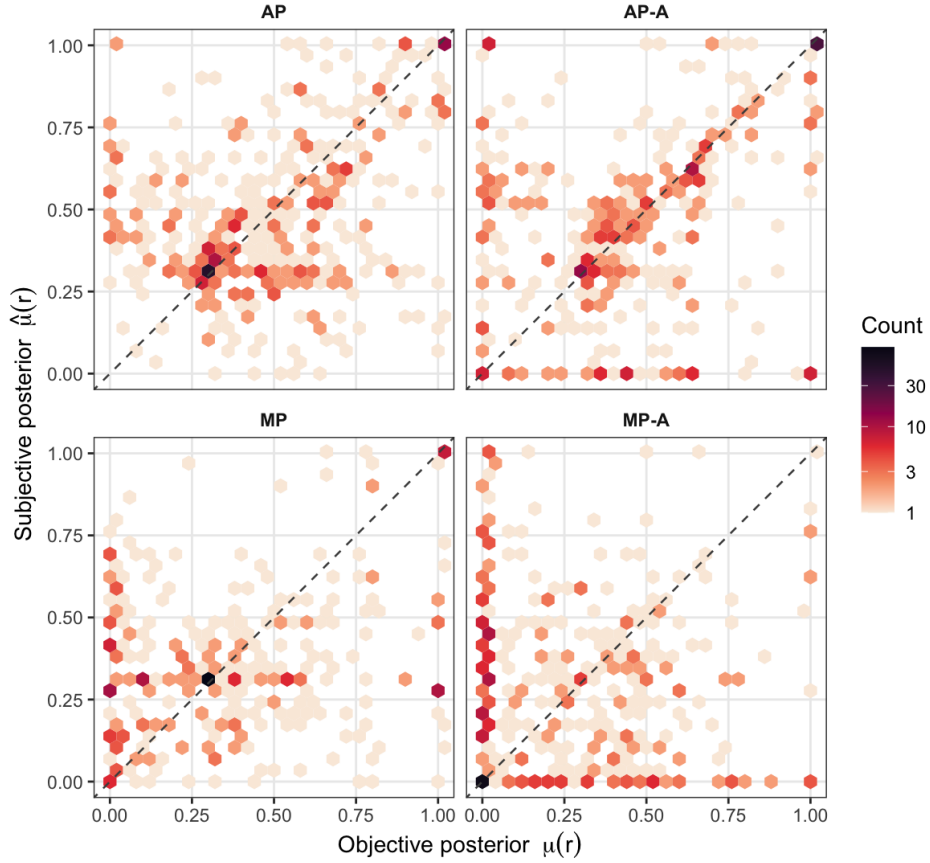


Figure D.4: Accuracy of receivers’ subjective beliefs

Notes: Each panel plots the objective posterior $\mu(r)$, implied by the sender’s actual information structure, against the subjective posterior $\hat{\mu}(r)$, implied by the receiver’s elicited belief about the sender’s information structure, for message r . The dashed line is the 45-degree line, along which the two posteriors coincide; the closer an observation lies to this line, the more accurate the receiver’s belief. Observations are aggregated into hexagonal bins, and the color of each bin indicates the number of observations on a logarithmic scale.

D.4 Individual Characteristics and Behavior in Games

Table D.4: Individual characteristics and behavior in games

	Senders' informativeness	Receiver Bayesian consistency	
	(1) AF, MF, AP, AP-A	(2) AF, MF	(3) AP, AP-A
Risk aversion	−0.031 (0.031)	−0.021 (0.016)	−0.018 (0.025)
Advantageous aversion	−0.055* (0.033)	0.016 (0.018)	−0.009 (0.027)
Disadvantageous aversion	−0.027 (0.033)	−0.009 (0.017)	−0.016 (0.027)
Numeracy	0.027 (0.031)	0.037** (0.017)	0.003 (0.026)
Treatment fixed effects	Yes	Yes	Yes
Observations	94	48	46
R^2	0.25	0.17	0.08

Notes: The dependent variable in column (1) is senders' informativeness $\phi(\theta, m)$, averaged across rounds at the sender level. The sample pools the four treatments in which the KG equilibrium is predicted (AF, MF, AP, AP-A). The dependent variable in columns (2) and (3) is the receiver-level Bayesian consistency rate: the share of choices (pooling both messages r and b across rounds) that coincide with the optimal threshold action given the posterior the receiver perceives. Under Full commitment (column 2), consistency is evaluated against the objective posterior $\mu(m)$ implied by the actual information structure π . Under Partial and A-Partial commitment (column 3), consistency is evaluated against the subjective posterior $\hat{\mu}(m)$ implied by the receiver's elicited belief $\hat{\pi}$ about the sender's information structure. Thus, column (2) measures Bayesian consistency with respect to the actual information structure, whereas column (3) measures Bayesian consistency with respect to the information structure believed by the receiver. All explanatory variables are standardized within each estimation sample. Risk aversion is measured by the switching point in Question 2 of Online Appendix E.2. Advantageous aversion is measured by the switching point in Question 1B of Online Appendix E.2, and Disadvantageous aversion by the switching point in Question 1A. Numeracy is measured by the number of correct answers to the two questions in Online Appendix E.3. All specifications are estimated by OLS with treatment fixed effects; standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

E Experimental Materials

This section provides English translations of the experimental materials. Online Appendix [E.1](#) presents the instructions for the sender-receiver game as implemented in AP-A. Online Appendix [E.2](#) presents the instructions for the common tasks on risk and social preferences administered after the sender-receiver game. Online Appendix [E.3](#) presents screenshots of the Berlin Numeracy Test.

E.1 Instruction for Sender-Receiver Game

Experiment Overview

- This experiment consists of two parts (Part 1 and Part 2). Instructions for Part 2 will be provided after Part 1 has been completed.
- At the end of the experiment, each participant will receive an individual payment.
- Your payment consists of a participation fee of 500 yen plus an additional payment determined by the outcomes of your decisions in the experiment.
- A portion of the points you earn during the experiment will later be converted into the additional payment according to the conversion rule explained later.

Experiment Part 1

- At the beginning of the experiment, all participants are randomly assigned by the computer to one of two roles: **Sender** or **Receiver**.
- Once assigned, your role will remain the same throughout the experiment.
- In each round, one Sender and one Receiver are randomly matched to form a pair and complete the task described below. The task is repeated for 20 rounds.
- Pairs are randomly rematched by the computer at the beginning of every round.

- At the beginning of each round, the computer randomly draws one ball from a box containing a total of 10 balls: 3 Red balls and 7 Blue balls.
 - Probability that a Red ball (Blue ball) is drawn: 30% (70%)
- Neither the Sender nor the Receiver directly observes the color of the ball drawn by the computer. Instead, they perform the following tasks:
 - Sender: Sends a message to the Receiver regarding the color of the ball drawn by the computer.
 - Receiver: Uses the Sender's message to guess the color of the ball that was drawn.
- The points earned by both the Sender and the Receiver depend on the Receiver's guess and the actual color of the ball drawn by the computer.
- Sender's payoff:
 - If the Receiver guesses red and the drawn ball is Red: 200 points.
 - If the Receiver guesses red and the drawn ball is Blue: 100 points.
 - If the Receiver guesses blue, regardless of the actual color: 80 points.
- Receiver's payoff:
 - If the Receiver guesses red and the drawn ball is Red: 150 points.
 - If the Receiver guesses blue and the drawn ball is Blue: 150 points.
 - In all other cases: 80 points.
- Thus, the Sender earns 100 points whenever the Receiver guesses red, and earns an additional 100 points if the drawn ball is actually Red. If the Receiver guesses blue, the Sender earns 80 points regardless of the actual color.
- In contrast, the Receiver earns 150 points for correctly guessing the color of the ball and 80 points otherwise.

- The payoff structure is also summarized in the table below.

	Receiver guesses red		Receiver guesses blue	
Red ball is drawn	Sender 200 pt	Receiver 150 pt	Sender 80 pt	Receiver 80 pt
Blue ball is drawn	Sender 100 pt	Receiver 80 pt	Sender 80 pt	Receiver 150 pt

- Each round consists of the following four stages:

Stage 1: Designing a Message Plan (Sender)

Stage 2: Revising a Message Plan (Sender)

Stage 3: Guessing the Color of the Ball (Receiver)

Stage 4: Guessing the Sender's Message Plan (Receiver)

- Each stage has a time limit of 1 minute.
- The page will not automatically advance when the time limit expires. However, please try to complete each task within the allotted time.
- If you exceed the time limit by a substantial amount, an experimenter may remind you to proceed.
- The following sections provide detailed instructions for each stage.

Stage 1: Designing a Message Plan

- In this stage, the Sender designs a **message plan**.
- A message plan specifies the probability with which the Sender will send either the message “**red**” or “**blue**” to the Receiver, conditional on the color of the ball drawn by the computer.
- Specifically, the Sender specifies the message sending probabilities for the following two cases:

If a **Red** ball is drawn:

- Probability of sending the message “**red**”: $X\%$
- Probability of sending the message “**blue**”: $(100 - X)\%$

If a **Blue** ball is drawn:

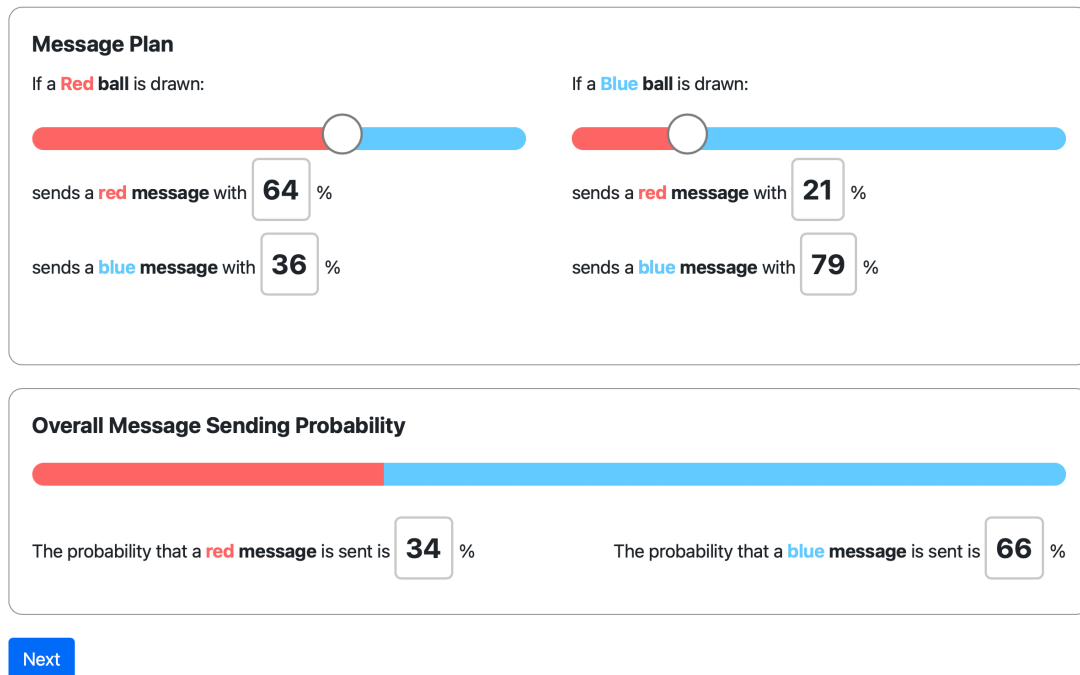
- Probability of sending the message “**red**”: $Y\%$
- Probability of sending the message “**blue**”: $(100 - Y)\%$

- The figure below shows part of the screen that the Sender uses in Stage 1.

The figure shows a user interface for setting message probabilities. It consists of two main panels. The top panel, titled "Message Plan", contains two sections. The left section is for the case "If a **Red** ball is drawn:" and features a horizontal slider with a red segment on the left and a blue segment on the right. Below the slider, there are two input fields: "sends a **red** message with -- %" and "sends a **blue** message with -- %". The right section is for the case "If a **Blue** ball is drawn:" and has a similar slider and two input fields: "sends a **red** message with -- %" and "sends a **blue** message with -- %". The bottom panel, titled "Overall Message Sending Probability", shows a single horizontal slider representing the overall probability distribution. Below it, there are two input fields: "The probability that a **red** message is sent is -- %" and "The probability that a **blue** message is sent is -- %". A blue "Next" button is located at the bottom left of the interface.

- The Sender sets these probabilities by adjusting the two sliders in the first panel, corresponding to the cases where a **Red** ball or a **Blue** ball is drawn.
- The probabilities selected using the sliders are displayed below them in real time as percentages.
- In addition, the second panel displays the overall probabilities that the messages “**red**” and “**blue**” will be sent under the entire message plan.
- We refer to these probabilities as the **overall message sending probabilities**.

- The overall message sending probabilities are calculated automatically from the message plan and are updated in real time whenever the sliders are adjusted.



- For example, suppose that the Sender designs the message plan shown above.

If a **Red** ball is drawn:

- Probability of sending the message “**red**”: 64%
- Probability of sending the message “**blue**”: 36%

If a **Blue** ball is drawn:

- Probability of sending the message “**red**”: 21%
- Probability of sending the message “**blue**”: 79%

- In this case, under the message plan as a whole, the message “**red**” will be sent with probability 34%, while the message “**blue**” will be sent with probability 66%.
- These are the overall message sending probabilities.
- The message plan designed in Stage 1 and the corresponding overall message sending probabilities are disclosed to the Receiver exactly as they are.

- Once you have finished designing your message plan, click the “Next” button in the lower-left corner of the screen. The message plan designed in Stage 1 will be announced to the Receiver, and the Sender will proceed to Stage 2.

Stage 2: Revising a Message Plan

- In this stage, the Sender may revise the message plan designed in Stage 1.
- The message actually sent to the Receiver will be determined according to the revised message plan created in this stage.
- The revised message plan will not be disclosed to the Receiver.
- However, when revising the message plan, the Sender cannot change the **overall message sending probabilities** implied by the message plan designed in Stage 1. The following example illustrates this restriction.
- Suppose that, as in the example on the previous page, the Sender designed and disclosed the following message plan in Stage 1:

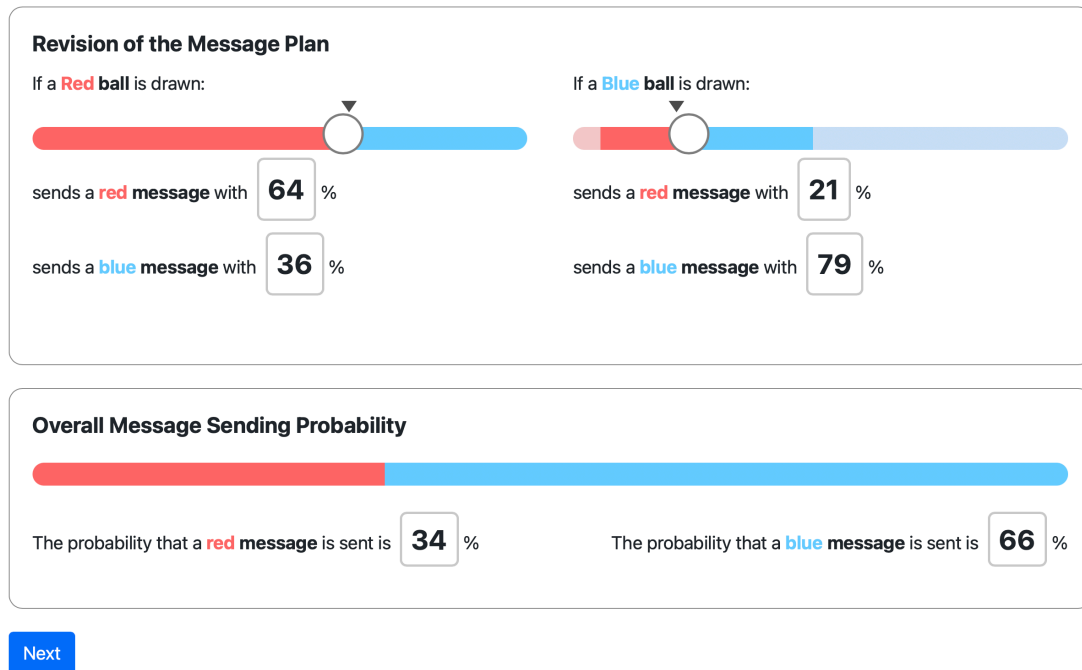
If a **Red** ball is drawn:

- Probability of sending the message “**red**”: 64%
- Probability of sending the message “**blue**”: 36%

If a **Blue** ball is drawn:

- Probability of sending the message “**red**”: 21%
- Probability of sending the message “**blue**”: 79%

- The figure below shows part of the screen that the Sender uses in Stage 2.



- Initially, the screen displays the message plan disclosed in Stage 1 together with the corresponding overall message sending probabilities.
- As in Stage 1, the Sender may revise the message plan by adjusting the sliders.
- However, the message plan may only be revised in ways that keep the overall message sending probabilities unchanged.
- When the Sender adjusts either slider, the other slider is automatically adjusted so that the overall message sending probabilities remain unchanged.
- The figure below illustrates an example of a revised message plan.

Revision of the Message Plan

If a **Red** ball is drawn:



sends a **red** message with **29** %

sends a **blue** message with **71** %

If a **Blue** ball is drawn:



sends a **red** message with **36** %

sends a **blue** message with **64** %

Overall Message Sending Probability



The probability that a **red** message is sent is **34** %

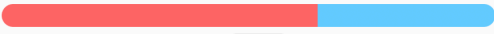
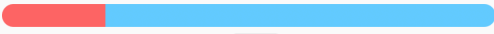
The probability that a **blue** message is sent is **66** %

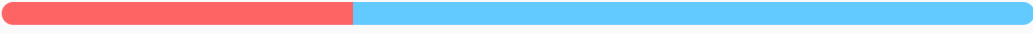
Next

- Note that although the individual message sending probabilities have changed, the overall message sending probabilities are identical before and after the revision.
- Also note that, in order to keep the overall message sending probabilities unchanged, part of each slider may be unavailable.
- In the example above, the slider corresponding to the case in which a **Blue** ball is drawn has a restricted range. The faded portion of the slider indicates probabilities that cannot be selected.
- The restricted range depends on the message plan designed in Stage 1.
- Once you have finished revising the message plan, click the “Next” button in the lower-left corner of the screen. This completes the Sender’s task.
- If you click “Next” without changing either slider from its initial position, the message plan designed in Stage 1 will be used without revision. However, the Receiver will not know whether the message plan was revised or left unchanged.

Stage 3: Guessing the Color of the Ball

- In this stage, the Receiver guesses the color of the ball drawn by the computer.
- The figure below shows part of the screen that the Receiver uses in Stage 3.

Message Plan
If a **Red** ball is drawn:

sends a **red** message with **64** %
sends a **blue** message with **36** %
If a **Blue** ball is drawn:

sends a **red** message with **21** %
sends a **blue** message with **79** %

Overall Message Sending Probability

The probability that a **red** message is sent is **34** % The probability that a **blue** message is sent is **66** %

Guess the Color of the Ball
If you receive a **red** message, your guess is
If you receive a **blue** message, your guess is

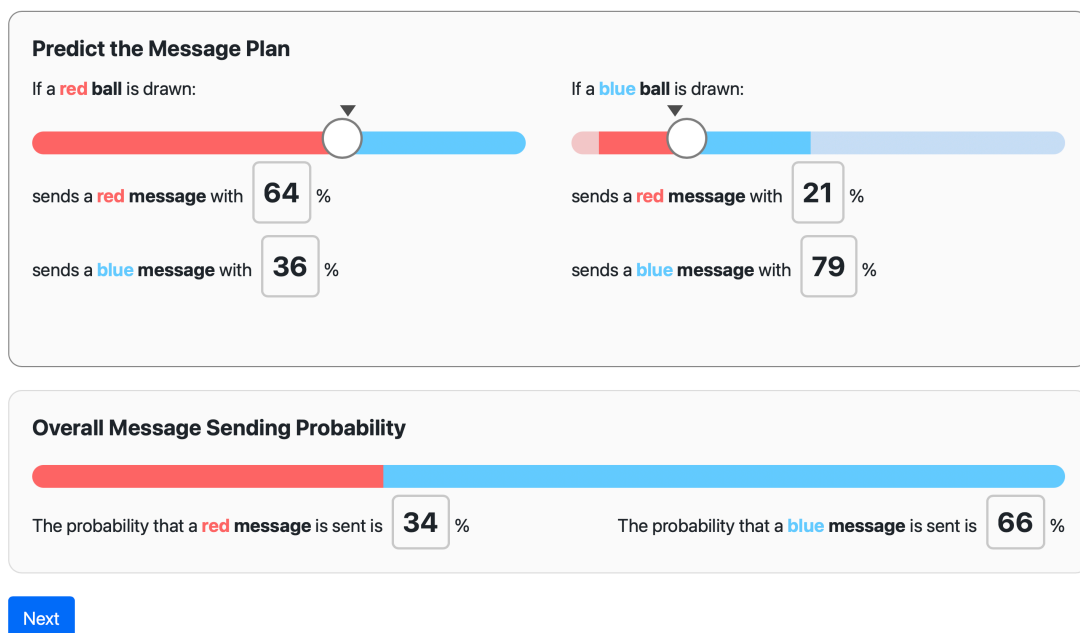
[Next](#)

- The top two panels display the message plan disclosed by the Sender in Stage 1 and the corresponding overall message sending probabilities.
- Note that the message plan disclosed by the Sender may have been partially revised in Stage 2.
- However, the overall message sending probabilities remain unchanged before and after the revision.
- After reviewing this information, the Receiver guesses the color of the ball in the third panel.

- The Receiver makes these guesses before observing the message sent by the Sender.
- Specifically, the Receiver guesses the color of the ball separately for each possible message that may be sent under the Sender's (revised) message plan: the message "red" and the message "blue."
- Once you have finished making your guesses, click the "Next" button in the lower-left corner of the screen. The Receiver will then proceed to Stage 4.

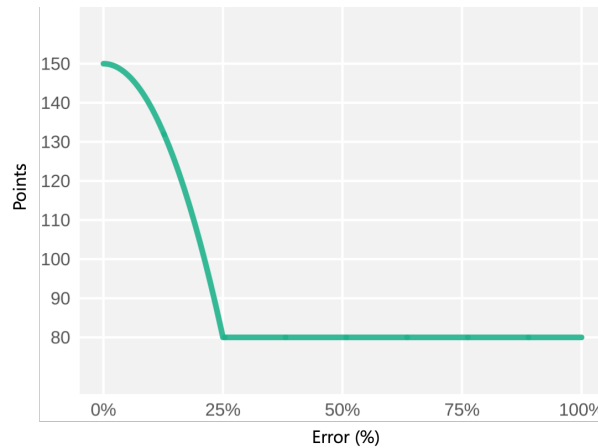
Stage 4: Guessing the Sender's Message Plan

- In this stage, the Receiver guesses the Sender's revised message plan.
- The figure below shows part of the screen that the Receiver uses in Stage 4.



- The sliders displayed in this stage function in the same way as those used by the Sender to revise the message plan in Stage 2.
- The Receiver guesses the Sender's revised message plan by adjusting the sliders.
- In addition to the points earned for guessing the color of the ball, the Receiver can earn points based on the accuracy of this guess.

- Specifically, the more accurately the Receiver guesses the probability with which the Sender sends the message “red” when a Red ball is drawn, the more points the Receiver earns.
- The Receiver earns a maximum of 150 points if the guess is exactly correct, and a minimum of 80 points if the prediction error is at least one quarter of the feasible range.
- Points are rounded up to the nearest integer.
- The relationship between the prediction error and the points earned is illustrated in the graph below.



- Once you have finished making your guess, click the “Next” button in the lower-left corner of the screen. This completes all tasks for the current round. At the end of the round, the results for that round will be displayed.

Results for Each Round

- The figure below shows a screenshot of the results screen displayed at the end of each round.

Ball Color:	Blue
Sender's Message:	Blue
Receiver's Guess:	Blue
Sender's Payoff	80 points
Receiver's Payoff	150 points or 145 points

Next

- From top to bottom, the screen displays:
 - the ball drawn by the computer,
 - the message actually sent under the (revised) message plan,
 - the Receiver's guess, and
 - the points earned by the Sender and the Receiver in that round.
- The points earned for the Sender and the Receiver are determined according to the payoff table on page 2 (kept as in the original instruction) based on the color of the drawn ball and the Receiver's guess.
- In addition, the Receiver earns points based on the accuracy of the message-plan guess made in Stage 4.
- Accordingly, the Receiver's results screen displays two types of points: the points shown on the left are those earned for guessing the color of the ball, while the points shown on the right are those earned for guessing the Sender's message plan.
- After reviewing the results, click the "Next" button in the lower-left corner of the screen. The current pair will then be dissolved, participants will be randomly re-matched by the computer, and the next round will begin (starting from Stage 1).

Conversion to Additional Payment

- We now explain how the points earned during the experiment are converted into your additional payment.
- First, the computer randomly selects one of the 20 rounds.
- The following points from the selected round are then used to determine the additional payment:
 - Sender: the points earned in that round.
 - Receiver: one of the two types of points earned in that round, selected at random by the computer.
- The selected points are converted into Japanese yen at the following rate:

$$1 \text{ point} = 10 \text{ yen}$$

- For example, suppose that a particular round is selected and its outcome is the same as shown in the figure on the previous page. Then the additional payment would be:
 - Sender: 800 yen.
 - Receiver: either 1,500 yen or 1,450 yen.
- This additional payment is added to the participation fee of 500 yen.

E.2 Individual Characteristics

Experiment Part 2

In Part 2, you will answer three questions (Questions 1A, 1B, and 2).

Questions 1A and 1B

- In Questions 1A and 1B, you will compare alternative monetary allocations between yourself and another participant in a hypothetical two-person pair and choose the allocation you consider more desirable.
- The figure below shows the actual decision screen for Question 1A.

1	You: 1,000 yen / Other: 1,000 yen	or	You: 800 yen / Other: 1,300 yen
2	You: 1,000 yen / Other: 1,000 yen	or	You: 900 yen / Other: 1,300 yen
3	You: 1,000 yen / Other: 1,000 yen	<u>or</u>	You: 1,000 yen / Other: 1,300 yen
4	You: 1,000 yen / Other: 1,000 yen	or	You: 1,100 yen / Other: 1,300 yen
5	You: 1,000 yen / Other: 1,000 yen	or	You: 1,200 yen / Other: 1,300 yen

- Question 1A consists of five pairs of allocation options. For each pair, please choose whether you prefer the allocation on the left or the one on the right.
- For example, in the first pair you choose between:
 - Left: You receive 1,000 yen and the other participant receives 1,000 yen.
 - Right: You receive 800 yen and the other participant receives 1,300 yen.
- In every pair, the left option is fixed at “You: 1,000 yen; Other participant: 1,000 yen.” The right option is “You: X yen; Other participant: 1,300 yen,” where X varies from 800 yen to 1,200 yen in increments of 100 yen.
- The figure below shows an example of the actual decision screen for Question 1B.

1	You: 1,000 yen / Other: 1,000 yen	or	You: 800 yen / Other: 700 yen
2	You: 1,000 yen / Other: 1,000 yen	or	You: 900 yen / Other: 700 yen
3	You: 1,000 yen / Other: 1,000 yen	or	You: 1,000 yen / Other: 700 yen
4	You: 1,000 yen / Other: 1,000 yen	or	You: 1,100 yen / Other: 700 yen
5	You: 1,000 yen / Other: 1,000 yen	or	You: 1,200 yen / Other: 700 yen

- Question 1B also consists of five pairs of allocation options. For each pair, please choose whether you prefer the allocation on the left or the one on the right.
- For example, in the first pair you choose between:
 - Left: You receive 1,000 yen and the other participant receives 1,000 yen.
 - Right: You receive 800 yen and the other participant receives 700 yen.
- In every pair, the left option is fixed at “You: 1,000 yen; Other participant: 1,000 yen.” The right option is “You: X yen; Other participant: 700 yen,” where X varies from 800 yen to 1,200 yen in increments of 100 yen.
- In all questions, please choose the option that you personally consider more desirable. There are no right or wrong answers.
- To simplify your decisions, your choices must be monotonic. If you select the left option for a given pair, all pairs above it will automatically be assigned the left option. Likewise, if you select the right option for a given pair, all pairs below it will automatically be assigned the right option.

Question 2

- The figure below shows the actual decision screen for Question 2.

1	10% chance of 820 yen, 90% chance of 660 yen	or	10% chance of 1590 yen, 90% chance of 40 yen
2	20% chance of 820 yen, 80% chance of 660 yen	or	20% chance of 1590 yen, 80% chance of 40 yen
3	30% chance of 820 yen, 70% chance of 660 yen	or	30% chance of 1590 yen, 70% chance of 40 yen
4	40% chance of 820 yen, 60% chance of 660 yen	or	40% chance of 1590 yen, 60% chance of 40 yen
5	50% chance of 820 yen, 50% chance of 660 yen	or	50% chance of 1590 yen, 50% chance of 40 yen
6	60% chance of 820 yen, 40% chance of 660 yen	or	60% chance of 1590 yen, 40% chance of 40 yen
7	70% chance of 820 yen, 30% chance of 660 yen	or	70% chance of 1590 yen, 30% chance of 40 yen
8	80% chance of 820 yen, 20% chance of 660 yen	or	80% chance of 1590 yen, 20% chance of 40 yen
9	90% chance of 820 yen, 10% chance of 660 yen	or	90% chance of 1590 yen, 10% chance of 40 yen
10	100% chance of 820 yen, 0% chance of 660 yen	or	100% chance of 1590 yen, 0% chance of 40 yen

- In Question 2, you will choose between two lotteries in each of ten pairs.
- For example, in the first pair you choose between:
 - Left lottery: a 10% chance of receiving 820 yen and a 90% chance of receiving 660 yen.
 - Right lottery: a 10% chance of receiving 1,590 yen and a 90% chance of receiving 40 yen.
- In every pair, the prize amounts remain fixed, while the probabilities of receiving those prizes change in increments of 10 percentage points.
- For each pair, please choose the lottery that you personally consider more desirable. There are no right or wrong answers.
- As in Questions 1A and 1B, your choices must be monotonic. If you choose the left lottery for one pair, all pairs above it will automatically be assigned the left lottery. If you choose the right lottery for one pair, all pairs below it will automatically be assigned the right lottery.

Additional Payment

- This section explains how your responses in Part 2 determine your additional payment.
- First, the computer will randomly divide all participants into two groups (Group 1 and Group 2).
- Participants assigned to Group 1 will then be randomly matched into pairs. Within each pair, one participant will be assigned the role of “You” and the other the role of “Other.”
- For each pair, one of the ten decision pairs from Questions 1A and 1B will be randomly selected.
- The choice made by the participant assigned the role of “You” in the selected decision pair will determine the payments for both members of the pair.
- For example, suppose the fifth decision pair from Question 1A is selected and the participant assigned the role of “You” chose the right option. In that case, the participant assigned the role of “You” will receive 1,200 yen, while the participant assigned the role of “Other” will receive 1,300 yen. These amounts constitute the additional payment for Part 2.
- Participants assigned to Group 2 will instead have their payment determined by their own responses to Question 2.
- The computer will randomly select one of the ten lottery pairs from Question 2, and the lottery you chose in that pair will be played.
- The amount won in the lottery will be paid as your additional payment for Part 2.

E.3 Screen Shots for Survey

In a small town of 1,000 residents, 500 are members of the local choir, including 100 men.
Among the remaining 500 residents who are not choir members, 300 are men.
What is the probability that a randomly selected man is a member of the choir?
Please answer with an integer from 0 to 100.

%

Next

Figure E.1: Berlin Numeracy Test 1

In a certain forest, 20% of the mushrooms are red, 50% are brown, and 30% are white.
Among the red mushrooms, 20% are poisonous. Among the non-red mushrooms, 5% are poisonous.
What is the probability that a randomly selected poisonous mushroom is red?
Please answer with an integer from 0 to 100.

%

Next

Figure E.2: Berlin Numeracy Test 2