Controlled Cursedness*

Wooyoung Lim[†] Chen Zhao[‡]

January 30, 2026

Abstract

We establish a strategic equivalence between cursed equilibrium and the introduction of fictitious players in Bayesian games, allowing for controlled manipulation of cursedness in lab settings. We consider a cheap-talk setting involving one sender and multiple receivers, one real and several fictitious. The sender's type is payoff-relevant to the real receiver but not to the fictitious receivers; however, his message is shared with all receivers. Uninformed of her being real or fictitious, the real receiver will neglect the correlation between the message and the sender's type—she has cursed beliefs. By adjusting the number of fictitious receivers, our lab results align with the comparative statics predicted by cursed equilibrium.

Keywords: Cursed Equilibrium, Communication Games, Laboratory Experiments

***JEL* classification numbers**: C72, C91, D82, D90

*We are grateful to Andreas Blume, Shengwu Li, and Joel Sobel for their valuable comments and suggestions. For helpful comments and discussions, we thank the seminar participants at the Chinese University of Hong Kong and Singapore Management University, and the conference and workshop participants at the 2023 Asia Pacific ESA meeting. Zhao gratefully acknowledges financial support by the Research Grants Council of Hong Kong (GRF17505020).

[†]Department of Economics, The Hong Kong University of Science and Technology. *Email:* `wooyoung@ust.hk`

[‡]Faculty of Business and Economics, University of Hong Kong. *Email:* `czhao@hku.hk`

1 Introduction

Cursed equilibrium, a concept introduced by Eyster and Rabin (2005), has emerged as a powerful framework for understanding deviations from standard equilibrium predictions in various strategic environments. The applicability of cursed equilibrium has been demonstrated in a wide range of settings, including common value auctions (Kagel and Levin, 1986), revelation games (Forsythe, Isaac, and Palfrey, 1989), jury voting (Guarnaschelli, McKelvey, and Palfrey, 2000), signaling games (Szembrot, 2018; Lin and Tan, 2025), and adverse selection (Wenner, 2019). The notion of cursedness captures the limited strategic sophistication of players who may not fully comprehend the correlation between the actions of other players and their private information, and has garnered attention for its empirical success in explaining laboratory data. Recent work has extended the notion of cursedness to extensive form games (Cohen and Li, 2022; Fong, Lin, and Palfrey, 2023), further broadening its applicability and relevance in understanding strategic decision-making.

The primary motivation for this boundedly rational equilibrium concept arises from substantial evidence that individuals often overlook correlations among information sources in various contexts, including belief formation (Eyster and Weizsacker, 2016), portfolio choices (Enke and Zimmermann, 2019), voting (Moser and Wallmeier, 2021), and school selection (Tergiman, 2024). Despite the abundant evidence for correlation neglect, the notion of cursed equilibrium is often regarded as an "as if" model. Specifically, there is no guarantee that individuals employ an algorithm where they first evaluate the extent of their opponents' cursedness and subsequently formulate their own strategies conditional on this assessment. Instead, they might be using entirely different strategies that roughly align with this calculation. Indeed, it is an empirical question whether individuals' behavior in these strategic settings systematically depends on the perceived cursedness of their opponents. However, the unobservable nature of the degree of cursedness in the lab makes investigation barely possible, and thus, the existing experimental literature has made limited progress in establishing the validity of cursed equilibrium as an "as is" model.

In this paper, we introduce a novel approach to comprehensively address this question through the *controlled manipulation of cursedness* within a laboratory setting. In particular, we investigate whether shifts in players' behavior in response to changes in the perceived cursedness of their opponents align with the comparative statics predicted by cursed equilibrium.

The key instrument that allows us to manipulate the perceived cursedness of players is the introduction of *fictitious players* into a game. Instead of the real Bayesian game, fictitious players play an auxiliary game in which each player's payoff does not directly depends on other players' types. Specifically, the auxiliary game and the Bayesian game share the same type space and action space. Given a *fixed* action profile, in the Bayesian game, a real player i's payoff depends both on her own type θ_i and her opponents' types θ_{-i}; in the auxiliary game, in contrast, a fictitious player i with the same type θ_i gets a payoff equivalent to the expected payoff of the real player i in the Bayesian game conditional on θ_i—i.e. the fictitious player behaves as if they were a real player in the Bayesian game who completely ignores the information content in her opponents' actions about their types. As a result, if a player is unsure about whether she is real or fictitious, she will behave as if she neglects the correlation between her opponents' actions and their types partially—i.e. she behaves as if she is cursed. In our implementation, the uncertainty as to whether a player is real or fictitious will be commonly known to all players in the game. Thus, by increasing the probability that a player is fictitious, the degree of perceived cursedness of the player is increased, and vice versa.

The formal procedure goes as follows: First, we randomly label a fraction, say χ, of the subjects as fictitious players and the rest as real players, without informing the subjects of their roles. Second, we conduct the Bayesian game among real players and the auxiliary game among fictitious players, without informing the subjects of the game they are playing. The procedure, including the fraction of fictitious players, is common knowledge to every subject. By design, each player will hold the belief that they may be fictitious with probability χ and thus appear to be cursed.[1] We introduce the formal apparatus and present the theoretical results regarding the strategic equivalence between the Bayesian game and the auxiliary game in Section 2.

While the procedure can be applied to any Bayesian game, our natural focus is on the class of games in which there exists an *informed* player whose payoff only depends on her own type. To such an informed player, the uncertainty as to whether she is real or fictitious does not matter. Thus, by varying the fraction of fictitious players in the game (χ) and examining the shifts in the informed player's strategy, we may obtain cleaner and sharper comparative statics for our purpose.

We implemented the controlled manipulation of cursedness in a canonical yet simplified cheap-talk environment, involving a privately informed sender and an uninformed receiver (Crawford and Sobel, 1982). There are several reasons why we adopt this specific environment. Firstly, it is one of the simplest Bayesian games with an informed player. On the one

[1]Note that it is crucial that the equilibrium actions in the real game and also the fictitious game are not informative about whether the game being played is real or fictitious; otherwise the agents may update their beliefs and thus perceive a level of cursedness different from χ in equilibrium. Thus, it helps if the players in the real game and those in the fictitious game are ex ante identical.

hand, the sender is informed and thus the degree of cursedness of the sender is irrelevant. On the other hand, the sender's strategy is a mapping from his private information to the message space, ensuring that the cursedness of the receiver matters. Moreover, theoretically, this environment yields non-trivial comparative statics regarding both players' strategies and welfare across varying degrees of cursedness, enabling us to investigate the intricate relationship between cursedness and players' strategies and welfare. Previous work by Lee, Lim, and Zhao (2023) has demonstrated that an appropriate level of cursedness can enhance the overall welfare of the game relative to a curse-free environment (0-cursed). We reproduce these theoretical predictions within our simplified communication environment in Section 3.

A cursed receiver in cheap-talk games behaves as if she incorrectly believes that there is a positive probability that the sender's messages are completely uninformative. To simulate the receiver's cursedness, we implement the above procedure in an economical manner: We do not introduce fictitious senders but let each fictitious receiver get a message randomly drawn from the set of messages sent in the real Bayesian game. By design, these messages are completely uninformative about the fictitious receiver's payoff, and thus, a real receiver who does not know whether she is real or fictitious will behave as if she is cursed. Our main focus is on how the informativeness of the senders' messages in the real Bayesian game shifts according to the fraction of fictitious receivers. By varying the fraction of fictitious receivers, we create three experimental treatments: High Cursed (HC), Low Cursed (LC), and Standard Talk (ST), where the real receivers are cursed to different extents. Additionally, we include a benchmark treatment without fictitious receivers but with optimally mediated communication (Mediated Talk, MT). This additional treatment enables us to compare the effects of cursedness with those of standard information garbling devices in strategic communication. Theoretically, the notion of cursed equilibrium suggests that the senders' messages should be less informative in ST than those in all other treatments, and should be equally as informative among LC, HC, and MT.

Our experimental data provide strong empirical support for the comparative statics predicted by cursed equilibrium theory. Firstly, session-level data reveals that senders' messages in ST are indeed less informative than those in all other treatments, while the differences in the informativeness of messages are not statistically significant among LC, HC, and MT. Moreover, individual-level data confirms that senders in ST use a pooling strategy more and a separating strategy less than those in all other treatments, and such dominance is not observed between any other pair of treatments. Our data suggests that senders do take the cursedness of the receivers into account when formulating their strategies and in the meantime are not nudged into choosing different strategies across Treatments LC, HC, and MT. Such behavior is consistent with the predictions of cursed equilibrium theory.

Encouragingly, the behavior of the receivers in our treatments qualitatively align with the predictions of cursed equilibrium theory, which indicates that our methodology effectively simulates cursed behavior. Specifically, session-level data suggests that receivers' strategies in Treatment HC exhibit less separation relative to all other treatments. In addition, individual-level analysis demonstrates that receivers in HC employ pooling strategies more frequently and separating strategies less frequently than those in other treatments, and such dominance is not observed between any other pairs of treatments. These behavioral patterns are consistent with the predictions of cursed equilibrium theory when receivers possess an intrinsic level of cursedness in addition to the simulated cursedness in our treatments.

Lastly, the outcome and welfare patterns obtained in our treatments also conforms to cursed equilibrium theory. We find that the outcome—i.e. the distribution of real receivers' actions conditioning on the state—in Treatment LC exhibits a higher degree of separation compared to the outcomes in Treatments ST and HC. In addition, the average earnings of the real receivers are significantly smaller in Treatment ST than in each of Treatments MT, LC, and HC, while the average earnings of the real receivers are higher in Treatment LC than in each of Treatments ST, MT, and HC, although these differences are not statistically significant.

We conduct a simple calibration exercise under the assumption that the intrinsic cursedness follows a truncated normal distribution among receivers. We find that concentrating the degree of intrinsic cursedness around 0.46 can simultaneously rationalize the observed behavior, outcome and welfare patterns in our data. Hence, our findings provide valuable empirical evidence that supports the predicted outcomes of the cursed equilibrium. Moreover, they illustrate the usefulness of our strategic equivalence and the incorporation of fictitious players as a robust framework for studying and manipulating cursedness in laboratory experiments.

The paper is organized as follows. Section 2 establishes the strategic equivalence between cursed equilibrium and the injection of fictitious players in a broad class of Bayesian games. In Section 3, we provide a detailed description of the simplified communication game environment. The experimental design is presented in Section 4, followed by a set of testable hypotheses in Section 5. The experimental results are reported in Section 6, and we present the results from the calibration exercise to measure the degree of intrinsic cursedness in Section 7.

2 Cursedness and Fictitious Players

Consider a standard Bayesian game $G = (N, \Theta, Y, u, \pi)$, in which $N = \{1, \dots, n\}$ is a finite set of players, $\Theta = \times_{i=1}^n \Theta_i$ is the set of possible type profiles, $Y = \times_{i=1}^n Y_i$ is the set of possible action profiles, $u = (u_1, \dots, u_n)$ is the profile of payoff functions, and $\pi \in \Delta(\Theta)$ is the (full support) common prior shared by all players. For each $i \in N$, Θ_i and Y_i are the finite set of possible types and actions of player i, respectively, and $u_i : \Theta \times Y \to \mathbb{R}$ is her payoff function. A (mixed) strategy for player i, $\sigma_i : \Theta_i \to \Delta(Y_i)$, specifies a probability distribution over actions for each type.

For each $i \in N$, let Θ_{-i} and Y_{-i} be respectively the possible type profiles and action profiles without player i. Player i's belief about other players' types given her own type $\theta_i \in \Theta_i$ is $\pi_i(\cdot|\theta_i) \in \Delta(\Theta_{-i})$, which is pinned down by π and Bayes' rule. For any strategy profile σ and each $i \in N$, let $\sigma_{-i}(\cdot|\theta_{-i}) \in \times_{j \neq i} \Delta(Y_j)$ be the distribution of actions that player i's opponents under σ when their types are $\theta_{-i} \in \Theta_{-i}$; let $\overline{\sigma}_{-i}(y_{-i}|\theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} \sigma_{-i}(y_{-i}|\theta_{-i}) \pi_i(\theta_{-i}|\theta_i)$ be the conditional probability that y_{-i} is played when player i is of type θ_i and all of player i's opponents follow σ.

To define the notion of cursedness, given any σ and $i \in N$, let $\pi_i(\cdot|\theta_i, y_{-i}, \sigma_{-i}) \in \Delta(\Theta_{-i})$ be the correct posterior belief player i should have if the rest of the players play $y_{-i} \in Y_{-i}$ under strategy σ, and $\hat{\pi}_i(\cdot|\theta_i, y_{-i}, \sigma_{-i}) \in \Delta(\Theta_{-i})$ be the posterior belief player i really has. In particular, for any $\chi \in [0, 1]$, we say that player i is χ-*cursed* if

$$\hat{\pi}_i(\cdot|\theta_i, y_{-i}, \sigma_{-i}) = \hat{\pi}_i^{\chi}(\cdot|\theta_i, y_{-i}, \sigma_{-i}) := \chi \pi_i(\cdot|\theta_i) + (1 - \chi) \pi_i(\cdot|\theta_i, y_{-i}, \sigma_{-i})$$

for all $\theta_i \in \Theta_i$, $y_{-i} \in Y_{-i}$, and strategy profile σ such that $\overline{\sigma}_{-i}(y_{-i}|\theta_i) > 0$. We say that a player is *Bayesian* if she is 0-cursed.

Now we define the χ-cursed equilibrium as in Eyster and Rabin (2005).

Definition 1. *For any $\chi \in [0, 1]$, a strategy profile σ is a χ-cursed equilibrium if for each $i \in N$, $\theta_i \in \Theta_i$, and each y_i^* such that $\sigma_i(y_i^*|\theta_i) > 0$,*

$$y_i^* \in \arg\max_{y_i \in Y_i} \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{y_{-i} \in Y_{-i}} u_i(\theta_i, \theta_{-i}, y_i, y_{-i}) \hat{\pi}_i^{\chi}(\theta_{-i}|\theta_i, y_{-i}, \sigma_{-i}) \overline{\sigma}_{-i}(y_{-i}|\theta_i).$$

In order to implement cursed beliefs among the players, we first introduce an auxiliary game as follows. For all $i \in N$, $(\theta_i, \theta_{-i}) \in \Theta$, and $(y_i, y_{-i}) \in Y$, define

$$\overline{u}_i(\theta_i, y_i, y_{-i}) = \sum_{\theta'_{-i} \in \Theta_{-i}} u_i(\theta_i, \theta'_{-i}, y_i, y_{-i}) \pi_i(\theta'_{-i}|\theta_i).$$

Thus, $\overline{u}_i$ the average payoff to player i if her type is θ_i and (y_i, y_{-i}) is played. In other words, $\overline{u}_i$ is independent of θ_{-i}. Given strategy profile σ of G, let the auxiliary game $G^* = (N, \Theta, Y, \overline{u}, \pi)$. Essentially, in G^*, each player behaves as if she learns nothing about her opponent's type from their actions. Thus, for every action profile (y_i, y_{-i}) and type θ_i, the expected payoff of player i is taken with respect to $\pi_i(\cdot|\theta_i)$, irrespective to the action profile played.

Now for any $\tilde{\chi} \in [0,1]$, let $G^{\tilde{\chi}} = \tilde{\chi}G^* + (1-\tilde{\chi})G$. Thus, in $G^{\tilde{\chi}}$, nature moves first to select the game to be played, of which the players are not informed: The auxiliary game G^* is played with probability $\tilde{\chi}$, and G is played with probability $1-\tilde{\chi}$. The following observation is immediate.

Proposition 1. *For any $\chi \in [0,1]$, a strategy profile σ is a Bayesian Nash equilibrium of G^{χ} if and only if it is a χ-cursed equilibrium of G.*

Thus, when all players are χ-cursed, G and G^{χ} are strategically equivalent in the sense of the proposition above. However, the welfare considerations of G and G^{χ} can be quite different. In G, the χ-cursed players behave as if they are Bayesian but incorrectly believe that G^{χ} is played. Thus, when evaluating the ex ante welfare of these players, the payoffs in the auxiliary game G^* should be irrelevant. To be more specific, given any strategy profile σ, the ex ante welfare of player i in G that is χ-cursed, denoted as $\mathrm{EU}_i^{\sigma}(G,\chi)$, should be given by

$$\mathrm{EU}_i^{\sigma}(G,\chi) = \sum_{(\theta_i,\theta_{-i})\in\Theta} \sum_{(y_i,y_{-i})\in Y} u_i(\theta_i,\theta_{-i},y_i,y_{-i})\pi_i(\theta_{-i}|\theta_i,y_{-i},\sigma_{-i})\overline{\sigma}_{-i}(y_{-i}|\theta_i)\sigma_i(y_i|\theta_i)\pi_i(\theta_i)$$

in which $\pi_i(\theta_i)$ denotes the prior probability that player i draws type θ_i. Note that in the evaluation of $\mathrm{EU}_i^{\sigma}(G,\chi)$, π_i is used instead of $\hat{\pi}_i$, and, as a result, $\mathrm{EU}_i^{\sigma}(G,\chi)$ does not depend on χ. Since π_i takes into account the correct correlation between θ_{-i} and y_{-i}, using π_i gives rise to the real average payoff that player i would get from the game.

By contrast, the ex ante welfare of a Bayesian player i in G^{χ} given σ, denoted as $\mathrm{EU}_i^{\sigma}(G^{\chi},0)$, should be given by

$$\mathrm{EU}_i^{\sigma}(G^{\chi},0) = \chi\mathrm{EU}_i^{\sigma}(G^*,0) + (1-\chi)\mathrm{EU}_i^{\sigma}(G,\chi),$$

in which

$$
\begin{aligned}
\mathrm{EU}_i^\sigma(G^*,0) &= \sum_{\theta_i\in\Theta_i}\sum_{(y_i,y_{-i})\in Y}\overline{u}_i(\theta_i,y_i,y_{-i})\overline{\sigma}_{-i}(y_{-i}|\theta_i)\sigma_i(y_i|\theta_i)\pi_i(\theta_i)\\
&= \sum_{(\theta_i,\theta_{-i})\in\Theta}\sum_{(y_i,y_{-i})\in Y}u_i(\theta_i,\theta_{-i},y_i,y_{-i})\pi_i(\theta_{-i}|\theta_i)\overline{\sigma}_{-i}(y_{-i}|\theta_i)\sigma_i(y_i|\theta_i)\pi_i(\theta_i)
\end{aligned}
$$

is the ex ante welfare of a Bayesian player i in G^* given strategy profile σ. It is clear that $\mathrm{EU}_i^\sigma(G^*,0)$ may be different from $\mathrm{EU}_i^\sigma(G,\chi)$, precisely because the correlation between θ_{-i} and y_{-i} via σ_{-i} is not payoff-relevant in G^*, as if the players have completely neglected such correlation.

The exercise above suggests that we may introduce fictitious roles into games to simulate the effects of cursedness. Consider the following implementation of G^χ with a population of subjects:

1. Randomly label χ fraction of the subjects as the *fictitious* players and label the rest as the *real* players. Subjects are not informed of their roles.

2. Divide the real players into groups of n and play G. Divide the fictitious players into groups of n and play G^*. Subjects are not informed of the game that they are playing.

Let the ex ante welfare for a real, Bayesian player i in the experiment above given strategy profile σ be $\mathrm{EU}_i^\sigma(G^\chi,0|\text{real})$. The following proposition enables us to analyze the welfare implications of cursedness in G by analyzing the real players in the experiment above.

Proposition 2. *Suppose all subjects are Bayesian. Then for all $i\in N$, $\chi\in[0,1]$, and strategy profile σ, we have $EU_i^\sigma(G^\chi,0|real) = EU_i^\sigma(G,\chi)$.*

Our goal is to implement cursedness in the lab by varying χ in G^χ. In reality, it is plausible that the subjects may bring some level of intrinsic cursedness, denoted as χ_0, to the lab. Thus, the following results will be useful in our calibration exercise.

Proposition 3. *The following statements are true for all $\chi_0,\chi\in[0,1]$ and $\hat{\chi}=\chi_0+(1-\chi_0)\chi$:*

(i) *A strategy profile σ is a χ_0-cursed equilibrium of G^χ if and only if it is a $\hat{\chi}$-cursed equilibrium of G.*

(ii) *Suppose all subjects are χ_0-cursed. Then for all $i\in N$ and strategy profile σ, we have $EU_i^\sigma(G^\chi,\chi_0|real) = EU_i^\sigma(G,\hat{\chi})$, in which $EU_i^\sigma(G^\chi,\chi_0|real)$ denotes the ex ante welfare for a real, χ_0-cursed player i in the experiment given strategy profile σ.*

To see how $\hat{\chi}$ is derived, note that in any game the posterior of a χ_0-cursed player is a χ_0-weighted average of her prior and the correct Bayesian posterior. In addition, a Bayesian player in G^χ behaves as if she is a χ-cursed player in G, whose posterior belief is a χ-weighted average of the prior and the posterior of a Bayesian player in G. Combining these two observations, a χ_0-cursed player in G^χ behaves as if she is playing G with a posterior belief that assigns a weight of $\chi_0 + (1 - \chi_0)\chi$ to the prior and the remaining weight to the posterior of a Bayesian player in G—the χ_0-cursed player in G^χ behaves as if she is a $\hat{\chi}$-cursed player in G with $\hat{\chi} = \chi_0 + (1 - \chi_0)\chi$.

Hence, using the experimental procedure described above, we will be able to obtain comparative statics regarding how cursedness affects welfare in G by analyzing how χ affects the welfare of the real players in the experiment.

3 A Simplified Strategic Communication Game

While the aforementioned procedure can be applied to any Bayesian game, our natural focus is on the class of games in which there exists an *informed* player whose payoff only depends on her own type. To such an informed player, the uncertainty as to whether she is real or fictitious does not matter. Thus, by varying the fraction of fictitious players in the game (χ) and examining the shifts in the informed player's strategy, we may obtain cleaner and sharper comparative statics for our purpose.

Consider a canonical yet simplified cheap-talk environment. There are two players, $N = \{S, R\}$, in which S is the sender and R is the receiver. The sender's type θ is drawn from $\Theta_S \in \{0, 1\}$. There is no uncertainty regarding the receiver's type and it is not payoff-relevant. Both players share the common prior $\pi \in \Delta(\Theta_S)$ such that that $\pi(1) = p \in (0, 1)$. The sender selects a message $m \in Y_S = \{0, 1\}$, and the receiver takes an action $y \in Y \subseteq [0, 1]$, after observing the message. Thus, the set of possible action plans for the receiver $Y_R = Y^{\{0,1\}}$, which is the set of all mappings from the message space $\{0, 1\}$ to Y. Given $\theta \in \{0, 1\}$, $y(\cdot) \in Y^{\{0,1\}}$, and $m \in \{0, 1\}$, The sender's payoff is $U_S(\theta, m, y(\cdot)) = -(y(m) - \theta - b)^2$, and the receiver's payoff is $U_R(\theta, m, y(\cdot)) = -(y(m) - \theta)^2$, in which $b \geq 0$.

3.1 Equilibrium Predictions when $Y = [0, 1]$

Now we present equilibrium predictions when $Y = [0, 1]$; that is, the receiver can choose any action from $[0, 1]$. Although we will further discretize Y in the actual experiment, these equilibrium predictions will serve as a guide for parameter selection.

3.1.1 Cursed Talk

To avoid issues related to off-equilibrium path beliefs, we will focus on equilibria in which all messages in Y_S are used. Fong, Lin, and Palfrey (2023) and Cohen and Li (2022) extend the idea of cursed equilibrium to extensive games. In each paper, a consistency requirement, similar to the one for standard sequential equilibrium, is imposed on beliefs off the equilibrium path. Fong et al. (2023) demonstrated that their consistency requirement mandates that the belief at any non-terminal history, regardless of whether it is on or off the equilibrium path, must assign a minimum weight of χ to the belief from the preceding period. In contrast, Cohen and Li (2022) allow each player to neglect the correlation between their opponents' actions and their private information, irrespective of whether this information is exogenously provided or endogenously acquired during the game. In our game, as is standard in cheap talk games, we can with out loss of generality assume that all messages are used in equilibrium so beliefs off the equilibrium path do not play a role. Since beliefs off the equilibrium path are irrelevant, our game can be treated as a simultaneous-move game and thus the original definition in Eyster and Rabin (2005) applies.

Formally, a strategy profile $\sigma = (\sigma_S, \sigma_R)$ is said to be *exhaustive* if

$$\overline{\sigma}_S(m) := \sum_{\theta \in \Theta_S} \sigma_S(m|\theta)\pi(\theta) > 0$$

for every $m \in Y_S$. Given an exhaustive strategy profile, all information sets in the game can be reached with positive probability. Thus, we get subgame perfection for free.

Definition 2. *A exhaustive strategy profile $\sigma = (\sigma_S, \sigma_R)$ is a χ-cursed equilibrium of the simplified strategic communication game if for each $\theta \in \Theta_S$, each $y^*(\cdot) \in Y_R$ such that $\sigma_R(y^*(\cdot)) > 0$, and each m^* such that $\sigma_S(m^*|\theta) > 0$,*

$$y^*(m) \in \arg\max_{y(\cdot) \in Y_R} \sum_{\theta \in \Theta_S} U_R(\theta, m, y(\cdot))\hat{\pi}^\chi(\theta|m, \sigma_S), \quad \forall m \in Y_S$$
$$m^* \in \arg\max_{m \in Y_S} \sum_{y \in Y_R} U_S(\theta, m, y(\cdot))\sigma_R(y(\cdot)),$$

in which $\hat{\pi}^\chi(\theta|m, \sigma_S) = \chi\pi(\theta) + (1-\chi)\frac{\sigma_S(m|\theta)\pi(\theta)}{\overline{\sigma}_S(m)}$.

It is easy to see that $Y = [0, 1]$ and quadratic utility together ensure that it is without loss of generality to focus on pure strategies for the receiver. In particular, in equilibrium, given the sender's strategy σ_S, the receiver's action plan $y(\cdot)$ must satisfy

$$y(m) = \sum_{\theta \in \Theta_S} \theta\hat{\pi}^\chi(\theta|m, \sigma_S) = \hat{\pi}^\chi(1|m, \sigma_S) = \chi p + (1-\chi)\frac{\sigma_S(m|1)p}{\overline{\sigma}_S(m)}$$

for every $m \in Y_S$.

As usual, the babbling equilibrium always exists, in which the receiver will take the ex ante optimal action p regardless of the sender's message. In this case, the receiver's ex-ante welfare is $-p(1-p)$. In the separating equilibrium, the sender's strategy is $\sigma_S(\cdot|0) = \delta_0$ and $\sigma_S(\cdot|1) = \delta_1$, in which δ_m is the Dirac measure at $m \in Y_S$. In this case, the equilibrium action plan of the receiver must be $y(0) = \chi p$ and $y(1) = (1-\chi) + \chi p$. Incentive compatibility on the sender's side then requires $U_S(0,0,y(\cdot)) \geq U_S(0,1,y(\cdot))$, which reads $b \leq \frac{1-\chi}{2} + \chi p$. Clearly, in the separating equilibrium, the ex-ante welfare of the receiver is $-\chi^2 p(1-p)$. When $b > \frac{1-\chi}{2} + \chi p$, it can be shown that no informative equilibrium can be sustained.

Similar to the standard cheap talk game with quadratic preferences, the ex ante welfare of the sender and the receiver is aligned in our game. Thus, it is without loss of generality to focus on the ex ante welfare of the receiver.

3.1.2 Mediated Cursed Talk

Lee et al. (2023) show that cursed beliefs mitigate the strategic tension between the sender and the receiver, but unlike standard information garbling devices that have a similar effect, it does so without contaminating the content of the messages.[2] As a result, cursedness may further enhance welfare beyond the bound achieved by standard garbling devices. In order to compare the welfare implication of cursedness with that of standard garbling devices, we now introduce a mediator in the spirit of Goltsman, Hörner, Pavlov, and Squintani (2009). To allow for intrinsic cursedness that subjects bring to the lab, we will characterize optimal mediation with a cursed receiver.

Given that the receiver is χ-cursed, the mediator chooses Borel probability measures $\mu(\cdot|\theta)$ on $Y = [0,1]$ for each $\theta \in \Theta_S$ that solve the following optimization problem:

$$\begin{aligned}
&\max_{\mu(\cdot|\theta),\, \theta\in\Theta_S} \quad \sum_{\theta\in\Theta_S} \mathbb{E}_\mu[-(y-\theta)^2|\theta]\pi(\theta) \\
&\text{subject to} \quad \mathbb{E}_\mu[-(y-\theta-b)^2|\theta] \geq \mathbb{E}_\mu[-(y-\theta-b)^2|\theta'], \quad \forall \theta,\theta' \in \Theta_S \\
&\qquad\qquad\quad y = \chi p + (1-\chi)\frac{p\mu(y|1)}{p\mu(y|1) + (1-p)\mu(y|0)}, \quad \text{if } p\mu(y|1) + (1-p)\mu(y|0) > 0,
\end{aligned}$$

in which $\mathbb{E}_\mu[\cdot|\theta]$ is the conditional expectation operator with respect to $\mu(\cdot|\theta)$. Basically, the mediator maximizes the receiver's ex ante welfare (which is aligned with the sender's) subject to the constraints that the sender will truthfully report her type to the mediator and

[2] Lee et al. (2023) frames cursedness as prior bias however the two concepts are observationally equivalent in cheap-talk games.

that the χ-cursed receiver will follow the mediator's recommendation.

It can be shown that when $b \leq \frac{1-\chi}{2} + \chi p$, optimal mediation induces the separating equilibrium; when $b \geq \frac{1-\chi}{2} + \frac{(1+\chi)p}{2}$, optimal mediation induces the babbling equilibrium; when $\frac{1-\chi}{2} + \chi p < b < \frac{1-\chi}{2} + \frac{(1+\chi)p}{2}$, it is optimal for the mediator to recommend $2b - 1 + \chi - \chi p$ if $\theta = 0$, and recommend $1 - \chi + \chi p$ with probability ϕ and $2b - 1 + \chi - \chi p$ with probability $1 - \phi$ if $\theta = 1$, in which

$$\phi = \frac{p - b}{2p(1 - \chi + \chi p - b)} + \frac{1}{2p}.$$

In this case, the ex ante welfare of the receiver under optimal mediation is $(1-p)[(1-\chi-2b)(1+\chi)+\chi(2+\chi)p]$. We provide the proof in Online Appendix A.

Thus, in any case, we can always restrict attention to the following type of mediation/information garbling: When the sender sends $m = 0$, the receiver always receives $m = 0$. When the sender sends $m = 1$, there is a certain probability that the receiver will receive $m = 0$; otherwise, with the remaining probability, the receiver will receive $m = 1$.

4 Experimental Design

A cursed receiver in cheap talk games behaves as if she incorrectly believes that with a positive probability, the sender's messages are completely uninformative about the type. Hence, controlling cursedness is equivalent to controlling the incorrect model in the receiver's head, which may seem impossible to implement in experiments without deception. With the guidance of Proposition 3, however, we can simulate cursedness by controlling the proportion of fictitious roles in the game. Since the sender's cursedness does not play a role in our game, we only have to introduce fictitious receivers.

There are three roles in our experiment: S, True R, and False R. At the beginning of the game, X participants are randomly assigned the role of S, another X the role of True R, and the rest (Y people) the role of False R. Each individual is informed whether the assigned role is S or not. If the role is not S, then no further information about whether the role is True R or False R is provided. Then, the state is drawn independently for each individual who is assigned True R or False R. One S and one True R (who does not know if him/herself is True R or False R) are randomly paired to play the cheap talk game. False R's are not paired with anyone, but still, receive a message randomly chosen from all messages sent by S participants such that they cannot tell they are not paired. Finally, each of the R participants takes an action that affects his/her payoff and, if he/she is paired, the corresponding S participant's payoff.

The procedure above is announced to all participants of the game in a Zoom meeting, and thus, the game setup is common knowledge to all participants. Each S participant knows that he/she observes the true state, while each R participant thinks that with probability $\frac{Y}{X+Y}$, he/she is False R and the message he/she receives is not informative. In addition, each S participant knows that the R paired with him/her exhibits mistrust. Thus, under this setup, the game played between each S and True R pair exactly implements (controlled) cursedness $\chi = \frac{Y}{X+Y}$.

It is worth emphasizing that under the setup above, if all S participants play the same strategy σ_S, the Bayesian posterior of an R participant upon receiving message m is exactly given by $\hat{\pi}^\chi(\cdot|m, \sigma_S)$ with $\chi = \frac{Y}{X+Y}$. For this to be true, it is crucial that each False R receives a message randomly drawn from all S participants. When all S participants are playing the same strategy, the ex ante distribution of messages that a False R may receive is the same as the ex ante distribution of messages that a True R may receive. This means that the message that each R participant receives is completely uninformative about her role, which gives rise to the posterior $\hat{\pi}^\chi(\cdot|m, \sigma_S)$.

One alternative way of implementing cursedness among receivers is to match fictitious receivers with preprogrammed machine senders.[3] However, if machines are perceived to behave differently from human senders, then the equilibrium actions in both the real game and the fictitious game will be informative about the receiver's role, and thus the level of cursedness implemented will be different from χ. Thus, implementing cursedness with machines requires careful calibration of the preprogrammed strategies. We opt not to go this route.

4.1 The Antidote Game

We implement the strategic communication game above as the Antidote Game in our experiments. Consider two treasure hunters S and R who are poisoned in the middle of their adventure. They have the recipe to make an antidote, according to which a certain kind of toxic herbal extract is the main ingredient. R has 5 grams of the herbal extract while S does not have any. The exact amount of the herbal extract needed to make a perfect antidote depends on the seriousness of the poison and whether the poison is type 0 or type 1. The situation of S is more serious than that of R, and thus S always needs 3 grams more of the herbal extract than R regardless of whether the poison is type 0 or type 1. That is, we implement $b = 3/5 = 0.6$.

[3] Note that adding machine receivers to the game does not affect the human subjects' beliefs at all.

If the poison is type 0, R does not need any herbal extract to make a perfect antidote. That means, S needs 3 grams for a perfect antidote. If the poison is type 1, R needs all 5 grams of the herbal extract to make a perfect antidote. That means, S needs 8 grams. Putting in too little or too much of the herbal extract will result in pain (a utility loss). The larger the difference between the exact amount needed and the actual amount taken, the more pain one will suffer. Precisely, the payoff becomes

$$300 - 4 \times [\text{The exact amount needed} - \text{The actual amount taken}]^2 \text{ for } S \text{, and}$$
$$300 - 12 \times [\text{The exact amount needed} - \text{The actual amount taken}]^2 \text{ for } R.$$

The following table illustrates the possible payoffs for each player in different scenarios.

Poison Type		Payoff for R		Payoff for S	
		Type 0	Type 1	Type 0	Type 1
The Actual Amount (gram) of Herbal Extract Taken	0	300	0	264	44
	1	288	108	284	104
	2	252	192	296	156
	3	192	252	300	200
	4	108	288	296	236
	5	0	300	284	264

Table 1: Payoffs

They know that the poison is type 0 with a 30% chance and type 1 with a 70% chance. Thus, we have $p = 0.7$. S has a pair of magic glasses that tell him privately whether the poison is type 0 or 1. They have limited time to make only one bowl of antidote, but after the antidote is prepared they can use a magic spell to quickly duplicate it such that each player can take a full cup (of identical antidote). The game unfolds as follows: First, for each round and for each group, the type of poison is randomly drawn. Second, S wears his magic glasses and privately learns the poison type. Third, S sends one of the following messages to R:

"The poison is Type 0."   "The poison is Type 1."

After receiving a message, R decides how many grams (between 0 and 5) of the herbal extract to put in to make an antidote. Lastly, the antidote created is duplicated. Each of R and S takes the antidote which exerts its effect fully on each user. Finally, the outcomes, depending on the exact amount of herbal extract needed and the actual amount taken, are realized.

4.2 Treatments and Procedure

Our experimental design is presented in Table 2 below which involves four treatments. The experimental environment is adopted from the antidote game described in Section 4.1.

Treatment	Simulated Cursedness (χ)	Garbling (ϕ)
Standard Talk (ST)	0	0
Low Cursed (LC)	0.2	0
High Cursed (HC)	0.5	0
Mediated Talk (MT)	0	0.9

Table 2: Experimental Treatments

The four treatments differ from each other with respect to 1) the degree of controlled cursedness injected to the game and 2) whether the message sent by the sender (S) is garbled or not.

Standard Talk (ST) Treatment. In this treatment, there were equal numbers of S and R. At the beginning of each round, one S and one R were randomly paired. They were randomly reshuffled after each round to form new pairs. The roles were fixed throughout the official rounds of the experiment. Participants knew their roles. The message sent by S was transmitted to the paired R without garbling.

Cursed (LC and HC) Treatments. The message transmission procedure in this treatment is the same as that of Treatment ST but the roles and matching procedure differ from it. There were three roles—S, True R, and False R—in the cursed treatments. At the beginning of the experiment, X participants were randomly assigned the role of S, another X the role of True R, and the rest (Y people) the role of False R, where $\chi = \frac{Y}{X+Y}$ captures the degree of cursedness. We set $\chi = 0.2$ for Treatment LC and $\chi = 0.5$ for Treatment HC such that truth-telling is incentive-compatible in both treatments. The roles were fixed throughout the experiment. At the beginning of the 1st official round, each individual was informed whether the assigned role was S or not. If the role was not S, then no further information about whether the role was True R or False R was provided. Thus, the only thing they knew was that the chances were $\frac{X}{X+Y}$ and $\frac{Y}{X+Y}$ that they were True R and False R, respectively. In each round, one S and one True R (who did not know if him/herself was True R or False R) were randomly paired to form a group of two. They were randomly reshuffled after each round to form new groups. Participants whose role was False R were not paired with anyone but still received a message randomly chosen from all S participants in the round of the session such that they could not tell they were not paired.

Mediated Talk (MT) Treatment. The roles and matching procedure in this treatment are the same as those in Treatment ST but the procedure of message transmission differs from it. When S sent the message "The poison is Type 0", the message was transmitted to the paired R. When S sends the message "The poison is Type 1", the message is garbled such that the paired R received the original message "The poison is Type 1" with $\phi = 90\%$ chance and the message "The poison is Type 0" with 10% chance.

All treatments shared the same feedback procedure. The end-of-each-round feedback for the subjects whose role was S included whether the poison was type 0 or 1, the exact amount of herbal extract needed for a perfect antidote, the message S sent, the message R received, the actual amount of herbal extract taken, and the earning. For the subjects whose role was R (True R or False R), the feedback was provided only at the end of the 20th round, but not of other rounds. This was to ensure that R had no way to update his/her belief about the actual role he/she was assigned to.

We elicited the ex-ante ideal action under the prior belief before the official rounds began. At Round 0, we asked every participant to play the role of R and to decide what action to take without receiving any message from S. This elicitation was fully incentivized as Round 0 can be chosen for the final payment. The individual outcome from Round 0 was revealed to each participant at the end of the round before he/she proceeded to the first official round.

We conducted four sessions for each of the ST and MT treatments, six sessions for the LC treatment, and seven sessions for the HC treatment. Thus we had 21 sessions in total. The number of True R participants was 31, 37, 48, and 42 in ST, MT, LC, and HC treatment, respectively. We used the random-matching protocol and between-subjects design. With three exceptions, every session had 18 participants.[4] Our experiment was conducted in English using Zoom and oTree (Chen, Schonger, and Wickens, 2016) via the real-time online mode at the Hong Kong University of Science and Technology (HKUST) where turning on their camera was a strict requirement. A total of 370 subjects who had no prior experience in our experiment were recruited from the graduate and undergraduate populations of the university.

One round out of 20 official rounds and Round 0 (for the elicitation of the ex-ante ideal action) was randomly chosen for the final payment. The total payment in HKD was the payoff each subject earned in the selected round plus an HKD 40 show-up fee. Subjects earned HKD 275 (≈ USD 35.25) by participating in a session that lasted 70 minutes on average. The final earnings were paid electronically via the HKUST Autopay System to the bank account each participant provided to the Student Information System (SIS).

[4]One session for Treatment MT had 20 participants whereas Treatment ST had one session with 12 participants and one session with 14 participants.

5 Hypotheses

Our first hypothesis focuses on the sender's strategy, aiming to determine whether senders consider the cursedness of the receivers when formulating their strategies. The unique equilibrium in Treatment ST is characterized by babbling. In Treatments MT, LC, and HC, the same babbling equilibrium also exists, but it is not neologism proof (Farrell, 1993) while the truth-telling equilibrium survives neologism proofness.[5] The theoretical predictions lead to the following hypothesis:

Hypothesis 1 (Sender Strategy)**.**

(a) *The proportion of message "Type 1" conditional on type 1 is higher than that conditional on type 0 in each of Treatments MT, LC, and HC. There is no such difference in Treatment ST.*

(b) *The proportion of sender subjects using the separating strategy is larger in each of Treatments MT, LC, and HC than that in ST.*

Next, we turn our attention to the receiver's strategy. The key question is whether our methodology of introducing fictitious receivers effectively simulates the cursed behavior of the receivers. Figure 1 illustrates the theoretical predictions of the cumulative distributions of the receiver's actions conditional on the state for each treatment. In Treatment ST, the two conditional cumulative distributions are identical because, in the unique equilibrium outcome, the receiver takes a pooling action regardless of the message, remaining indifferent between actions 3 and 4. In contrast, Treatments MT, LC, and HC feature a truth-telling equilibrium that results in varying degrees of separation in the receiver's action choices, depending on the specific communication channel. As previously noted, the truth-telling equilibrium is the unique robust equilibrium that survives the NITS condition.

The degree of separation in the receiver strategy can be captured by the degree of the first-order-stochastic-dominance (FOSD) relationship between the two conditional cumulative distributions. To set a concrete testable hypothesis, we adopt the following measure of the degree of separation. Let F, G be two cumulative distributions over the space of actions

[5] A neologism-proof equilibrium is one where no credible neologism exists. Farrell (1993) defines credible neologisms by stating that for every putative equilibrium, a neologism indicating "my type is in K" exists for each non-empty subset K of the type space. A neologism is credible if the sender's types in K strictly prefer the neologism's outcome over the proposed equilibrium outcome, and the types not in K weakly prefer to stay in the proposed equilibrium. In the babbling equilibrium in Treatments MT, LC, and HC, there always exists a single type (either Type 0 or Type 1) who wants to create and send a credible neologism revealing his type.

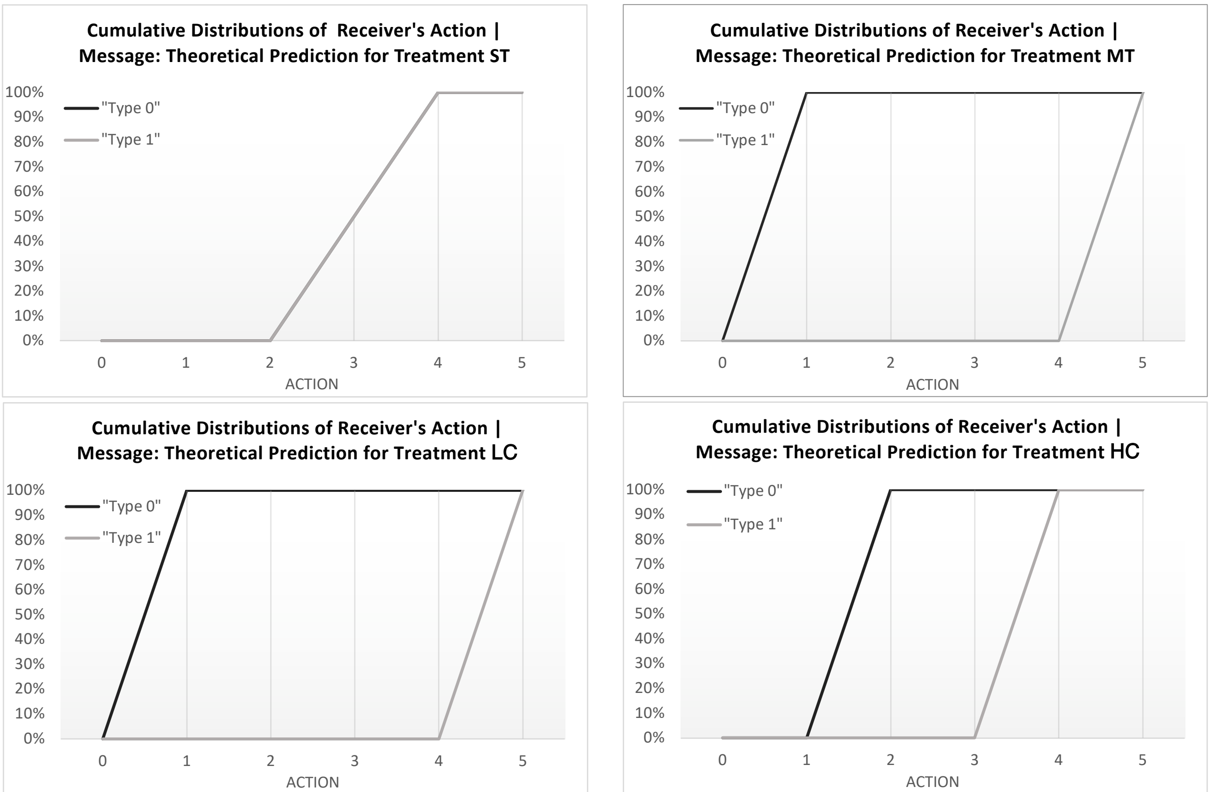


Figure 1: Receiver Strategy (CDF) - Theoretical Predictions

$Y = \{0, 1, 2, 3, 4, 5\}$. The degree of separation from F to G, denoted as $\Delta(F, G)$, is given by

$$\Delta(F, G) = \sum_{y \in Y} (F(y) - G(y))/5, \tag{5.1}$$

In case that G first-order stochastically dominates F, then (i) $\Delta(F, G)$ is simply the L^1-metric between F and G normalized by a factor of 1/5, and (ii) $\Delta(F, G) \in [0, 1]$, with $\Delta(F, G) = 1$ if and only if F assigns all probability mass to action 0 and G assigns all probability mass to action 5.

We define the degree of separation in the receiver's strategy, Δ^R, as follows:

$$\Delta^R = \Delta(F_0^R, F_1^R)$$

where F_m^R is the cumulative distribution of actions conditional on the message "Type m" for $m \in \{0, 1\}$. We will use Δ_j^R to denote the degree of separation in the receiver's strategy for Treatment j, for $j \in \{\text{ST, MT, LC, HC}\}$. Thus, the receiver's strategy of Treatment j is said to be more (less) separating than that of Treatment k if Δ_j^R is larger (smaller) than Δ_k^R. It is straightforward to calculate that $\Delta_{\text{ST}}^R = 0$, $\Delta_{\text{MT}}^R = \Delta_{\text{LC}}^R = 0.8$, and $\Delta_{\text{HC}}^R = 0.4$. Our second hypothesis is as follows:

Hypothesis 2 (Receiver Strategy)**.**

(a) *In each of Treatments MT, LC, and HC, the cumulative distribution of receiver's action conditional on "Type 0" first-order stochastically dominates (FOSD) that conditional on "Type 1". There is no such FOSD relationship in Treatment ST.*

(b) $\Delta^R_{ST} = 0 < \Delta^R_{HC} < \Delta^R_{MT} = \Delta^R_{LC}$.

(c) *The proportion of receiver subjects using the separating strategy is larger in each of Treatments MT, LC, and HC than that in ST.*

In cheap-talk games, the outcome is defined as a joint distribution over the action space and the state (type) space. Figure 2 presents the theoretical predictions of the cumulative distributions of the receiver's action conditional on the state for each treatment. In Treatment ST, the two conditional cumulative distributions are identical because pooling is the unique equilibrium outcome where the receiver is indifferent between actions 3 and 4. In each of the Treatments MT, LC, and HC, there is a robust truth-telling equilibrium that leads to some degrees of separation in outcome depending on the exact communication channel.

We define the degree of separation in outcome, Δ^O, as follows:

$$\Delta^O = \Delta(F^O_0, F^O_1)$$

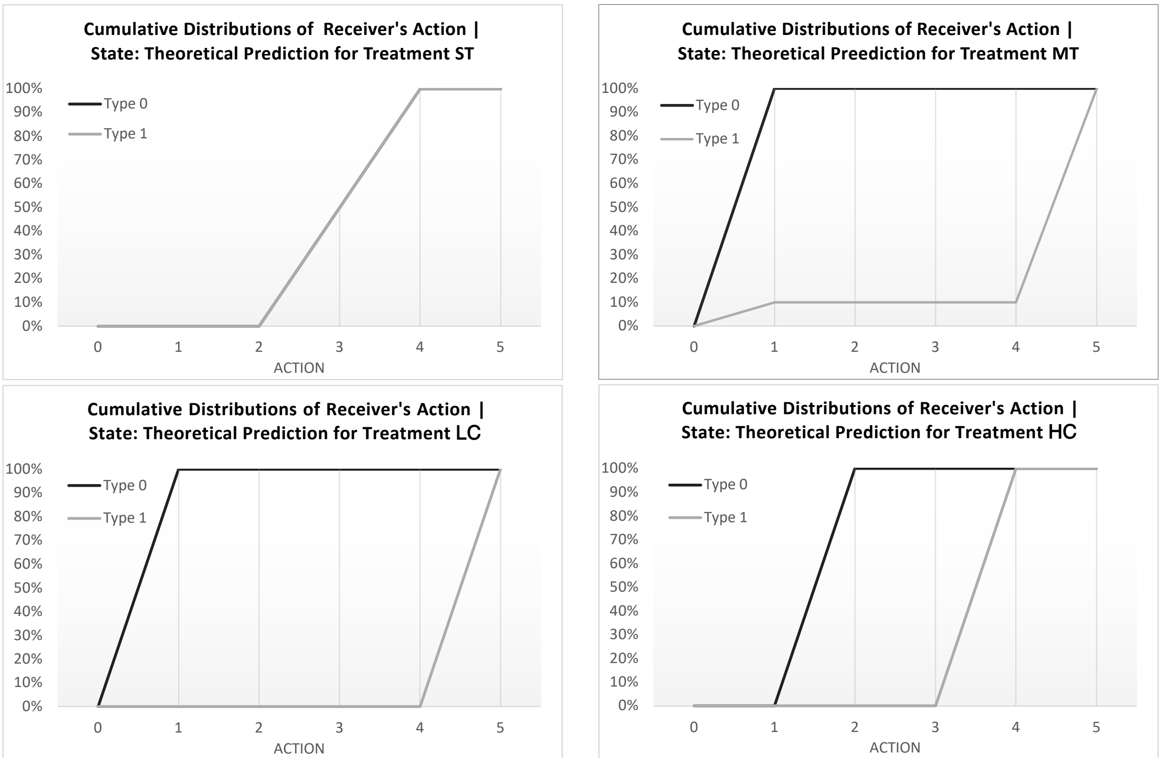
where F^O_θ is the cumulative distribution of actions conditional on type $\theta \in \{0, 1\}$. We will use Δ^O_j to denote the degree of separation in outcome for Treatment j, for $j \in \{\text{ST, MT, LC, HC}\}$. Thus, the outcome of Treatment j is said to be more (less) separating than that of Treatment k if Δ^O_j is larger (smaller) than Δ^O_k.[6] It is straightforward to calculate that $\Delta^O_{\text{ST}} = 0$, $\Delta^O_{\text{MT}} = 0.72$, $\Delta^O_{\text{LC}} = 0.8$ and $\Delta^O_{\text{HC}} = 0.4$. Our third hypothesis is as follows:

Hypothesis 3 (Outcome)**.**

(a) *In each of Treatments MT, LC, and HC, the cumulative distribution of receiver's action conditional on type 0 first-order stochastically dominates (FOSD) that conditional on type 1. There is no such FOSD relationship in Treatment ST.*

(b) $\Delta^O_{ST} = 0 < \Delta^O_{HC} < \Delta^O_{MT} < \Delta^O_{LC}$.

[6] A higher degree of separation does not necessarily imply a higher welfare measured by the receiver's payoff due to the larger (than the binary) action space and the quadratic loss function we assumed. We will have a separate hypothesis on the welfare ranking later.



■ For Treatments ST and LC, the cumulative distributions are drawn based on the assumption that the receiver is uniformly randomizing between the two indifferent actions, 3 and 4.

Figure 2: Outcome - Theoretical Predictions

Note that neither Δ^R nor Δ^O fully captures the welfare generated by the information transmission. The right measure is the payoff of the receiver as presented in Table 2. Our last hypothesis thus states:

Hypothesis 4 (Welfare)**.** *True Rs' average earnings are ranked as follows:*

$$EU_{ST} < EU_{MT} < EU_{HC} < EU_{LC}.$$

6 Experimental Results

We begin by reporting our experimental findings related to sender behavior, followed by an analysis of receiver behavior. In our examination of receiver behavior, we utilize data from both False Rs and True Rs (when available), as they are ex-ante identical. We then compare the outcomes across the four treatments. Since False Rs in the Cursed treatments (LC and HC) were introduced to accurately replicate the environment in which True Rs exhibit cursed behavior, we focus exclusively on the outcomes generated by True Rs in these two treatments. Finally, we will present the results concerning welfare.

All analyses presented in this section used data aggregated across all 20 rounds of all sessions of each treatment.[7] We compare distributions for first-order stochastic dominance via the non-parametric Barrett-Donald (BD) test procedure proposed by Barrett and Donald (2003).[8] Tables 4-5 in Online Appendix D present all non-parametric test results reported in this section.

6.1 Sender Behavior

Figure 3 reports the treatment-level aggregated proportions for a sender to send the message "The poison is Type 1" (shortly, "type 1" hereafter) conditional on each type.[9] Figure 13 in Online Appendix C provides the session-level data.

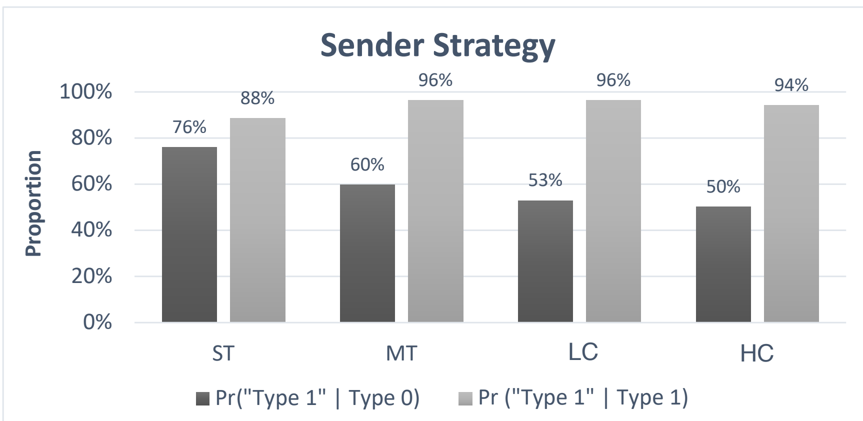


Figure 3: Sender Strategy

Notice that the proportion of senders sending "type 1" conditional on type 1 is higher than that conditional on type 0 in all treatments. This observation is not only true at the treatment level but also true at the session level. The paired sample Wilcoxon signed-rank tests using session-level data as independent observations reveal that the differences are statistically significant in all treatments (one-sided, p-value= 0.0625 for Treatments ST and

[7] We will use three types of aggregations of data. First, *treatment level* means that the data is aggregated over all 20 rounds across all sessions for each treatment. Second, *session level* means that the data is aggregated over all 20 rounds for each session. Third, *individual level* means that the data is aggregated over all 20 rounds for each individual. Given that we did not provide any feedback to the receiver subjects in our experiments, it is appropriate to look at the data aggregated across all 20 official rounds.

[8] Given two distributions F and G, the BD procedure requires to test both Null hypotheses $F \geq G$ and $G \geq F$. F first-order stochastically dominates G if and only if $F \geq G$ but not $G \geq F$ is accepted. For the p-values in each test, we employ a bootstrap of size 10,000.

[9] We follow the convention of adding double-quotation marks to distinguish states (type 0 and type 1) from messages ("type 0" and "type 1").

MT, 0.0156 for Treatment LC, and 0.0078 for Treatment HC).[10] This observation implies that senders on average sent informative signals to receivers in all four treatments. Thus, we have to partially reject Hypothesis 1(a) for Treatment ST.

The informativeness of senders' messages, however, is not the same across treatments. The Mann-Whitney (henceforth MW) tests using session-level data as independent observations indicate that the proportions of senders sending "type 1" conditional on type 0 are significantly larger in Treatment ST than in Treatments MT, LC, and HC (one-sided, p-values= 0.0571, 0.0048 and 0.0212, respectively). At the same time, the proportions of senders sending "type 1" conditional on type 1 across all four treatments are not statistically different from each other (two-sided, p-values$>$ 0.1714). Combining these two observations implies that the senders' messages are significantly less informative in Treatment ST than in each of Treatments MT, LC, and HC.

We conduct an additional analysis on the informativeness of senders' messages using individual-level data as follows. First, we calculate the proportion for each individual sender subject to send "type 1" conditional on type 1 and that conditional on type 0. The larger the difference, the more separating the strategy the sender subject is using. For example, the truth-telling (separating) strategy should generate a difference of 1 while the fully babbling (pooling) strategy should generate a difference of 0. Second, we classify each individual sender subject as a separating category if the calculated difference is larger than a cutoff. Similarly, we classify each individual sender subject as a pooling category if the calculated difference is smaller than a cutoff. As the cutoff for the separating (pooling) category becomes higher, the proportion of sender subjects in the separating (pooling) category weakly decreases (increases).

Figure 4 reports results from the individual-level analysis. The horizontal axis of each panel indicates the different cutoffs employed for the separating category (first row) and the pooling category (second row). The black area at the top shows the proportion of individual sender subjects who belong to the separating category. The dark gray area at the bottom shows the proportion of individual sender subjects who belong to the pooling category. The light-gray area in-between indicates the proportion of subjects who belong to neither of them. For example, looking at the top-left panel of Figure 4 for Treatment ST, when the cutoff values are 0.5 and 0.5 for both separating and pooling categories, the figure shows that the proportions are 13% for separating and 87% for pooling. As we move to the right, the cutoff for separating increases, and that for pooling decreases with the increment of 0.1. As a result, the proportions for both separating and pooling categories decrease (weakly) monotonically

[10] For the Wilcoxon signed-rank tests, the one-sided p-value= 0.0625 is the lowest possible value for four paired observations. So we adopt it as the threshold for the statistical significance.

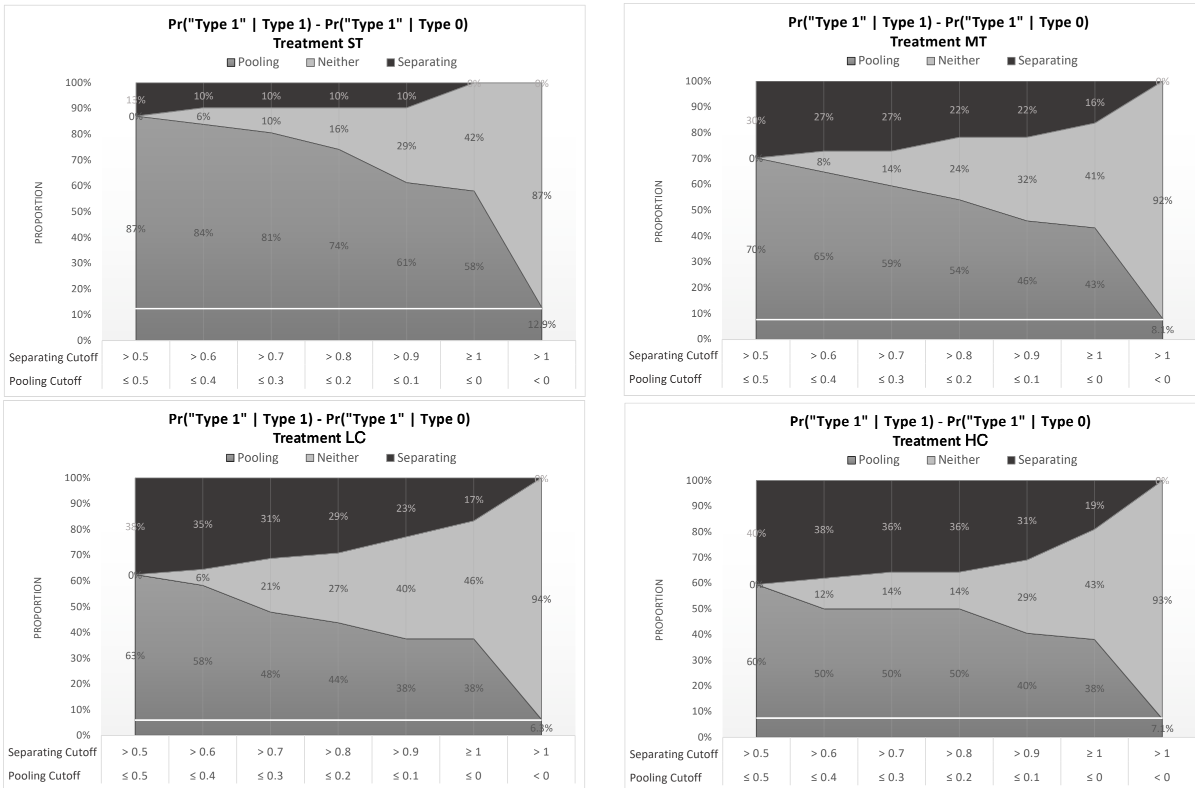


Figure 4: Sender Strategy - Individual Classification

while more subjects are classified as "neither". Note that the difference between the two rightmost columns in each panel is whether to include 1 for separating and 0 for pooling (the second rightmost column) or not (the first rightmost column). Thus, the difference in proportion between these two columns reveals the proportions of subjects using the fully truthful strategy and the complete babbling strategy, respectively.[11]

We say that senders in Treatment A use a separating strategy more than those in Treatment B if 1) the proportion of sender subjects classified as separating is larger in Treatment A than in Treatment B and 2) the proportion classified as pooling is smaller in Treatment A than in Treatment B *regardless of* the cutoff one adopts. For a statistical test, we take the distributions of the proportions of sender subjects classified as separating/pooling with the support $\{0, 0.1, .., 0.9, 1\}$ for different cutoffs employed ($\{0, ..., 0.5\}$ for pooling and $\{0.5, ..., 1\}$

[11]By definition, the area below the white horizontal line indicates the proportion of those who use the non-monotonic strategy, i.e., Pr("Type 1"|Type 1)<Pr("Type 1"|Type 0). In our classification exercise, we classify those subjects using a non-monotonic strategy as pooling, because the vast majority of senders are following the literal meaning of the message in their message strategy and the modal receiver's response is also monotonic; the number of sender subjects who use a reasonably informative non-monotonic strategy (defined as Pr("Type 1"|Type 1)-Pr("Type 1"|Type 0)< -0.25) is 2 in ST, 1 in MT, and 0 in both LC and HC treatments.

for separating)[12] and compare them with respect to the first-order stochastic-dominance using the BD test. The test confirms that senders in Treatment ST use a pooling strategy significantly more and a separating strategy significantly less than those in Treatments LC and HC (p-values = 0.0038 and 0.0030 for the Null and p-value = 0.9317 and 0.9207 for the reversed Null hypothesis). Moreover, senders in Treatment ST use a pooling strategy more and separating strategy less than those in Treatments MT with marginal statistical significance (p-values = 0.0741 for the Null and p-values = 0.9236 for the reversed Null hypothesis). The BD tests further reveal that no other pairs have a FOSD relationship. Partially accepting Hypothesis 1(b), these results are summarized as follows.

Result 1 (Sender Strategy)**.**

(a) *In all treatments, the proportion of message "Type 1" conditional on type 1 is higher than that conditional on type 0.*

(b) *Senders in Treatments LC, MT and HC use a separating strategy more and pooling strategy less than those in Treatment ST.*

Our results suggest that senders largely behave in line with cursed equilibrium theory, with the caveat that they tend to over-communicate in ST. This indicates the presence of residual behavioral effects that are not fully explained by the simulated level of cursedness.

6.2 Receiver Behavior

Figure 5 presents the cumulative distributions of the receiver's action conditional on the message, aggregated across all 20 rounds of all sessions of each treatment. The versions with probability distributions are provided in Figure 14 (treatment level) and Figure 16 (session level) in Online Appendix C.[13]

One immediate observation across all treatments is that the cumulative distribution conditional on Type 0 is first-order stochastically dominated by that conditional on Type 1. The BD test reveals that the first-order stochastic dominance relationship is significant in all treatments (p-values < 0.0001 for the Null and p-values > 0.8761 for the reversed Null hypothesis). This observation indicates that substantial proportions of receiver subjects may

[12] Graphically, the distribution can be obtained by flipping the pooling curve in Figure 4 around the vertical axis and combine it with the separating curve.

[13] In Treatments LC and HC, there are two types of receivers, True R and False R. When making their decisions, they did not know if their role was True R or False R. Thus, we do not distinguish between the two roles for our data analysis in this section.

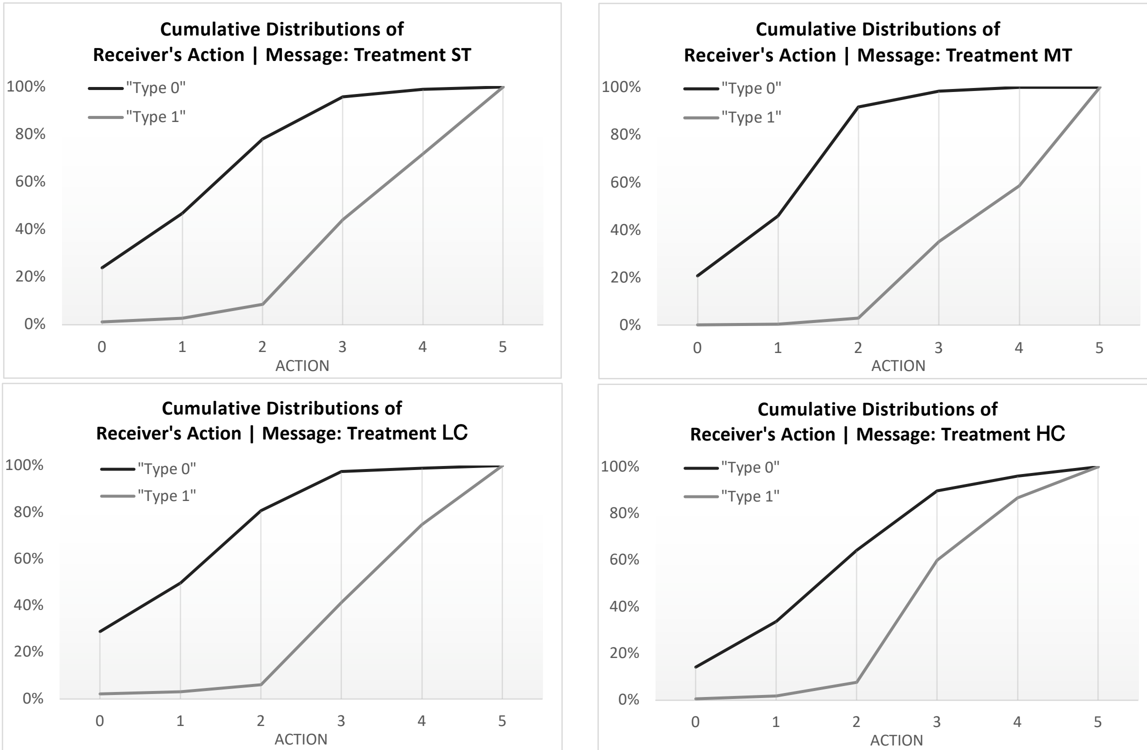


Figure 5: Receiver Strategy

use separating strategies in all our treatments. We thus have to reject Hypothesis 2(a) for Treatment ST.

It is apparent that the gap between the two conditional distributions is larger in some treatments than in others. The average Δ^R scores are 0.431, 0.519, 0.456, and 0.282 for Treatments, ST, MT, LC, and HC, respectively. The Mann-Whitney tests using session-level data as independent observations reveal that Δ^R_{HC} are marginally or significantly smaller than each of Δ^R_{ST}, Δ^R_{MT}, and Δ^R_{LC} (one-sided, p-values= 0.0818, 0.0303, and 0.0023, respectively).

We conduct additional analysis on the degree of separation in the receiver's strategy using individual-level data as follows. First, we calculate the average action taken by each individual receiver subject conditional on receiving the message "type 1" and that conditional on the message "type 0". The larger the difference, the more separating the strategy the receiver subject is using. For example, the fully separating strategy should generate a difference of 5 while the fully pooling strategy should generate a difference of 0. Second, we classify each individual receiver subject as a separating category if the calculated difference is larger than a cutoff. Similarly, we classify each individual receiver subject as a pooling category if the calculated difference is smaller than a cutoff. As the cutoff for the separating (pooling) category becomes higher, the proportion of receiver subjects in the separating (pooling) category weakly decreases (increases).

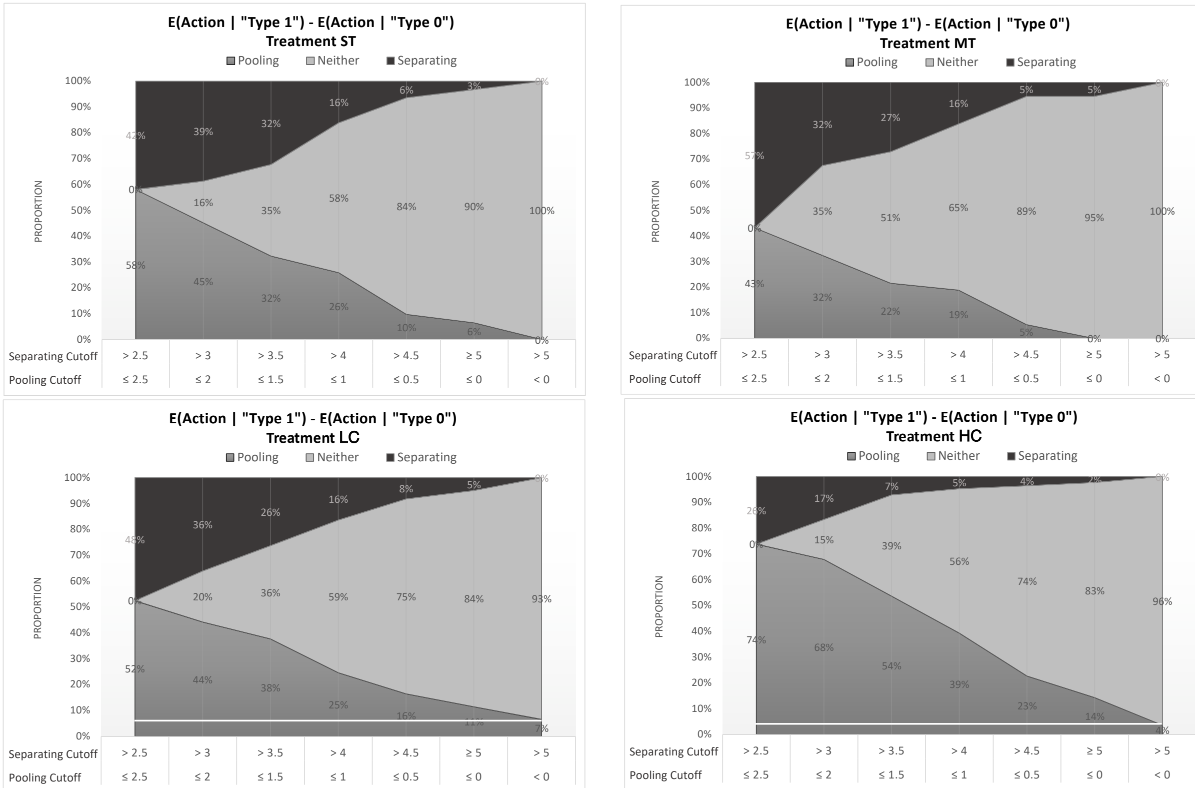


Figure 6: Receiver Strategy - Individual Classification

Figure 6 presents the results from the individual-level analysis. The horizontal axis of each panel indicates the different cutoffs for the separating category (first row) and the pooling category (second row). The black area at the top shows the proportion of individual receiver subjects who belong to the separating category. The dark gray area at the bottom shows the proportion of individual receiver subjects who belong to the pooling category. The light-gray area in-between indicates the proportion of subjects who belong to neither of them. For example, looking at the bottom-right panel of Figure 6 for Treatment HC, when the cutoff values are 2.5 and 2.5 for both separating and pooling categories, the figure shows that the proportions are 26% for separating and 74% for pooling. As we move to the right, the cutoff for separating increases, and that for pooling decreases with the increment of 0.5. As a result, the proportions for both separating and pooling categories decrease (weakly) monotonically while more subjects are classified as "neither". Note that the difference between the two rightmost columns in each panel is whether to include 5 for separating and 0 for pooling (the second rightmost column) or not (the first rightmost column). Thus, the difference in proportion between these two columns reveals the proportions of subjects using a fully separating strategy and a fully pooling strategy.[14]

[14] By definition, the area below the white line reveals the proportion of those who use the non-monotonic

The BD tests confirm that receivers in Treatment HC use a separating strategy significantly less and a pooling strategy significantly more than those in Treatments ST, MT, and LC (p-values = 0.0119, 0.0001, and 0.0076 for the Null and p-values = 0.9558, 0.9665, and 0.8825 for the reversed Null hypothesis).[15] The BD tests further reveal that no other pairs have a FOSD relationship. Partially accepting Hypothesis 2, these results are summarized as follows:

Result 2 (Receiver Strategy)**.**

(a) *In all treatments, the cumulative distribution of receiver's action conditional on "Type 0" first-order stochastically dominates (FOSD) that conditional on "Type 1".*

(b) $\Delta^R_{HC} < \Delta^R_{ST}$, Δ^R_{MT}, *and* Δ^R_{LC}.

(c) *Receivers in Treatment HC use a pooling strategy more and separating strategy less than those in Treatments ST, MT and LC.*

This result is not driven by a difference in subjects' degree of understanding/comprehension of the instructions. Figure 17 in Online Appendix C shows that the distributions of actions taken by participants in Round 0 under the prior belief (when receiving no message from a sender) are almost identical across treatments. Indeed, the Kolmogorov-Smirnov (KS) tests using the individual data as independent observations confirm that the distributions are not significantly different from each other (two-sided, p-values > 0.6227). This finding indicates that the treatment effects observed in receiver behavior cannot be attributed to subjects misunderstanding the strategic environment.

Compared with our hypotheses on receiver behavior, we observe two deviations: over-communication in ST and a reversal of the expected separation ranking between HC and ST. Because HC senders send more informative messages than ST senders, this pattern suggests that HC receivers discount messages more heavily than is implied by the simulated level of cursedness.

---

strategy, i.e., $\mathbb{E}[\text{Action}|\text{"Type 1"}] < \mathbb{E}[\text{Action}|\text{"Type 0"}]$. In our classification exercise, we classify those subjects using a non-monotonic strategy as pooling, because the vast majority of Receivers are following the literal meaning of the message in their action strategy and the modal sender strategy is also monotonic.

[15] This result remains robust regardless of how we define the fully separating strategy of the receiver. For instance, one could define a fully separating strategy as one in which an individual receiver consistently takes a higher action upon receiving the message "Type 1" compared to when they receive the message "Type 0" such that the action distribution conditional on receiving message "Type 0" has no overlap with that on receiving "Type 1." According to this definition, the proportion of receivers employing the fully separating strategy in Treatment HC is 44.7%, while this proportion increases to 60.7%, 62.2%, and 69.2% in Treatments ST, MT, and LC, respectively.

6.3 Outcome

Figure 7 presents the cumulative distributions of the receiver's action conditional on the state, aggregated across all 20 rounds of all four sessions of each treatment. The versions with probability distributions are provided in Figure 10 (treatment level) and Figure 12 (session level) in Online Appendix C.

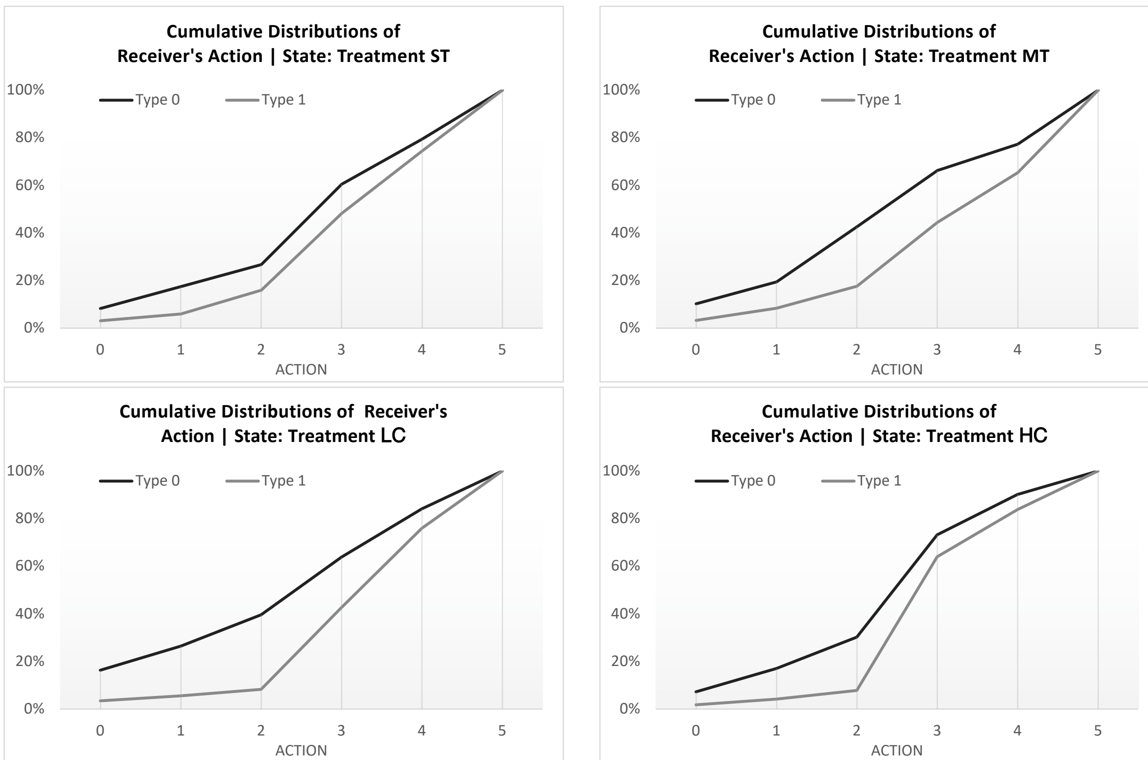


Figure 7: Outcome Comparison

Consistent with the receiver strategy data, we also observe a clear first-order-stochastic dominance relationship between the two conditional distributions in all treatments. For each treatment, we cannot reject the Null that the cumulative distribution conditional on Type 1 first-order stochastically dominates that conditional on Type 0 (p-values < 0.003), while rejecting the reversed Null (p-values > 0.8601). It implies that a strictly higher action is taken in type 1 than in type 0. This observation is true not only at the aggregated treatment level but also at the session level. It implies that we have "over-communication" in Treatment ST where theory predicts no first-order stochastic dominance relationship between the two conditional distributions.[16] Thus, we have to reject Hypothesis 3(a) partially.

In spite of the first-order-stochastic-dominance relationship observed in all treatments,

[16]Over-communication is well documented in the cheap-talk literature. See Blume, Lai, and Lim (2019b) for the most recent survey of the literature.

it is apparent that the gap between the two conditional distributions is larger in some treatments than others, implying that the degrees of separation in outcome differ across treatments. The average Δ^O scores are 0.0896, 0.154, 0.189, and 0.112 for Treatments, ST, MT, LC, and HC, respectively. Moreover, the Mann-Whitney tests using session-level data as independent observations reveal that Δ^O_{LC} is larger than Δ^O_{ST} and Δ^O_{HC} with the difference marginally significant (one-sided, p-values= 0.05714 and 0.0688, respectively). It turns out that Δ^O is strictly below the theoretical values reported in the previous section for Treatments MT, LC, and HC, implying that we have "under-communication" in these three treatments. This observed under-communication phenomenon is in sharp contrast to the over-communication phenomenon observed in ST. It is challenging to reconcile these two contradictory observations.[17]

Result 3 (Outcome)**.**

(a) *In all four treatments, the cumulative distribution conditional on type 0 is first-order stochastically dominated by that conditional on type 1.*

(b) *The outcome obtained in Treatment LC is more separating than those obtained in Treatments ST and HC.*

6.4 Welfare

Figure 8 reports the average earnings of True Rs in each treatment. Figure 19 in Online Appendix C provides the session-level earning data. The average earning for True Rs is highest in Treatment LC and lowest in Treatment ST. The MW tests show that the average earning for True Rs is significantly lower in Treatment ST than that in each of Treatments MT, LC, and HC (one-sided, p-values = 0.0143, 0.0571, and 0.0121, respectively). The MW tests further show that the average earning for True Rs is higher in Treatment LC than that in each of Treatments ST, MT, and HC (one-sided, p-values = 0.0571, 0.0857, and 0.1474, respectively) although the difference is either only marginally significant or statistically insignificant. As presented in Figure 18 in Online Appendix C, for each of

[17]In particular, it is surprising to see that the outcome obtained in Treatment HC is almost perfect babbling because the theoretical environment guarantees the existence of truth-telling equilibrium. The session-level outcome data presented in Figure 12 (Online Appendix C) further illustrates that an almost perfect babbling outcome was obtained in Session 1 and an outcome reasonably close to pooling was obtained in Sessions 3 and 4. To our knowledge, we are one of a few exceptions in the literature to report behavior observed in the laboratory that is more consistent with the babbling equilibrium prediction in the presence of a truth-telling equilibrium. Blume, Lai, and Lim (2019a) report the laboratory data that the observed behavior in the lab is more consistent with the non-truthful but informative equilibrium predictions even when there is a truthful equilibrium.

Treatments LC and HC, the average earnings for False Rs seem substantially smaller than those for True Rs, although the differences are not statistically significant.

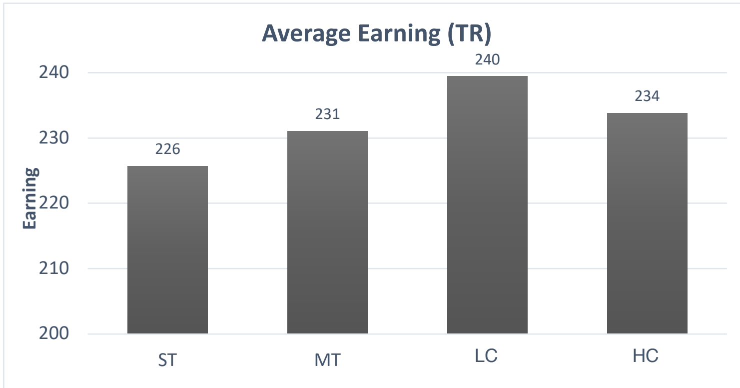


Figure 8: Average Earning

Result 4 (Earnings). *True Rs' average earning is significantly smaller in Treatment ST than in each of Treatments MT, LC, and HC. Moreover, True Rs' average earning is higher in Treatment LC than in each of Treatments ST, MT, and HC, although the differences are not significant.*

To summarize, our experimental data confirm that senders adapt their behavior based on the perceived cursedness of the receiver. We further observe two contradictory phenomena: the typical over-communication in Treatment ST and the unfamiliar under-communication in Treatments MT, LC, and HC, in which over- and under-communication are defined relative to the most informative equilibrium in the corresponding environment. Moreover, receivers in HC use a separating strategy less frequently and a pooling strategy more frequently than those in ST, despite the fact that the truth-telling equilibrium exists in HC but not in ST. The observed under-communication cannot be rationalized by merely introducing a truth-telling preference or lying cost. A level-k model a la Crawford (2003) in which the level-0 sender is assumed to be truthful also fails to rationalize our experimental data.[18] In the next section, we assume that our experimental participants may bring some intrinsic cursedness to the lab and explore whether a sensible distribution of intrinsic cursedness could account for the observed departure from the theoretical predictions.

[18]The assumption that the naivety of level-0 senders is modeled as truthful is well-accepted in the cheap-talk literature. Online Appendix E presents predictions from the level-k model. In any level above zero, the unique prediction is that the sender babbles in Treatment ST and tells the truth in all three other treatments. With the assumption of truthful level-0 senders, the best-response structure of the level-k model leaves no room for the receiver to not fully comprehend the information provided by the sender.

7 Intrinsic Cursedness

Assume that subjects may have brought some intrinsic cursedness to the experiments in addition to the cursedness induced by introducing fictitious R's. Let χ_0 denote the subject's intrinsic cursedness. Then, the effective cursedness subjects exhibit in the experiments, denoted as $\hat{\chi}$, would be given by

$$\hat{\chi} = 1 - (1 - \chi)(1 - \chi_0) = \chi + \chi_0 - \chi\chi_0,$$

in which χ is the induced cursedness in the experiment. Table 3 summarizes the receiver's optimal strategy in the separating equilibrium as a function of the intrinsic cursedness χ_0 in the strategic communication game. In addition, the range of χ_o that supports the separating equilibrium in each treatment is listed in the last column.[19] In the table, $\lfloor x \rfloor$ denotes the largest integer that does not exceed x. We need this operator since the action space is discretized.

Treatment	$\hat{\chi}$	ϕ	Given "1"	Given "0"	Range of χ_0
ST	χ_0	0	$\lfloor 5.5 - 1.5\chi_0 \rfloor$	$\lfloor 0.5 + 3.5\chi_0 \rfloor$	$[\frac{1}{7}, \frac{1}{3}] \cup [\frac{3}{7}, 1)$
MT	χ_0	0.9	$\lfloor 5.5 - 1.5\chi_0 \rfloor$	$\lfloor \frac{107}{74} + \frac{189}{74}\chi_0 \rfloor$	[0,1)
LC	$0.2 + 0.8\chi_0$	0	$\lfloor 5.2 - 1.2\chi_0 \rfloor$	$\lfloor 1.2 + 2.8\chi_0 \rfloor$	$[0, \frac{1}{6}] \cup [\frac{2}{7}, 1)$
HC	$0.5 + 0.5\chi_0$	0	$\lfloor 4.75 - 0.75\chi_0 \rfloor$	$\lfloor 2.25 + 1.75\chi_0 \rfloor$	[0,1)

Table 3: Receiver's Optimal Responses Given the Messages Received in the Most Informative Equilibrium

Based on Table 3, we conduct a simple calibration exercise as follows. We assume that χ_0 follows a truncated normal distribution over $[0, 1]$. We say that a distribution of the intrinsic cursedness χ_0 *rationalizes* our data if its theoretical predictions meet the following four criteria: 1) the proportion of senders who use the separating strategy is lowest in Treatment ST (ST1 Criterion); 2) welfare is lowest in Treatment ST (ST2 criterion); 3) welfare is highest in Treatment LC (LC criterion); and 4) the degree of separation in the receiver's actions is lowest in Treatment HC (HC criterion). The left panel of Figure 9 illustrates the range of mean (horizontal axis) and variance (vertical axis) of the truncated normal distribution, which χ_0 follows, that rationalizes our data. This graph demonstrates that a concentration of χ_0 around 0.46 can rationalize our data. Furthermore, when the variance exceeds a certain threshold (greater than 0.46), the range of means that can rationalize our data consistently expands. This finding suggests that incorporating individual heterogeneity in the degree of intrinsic cursedness makes it easier to rationalize our data. The right panel of Figure 9

[19] We present the exact calculations in Online Appendix B.

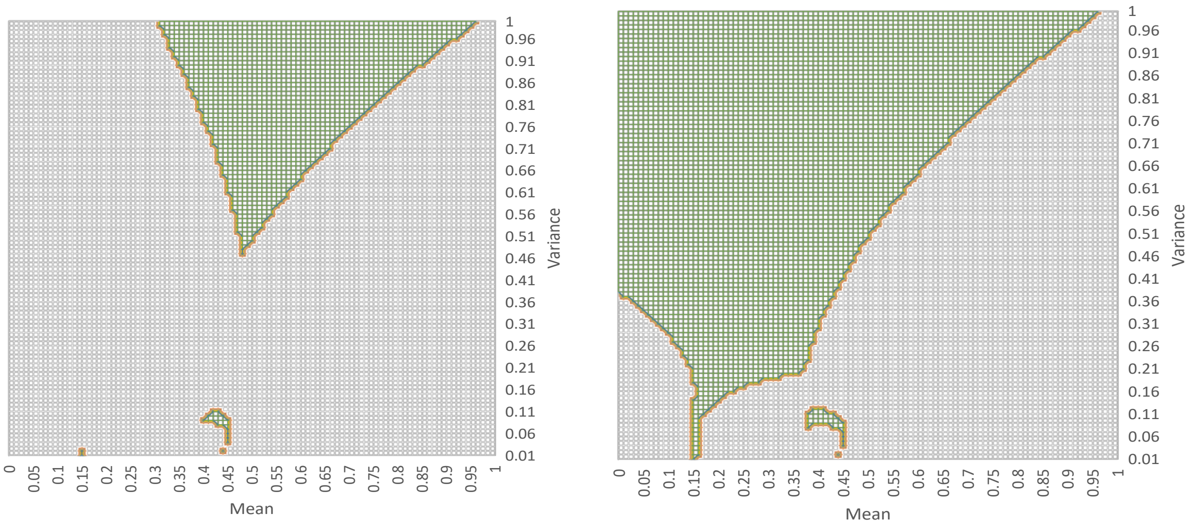


Figure 9: Compatible Normal Distributions with (Left) and without (Right) the LC Criterion

exhibits the range of mean and variance values for χ_0 that can rationalize our data while disregarding the LC criterion. This is motivated by the fact that the welfare in Treatment LC is not statistically different from that in Treatment MT. The results demonstrate a considerably wider range, indicating that it is much easier to rationalize our data without applying the LC criterion.

Our finding that χ_0 concentrated around 0.46 can rationalize the experimental data aligns with the results of Szembrot (2018), who estimated a similar degree of intrinsic cursedness to be 0.45. However, the empirical analysis conducted by Eyster and Rabin (2005) suggests that the estimated degrees of intrinsic cursedness in trading games and common value auctions are generally higher than 0.45, displaying a significant level of individual heterogeneity.

We believe that this discrepancy may be attributed to the explicit nature of the correlation between the sender's strategy and their private information in both our communication game and the signaling game examined in Szembrot (2018), compared to other types of games.[20] In our view, the participants' belief hierarchy about how cursed each of them is can drastically affect their behavior in the game. For example, a bidder in a common value English auction may perceive other bidders to be less cursed than they really are, and thus, she may bid higher and appear to be more cursed.

[20]This intuition was shared by Shengwu Li during an early-stage discussion, and we are grateful for his valuable input.

References

Barrett, G. F. and S. G. Donald (2003). Consistent tests for stochastic dominance. *Econometrica 71*(1), 71–104.

Blume, A., E. K. Lai, and W. Lim (2019a). Eliciting private information with noise: the case of randomized response. *Games and Economic Behavior 113*, 356–380.

Blume, A., E. K. Lai, and W. Lim (2019b). Strategic information transmission: A survey of experiments and theoretical foundations. In M. R. C. Monica Capra, Rachel Croson and T. Rosenblat (Eds.), *Handbook of Experimental Game Theory*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

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.

Cohen, S. and S. Li (2022). Sequential cursed equilibrium. *arXiv preprint arXiv:2212.06025*.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review 93*(1), 133–149.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica*, 1431–1451.

Enke, B. and F. Zimmermann (2019). Correlation neglect in belief formation. *The Review of Economic Studies 86*(1), 313–332.

Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica 73*(5), 1623–1672.

Eyster, E. and G. Weizsacker (2016). Correlation neglect in portfolio choice: Lab evidence. *Available at SSRN 2914526*.

Farrell, J. (1993). Meaning and credibility in cheap-talk games. *Games and Economic Behavior 5*(4), 514–531.

Fong, M.-J., P.-H. Lin, and T. R. Palfrey (2023). Cursed sequential equilibrium. *arXiv preprint arXiv:2301.11971*.

Forsythe, R., R. M. Isaac, and T. R. Palfrey (1989). Theories and tests of" blind bidding" in sealed-bid auctions. *The Rand Journal of Economics*, 214–238.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009). Mediation, arbitration and negotiation. *Journal of Economic Theory 144*(4), 1397–1420.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000). An experimental study of jury decision rules. *American Political science review 94*(2), 407–423.

Kagel, J. H. and D. Levin (1986). The winner's curse and public information in common value auctions. *The American economic review*, 894–920.

Lee, Y.-J., W. Lim, and C. Zhao (2023). Cheap talk with prior-biased inferences. *Games and Economic Behavior 138*, 254–280.

Lin, P.-H. and Y. L. Tan (2025). Cursed job market signaling. *arXiv preprint arXiv:2503.19089*.

Moser, J. and N. Wallmeier (2021). Correlation neglect in voting decisions: An experiment. *Economics Letters 198*, 109656.

Szembrot, N. (2018). Experimental study of cursed equilibrium in a signaling game. *Experimental Economics 21*(2), 257–291.

Tergiman, C. J. (2024). Correlation neglect in student-to-school matching. *American Economic Journal: Microeconomics*.

Wenner, L. (2019). Cursedness in markets with asymmetric information: Theory and experimental evidence. *Available at SSRN 3439387*.

Online Appendices

A Optimal Mediation with a Cursed Receiver

Simple algebra yields that under any feasible $\mu(\theta|0)$ and $\mu(\theta|1)$,

$$\sum_{\theta\in\Theta_S} \mathbb{E}_\mu[-(y-\theta)^2|\theta]\pi(\theta) = -(1+\chi)(1-p)\mathbb{E}_\mu[y|0] + \chi p(1-p). \tag{A.1}$$

Thus, the optimal mediation problem can be reduced to the following:

$$\begin{aligned} &\min_{\mu(\cdot|0),\mu(\cdot|1)} \mathbb{E}_\mu[y|0] \\ &\text{subject to} \quad 0 \le \mathbb{E}_\mu[(y-b)^2|1] - \mathbb{E}_\mu[(y-b)^2|0] \le 2(\mathbb{E}_\mu[y|1] - \mathbb{E}_\mu[y|0]) \\ &\qquad\qquad y = \chi p + (1-\chi)\frac{p\mu(y|1)}{p\mu(y|1)+(1-p)\mu(y|0)}, \quad \text{if } p\mu(y|1)+(1-p)\mu(y|0) > 0. \end{aligned}$$

Note that since

$$\mathbb{E}_\mu[y|0] = \chi p + (1-\chi)\mathbb{E}_\mu\left[\frac{p\mu(y|1)}{p\mu(y|1)+(1-p)\mu(y|0)}\middle| 0\right] \ge \chi p,$$

the separating equilibrium, if exists, is a feasible mediated equilibrium and achieves the welfare bound. Hence, if $b \le \frac{1-\chi}{2} + \chi p$, no additional mediation is needed and the ex-ante welfare under optimal mediation is $-\chi^2 p(1-p)$.

Now suppose $b > \frac{1-\chi}{2} + \chi p$. To further analyze the optimal mediation problem, we need a series of lemmas. Let $\overline{y} = \chi p + (1-\chi)$ and $\underline{y} = \chi p$. For the simplicity of exposition, we will write $\mu_0, \mu_1, \mathbb{E}_{\mu_0}[\cdot]$, and $\mathbb{E}_{\mu_1}[\cdot]$ instead of $\mu(\cdot|0), \mu(\cdot|1), \mathbb{E}_\mu[\cdot|0]$, and $\mathbb{E}_\mu[\cdot|1]$, respectively.

Lemma 1. *If (μ_0, μ_1) satisfies the receiver's IC constraint, so does $\alpha(\mu_0,\mu_1) + (1-\alpha)(\delta_{\underline{y}}, \delta_{\overline{y}})$ for any $\alpha \in [0,1]$.*

Proof. When $\alpha = 1$ the claim is trivial. Suppose $\alpha \in [0,1)$. Let $(\nu_0,\nu_1) = \alpha(\mu_0,\mu_1) + (1-\alpha)(\delta_{\underline{y}}, \delta_{\overline{y}})$. Suppose $p\mu_1(\underline{y}) + (1-p)\mu_0(\underline{y}) > 0$. Then for (μ_0,μ_1) to satisfy the receiver's IC, we need

$$0 = \frac{p\mu_1(\underline{y})}{p\mu_1(\underline{y}) + (1-p)\mu_0(\underline{y})},$$

which implies that $\mu_1(\underline{y}) = 0$. When $p\mu_1(\underline{y}) + (1-p)\mu_0(\underline{y}) = 0$, we also have $\mu_1(\underline{y}) = 0$.

Thus, by construction, $\nu_1(\underline{y}) = 0$, which implies that

$$\underline{y} = \chi p + (1-\chi)\frac{p\nu_1(\underline{y})}{p\nu_1(\underline{y}) + (1-p)\nu_0(\underline{y})}.$$

Similarly, we have

$$\overline{y} = \chi p + (1-\chi)\frac{p\nu_1(\overline{y})}{p\nu_1(\overline{y}) + (1-p)\nu_0(\overline{y})}.$$

Thus (ν_0, ν_1) satisfies the receiver's IC. □

Lemma 2. *Suppose $b > \frac{1-\chi}{2} + \chi p$. If (μ_0, μ_1) is a solution to the optimal mediation problem, then $\mathbb{E}_{\mu_1}[(y-b)^2] = \mathbb{E}_{\mu_0}[(y-b)^2]$.*

Proof. Let (μ_0, μ_1) be a solution to the problem. When $b > \frac{1-\chi}{2} + \chi p$, the separating equilibrium cannot be sustained. It follows that $\mathbb{E}_{\mu_0}[y] > \underline{y} = \chi p$. By way of contradiction assume $\mathbb{E}_{\mu_1}[(y-b)^2] > \mathbb{E}_{\mu_0}[(y-b)^2]$. Consider $\nu_0 = \alpha\mu_0 + (1-\alpha)\delta_{\underline{y}}$ and $\nu_1 = \alpha\mu_1 + (1-\alpha)\delta_{\overline{y}}$ for some $\alpha \in (0,1)$.

Since $b > \frac{1-\chi}{2} + \chi p$, we have

$$(\overline{y} - b)^2 - (\underline{y} - b)^2 < 0 < 2(\overline{y} - \underline{y}).$$

Thus, since (μ_0, μ_1) is feasible, we have

$$\mathbb{E}_{\nu_1}[(y-b)^2] - \mathbb{E}_{\nu_0}[(y-b)^2] \leq 2(\mathbb{E}_{\nu_1}[y] - \mathbb{E}_{\nu_0}[y]).$$

Thus, if $\mathbb{E}_{\mu_1}[(y-b)^2] > \mathbb{E}_{\mu_0}[(y-b)^2]$, we can pick $\alpha \in (0,1)$ such that

$$\alpha(\mathbb{E}_{\mu_1}[(y-b)^2] - \mathbb{E}_{\mu_0}[(y-b)^2]) + (1-\alpha)((\overline{y}-b)^2 - (\underline{y}-b)^2) = 0.$$

Furthermore, we know that

$$\mathbb{E}_{\nu_0}[y] = \alpha\mathbb{E}_{\mu_0}[y] + (1-\alpha)\underline{y} < \mathbb{E}_{\mu_0}[y].$$

The corresponding (ν_0, ν_1) is thus feasible and achieves a higher level of welfare than (μ_0, μ_1), a contradiction. □

By the previous lemma, if $b > \frac{1-\chi}{2} + \chi p$, the optimal mediation problem reduces to

$$\begin{aligned}
\min_{\mu_0,\mu_1} \quad & \mathbb{E}_{\mu_0}[y] \\
\text{subject to} \quad & \mathbb{E}_{\mu_0}[(y-b)^2] = \mathbb{E}_{\mu_1}[(y-b)^2] \\
& \mathbb{E}_{\mu_0}[y] \leq \mathbb{E}_{\mu_1}[y] \\
& y = \chi p + (1-\chi)\frac{p\mu_1(y)}{p\mu_1(y) + (1-p)\mu_0(y)}, \quad \text{if } p\mu_1(y) + (1-p)\mu_0(y) > 0.
\end{aligned}$$

The following proposition summarizes the solution to the optimal mediation problem.

Proposition 4. *When* $b \le \frac{\underline{y}+\overline{y}}{2}$, *optimal mediation induces the separating equilibrium. When* $b \ge \frac{\overline{y}+p}{2}$, *optimal mediation induces the babbling equilibrium. When* $\frac{\underline{y}+\overline{y}}{2} < b < \frac{\overline{y}+p}{2}$, *it is optimal for the mediator to recommend* $2b-\overline{y}$ *when* $\theta = 0$*; recommend* $\overline{y}$ *with probability* ϕ *and* $2b-\overline{y}$ *with probability* $1-\phi$, *in which*

$$\phi = \frac{p - 2b + \overline{y}}{2p(\overline{y} - b)}. \tag{A.2}$$

In this case, the ex-ante welfare of the receiver under optimal mediation is $(1-p)[(1-\chi-2b)(1+\chi)+\chi(2+\chi)p]$.

Proof. From the sender's point of view, if (μ_0, μ_1) is feasible, the ex ante welfare is

$$\begin{aligned}
&(1-p)\mathbb{E}_{\mu_0}[-(y-b)^2] + p\mathbb{E}_{\mu_1}[-(y-1-b)^2] \\
=\ &\mathbb{E}_{\mu_0}[-(y-b)^2] + p(\mathbb{E}_{\mu_1}[-(y-1-b)^2] - \mathbb{E}_{\mu_1}[-(y-b)^2]) \\
=\ &\mathbb{E}_{\mu_0}[-(y-b)^2] + 2p\mathbb{E}_{\mu_1}[y] - p(2b+1) \\
=\ &\mathbb{E}_{\mu_0}[-(y-b)^2] + 2(p-(1-p)\mathbb{E}_{\mu_0}[y]) - p(2b+1) \\
=\ &\mathbb{E}_{\mu_0}[-(y-b)^2] - 2(1-p)\mathbb{E}_{\mu_0}[y] - p(2b-1),
\end{aligned}$$

which, together with (A.1) and the fact that the sender's ex ante welfare will be b^2 lower than the receiver's, yields

$$-(1+\chi)(1-p)\mathbb{E}_{\mu_0}[y] + \chi p(1-p) - b^2 = \mathbb{E}_{\mu_0}[-(y-b)^2] - 2(1-p)\mathbb{E}_{\mu_0}[y] - p(2b-1). \tag{A.3}$$

Let $y_\theta = \mathbb{E}_{\mu_\theta}[y]$ and $\sigma_\theta^2 = var_{\mu_\theta}(y)$. The equation above reads

$$\sigma_0^2 = (p - y_0)(y_0 + 1 - \chi + \chi p - 2b). \tag{A.4}$$

It follows from (A.3), $y_1 \ge y_0$, $p = py_1 + (1-p)y_0$, and $\sigma_0^2 \ge 0$, that

$$y_0 \ge 2b - 1 + \chi - \chi p = 2b - \overline{y}. \tag{A.5}$$

Since we need $y_0 \le p$, if $2b - \overline{y} \ge p$, i.e. $b \ge \frac{\overline{y}+p}{2}$, then optimal mediation induces the babbling equilibrium.

Suppose $\frac{\overline{y}+\underline{y}}{2} < b < \frac{\overline{y}+p}{2}$. Then by (A.1) and (A.5), the receiver's ex ante welfare cannot be strictly larger than

$$-(1+\chi)(1-p)(2b-\overline{y}) + \chi p(1-p) = (1-p)[(1-\chi-2b)(1+\chi)+\chi(2+\chi)p]. \tag{A.6}$$

To show that the upper bound is tight, we now construct the optimal policy. Let $y_0 = 2b - \overline{y}$. It follows that $\sigma_0^2 = 0$. Hence if $\theta = 0$, the mediator always recommends y_0. If $\theta = 1$, for the receiver's IC to hold, the mediator can only recommend y_0 or $\overline{y}$. Suppose the mediator recommends $\overline{y}$ with probability ϕ, we need

$$p((1-\phi)y_0 + \phi\overline{y}) + (1-p)y_0 = p$$

which reads

$$\phi = \frac{p - y_0}{p(\overline{y} - y_0)}.$$

Verify that

$$\chi p + (1-\chi)\frac{p(1-\phi)}{p(1-\phi)+1-p} = 2b - \overline{y} = y_0.$$

Thus, the receiver's IC is satisfied. It suffices to check that the sender's IC is satisfied, which is trivial. Thus, the bound is tight. □

B Calculations for Table 3

Given $x \in \mathbb{R}$, let $\lfloor x \rfloor$ denote the greatest integer less than or equal to x. When there is a tie from the receiver's point of view, assume that the larger action will be chosen for simplicity.

We consider ST, LC, and HC first. In the separating equilibrium, given message "1", the receiver's posterior is $1 \cdot (1-\hat{\chi}) + 0.7 \cdot \hat{\chi} = 1 - 0.3\hat{\chi}$, and thus action $\lfloor 5.5 - 1.5\hat{\chi} \rfloor$ will be chosen; given message "0", the receiver's posterior is $0 \cdot (1-\hat{\chi}) + 0.7 \cdot \hat{\chi} = 0.7\hat{\chi}$, and thus action $\lfloor 3.5\hat{\chi} + 0.5 \rfloor$ will be chosen. The separating equilibrium can be sustained if $\lfloor 5.5 - 1.5\hat{\chi} \rfloor + \lfloor 3.5\hat{\chi} + 0.5 \rfloor \geq 6$ and $\lfloor 5.5 - 1.5\hat{\chi} \rfloor > \lfloor 3.5\hat{\chi} + 0.5 \rfloor$, which yields $\hat{\chi} \in [\frac{1}{7}, \frac{1}{3}] \cup [\frac{3}{7}, 1)$. Outside this range, only the pooling equilibrium can be sustained.

Now consider MT. The Bayesian updates conditioning on the messages are respectively $\Pr(\theta = 1|m = 1) = 1$ and $\Pr(\theta = 1|m = 0) = \frac{7}{37}$. Thus, given message "1", the receiver's posterior is $1 \cdot (1-\hat{\chi}) + 0.7 \cdot \hat{\chi} = 1 - 0.3\hat{\chi}$, and thus action $\lfloor 5.5 - 1.5\hat{\chi} \rfloor$ will be chosen; given message "0", the receiver's posterior is $\frac{7}{37} \cdot (1-\hat{\chi}) + 0.7 \cdot \hat{\chi} = \frac{70+189\hat{\chi}}{370}$, and thus action $\lfloor \frac{107+189\hat{\chi}}{74} \rfloor$ will be chosen. Incentive compatibility is always ensured. We just need $\hat{\chi} < 1$.

C Additional Figures and Tables

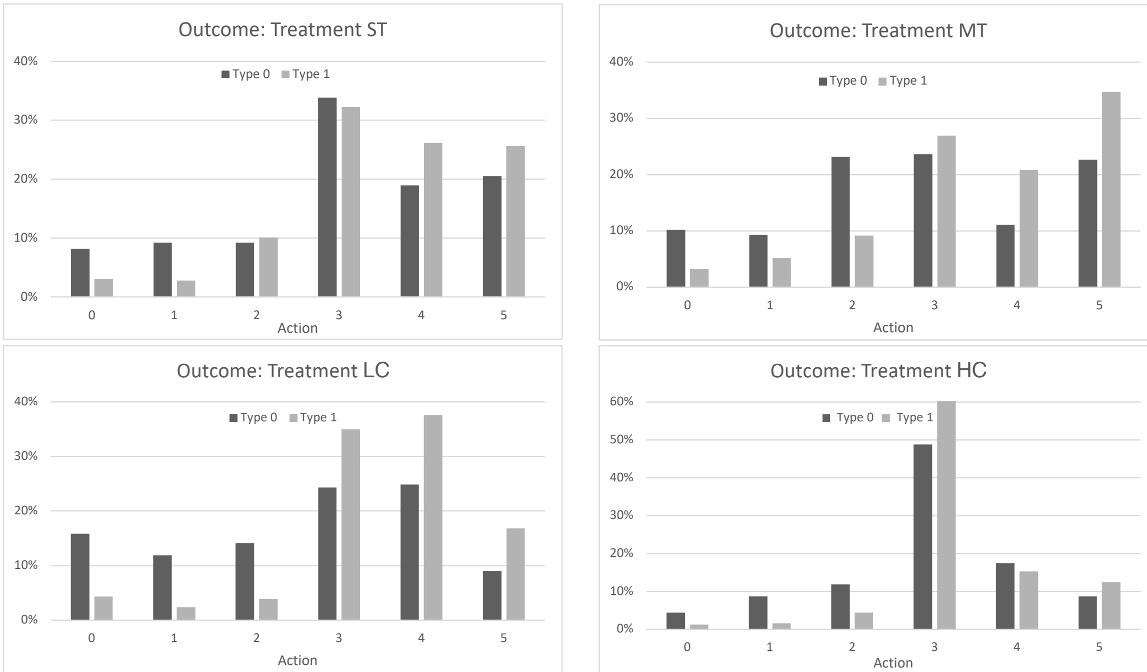


Figure 10: Outcome Comparison

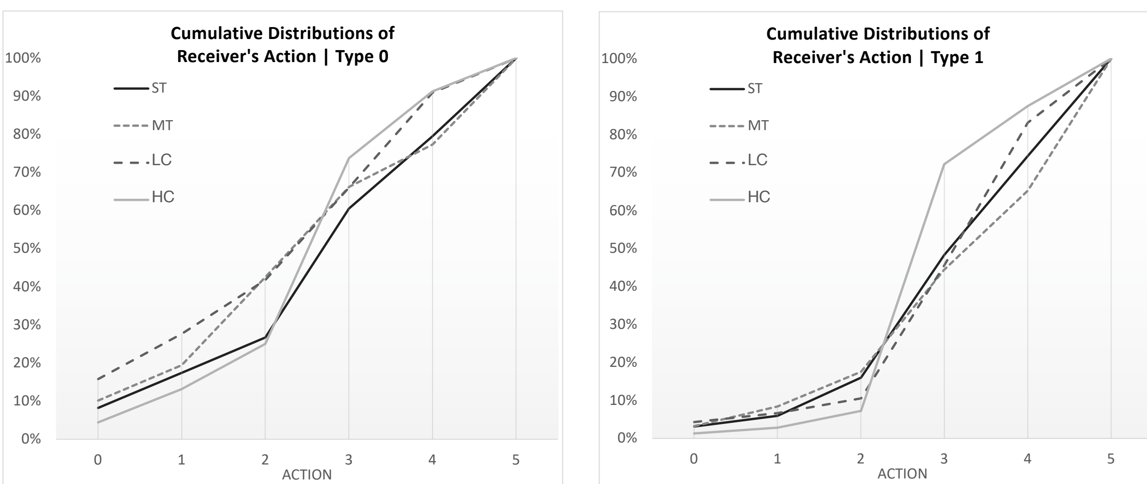


Figure 11: Outcome (CDF): Aggregate Level

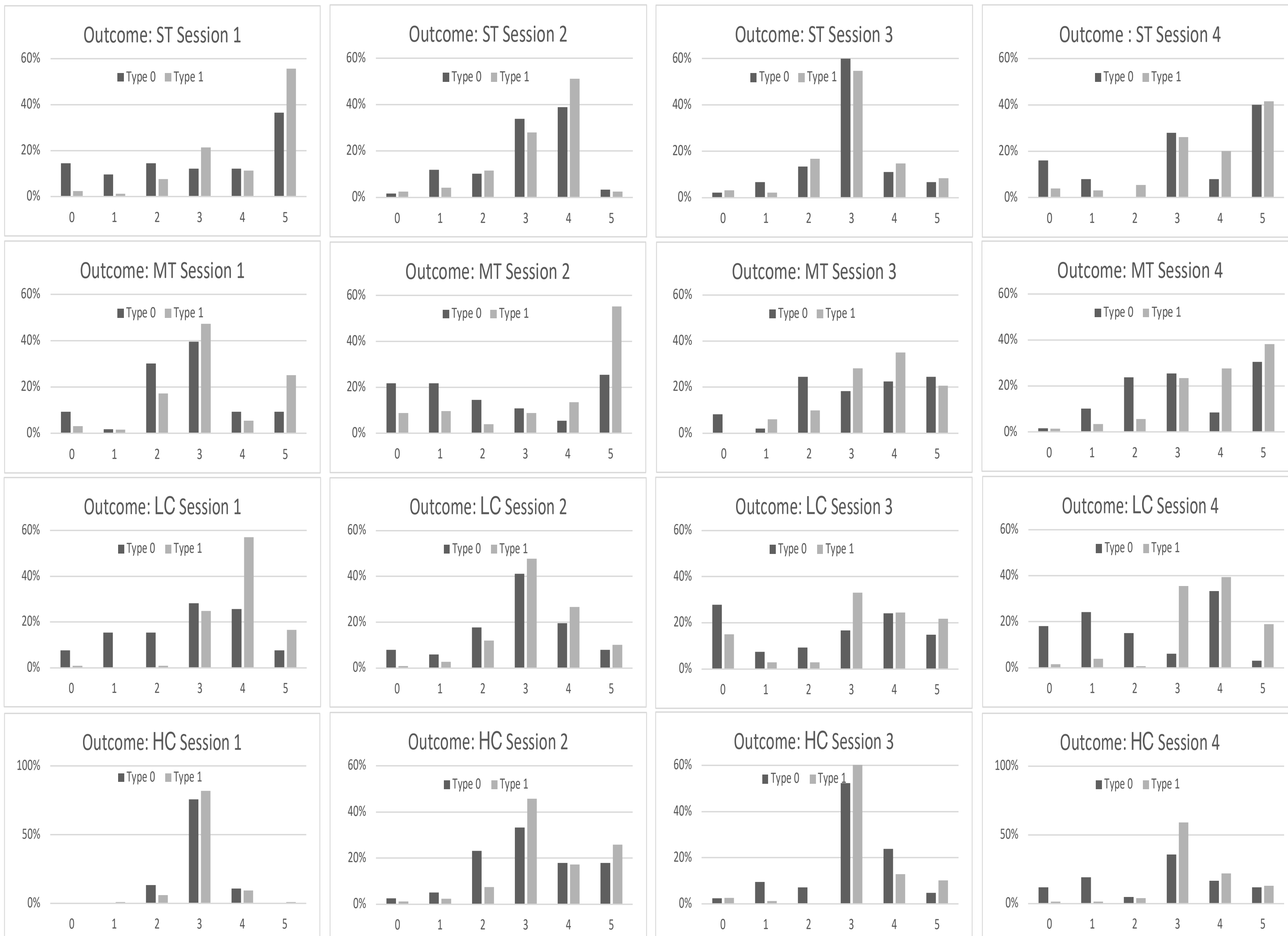


Figure 12: Outcome Comparison: Session Level

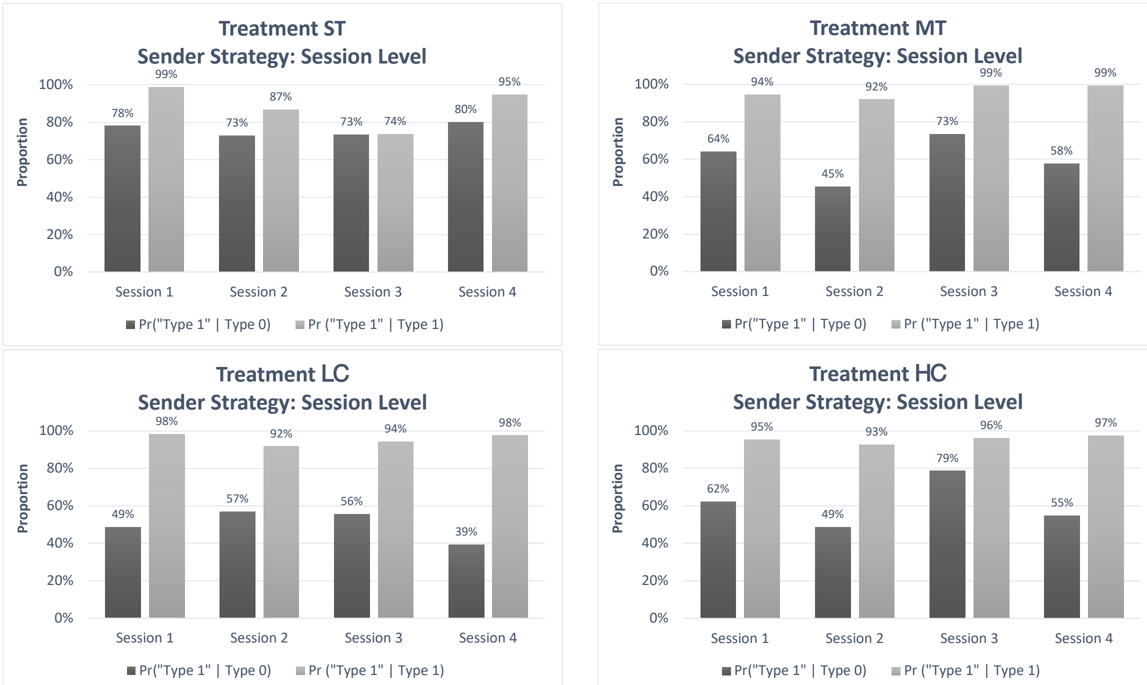


Figure 13: Sender Strategy: Session Level

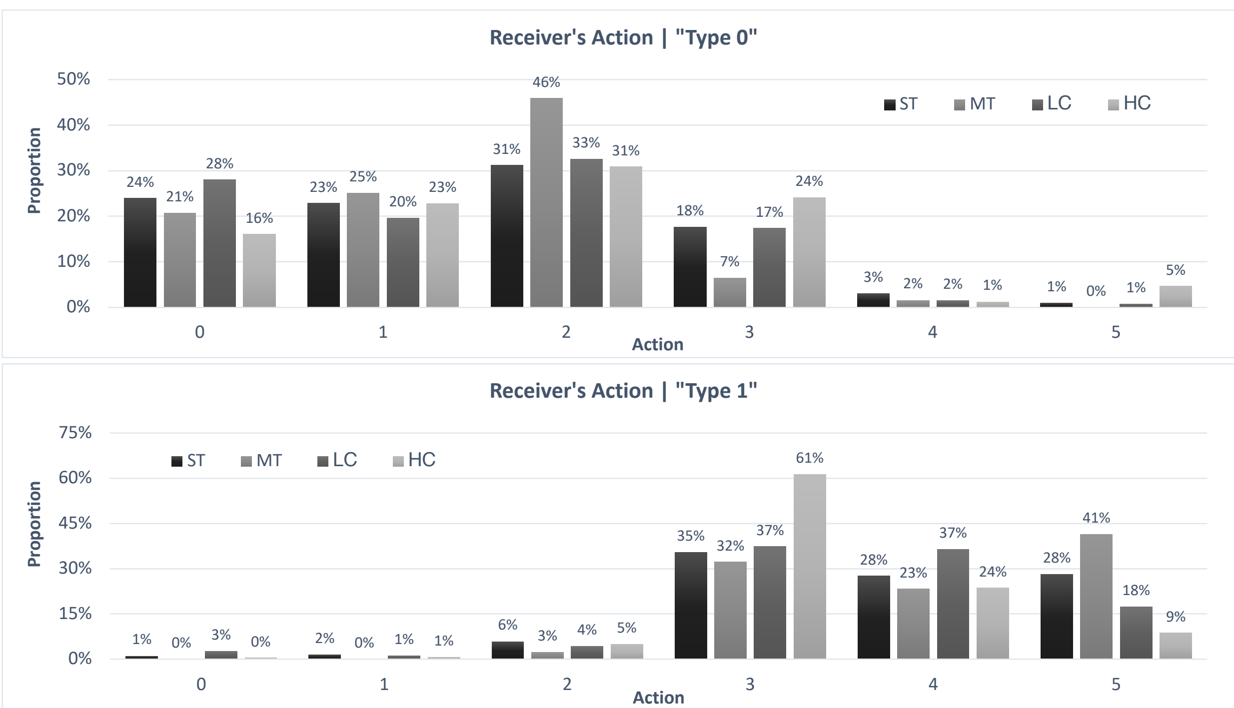


Figure 14: Receiver Strategy: Aggregate Level

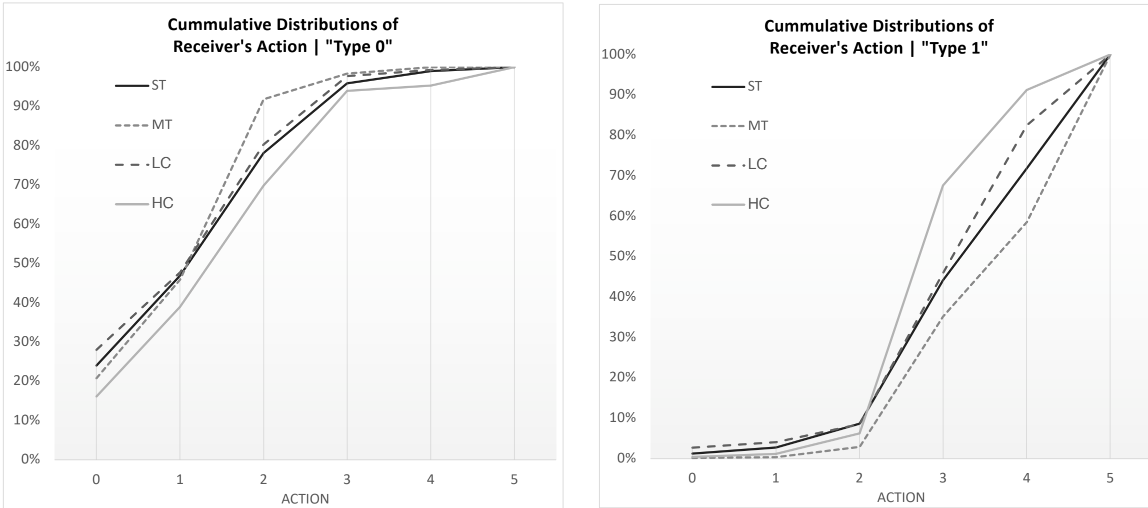


Figure 15: Receiver Strategy (CDF): Aggregate Level

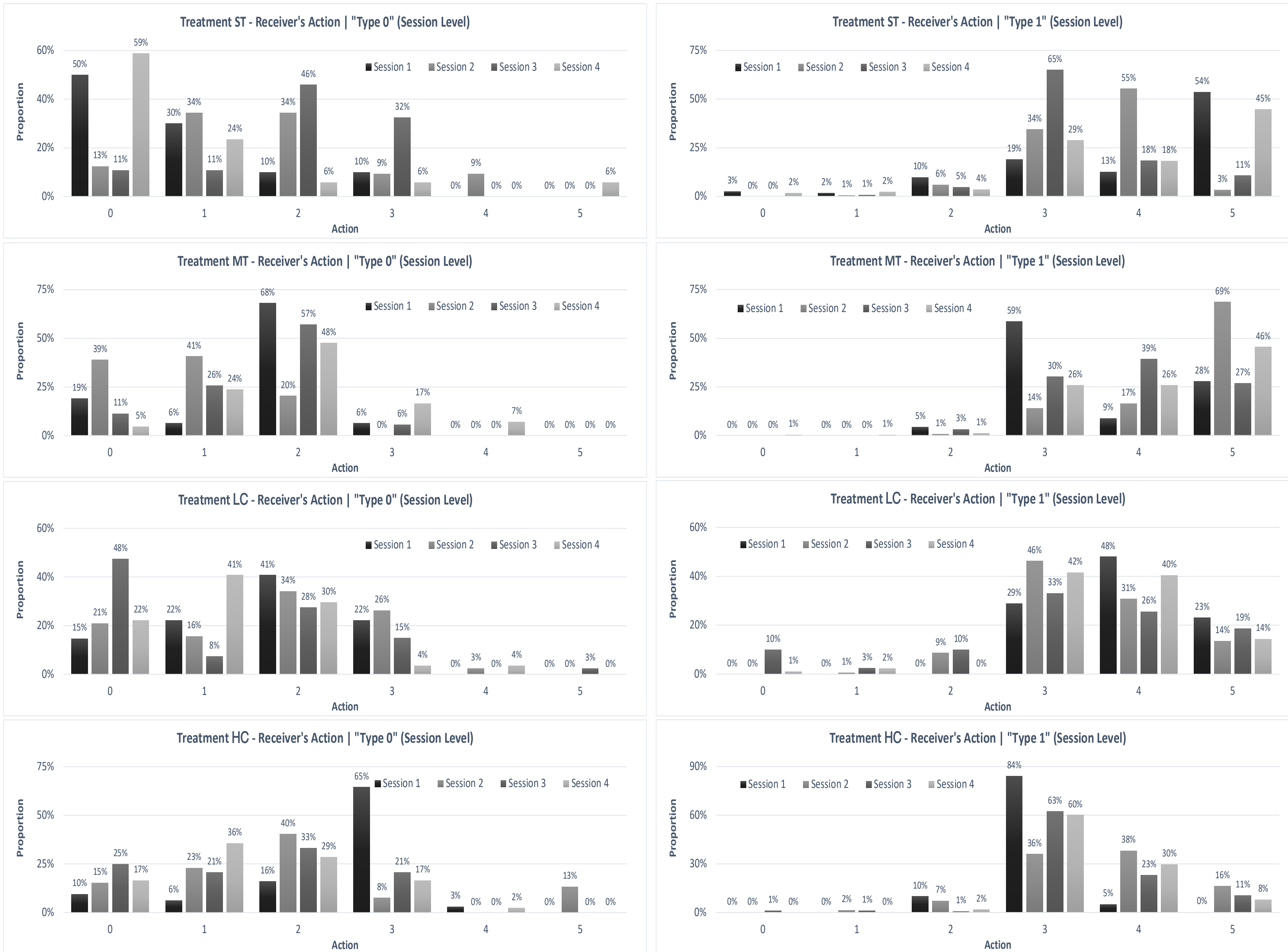


Figure 16: Receiver Action (Session Level) Conditional Upon “Type 0” (Left) and Upon “Type 1” (Right)

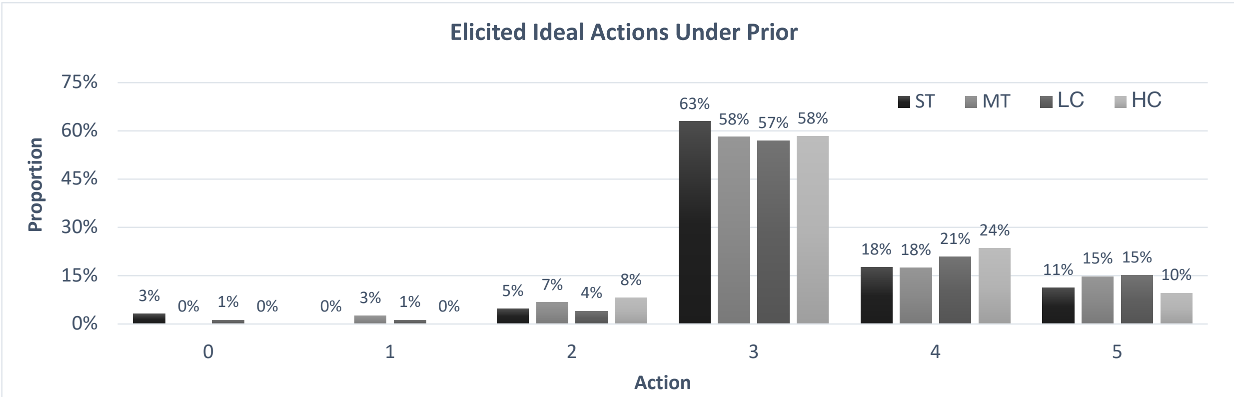


Figure 17: Elicited Action Under the Prior Belief

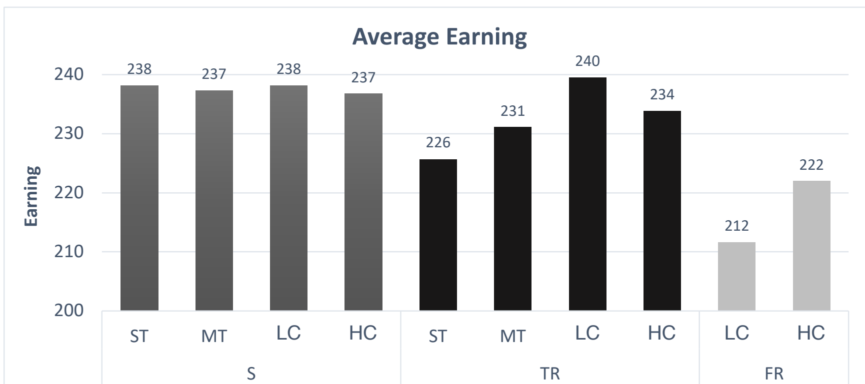


Figure 18: Average Earning

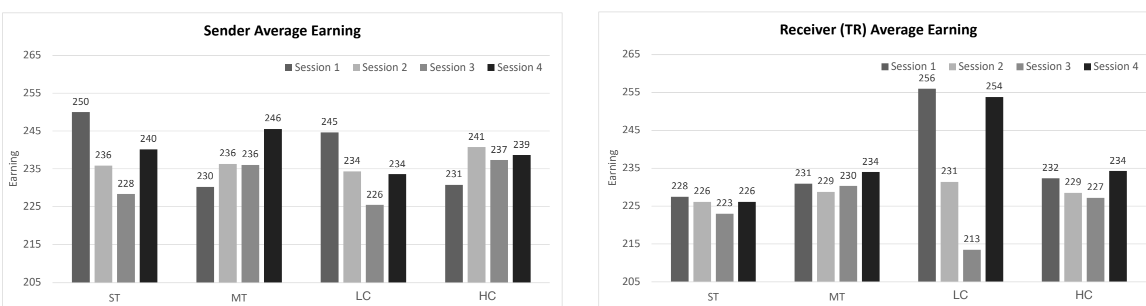


Figure 19: Average Earning Session Level

D Non-parametric Tests Results

Test	1/2-sided?	Null Hypothesis	Treatment Averages	p-values
BD	1	In Treatment ST, the cumulative distribution of receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.003 (Null) / 0.8727 (Reversed Mull)
		In Treatment MT, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.0030 (Null) / 0.8601 (Reversed Mull)
		In Treatment LC, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.000 (Null) / 0.8658 (Reversed Mull)
		In Treatment HC, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.0019 (Null) / 0.8634 (Reversed Mull)
MW	1	$\Delta^O_{ST} < \Delta^O_{MT}$	$\Delta^O_{ST} = 0.0075$ $\Delta^O_{MT} = 0.1268$ $\Delta^O_{LC} = 0.1540$ $\Delta^O_{HC} = 0.0603$	0.17143
		$\Delta^O_{ST} < \Delta^O_{LC}$		0.1000
		$\Delta^O_{ST} < \Delta^O_{HC}$		0.44286
		$\Delta^O_{MT} < \Delta^O_{LC}$		0.82857
		$\Delta^O_{MT} > \Delta^O_{HC}$		0.17143
		$\Delta^O_{LC} > \Delta^O_{HC}$		0.05714
Wil	1	In Treatment ST, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	76% vs. 88%	0.0625
		In Treatment MT, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	60% vs. 96%	0.0625
		In Treatment LC, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	51% vs. 96%	0.0625
		In Treatment HC, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	61% vs. 95%	0.0625
MW	1	The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.	76% (ST) 60% (MT) 51% (LC) 61% (HC)	0.05714
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.01429
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.10000
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.10000
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.55714
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.19182

■ BD, MW, Wilc, and KS refer to the Barett-Donald test, Mann-Whitney U (rank-sum) test, one-sample Wilcoxon (signed rank) test, and Kolmogorov-Smirnov test, respectively.
■ The test statistic of the BD test

Table 4: Non-parametric Tests Result I

Test	1/2-sided?	Null Hypothesis	Treatment Averages	p-values
BD	1	In Treatment ST, the cumulative distribution of receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.003 (Null) / 0.8727 (Reversed Mull)
		In Treatment MT, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.0030 (Null) / 0.8601 (Reversed Mull)
		In Treatment LC, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.000 (Null) / 0.8658 (Reversed Mull)
		In Treatment HC, the cumulative distribution of the receiver's action conditional on Type 0 (state) 1st-order stochastically dominates that conditional on Type 1.	–	0.0019 (Null) / 0.8634 (Reversed Mull)
MW	1	$\Delta^O_{ST} < \Delta^O_{MT}$	$\Delta^O_{ST} = 0.0075$ $\Delta^O_{MT} = 0.1268$ $\Delta^O_{LC} = 0.1540$ $\Delta^O_{HC} = 0.0603$	0.17143
		$\Delta^O_{ST} < \Delta^O_{LC}$		0.1000
		$\Delta^O_{ST} < \Delta^O_{HC}$		0.44286
		$\Delta^O_{MT} < \Delta^O_{LC}$		0.82857
		$\Delta^O_{MT} > \Delta^O_{HC}$		0.17143
		$\Delta^O_{LC} > \Delta^O_{HC}$		0.05714
Wil	1	In Treatment ST, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	76% vs. 88%	0.0625
		In Treatment MT, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	60% vs. 96%	0.0625
		In Treatment LC, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	51% vs. 96%	0.0625
		In Treatment HC, the proportion of senders sending "Type 1" conditional on Type 0 is smaller than that conditional on Type 1.	61% vs. 95%	0.0625
MW	1	The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.	76% (ST) 60% (MT) 51% (LC) 61% (HC)	0.05714
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.01429
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.10000
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.10000
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.55714
		The proportion of senders sending "Type 1" conditional on Type 0 is larger in Treatment ST than in Treatment MT.		0.19182

■ BD, MW, Wilc, and KS refer to the Barett-Donald test, Mann-Whitney U (rank-sum) test, one-sample Wilcoxon (signed rank) test, and Kolmogorov-Smirnov test, respectively.
■ The test statistic of the BD test

Table 5: Non-parametric Tests Result II

E Level-k Predictions

This section is devoted to providing predictions from the level-k model a la Crawford (2003).

- Standard Talk (ST) Prediction: Sender types pool with "Type 1" and Receiver takes the ex-ante ideal action.

k	Sender		Receiver	
	Type 0	Type 1	"Type 0"	"Type 1"
0	"Type 0"	"Type 1"	0g	5g
1 or higher	**"Type 1"**	**"Type 1"**	**0g (off-path)**	**3g or 4g**

- Mediated Talk (MT) Prediction: Sender types separate by sending truthful messages and Receiver takes 1g and 5g, respectively.

k	Sender		Receiver	
	Type 0	Type 1	"Type 0"	"Type 1"
0	"Type 0"	"Type 1"	1g	5g
1 or higher	**"Type 0"***	**"Type 1"**	**1g**	**5g**

*Indifference is broken by a lexicographical preference for truth-telling.

- Low Cursedness (LC) Prediction: Sender types separate by sending truthful messages and Receiver takes 1g and 5g, respectively.

k	Sender		Receiver	
	Type 0	Type 1	"Type 0"	"Type 1"
0	"Type 0"	"Type 1"	1g	5g
1 or higher	**"Type 0"***	**"Type 1"**	**1g**	**5g**

*Indifference is broken by a lexicographical preference for truth-telling.

- High Cursedness (HC) Prediction: Sender types separate by sending truthful messages and Receiver takes 2g and 4g, respectively.

k	Sender		Receiver	
	Type 0	Type 1	"Type 0"	"Type 1"
0	"Type 0"	"Type 1"	2g	4g
1 or higher	**"Type 0"***	**"Type 1"**	**2g**	**4g**

*Indifference is broken by a lexicographical preference for truth-telling.