Preferences vs. Strategic Thinking: An Investigation of the Causes of Overcommunication*

Jonathan Lafky[†] Ernest K. Lai[‡] Wooyoung Lim[§]

June 13, 2022

Abstract

The extent of information sharing in strategic communication experiments persistently exceeds what theory predicts. The literature identifies homegrown preferences and heterogeneity in strategic thinking as two major causes of overcommunication. We design an experiment that features team decisions and combines strategic and non-strategic communication to evaluate these competing explanations. We find that (a) the vast majority of strategic truthful behavior in communication of private information coincides with best responses to beliefs about opponents, (b) truthful behavior observed in non-strategic communication has limited ability to predict behavior in strategic communication, and (c) other-regarding preferences play a minimal role in influencing communication behavior in our strategic environments. Our findings favor strategic thinking as the primary explanation of overcommunication and caution that truthful preferences documented in non-strategic settings may not readily explain strategic truthful behavior.

Keywords: Strategic Communication; Overcommunication; Truthful Preferences; Level-k Models; Team Design; Laboratory Experiment

JEL classification: C72; C92; D82; D83

*We are grateful to Vince Crawford, Xun Lu, Joel Sobel, the editor, and two anonymous referees for valuable comments and suggestions. We thank participants at the LACBEE 2019 workshop and an online seminar jointly organized by Hong Kong Baptist University and National Taiwan University for useful feedback. Iris Arbogast, Eli Inkelas, and Ilya Kukaev provided excellent research assistance. This study is supported by a grant from the Research Grants Council of Hong Kong (Grant No. GRF-16502015).

†Carleton College, Department of Economics, Northfield, MN 55057. Email: jlafky@carleton.edu. Phone: (507) 222-4103.

‡Lehigh University, Department of Economics, Bethlehem, PA 18015. Email: kwl409@lehigh.edu. Phone: (610) 758-5726. Fax: (610) 758-4677.

§The Hong Kong University of Science and Technology, Department of Economics, Kowloon, Hong Kong. Email: wooyoung@ust.hk. Phone: (852) 2358-7628. Fax: (852) 2358-2084.

"I always tell the truth. Even when I lie."

—Al Pacino as "Tony Montana" in the movie *Scarface* by De Palma and Stone

1 Introduction

The extent of information sharing observed in strategic communication experiments consistently exceeds what equilibrium theory with conventional preferences predicts (two early studies are Forsythe et al., 1999; Blume et al., 2001). This empirical regularity presents a puzzle to theorists and experimentalists, reminiscent of the overgenerous sharing behavior observed in experimental studies of bargaining games. Similar to how those anomalous findings have inspired the incorporation of behavioral motives into economic theory (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), researchers have attempted to rationalize why people reveal information when equilibrium predicts otherwise. Two major approaches, one focused on *preferences* and the other on *strategic thinking*, have been proposed. We design an experiment to evaluate the explanatory powers of these alternative approaches in explaining overcommunication in strategic environments.

The preference-based approach departs from the strictly self-interested model with preferences defined over outcomes. It posits that agents excessively share information either due to altruistic preferences or innate preferences for being truthful. In a seminal experimental study, Gneezy (2005) finds evidence of both types of preferences at work. Subjects refrain from lying because they take into consideration the welfare consequences lying causes to others. Given the same social allocations, subjects also avoid taking self-beneficial actions if the action involves lying. This second finding establishes a preference for being truthful that is independent of its consequences—the means to achieve a given end matter. On the theoretical front, Kartik et al. (2007) introduce lying costs—a manifestation of truthful preferences—into the kind of cheap-talk model of Crawford and Sobel (1982). The lying costs in their communication model of private information generate "overcommunication," admitting a separating equilibrium otherwise not attainable in the canonical environment.

The strategic-thinking approach, while preserving standard preference maximization, allows beliefs to depart from the equilibrium requirement that they be consistent with the behavior they generate. Players hold and best respond to heterogenous beliefs about their opponents, which is operationalized by positing that they interact in a cognitive hierarchy of strategic thinking. A player of certain level in the hierarchy believes that others are of lower level and responds accordingly. The profile of behavior generated from the hierarchy of a given game is then applied to rationalize observed plays of the game. The non-equilibrium structures, commonly referred to as the level-k models, have been successfully applied to account for systematic departures from equilibrium observed in experimental games (see, e.g., the survey by Crawford et al., 2013).

Crawford (2003) is the first to apply the level-k paradigm to study communication. Using communication of intentions as the vehicle, he provides a formal framework for us to systematically think about variations in strategic sophistication and their consequences on communication behavior.[1] While information sharing may be at odds with equilibrium, it may be consistent with the behavior of some level types in the hierarchy of strategic sophistication. Level-k models of communication typically anchor in a lowest level type that exhibits truthful behavior. In contrast to an equilibrium reached by an iterative reasoning process, which would entirely remove the effects of this truthful anchor, the iterated best responses in a level-k model chip away some but not all of the effects. This allows the non-equilibrium model to generate a structure to explain observed patterns of overcommunication such as those documented in Cai and Wang (2006), Wang et al. (2010), and Lai and Lim (2012).[2]

The preference-based and strategic-thinking approaches, though complementary in their objectives, are non-intersecting in their lines of attack, each dealing with a different component of the theory of strategic behavior. An experiment that "searches for facts" (Roth, 1987) in this regard, documenting the relative importance of these components in driving overcommunication, provides inputs for "speaking to theorists" (Roth, 1987) about how theories of communication may be systematically improved to admit the anomalies into its domain, a goal of Sobel (2013) in suggesting unexplored experiments on communication. We attempt to achieve this objective with three features of our experimental design; we examine the roles of truthful preferences, other-regarding preferences, and heterogeneity in strategic thinking in driving overcommunication through the uses of a within-subject variation of communication tasks, asymmetric knowledge of payment rounds, and, most importantly, a team design.

Overcommunication often manifests as excessive truthful behavior. Recent studies of truthful behavior have directed their attention to non-strategic environments. A notable example is the self-reporting game developed by Fischbacher and Föllmi-Heusi (2013), in which sub-

[1] Though not in a level-k framework, Kartik et al. (2007) also consider the effects of introducing naive players into strategic communication models.

[2] In addition to the general survey by Crawford et al. (2013), refer to Crawford's presentation slides "Deceived while Rational? Game-theoretic Models of Deception and Gullibility" (available at http://econweb.ucsd.edu/~v2crawford/Deceived%20while%20Rational.pdf; last accessed June 26, 2021) for a more devoted treatment of the applications of level-k reasoning to strategic communication.

jects report private information to the experimenter. This specialization has the advantage of isolating truthful preferences from motives that operate in strategic or social contexts, such as other-regarding preferences. The absence of a participating counter-party, however, may limit the external validity to explain truthful behavior observed in strategic settings.

The first feature of our design is a response to this concern. We gauge the extent to which strategic and non-strategic communication behavior is related via a within-subject experimental design. Each subject takes part in three communication tasks, two being strategic and one non-strategic. One of the strategic games involves communication of private information and the other communication of intentions. The non-strategic task is a variant of the self-reporting game in Fischbacher and Föllmi-Heusi (2013). It is imperative to recognize that our objective is not to establish the phenomenon of overcommunication, which has been well documented. Instead, our goal is to examine the underlying causes of overcommunication. The two strategic games we deploy, in which the most informative communication supported by equilibrium is no communication, provide us with the "best environments" for overcommunication to occur.

The second feature of our design leverages asymmetric knowledge of payment rounds to isolate the effects of social preferences. While subjects are uninformed of their own randomly selected payment rounds until the end of the experiment, they are told whether the current round is their opponents' payment round. This "amended random payment method," which is developed by Lim and Xiong (2021) and similar to that of Lafky (2014), allows us to attribute any significant variation in communication behavior across payment and non-payment rounds of opponents to the influences of other-regarding preferences.

A strong correlation of truthful behavior across the strategic and non-strategic games would lend force to truthful preferences being the culprit behind overcommunication. Similarly, any truthful behavior observed in opponent payment rounds but not in non-payment rounds would indicate altruistic preferences at work. The absence of these findings, however, does not by itself point to heterogeneity in strategic thinking as the cause of overcommunication. The third feature of our design, which we consider to be the most important feature highlighting the methodological contribution of our communication experiment, fills the void by utilizing the team design developed by Burchardi and Penczynski (2014) to elicit the reasoning behind subject decisions in communication.[3] In each of our two strategic communication games, two subjects in the same role form a team and engage in an exchange of opinions on how to play. The opinions, which are elicited in an incentive compatible fashion, offer us a window into subject beliefs about their opponents, how they make decisions given those beliefs, and what

[3] The team design developed by Burchardi and Penczynski (2014) in turn builds on Cooper and Kagel (2005).

other motives might drive their decisions.

The team design generates a rich set of data on beliefs and motives that is indispensable to our inquiry into the causes of overcommunication. While the team design for the strategic games allows us to match subjects' strategic motives with their individual behavior in the non-strategic task, the design necessarily confounds moving between strategic and non-strategic environments with moving between team and individual decision making. To address the possibility of these confounds, we also conduct a robustness treatment. Otherwise identical to the main treatment, this robustness treatment omits the team design and belief elicitations.

The principal findings of our experiment is that senders and receivers best respond to their beliefs, but those beliefs depart from equilibrium. We observe overcommunication in the strategic game of communicating private information but not in the game of communicating intentions. The elicited opinions reveal, however, that in both games subjects hold heterogeneous beliefs about their opponents. For senders, only a small number of cases are consistent with equilibrium beliefs. There are more equilibrium beliefs documented for receivers, but they comprise at most half of the cases. The vast majority of senders best respond to their beliefs in communication of private information, which suggests that some of the overcommunication observed in the form of truthful behavior may indeed be best responses to non-equilibrium beliefs, consistent with the founding premise of the level-k paradigm. There are fewer best responses in communication of intentions, but those still make up the majority of cases.

Regarding the two preference motives we investigate, we find, first, that truthful behavior in the non-strategic task has limited ability to predict behavior in the strategic games; second, we find no evidence that communication in our strategic games is impacted by other-regarding preferences. These results continue to hold in the robustness treatment. Our findings on subjects' beliefs and motives made possible by the team design provide clear evidence that heterogeneity in strategic thinking plays a major role in the observed overcommunication, and our cross-task comparisons caution that truthful preferences documented in non-strategic settings may not readily explain strategic truthful behavior.

Related Literature

The laboratory phenomenon of excessive transmission of information has been widely documented in the two constituent strands of the experimental literature on communication, strategic communication (e.g., Forsythe et al., 1999; Blume et al., 2001; Gneezy, 2005; Cai and Wang, 2006; Sánchez-Pagés and Vorsatz, 2007; Kawagoe and Takizawa, 2009; Wang et al., 2010; Lai and Lim, 2012; Gneezy et al., 2013) and non-strategic communication (e.g., Gibson et al., 2013; Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2018; Abeler et al., 2019).

For strategic communication of private information, Gneezy (2005), Hurkens and Kartik (2009), and Sánchez-Pagés and Vorsatz (2007), among others, attribute overcommunciation to truthful or social preferences.[4] Preference-based explanations are also offered to account for behavior observed in communication of intention. Charness and Dufwenberg (2006), e.g., find evidence of "guilt aversion," where subjects make choices in order to avoid the guilt from not living up to others' expectations shaped by their communicated intention. Vanberg (2008) evaluate the relative explanatory powers of a variant of truthful preferences—preferences for keeping promises—and guilt aversion. The evidence suggests that subjects have an intrinsic preference for keeping their word.

Cai and Wang (2006) are the first to adopt the level-k approach to communication pioneered by Crawford (2003) to explain overcommunication observed in the laboratory. The approach has since become one of the dominant avenues to rationalize non-equilibrium behavior, which often takes the form of truth telling, observed in communication games of private information. Subsequent studies in this vein include Kawagoe and Takizawa (2009), Wang et al. (2010), and Lai and Lim (2012).[5]

In the absence of a participating counter-party, preferences have predictably been the sole explanation for non-optimal truthful behavior observed in non-strategic settings. Recognizing the difficulty of isolating truthful preferences from social preferences in strategic environments, Gibson et al. (2013) design a decision-theoretic experiment to understand the nature of the former preferences. They find that subjects occupy a wide spectrum of heterogeneous truthful preferences rather than a dichotomy of truthful and economic types. The self-reporting game of Fischbacher and Föllmi-Heusi (2013) represents another contemporaneous attempt to study non-strategic truthful behavior and has become a workhorse for this line of studies. Gneezy et al. (2018) demonstrate that the partial lying observed in this kind of self-reporting games can be explained by combining intrinsic truthful preferences and image concern. Abeler et al. (2019) arrive at the same conclusion with a new set of experiments and a meta-study of prior experiments in economics, psychology, and sociology.[6]

Our contribution can be seen as answering two questions that are largely unaddressed in the experimental literature on communication reviewed above. First, within the strand of strategic communication, the two approaches to account for overcommunication, truthful preferences and heterogeneity in strategic thinking, are studied in separate experiments and

[4]In addition to Kartik et al. (2007), theoretical studies that incorporate truthful preferences or lying costs into models of strategic communication include Kartik (2009), Chen et al. (2008), and Chen (2011).

[5]Refer to Blume et al. (2020) for the most recent survey of this literature.

[6]Dufwenberg and Dufwenberg (2018) and Khalmetski and Sliwka (2019) provide pure theoretical analysis of truth-telling preferences and image concern operating in the self-reporting game environment.

appear to be disconnected with no dialogue between them.[7] We bridge this gap by using *a single experiment* to examine the relative importance of the two approaches in explaining departures from equilibrium predictions.[8]

Second, as alluded to above, there is a tendency in the non-strategic strand of the literature to extrapolate the truthful preferences documented therein to explain overcommunication in strategic settings. While it is a fruitful direction to use decision-theoretic experiments to disentangle truthful and other-regarding preferences, strategic overcommunication may be driven by reasons other than social preferences. By implementing a within-subject design that combines strategic and non-strategic communication with asymmetric knowledge of payment rounds, we can observe the correlation between behavior in the different communication environments while controlling for other-regarding preferences.[9]

The rest of the paper proceeds as follows. Section 2 analyzes the equilibrium or optimal behavior in the three communication games that serve as our experimental tasks. Section 3 describes our experimental design. Section 4 presents the findings. Section 5 concludes.

2 Communication Games

Subjects participate in three tasks presented to them as Task I, Task II, and Task III. Task I is a strategic communication game of private information, Task II is a strategic communication game of intention, and Task III is a non-strategic, self-reporting game of private information. In Sections 2.1–2.3, we describe each game in turn and characterize the equilibrium or optimal behavior under the standard paradigm that players are strictly self-interested and fully strategic. The characterizations provide benchmarks with which we define overcommunication.

Successful communication in a strategic setting depends upon the encoding and decoding of messages but does not necessarily require literally truthful behavior on the sender's part. By contrast, in a non-strategic setting, the extent of communication amounts to the extent to which truth is told. In Section 2.4, we discuss the relationship between overcommunication,

[7] An exception is Wood (2016), who structurally estimates subjects' cognitive levels and the fraction of honest subjects in his experiment on the role of vague language in communication. Consistent with our findings, his estimation shows that bounded rationality trumps honesty in explaining the observed communication.

[8] In suggesting ten possible experiments on communication and deception, Sobel (2013) states that "[t]he mechanism underlying this [overcommunication] phenomenon is uncertain" and that "[i]t would be useful to construct tests that distinguish between lying cost models and level-k models."

[9] Sutter (2009) also utilizes teams to study strategic communication of private information. In the communication environment of Gneezy (2005), he finds that a significant proportion of senders deceives by literally telling the truth in anticipation of receivers' mistrust. Our experiment differs in terms of the question addressed and the scope of the design as described above.

truthful behavior, and lying, explaining how subtly different preferences for truthfulness can influence the interpretation of observed communication behavior.

2.1 Task I: Strategic Communication of Private Information

A sender privately observes the state of the world $\theta^{\mathrm{I}} \in \Theta^{\mathrm{I}} = \{\mathrm{Blue}, \mathrm{Green}\}$. Blue and Green are commonly known to be equally likely. After observing θ^{I}, the sender sends a cheap-talk message $m^{\mathrm{I}} \in M^{\mathrm{I}}$ to a receiver, where M^{I} contains three literal messages, "The color is Blue," "The color is Green," and "The color is either Blue or Green."[10] After receiving a message, the receiver takes an action $a^{\mathrm{I}} \in \{\mathrm{X}, \mathrm{Y}, \mathrm{Z}\}$. Table 1(a) presents the payoffs. In each cell, the numbers represent (sender payoff, receiver payoff) for the corresponding state-action pair.

	X	Y	Z
Blue	80, 80	10, 10	50, 50
Green	30, 10	10, 80	50, 50

(a) Payoffs

	X	Y	Z
Blue	0%	0%	50%
Green	0%	0%	50%

(b) Equilibrium Outcome

Table 1: Task I

An outcome of the game is a joint distribution over states and actions. In cheap-talk games, there are multiple equilibria with different encodings and decodings of messages that are associated with a given outcome. We consider *overcommunication as a phenomenon in communication outcomes.* The theoretical benchmark for establishing the occurrence of over-communication is the most informative equilibrium outcome of the game.[11]

Separation, in which distinct messages are sent for Blue and Green, cannot be an equilibrium outcome. To see this, note that with separation the receiver takes ideal actions X in state Blue and Y in state Green. The sender in state Green would then receive the lowest possible payoff, 10, and has a strict incentive to mimic being the sender who has observed Blue to obtain a payoff of 30. On the other hand, as in any cheap-talk game, pooling is always an equilibrium outcome. In Appendix A, we show that pooling is indeed the unique—and thus the most informative—equilibrium outcome of the game. Regardless of the state, the receiver takes the action that is optimal under the prior, which is Z. Table 1(b) shows the

[10] The messages stated in complete sentences are the actual choices provided to subjects in the experiment. For brevity, from now on we refer to them as "Blue," "Green," and "Either."

[11] The equilibrium concept is perfect Bayesian equilibrium, in which the sender best responds to the receiver, the receiver best responds to beliefs, and those beliefs are derived from Bayes' rule whenever possible.

joint distribution over states and actions of this pooling equilibrium outcome.

The maximal communication predicted by equilibrium is *no communication*. Any communication outcome that gravitates toward separation therefore indicates overcommunication. Applying the joint distribution in Table 1(b) to the receiver payoff in Table 1(a), we compute the receiver expected payoff in equilibrium to be 50. On the other hand, when outcomes gravitate toward separation, the receiver expected payoff approaches 80. These expected payoffs give us an operationalizable definition of overcommunication and an empirical measure based on average payoff that can be readily applied to the data:

Definition 1. *Overcommunication in Task I occurs when average receiver payoffs are significantly greater than 50.*

2.2 Task II: Strategic Communication of Intention

A sender communicates to a receiver about the sender's intended action in a matching pennies game. Table 2(a) presents the normal-form game, where each player chooses an action $a^{\text{II}} \in \{\text{H}, \text{T}\}$. The sender prefers to match the pennies, while the receiver prefers to mismatch. Before the matching pennies is played, the sender sends a cheap-talk messages $m^{\text{II}} \in M^{\text{II}}$, where M^{II} contains two messages each indicating the intention to take one of the two actions. In the experiment, the two literal messages provided under the team design are "We will choose H" and "We will choose T."[12] The players choose their actions simultaneously after the receiver receives m^{II}.

Sender \ Receiver	H	T
H	$40, 20$	$20, 40$
T	$20, 40$	$40, 20$

(a) Normal Form

Sender \ Receiver	H	T
H	25%	25%
T	25%	25%

(b) Equilibrium Outcome

Table 2: Task II

The game has a unique mixed-strategy Nash equilibrium, in which both players uniformly randomize between H and T. The equilibrium is *independent of* the pre-play communication: if the receiver believes that the sender will choose a particular action, the sender's best response

[12] For brevity, from now on we refer to the messages as "H" and "T."

is to choose a different action, which implies that in equilibrium the receiver must ignore the sender's message and babbling by the sender is a best response. Table 2(b) shows the unique equilibrium outcome of the game as a distribution over the four action profiles.

Given that in the equilibrium the players are equally likely to face their favorable action profiles with payoff 40 and unfavorable action profiles with payoff 20, their equilibrium expected payoffs are 30. If the receiver faces the favorable mismatching action profiles more often than the unfavorable matching action profiles, overcommunication is implied. Similar to the definition for Task I, we define overcommunication in terms of payoffs, which give us a handy summary variable as an empirical barometer to determine the occurrence of the phenomenon:

Definition 2. *Overcommunication in Task II occurs when average receiver payoffs are significantly greater than 30.*[13]

2.3 Task III: Non-Strategic Reporting of Private Information

A single decision maker observes the realizations of two independent and uniformly distributed random variables, each taking an integer value from one to six. The decision maker is then asked to report the lower of the two realized numbers, θ^{III}, by sending m^{III}. There is no player on the receiving end, hence the non-strategic nature of the communication. Table 3 presents the payoffs, which depend only on the number reported.

Choice of Report	1	2	3	4	5	6
Payoff	10	20	30	40	50	60
Optimal Report	0%	0%	0%	0%	0%	100%

Table 3: Task III

Optimal reporting by a strictly self-interested decision maker involves choosing the highest number, six, 100% of the time irrespective of the realized numbers, resulting in a payoff of 60. Overcommunication, which has so far been defined in terms of outcomes of the games, has its counterpart in Task III as truthful reports. The decision maker is truthful if the reported

[13] From the perspective of information theory, the equilibrium outcome, where H and T are chosen with uniform probabilities, has the maximal level of entropy, and any departure from it, be it favorable to the receiver or the sender, is a manifestation of "order." Linking this to communication, one may argue that overcommunication occurs not only when the receiver payoff is greater than 30 but also when it is less than 30. As will become clear below, however, adopting either definition is not consequential, as we do not observe average receiver payoffs that are below 30.

number is the same as the lower of the two realized numbers. In inferring preferences from observed behavior, a reported number less than six with an associated payoff less than 60 indicates the presence of at least some level of truth-telling preferences. This motivates the following empirical measure that parallels Definitions 1 and 2 to establish the presence of truthful behavior in Task III:

Definition 3. *Truthful behavior occurs in Task III when average payoffs are significantly less than 60.*

While average payoffs show whether truth-telling preferences are at work based on aggregate behavior, reporting behavior by individual subjects is essential to our analysis. We cannot tell whether truth-telling preferences are present for a subject when the reported number and the underlying realized number are both six. We use two random variables instead of one, effectively rolling two dice, as an attempt to decrease the likelihood of this ambiguous event.

2.4 Overcommunication, Truthful Behavior, and Lying

Our goal in identifying the causes of strategic overcommunication is to understand the primary *behavior* that leads to the communication *outcome*. With communication in Tasks I and II being conducted in cheap-talk environments, where meanings are endogenous, overcommunication can result from both truthful or non-truthful revealing behavior. For example, revealing behavior with full separation in Task I could involve truthfully sending "Blue" in state Blue and "Green" in state Green or non-truthfully sending "Green" in Blue and "Blue" in Green. On the other hand, "overcommunication" in Task III is synonymous with truth being told, and there is virtually no difference between behavior and outcome in the non-strategic task.

As we explain below, part of our empirical strategy is to compare communication behavior across strategic and non-strategic tasks. Given that overcommunication in Task III amounts to truth telling, comparing like with like calls for carefully defining truthful behavior in the strategic tasks to supplement Definitions 1 and 2. To this end, we adopt the approach of Sobel (2020), who leverages exogenous literal meanings in defining truth and lies. The following is a slight variant of his Definition 1:

Definition 4. *For a subset of the possible states of the world $\Theta_0 \subseteq \Theta$, denote by $m_{\Theta_0} \in M$ the message with exogenous literal meaning that m_{Θ_0} means $\theta \in \Theta_0$.*

(a) The message m is a lie given θ if $m = m_{\Theta_0}$ and $\theta \notin \Theta_0$.

(b) The message m is incompletely true given θ if $m = m_{\Theta_0}$, $\theta \in \Theta_0$, and $|\Theta_0| > 1$.

(c) The message m is completely true given θ if $m = m_{\Theta_0}$, $\theta \in \Theta_0$, and $|\Theta_0| = 1$.

We refer to both completely and incompletely true messages as "true."[14] Applying Definition 4 to Task I, we see that (a) in state Blue, "Blue" is completely true and "Green" is a lie, (b) in state Green, "Green" is completely true and "Blue" is a lie, and (c) in both states, "Either" is incompletely true. Considering sender intention in Task II as a state of the world, it is straightforward that "H" ("T") is completely true and "T" ("H") is a lie when the sender intends to take action H (T). There is no incompletely true message in Task II.[15] Similarly, there is no incompletely true report in Task III. For brevity, we generally refer to true messages or reports for Tasks II and III without "completely."

Distinguishing between completely and incompletely true messages introduces some nuance in how we analyze and interpret the motivations behind truthful behavior. Specifically, how one posits the direct payoffs from sending completely versus incompletely true messages affects the interpretation of truthful behavior in Task I and its comparisons to the other tasks. If a sender's motivation for truthfulness is considered an attempt to provide as full and accurate a description as possible, then the highest truthful payoff is obtained when the sender is completely truthful, and the incompletely true "Either" should yield a lower payoff than a completely true "Blue" or "Green." By contrast, if the sender is unconcerned with the informativeness of the message but simply seeks to avoid sending a message that is literally false, then "Either" should be functionally the same as a true "Blue" or "Green."[16]

Truthful or truth-telling preferences and lying aversion are commonly considered synonyms in the literature. The two motivations described above suggest, however, that there is a benefit from refining the terminology by considering lying aversion as one subtype of truthful preferences and using "truth affinity," the desire to communicate as accurately as possible, as another distinct subtype. We explore both motivations in our evaluation of the preference-based approach to overcommunication and revisit the issue in Section 4 when analyzing the experimental findings.[17]

[14] Sobel (2020) defines lies, truth, and whole truth, in which whole truth, which corresponds to our complete truth, is a special case of truth. Our incomplete truth, which is not explicitly defined in his Definition 1, is a truth that is not a whole truth by his definition. We adopt a different anchor for our definition to better describe truth and lies for our specific experimental games.

[15] While both Tasks I and II have two states of the world, we provide three messages for Task I and only two messages without an incompletely true message for Task II because there are three receiver actions in the former and only two in the latter.

[16] Sánchez-Pagés and Vorsatz (2009) and Khalmetski et al. (2017), e.g., demonstrate experimentally that the provisions of ambiguous messages like our incompletely true "Either" allow senders to avoid the cost associated with overt lying by remaining silent or engaging in evasive lying.

[17] We thank an anonymous referee for pointing out the subtly different types of truthful behavior in our Task I and urging us to clarify the issue, which has led to this subsection.

3 Experimental Design

3.1 Treatments and Procedures

Our experiment was implemented in English using z-Tree (Fischbacher, 2007) and networked computers at The Hong Kong University of Science and Technology. We conducted two treatments, labeled TEAM and SOLO, which differ with respect to whether a team design is used. All subjects in both treatments participated in the three tasks described in Section 2. In the TEAM treatment, which is our main treatment, subjects performed the strategic Tasks I and II in teams and the non-strategic Task III individually.

The team design, which will be described in details in Section 3.2, elicits information about how subjects approach the strategic tasks. It forms an indispensable part of our investigation and underlies the main contribution of our study. However, the fact that the within-treatment variation from strategic to non-strategic tasks coincides with the variation from team to individual decisions introduces potential confounds to identifying the causes of differences in strategic and non-strategic behavior. We introduce the SOLO treatment to address this concern. Otherwise identical to the main treatment, this secondary treatment has subjects perform all three tasks without being in teams and thus without the elicitations associated with the team design. By examining the choices made in the two treatments, we are able identify if the team design itself leads to significant differences between strategic and non-strategic communication behavior.

A total of 240 subjects, who had no prior experience with the experiment, were recruited from the undergraduate population of the university; 160 participated in TEAM and 80 in SOLO. For TEAM, ten experimental sessions were conducted. In each session, 16 subjects were equally divided into two independent cohorts for matching purposes, where a subject only interacted with others in the same cohort. For SOLO, nine sessions, each with either eight or ten subjects, were conducted. A subject interacted with others in the same session, meaning that in terms of independent observations each SOLO session is equivalent to a matching cohort in TEAM. For expositional convenience, hereafter we refer to each session in SOLO and each matching cohort in TEAM both as matching cohorts.

In all treatments and all sessions, the experimental tasks were performed in the order of Task I (*two rounds*), Task II (*two rounds*), and then Task III (*one round*). Given the similar strategic nature of Tasks I and II, there is no *a priori* reason to expect that experience with one would influence decisions in the other. On the other hand, with the obvious incentive to lie in the very simple decision environment in Task III, it is conceivable that experience with the non-strategic task may leave an impression on subjects that lying is the goal. We therefore implemented Task III as the last task and did not administer alternative sequences to control for any potential order effect.

Upon arrival at the laboratory, subjects were instructed to sit at separate computer terminals. Each received a copy of the experimental instructions for Task I, which were then read aloud by the experimenter using slide illustrations as an aid. Subjects were told that additional instructions for Tasks II and III would be forthcoming.[18]

The procedure for determining subject payments features an *asymmetric knowledge of payment rounds*, which allows us to control for the potential effects of other-regarding preferences in Tasks I and II. In translating round rewards in these two strategic tasks to cash payments, we randomly selected one round from each task. Subjects did not learn which round would be their own payment round until the end of the experiment. They were, however, told whether the current round was the payment round of their opponents. If, for instance, altruistic preferences were driving overcommunication, we would expect to see more informative messages sent in the opponent payment round than in the opponent non-payment round.

The final cash payment of a subject was made up of earnings from the two randomly selected rounds of Tasks I and II, earning from the single round of Task III, and a show-up fee. For TEAM, the show-up fee was HKD 30, and the payoff points of the tasks were converted into earnings in HKD at an exchange rate of 1 payoff point = HKD 0.7. For SOLO, the show-up fee was HKD 40, and the conversion rate was 1 payoff point = HKD 1. The average cash payments per subject were HKD 127.90 (≈ USD 16.40) for TEAM and HKD 181.5 (≈ USD 23.27) for SOLO. A session lasted on average two hours for TEAM and one and a half hour for SOLO.

3.2 Experimental Tasks and Team Design

In the TEAM treatment, subjects participated in each round of the two strategic communication tasks as teams. Figure 1 presents the team structure in a matching cohort with eight subjects. Subjects were randomly placed into four teams, each with two members.[19] The four teams were randomly matched into two pairs of teams. Each pair of interacting teams is referred to as a decision group. The two teams in a decision group were randomly designated as Team A (senders) and Team B (receivers).

[18] The full experimental instructions for the TEAM treatment can be found in Appendix D.1.

[19] All the randomizations in the experiment, including random matchings of subjects and random draws of private information, were based on uniform distributions.

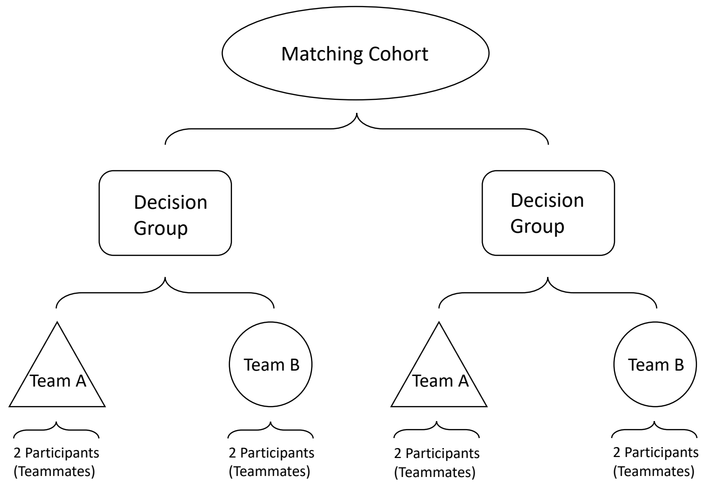


Figure 1: Matching Cohort, Decision Group, and Team

Task I

The random formations of teams and decision groups took place at the beginning of Task I. In each of the two rounds of Task I and for each decision group, the computer randomly selected the color to be either Blue or Green. The selected color was revealed to Team A, and the two members of Team A then engaged in an "exchange of opinions."

Each Team A member was prompted to independently answer five survey questions. Four questions were multiple choice, asking the sender-subjects to indicate (a) what action they anticipated the paired Team B would take in response to each possible message, and (b) which message they should send. For each question, they were also asked to give a free-form written justification for their selection of the multiple-choice option. The fifth question, which took only free-form responses, asked them for any other comment or opinion they wished to share with their teammate.[20] The teammates' answers and justifications were then revealed to each other, after which they independently chose one of three messages, "Blue," "Green," or "Either." The computer randomly selected the message chosen by one member and transmitted it to the paired Team B.

Our team design follows that of Burchardi and Penczynski (2014), who asked subjects to exchange opinions in open-ended written messages only. In view of the subjective nature of text responses, we extend on their design by asking subjects to answer multiple-choice questions in addition to open-ended responses. To the subjects, the exchange of opinions provides an opportunity for them to suggest a choice to their teammate, backed by their written analysis of opponent behavior. To the experimenter, the elicited opinions provide a

[20]The full list of survey questions for Tasks I and II can be found in Appendix D.2.

window into subjects' beliefs about their opponents, how they made decisions given those beliefs, and what motives might have driven their choices. The random execution of choices provides *incentive compatibility*: the fact that one of the teammate choices becomes the team choice incentivizes both members to make and justify suggestions to their teammate based on what they believe to be the "right" choice.

Both members of Team B learned the message from Team A. They then engaged in a similar exchange of opinions, independently answering four survey questions. In addition to open-ended text responses, they were also asked via multiple-choice questions to indicate their opinions on the strategy of Team A and what action they should choose. Their answers were then revealed to each other, after which they independently chose one of three actions, X, Y, or Z. One of the teammate choices was randomly selected to be the team action. Together with the randomly selected color, this team action determined the current-round reward for every subject in the decision group in accordance with the payoff numbers in Table 1(a).

After the first round, subjects were randomly reassigned teammates, but their roles remained the same. Team A (B) subjects were randomly rematched with other Team A (B) subjects. The new teams were then randomly matched to form new decision groups to play the second round. In order not to interfere with subject opinions of how to play the games, *no information feedback* was provided between rounds and between tasks.

Task II

After completing the second round of Task I, subjects received new instructions for Task II and participated in two rounds of the game. Subjects who were in Team A (B) in Task I were again randomly assigned to Team A (B) in Task II, i.e., senders continued to be senders, and receivers continued to be receivers, though teammates were randomly reassigned. The rematching between the first and second round of Task II was performed similarly.

Each round began with an exchange of opinions among Team A members. Each member independently answered five questions regarding (a) the anticipated action choices of the paired Team B, (b) their own choice of action and message, and (c) any additional comment they wished to share with their teammate. Their answers were then revealed to each other, after which they independently chose a message, either "H" or "T," and an action, either H or T. One of the teammate choices was randomly selected to be the team message and action.

After learning the message transmitted from Team A, the two members of Team B exchanged opinions themselves. They independently answered four questions regarding (a) the strategy of Team A, (b) their own action choice, and (c) any additional comments. Each member independently chose an action after learning their teammate's answers, and the computer randomly selected the action chosen by one teammate to be the team action. The actions of Teams A and B then determined the current-round reward for every subject in the decision group in accordance with the payoff numbers in Table 2(a).

Task III

All teams were disbanded after the second round of Task II. Subjects received additional instructions for Task III, in which they participated in one round of the task without any matching. Each subject observed rolls of two computerized six-sided fair dice and was asked to report the smaller of the two rolls. Payoffs were determined solely based on the reports, as described in Table 3. The single round of Task III concluded the experiment.

Task III is similar to the self-reporting game in Fischbacher and Föllmi-Heusi (2013), with two differences. The first difference is the additional roll of dice and the report of the smaller roll for reason explained in Section 2.3. As an essential part of our investigation, we need to know how each subject's behavior in Task III relates to that subject's behavior in Tasks I and II; the second difference is that we associate each realized draw with each subject's report in Task III to examine the correlation, while in Fischbacher and Föllmi-Heusi (2013) the draws are not observed by the experimenter and lying is only detected at the aggregate level. To minimize any experimenter demand effect, we nevertheless reminded subjects that their real identity would not be matched to their reporting behavior.[21]

Tasks in the SOLO Treatment

For the SOLO treatment, subjects completed the same strategic Tasks I and II, but without any teams or exchanges of opinions. In each round, a single Member A (sender) was randomly matched to play with a single Member B (receiver). All other aspects of the tasks remain the same. The non-strategic Task III was implemented in exactly the same way as that in TEAM.

4 Experimental Findings

4.1 Analysis Outline and Summary Statistics

We conduct our primary analysis in Sections 4.2 – 4.6, using data from our main TEAM treatment, and then summarize the key findings from our robustness SOLO treatment in

[21] Kajackaite and Gneezy (2017) and Abeler et al. (2019) obtain evidence that observation of realized states by experimenters in self-reporting games discourages but does not eliminate lying. Given that our focus is not the extent of truthful behavior in non-strategic communication itself but its correlation with that in strategic communication, any discouragement of lying would not undermine our design vis-à-vis its objective, especially since the experimenter's observations of states are controlled for across our three communication tasks.

Section 4.7. To set the stage for investigating the causes of overcommunication, we begin in Section 4.2 with an analysis of aggregate communication outcomes and behavior, evaluating if overcommunication or truthful behavior occurs in the three communication tasks. In Sections 4.3 and 4.4 we analyze the motives behind the observed behavior of senders and receivers in Tasks I and II. We further explore the effects of other-regarding preferences in Section 4.5 by comparing findings from the opponent payment and non-payment rounds. Section 4.6 examines in more detail the non-strategic reporting behavior in Task III and its relation to the strategic communication behavior in Tasks I and II.

The motive analysis in Sections 4.3 and 4.4 represents our main contribution. Key to this analysis is the exchanges of opinions, which reveal otherwise unobservable variables—beliefs and homegrown preferences—that drive subject behavior. The multiple-choice portion of the survey questions effectively elicits subject beliefs about their opponents' strategies, allowing us to scrutinize best-responding behavior given those beliefs. The open-ended text responses provide a further window for probing into subject motives based on their own words. To quantify the open-ended responses, we hired three research assistants, one PhD student and two undergraduates, to classify the response transcripts. The classification focuses on concerns for payoffs and truthfulness as motives but also include other plausible impetuses.

For senders, our RAs independently determine whether each transcript indicates evidence of any of four non-mutually-exclusive attributes, briefly described as follows:[22]

(a) *Truth-telling preferences*: any desire to be truthful for its own sake.

(b) *Payoff concern*: any concern for monetary rewards, including concern for other subjects' reward.

(c) *Image concern*: any concern about one's image to another subject or the experimenter.[23]

(d) *Confusion*: any fundamental misunderstanding of the structure of the game.

For analyzing receiver transcripts, truth-telling preferences are replaced by two related but conceptually distinct attributes, *taste for trust*: any desire to trust others for its own sake; and *credulity*: believing the message to be true without regard to the sender's motives.

Each attribute is recorded by a separate indicator variable, taking the value of one if the attribute is identified to be present and zero otherwise. It is possible for more than one

[22] The detailed definitions of these attributes can be found in the Classification Instructions in Appendix E.

[23] Image concern combined with truth-telling preferences serve to explain the lying behavior observed in self-reporting games (Gneezy et al., 2018; Abeler et al., 2019). Even though we are mainly concerned with strategic communication, image concern is a natural motive to consider given the inclusion of truth-telling preferences.

attribute or none of them to be considered present in a subject's transcript. Our RAs also specify how confident they are with each classification, on a scale from zero to three with three being the most confident. The three RAs agree 96% of the time in their classifications, with an average confidence score of 2.95. In case of disagreements, we use the majority classification.

The motive analysis concerns individual subject behavior. To maintain statistical independence, however, except for the regression analysis in Section 4.6, all our data analyses and statistical tests treat a matching cohort as an independent observation. We therefore have 20 independent observations for the TEAM treatment and nine for the SOLO treatment. All the reported treatment-level figures are obtained after two steps of aggregation, first within a matching cohort to obtain an independent observation and then across the matching cohorts to obtain the reported treatment-level values.[24]

For Tasks I and II, the TEAM treatment generates three alternative sets of choice data that can all potentially be used in the analysis. Each subject is associated with three levels of choices: (a) suggested choice indicated in the exchange of opinions, (b) member choice made following the exchange, and (c) team choice randomly selected by the computer from the two member choices. We use *member choices* in our analysis. We do not use team choices, which are solely for incentive compatibility, because they are random subsets of member choices and thus dominated in usable observations. We do not use suggested choices in case subjects are concerned about how their shared opinions will be perceived by their teammates, although results based on suggested choices are very similar.[25]

Table 4 provides summary statistics of the three sets of choices made in Tasks I and II in TEAM. Statistically, focusing on member choices does not impose consequential restrictions; there are no significant differences in any pairwise comparison of the three choices for each of the 12 messages or actions in the two tasks (average two-sided $p = 0.60$, ranging from 0.13 to 1.00). Table 4 also includes summary statistics of the choices made in the SOLO treatment. There is only one set of choices for each subject in SOLO, which parallels the member choices in TEAM. Using members choices in the analysis of TEAM therefore also gives us consistency

[24]We distinguish between aggregate and individual analysis based on whether any analytical procedure is applied prior to aggregation to the matching-cohort level. For example, for an independent observation of the aggregate communication outcome (Blue, X) in Task I (Section 4.2), we count the number of times this outcome occurs in a matching cohort to obtain its relative frequency; for an independent observation about individual sender best responses (Section 4.3), we first determine whether each sender in a round is best responding and then aggregate across all subjects within the matching cohort to obtain the relative frequency of best-responding behavior for that cohort. For non-parametric tests, we report one-sided p-values from Wilcoxon signed rank tests if the objective is to evaluate whether one median is greater than the other and two-sided p-values if the objective is to evaluate whether the two medians are different.

[25]Using suggested choices has some potential advantages as well. Since observing a teammate's opinion may alter a subject's beliefs about the opponents, suggested choices may better reflect initial beliefs, while member choices better reflect payoff-relevant decisions that may also include the effects of the opinion exchanges.

Task I				
	TEAM			SOLO
Sender Message	Suggested Choice	Member Choice	Team Choice	Member Choice
"Blue"	69%	70%	70%	56%
"Green"	7%	4%	5%	17%
"Either"	24%	26%	25%	27%
Receiver Action	Suggested Choice	Member Choice	Team Choice	Member Choice
X	43%	44%	43%	38%
Y	17%	19%	23%	23%
Z	40%	37%	34%	39%
Task II				
	TEAM			SOLO
Sender Message	Suggested Choice	Member Choice	Team Choice	Member Choice
"H"	61%	64%	61%	64%
"T"	39%	36%	39%	36%
Sender Action	Suggested Choice	Member Choice	Team Choice	Member Choice
H	74%	74%	76%	58%
T	26%	26%	24%	42%
Receiver Action	Suggested Choice	Member Choice	Team Choice	Member Choice
H	47%	41%	41%	36%
T	53%	59%	59%	64%

Note: For TEAM, "Suggested Choice" refers to the choices subjects suggested to teammates in opinion exchanges; "Member Choice" refers to the choices made after opinion exchanges; and "Team Choice" refers to the randomly selected member choice that is implemented for the team. For SOLO, "Member Choice" refers to the only set of choices subjects made. The sample sizes are 20 for TEAM and 9 for SOLO.

Table 4: Summary Statistics (Tasks I and II)

with the analysis of SOLO.

Our design to control for the influences of other-regarding preferences also elicits two sets of choice data from each subject, one when it is the opponent payment round and one when it is not. We separate the two sets of data in Section 4.5. For other sections and unless otherwise indicated, hereafter all references to messages and actions in Tasks I and II pertain to member choices, and the aggregations include both payment and non-payment rounds of opponents.[26]

[26] Fully controlling for other-regarding preferences by using only data from opponent non-payment rounds yields similar findings. The aggregate comparison in Section 4.5 and the Chow tests reported in Section 4.6 show that varying the opponent-payment-round type has minimal effect on subject behavior.

4.2 Aggregate Communication Outcomes and Behavior

We analyze the aggregate communication outcomes utilizing Definitions 1 – 3 to determine if overcommunication or, for Task III, truthful behavior is present in the communication tasks.[27] For Tasks I and II, we further analyze the sender and receiver aggregate behavior that contributes to the communication outcomes.

Task I

Table 5 presents the joint frequencies over states and actions in Task I, alongside the joint distribution predicted by equilibrium. The unique pooling equilibrium outcome predicts that (Blue, Z) and (Green, Z) account for 100% of the observed state-action pairs. In the data, however, they occur in only 36% of the observations. On the other hand, the separating state-action pairs, (Blue, X) and (Green, Y), occur in 47% of the observations, which, while only moderately more common than the observed pooling pairs, are dramatically more frequent than the theoretical prediction of 0%. We also observe a small amount of "miscommunication," represented by (Blue, Y) and (Green, X), in 17% of the observations.

	X	Y	Z
Blue	**35%**	7%	15%
Green	10%	**12%**	21%

(a) Observed Outcome

	X	Y	Z
Blue	0%	0%	50%
Green	0%	0%	50%

(b) Predicted Outcome

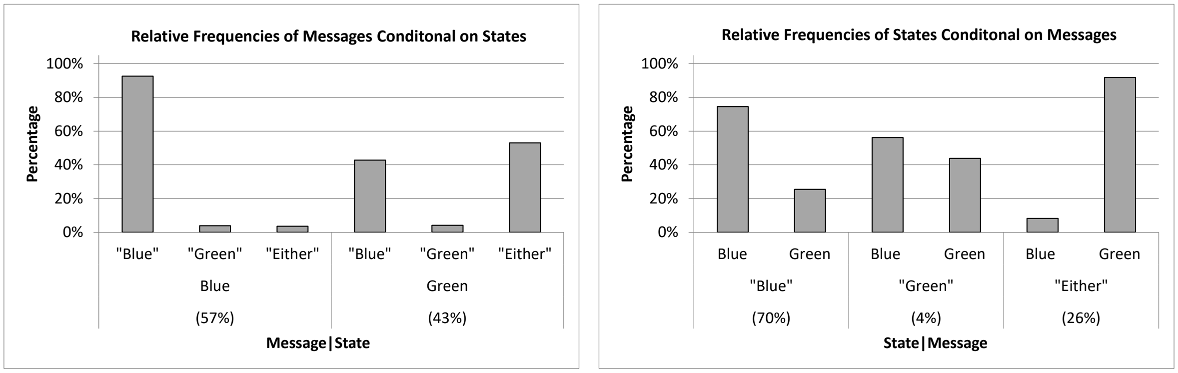
Table 5: Observed and Predicted Outcomes in Task I

The higher than predicted frequencies of the separating pairs (Blue, X) and (Green, Y), both bringing the receiver the highest possible payoff, translate into average receiver payoffs of 57.56, which are significantly greater than the equilibrium benchmark of 50 of the pooling outcome ($p < 0.01$). Such high payoffs indicate that, in contrast to the equilibrium prediction, receivers are sometimes able to take their ideal actions, thereby suggesting the presence of overcommunication. We summarize this in our first finding:

Finding 1. *Average receiver payoffs in Task I are significantly greater than the equilibrium level, indicating the presence of overcommunication.*

We next analyze the sender and receiver aggregate behavior that contributes to the observed overcommunication. For senders, Figure 2(a) presents the relative frequencies of the

[27] All of the analyses starting from this section to Section 4.6 are based on data from the main TEAM treatment. For brevity, we omit references to "TEAM" until the very end of Section 4.6.



(a) Message Behavior

(b) Implied Meanings of Messages

Figure 2: Aggregate Message Behavior and Implied Meanings in Task I

three messages conditional on the two randomly-determined states. As discussed in Section 2.4, because meanings of cheap-talk messages are endogenous, a given communication outcome can be attained by truthful or non-truthful revealing behavior. For messages that are sent truthfully, we can further divide them into completely and incompletely true messages.

In aggregate, senders are completely truthful most of the time in state Blue but are rarely so in state Green. Conditional on state Blue, the relative frequency of the completely true "Blue" is 93%. Conditional on state Green, the two most frequent messages, which together account for 96% of the messages sent, are the incompletely true "Either" (53%) and the lie "Blue" (43%). Grouping both states, senders are completely truthful slightly more often than not at 54%; including the incompletely true "Either," senders are truthful 80% of the time.

As long as receivers understand how senders encode messages, those messages may transmit information even when literally they are lies, which we have referred to as non-truthful revealing behavior. To examine whether and how sender aggregate behavior may be revealing, we derive the empirically implied meaning of each message by applying Bayes' rule to the relative frequencies reported in Figure 2(a). The resulting relative frequencies of the two states conditional on the three messages are presented in Figure 2(b).

The empirically implied meanings show an aggregate tendency to separate, indeed not always via completely true messages. The three messages can be divided into two groups according to their most frequent endogenous meanings. The modal meanings of "Blue" and "Green" are both Blue, with 75% of all cases of "Blue" and "56%" of all cases of "Green" sent in state Blue. The modal meaning of "Either" is Green, with 92% of all cases of the incompletely true message sent in state Green.

While the endogenous meaning of "Blue" is in line with its exogenous literal meaning,

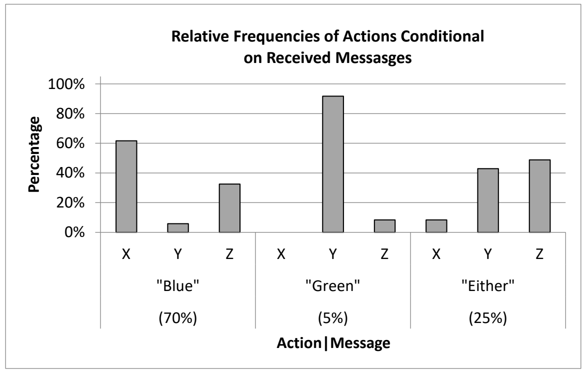


Figure 3: Aggregate Receiver Behavior in Task I

for "Green" and "Either" the endogenous meanings depart from the exogenous, in which endogenously "Green" acts more like "Either" and "Either" acts more like "Green."[28] The aggregate tendency to separate, which is the key component of sender behavior enabling the overcommunication we observe, occurs primarily via messages "Blue" used to indicate Blue and "Either" to indicate Green. We summarize the above with our next finding:

Finding 2. *Sender aggregate behavior in Task I indicates a tendency to separate. The separation occurs primarily via the messages "Blue" and "Either," where the modal endogenous meaning of "Blue" is Blue and that of "Either" is Green.*

To understand how messages are actually interpreted we next turn to aggregate receiver behavior, exploring the other half of the interaction that causes the observed overcommunication. Figure 3 presents the relative frequencies of receiver actions conditional on received messages.[29] The modal behavior of receivers involves taking actions X, Y, and Z in response to messages "Blue," "Green," and "Either" respectively. The relative frequency of X conditional on "Blue" is 62%, that of Y conditional on the rarely received "Green" is 92%, and that of Z conditional on "Either" is 49%. Note that X, Y, and Z are optimal with respect to the literal meanings of "Blue," "Green," and "Either" respectively, and we call them *naive responses* of receivers.[30] Grouping three messages, naive responses occur with a relative frequency of 59%.

[28] We note, however, that "Green" is rarely sent (in only 4% of all messages), and thus its modal meaning may not be a very dependable indicator.

[29] For ease of reference, we include in our figures the relative frequencies of the variables being conditioned on. Note that the frequencies of messages ("Green" and "Either") in Figure 2(b) are slightly different from those in Figure 3. This is because the former, being about messages chosen by senders, is based on member choices as discussed in Section 4.1, and the latter, being about messages actually received by receivers, is based on team choices. The same applies below to Figures 4(b) and 5 for Task II.

[30] Formally in reference to Definition 4, an action in response to m_{Θ_0} is a naive response if the action is optimal given the beliefs that θ is fully supported on Θ_0 with distribution consistently derived from the prior. Note that we use the term "naive responses," which refers to *observed credulous behavior*, to distinguish from credulity as an *identified motive* in the classification of text-response transcripts.

We summarize this with our next finding:

Finding 3. *The modal behavior of receivers in Task I is consistent with naive responses.*

Given the naive responses, the sender modal behavior, in which "Blue" is the most frequent message in state Blue and "Either" in state Green, constitutes best responses to receiver modal behavior. To evaluate if receiver modal behavior constitutes best responses, it is useful to distinguish between (a) best response to a particular message given its modal endogenous meaning, and (b) best responses to sender modal behavior, which require best responses to all three messages.

In the sense of (b), the naive responses of receivers fall short of being best responses to sender modal behavior. Considered separately in terms of (a), however, the modal response to "Blue" does coincide with the best response to "Blue," X, and the best response to "Either," Y, is observed as the second most frequent action. These suggest that receivers indeed pick up some of the information transmitted under sender separation, which contributes to explain the overcommunication reported in Table 5(a) and Finding 1. There is yet another type of receiver best responses—the one that is to their own beliefs. As we discuss below in Section 4.4, when best responses are determined using belief data from the opinion exchanges, best responding to sender separation is indeed the most common type of receiver behavior.

Task II

Table 6 presents the aggregate outcomes in Task II—the relative frequencies of action profiles that could potentially be realized based on member choices—alongside the equilibrium outcome. The unique mixed-strategy equilibrium with or without pre-play communication predicts that the receiver's two favorable action profiles, (H, T) and (T, H), account for 50% of the observations. The sum of the observed frequencies of these two mis-matching action profiles is only slightly higher than the prediction at 53%.

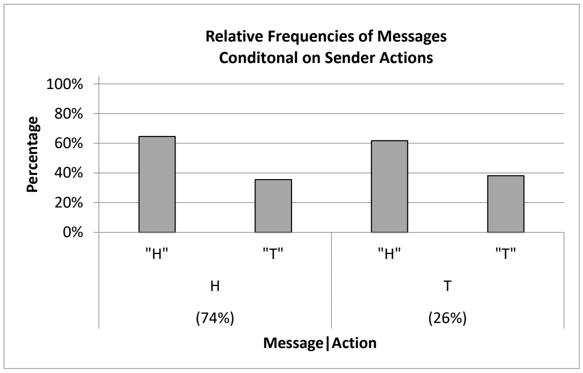
Sender \ Receiver	H	T
H	31%	**43%**
T	**10%**	16%

(a) Observed Outcome

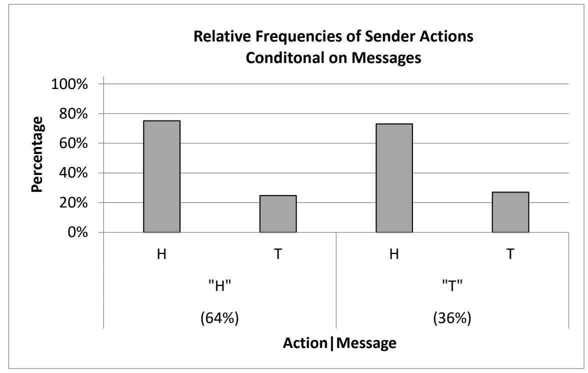
Sender \ Receiver	H	T
H	25%	25%
T	25%	25%

(b) Predicted Outcome

Table 6: Observed and Predicted Outcome in Task II



(a) Message Behavior



(b) Implied Meanings of Messages

Figure 4: Aggregate Message Behavior and Implied Meanings in Task II

The fact that the favorable action profiles of receivers occur slightly more often than their unfavorable action profiles translates into average receiver payoffs of 30.63, which are not significantly different from the equilibrium level of 30 (two-sided $p = 0.30$). This gives us our first finding for Task II:

Finding 4. *Average receiver payoffs in Task II are not significantly different from the equilibrium level, indicating the absence of overcommunication.*

While the average receiver payoffs are consistent with the level predicted by the unique mixed-strategy equilibrium, the frequencies reported in Table 6 indicate that subjects on average are not playing the equilibrium. The action-profile frequencies deviate from the predicted uniform distribution. For example, the favorable action profile of receivers, (H, T), is four times as frequent as their other favorable action profile, (T, H). To better understand the absence of aggregate overcommunication and how the non-equilibrium plays lead to equilibrium payoffs, we examine separately the aggregate behavior of senders and receivers.

The member choice column in Table 4 shows that the unconditional relative frequency with which senders send message "H" is 64%. For the conditional counterparts, Figure 4(a) presents the relative frequencies of sender messages conditional on their actions. The frequencies of "H" conditional on actions H and T are 64% and 62% respectively, which are not significantly different from each other and from the unconditional 64% (two-sided $p \geq 0.42$). This absence of differences suggests that messages are not correlated with actions and thus are not informative about them. In terms of truthful—but not necessarily revealing—behavior, more true messages are sent when the action is H (64%) than when it is T (38%), with an overall frequency of 59% grouping both actions.

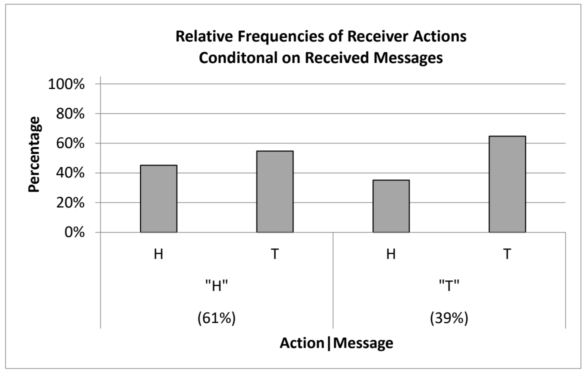


Figure 5: Aggregate Action Behavior in Task II

Equilibrium predicts that the pre-play communication is uninformative and can take the form of babbling. Any observed frequencies of messages, so long as they do not vary across actions as is the case in our data, are thus consistent with equilibrium. To further substantiate the babbling nature of the aggregate pre-play communication, we derive the empirically implied meanings of messages using Bayes' rule as we have done for Task I. Figure 4(b) presents the relative frequencies of sender actions conditional on sender messages. The two messages carry virtually the same endogenous meanings, providing effectively no information in aggregate terms about intended actions. The frequencies of action H conditional on messages "H" and "T" are, respectively, 75% and 73%.

While sender aggregate communication behavior is consistent with the predicted babbling, their actions, which lean toward H, are not consistent with the uniform equilibrium mixed strategy. The relative frequency of sender action H is 74%, significantly higher than the equilibrium prediction of 50% ($p < 0.01$). To analyze how the non-equilibrium plays by senders result in the equilibrium level of payoffs, we first examine the other contributing factor: receiver aggregate behavior.

Similar to the case of sender actions, receiver aggregate actions depart from the uniform equilibrium mixed strategy and exhibit a proclivity toward T. The relative frequency of receiver action T is 59%, significantly higher than the equilibrium prediction of 50% ($p = 0.03$). In terms of communication, however, receiver actions are not significantly influenced by sender messages, which is consistent with the predicted babbling; the proclivity toward T is largely message-independent. Figure 5 presents the relative frequencies of receiver actions conditional on received messages. The frequency of action T conditional on message "T" is 65%, which is higher, but not significantly so, than the 55% conditional on "H" ($p = 0.10$). We summarize the analysis of sender and receiver aggregate behavior in Task II with our next finding:

Finding 5. *Aggregate behavior in Task II indicates that (a) the pre-play communication is consistent with equilibrium babbling, in which sender intended actions are not revealed and receiver actions are not significantly influenced by the communication, and (b) the actions chosen are not consistent with the uniform-mixing uniquely predicted by equilibrium, in that sender actions gravitate toward H and receiver actions toward T.*

The non-equilibrium gravitations toward different actions, 74% H by senders and 59% T by receivers, combine to generate the frequencies of action profiles reported in Table 6, in which receivers' favorable action profiles occur roughly 50% of the time. This results in average payoffs that are virtually at the mixed-strategy equilibrium level as reported in Finding 4. While the details may differ, this is reminiscent of the boundedly rational matching-pennies communication games studied by Crawford (2003), in which he shows that the behavior of some "sophisticated" players can offset that of "mortal" players in a level-k structure to replicate the mixed-strategy equilibrium. As we report in the motive analysis in Sections 4.3 and 4.4 below, there are a sizable number of subjects in our experiment who are like the mortal players in Crawford (2003) and best respond to beliefs that differ from equilibrium beliefs.

Task III

We conclude our aggregate analysis by examining the aggregate reporting behavior in Task III. Table 7 presents the observed frequency of each reported number, together with the predicted distribution of reports and the associated payoffs. In contrast to the prediction that strictly self-interested decision makers always report maximally with a resulting payoff of 60, only 60% of observed reports are sixes. The profile of observed reports translates into average payoffs of 50.06, significantly lower than the predicted level of 60 ($p < 0.01$).

Choice of Report	1	2	3	4	5	6
Payoff	10	20	30	40	50	60
Optimal Report	0%	0%	0%	0%	0%	100%
Observed Report	**6%**	**5%**	**7%**	**8%**	**14%**	**60%**

Table 7: Observed Reports in Task III

We revisit Task III in Section 4.6, examining in more detail the non-strategic reporting behavior and its relation to the strategic communication behavior observed in Tasks I and II. To conclude this subsection, we report the following finding:

Finding 6. *Average payoffs in Task III are significantly less than the predicted level, indicating the presence of truthful behavior.*

4.3 Sender Motives

In this subsection, we utilize inputs from the opinion exchanges to analyze sender beliefs, best responses, and motives, examining the constituents that make up the aggregate sender behavior documented above for Tasks I and II.

Task I

Each sender's answers to the multiple-choice questions in the opinion exchanges allow us to construct the sender's beliefs about how the receiver responds to messages, i.e., the receiver's strategy. Beliefs so constructed are heterogeneous across senders and also within some senders over rounds. In addition to naive responses introduced in Section 4.2 to characterize receiver aggregate behavior, we consider three other distinct types of responses to messages as candidates for sender beliefs. Together, these four types of responses are believed by senders in over 80% of the cases to be the strategy of the receiver they face.

The three additional responses besides naive responses are (a) *separating responses*, taking actions that are best responses to the primary separation observed in sender aggregate behavior (Finding 2), i.e., taking X and Y in response to, respectively, "Blue" and "Either," (b) *pooling responses*, taking Z irrespective of the message received, and (c) *hybrid responses*, taking X, Z, and Z in response to "Blue," "Green," and "Either" respectively.[31] Note that pooling responses are the equilibrium strategy of the receiver, and hybrid responses combine elements of naive and separating responses (with regard to the response to "Blue") and pooling responses (with regard to the responses to "Green" and "Either").

While as a category hybrid responses are distinct from naive responses, given that the two responses differ by the response to the rarely sent "Green," the behavioral significance of the distinction may be limited. We find that these two sets of practically similar beliefs account for the majority of the constructed beliefs, with naive and hybrid responses accounting for 42% and 23% of the cases respectively. The next most frequent beliefs are separating responses (13%), and the least frequent are pooling responses (7%). Since pooling responses are the receiver equilibrium strategy, anticipated pooling responses are the sender beliefs in equilibrium. Thus,

[31] Separating responses are defined based on the best responses to the modal endogenous meanings of messages. As stated in footnote 28, however, the modal meaning of "Green" may not be an dependable indicator given that the message is rarely sent. We therefore do not restrict the response to "Green" in our definition for separating responses.

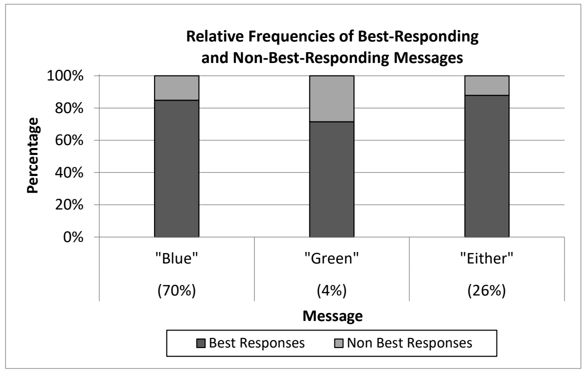


Figure 6: Individual Sender Best Responses in Task I

only a small number of senders hold the equilibrium beliefs.

Using these constructed beliefs, we further analyze individual best responses. We say that a sender best responds if the message chosen maximizes the sender's own monetary reward given that sender's constructed beliefs. Figure 6 presents, for each of the three messages, the conditional relative frequency of cases in which the message is sent as a best response.[32] The analysis reveals that 85% of "Blue," 71% of "Green," and 88% of "Either," or a total of 85% of all messages, are chosen as best responses. Combining this with Finding 2 from the aggregate analysis in Section 4.2, we obtain the picture that the separation observed in sender aggregate behavior is in most cases made up of individual best responses to the beliefs that the receiver would respond to "Blue" with X and to "Either" with Z.

Other than the separation, our aggregate analysis also indicates that senders are truthful as frequently as 80% of the time. It will be constructive to examine the interaction between truthful behavior and individual best responses. Table 8(a) reports the relative frequencies of six categories of sender behavior in a 3 × 2 matrix, where the rows correspond to completely truthful, incompletely truthful, and lying behavior, and the columns correspond to best-responding and non-best-responding behavior. Nearly all true messages are sent as best responses: the relative frequency of best-responding messages conditional on being true is 93%, with 94% for completely true and 88% for incompletely true messages. Reciprocally, there are comparatively fewer best-responding messages that are true: the relative frequency of true messages conditional on being best responses is 87%, and the frequency drops to 60% if one restricts to completely true messages.

Note that the sender's and the receiver's ideal actions align in state Blue. Table 8(b) presents the same categories of sender behavior but focuses on state Green only, where interests

[32] If the sender's constructed beliefs admit multiple best responses, any choice that coincides with a best response is counted as a qualifying case.

	Best-Responding (85%)	Non-Best-Responding (15%)
Completely Truthful (54%)	51%	3%
Incompletely Truthful (26%)	23%	3%
Lying (20%)	11%	9%

(a) States Blue and Green

	Best-Responding (70%)	Non-Best-Responding (30%)
Completely Truthful (4%)	1%	3%
Incompletely Truthful (53%)	50%	3%
Lying (43%)	19%	24%

(b) State Green Only

Table 8: Sender Best-Responding and Truthful Behavior in Task I

are not aligned. A similar pattern in the reciprocal conditional relative frequencies is observed: 89% of true messages are best responses, while 74% of best-responding messages are true. Overall, there is a positive correlation between best-responding and truthful behavior observed in the full as well as the restricted state-Green samples.

Truth-Telling Preferences	Payoff Concern	Image Concern	Confusion	None
6%	95%	0%	6%	4%

Note: The four motive attributes are not mutually exclusive. "None" refers to the cases in which none of the four attributes are identified.

Table 9: Sender Motive Attributes in Task I

Further evidence of the primacy of payoffs in motivating sender behavior comes from the classifications of senders' written responses to the survey questions. Table 9 presents the relative frequencies of the classified motive attributes. Payoff concern is by far the most common attribute, with 95% of the written transcripts indicating payoff-related motives. Only 6% of transcripts indicate any desire to be truthful for its own sake.[33] We summarize our

[33]Here is a sample of representative transcripts for each type of motives, edited slightly for grammar: (a) truth-telling preferences: "Be honest"; (b) payoff concern: "Give them an unclear message to choose Z which

analysis of sender best responses and motives in Task I:

Finding 7. *In the large majority of cases, senders in Task I best respond to their beliefs about receiver strategy. Analysis of written transcripts reveals that concern for payoffs is by far the most common sender motive.*

Task II

Although our aggregate analysis indicates that there is no overcommunication in Task II and the pre-play communication is babbling, an analysis of sender beliefs and best responses helps understand what contribute to this aggregate communication outcome.

The most common sender beliefs are that the receiver is *naive*, best responding to messages as if they were true. These beliefs, which parallel the anticipated naive responses in Task I, account for 64% of the cases. The next most common sender beliefs are that the receiver is *skeptical*, best responding to the messages as if they were lies, with 30% of the constructed beliefs showing this property. In the remaining 6% of the cases, the senders believe that the receiver chooses a fixed action irrespective of the message. Since the receiver ignores the sender's message in equilibrium, similar to the finding from Task I, this suggests that only a small number of senders hold the equilibrium beliefs in Task II.

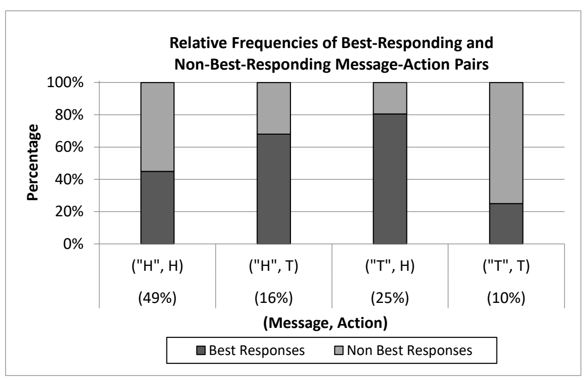


Figure 7: Individual Sender Best Responses in Task II

Based on these constructed beliefs, we further analyze sender best responses, which for Task II are defined jointly in terms of messages and actions: a choice of message-action pair is a best response if it maximizes the sender's own monetary reward given the sender's beliefs about the strategy of the receiver. Figure 7 presents the conditional relative frequencies with which each of the four message-action pairs constitutes a best response. We find that 80% of ("T", H), 68% of ("H", T), 45% of ("H", H), and 25% of ("T", T) are chosen as best responses.

maximizes our earnings"; (c) confusion: "If they believe us its blue, they will take either X or Y, which is good for us"; and (d) none: "Tell them it is blue or green."

	Best-Responding (56%)	Non-Best-Responding (44%)
Truthful (59%)	25%	34%
Lying (41%)	31%	10%

Table 10: Sender Best-Responding and Truthful Behavior in Task II

We further examine the interaction between truthful behavior and individual best responses as we have done for Task I. Since there is no incompletely true message in Task II, there are only four categories of sender behavior, the relative frequencies of which are reported in Table 10. There are slightly more best-responding lies than best-responding true messages: the relative frequencies of the two types of message-action pairs conditional on being best responses are 55% and 45% respectively.[34] By contrast, when senders are not best responding, their behavior is predominately truthful: conditional on not being best responses, the relative frequency of true messages is 77%, more than three times that of lies.

Table 10 reveals that payoff considerations given beliefs drive more than half of sender behavior whether it is truthful or not, but when not being driven by payoffs senders are mostly truthful. With 34% cases of non-best-responding true messages, the analysis suggests, however, that in about two-third of the observations senders are driven by motives other than truthful preferences. We seek further evidence on sender motives from the classifications of the opinion-exchange transcripts.

Truth-Telling Preferences	Payoff Concern	Image Concern	Confusion	None
8%	83%	0%	0%	14%

Note: The four motive attributes are not mutually exclusive. "None" refers to the cases in which none of the four attributes are identified.

Table 11: Sender Motive Attributes in Task II

Table 11 reports the relative frequencies of the classified sender motive attributes in Task II. The findings are broadly similar to those in Task I, except that payoff concern is modestly less frequent at 83% and truth-telling preferences are slightly more common at 8%. There are

[34] Recall that the sender prefers to match the pennies, while the receiver prefers to mismatch. For the true ("H", H) and ("T", T) to be best responses, the underlying sender beliefs must be that the receiver is skeptical; for the lying ("T", H) and ("H", T) to be best responses, the sender beliefs must be that the receiver is naive. Senders thus best respond slightly more often to the beliefs that the receiver is naive than to the beliefs that the receiver is skeptical.

also more cases of "None." Payoff concern remains far and away the most common attribute.[35] We summarize our analysis of sender best responses and motives in Task II:

Finding 8. *In the slight majority of cases, senders in Task II best respond to their beliefs about receiver strategy. When not best responding, senders are predominantly truthful. Analysis of written transcripts reveals that payoff consideration is by far the most common sender motive.*

4.4 Receiver Motives: A Summary

While we have provided a detailed analysis of sender motives, for brevity and for the reason that receiver behavior is arguably less primary than sender behavior in driving communication outcomes, we briefly summarize in this subsection our analysis of receiver motives, leaving the details to Appendix B.

Task I

Of the actions available for receivers, X and Y are the separating actions, while Z is the pooling action. The opinion exchanges provide us with inputs to construct receiver beliefs, which in turn help us determine the extent to which receivers choose the separating and the pooling actions as best responses. The written transcripts further allow us to identify the relevant motives. We summarize our finding in this regard:

Finding 9. *Separating best responses are the most common type of receiver behavior in Task I, while pooling best responses are the least common; best responses to messages as if they were true, taking X, Y, and Z in response to "Blue," "Green," and "Either" respectively without regard to beliefs about the sender's strategy, account for half of the non-best responses. Analysis of written transcripts reveals that payoff consideration is by far the most common receiver motive.*

Finding 9 indicates that equilibrium behavior exists among receivers but is outnumbered by non-equilibrium best responses based on beliefs that the sender separates. Given the separation observed as modal behavior among senders (Finding 2), this suggests that the most common type of receiver behavior at the individual level consists of best responses to sender aggregate behavior.

[35] Here is a sample of representative transcripts for each type of motives, edited slightly for grammar: (a) truth-telling preferences: "My mum told me that I have to be honest"; (b) payoff concern: "By telling them we choose T, they will also choose T as they think they will earn 40"; and (c) none: "It depends on others' personality. Both is possible."

Task II

The opinion exchanges again provide us with inputs to evaluate receiver best-responding behavior and motives in Task II. We summarize our finding in this regard:

Finding 10. *In the majority of cases, receivers in Task II best respond to their beliefs about sender actions, independent of whether they believe the sender adopts a separating or a pooling strategy. Analysis of written transcripts reveals that payoff consideration is by far the most common receiver motive.*

4.5 Opponent Payment Rounds

To explore the potential effects of other-regarding preferences, we divide our sample by the type of opponent payment rounds. We compare between the subsamples the key variables

	Opponent Payment Round	Opponent Non-Payment Round	Wilcoxon Signed Rank Test
Task I			
Senders			
True Messages	81%	79%	$p = 0.83$
Completely True	60%	49%	$p = 0.39$
Incompletely True	21%	30%	$p = 0.43$
Best Responses	90%	80%	$p = 0.12$
Receivers			
Payoffs	55.00	60.13	$p = 0.16$
Best Responses	59%	61%	$p = 0.53$
Task II			
Senders			
True Messages	65%	53%	$p = 0.19$
Best Responses	53%	59%	$p = 0.46$
Receivers			
Payoffs	30.00	31.25	$p = 0.52$
Best Responses	71%	65%	$p = 0.46$

Note: Opponent payment round is relative to the role listed for each set of rows, i.e., the "Senders" values are based on receiver payment and non-payment rounds, while the "Receivers" values are based on sender payment and non-payment rounds. Reported p-values are from two-sided tests. All differences between the two payment-round types remain insignificant at the 5% level using one-sided tests.

Table 12: Comparisons between Opponent Payment and Non-Payment Rounds

analyzed in Tasks I and II. Table 12 presents, for each subsample, the relative frequencies of sender true messages, sender and receiver best responses, and average receiver payoffs.

None of the differences between the two payment-round types with regard to these variables are statistically significant. The patterns of differences are also not all systematic across the tasks. For sender best responses, e.g., in Task I there are more cases in opponent payment rounds than in non-payment rounds, but in Task II there are more cases in non-payment rounds than in payment rounds. We summarize the effects of other-regarding preferences:

Finding 11. *Comparisons of aggregate behavior between opponent payment and non-payment rounds indicate that other-regarding preferences have minimal effects on the key strategic communication behavior in Tasks I and II.*

One of the findings from our regression analysis in the next subsection shows that the same conclusion is reached based on subject-level data.[36]

4.6 Non-Strategic Reporting Behavior in Task III

As reported in the aggregate analysis of Task III in Section 4.2, the predicted report of six occurs in 60% of all observed reports, with the remaining 40% distributed over one to five. In this subsection, we first provide a more detailed account of the non-strategic reporting behavior in Task III. We then restrict our attention to subjects who play the role of senders in Tasks I and II and examine the relationship between their strategic communication and non-strategic reporting behavior.

While there is only one kind of true report in Task III, we follow Gneezy et al. (2018) and classify a lie as either a *maximal lie* for a report of six or a *partial lie* otherwise. Among all observed reports, 28% are true, 14% are partial lies, and 58% are maximal lies.[37] Figure 8 presents the relative frequencies of the three types of reports conditional on each realized lower number. Maximal lies, when defined, are the most frequent types of reports for each realized number, with relative frequencies ranging from 56% to 63%.[38] True reports come in second,

[36]It is constructive to discuss our findings on other-regarding preferences in light of Gneezy (2005), who find among other things that the payoff consequences to the parties being lied to matter. Our findings should not be interpreted as necessarily incompatible with his findings. It is conceivable that by magnifying the payoffs in our environments while maintaining the ordinal preferences so that the equilibrium predictions remain intact, we would reach a point where other-regarding preferences play a more significant role. Given our current experimental parameters, however, we do not find a first-order effect of other-regarding preferences, and this allows us to focus on truthful preferences and induced best-responding motives, which are arguably more portable across payoff environments, as the two major motives of communication we attempt to disentangle.

[37]Not all 60% of reported sixes are maximal lies. There is a small fraction (2%) of reported sixes that are true, which occur when both dice roll six.

[38]Maximal lies are not defined when the realized lower number is six. Note also that partial lies are not

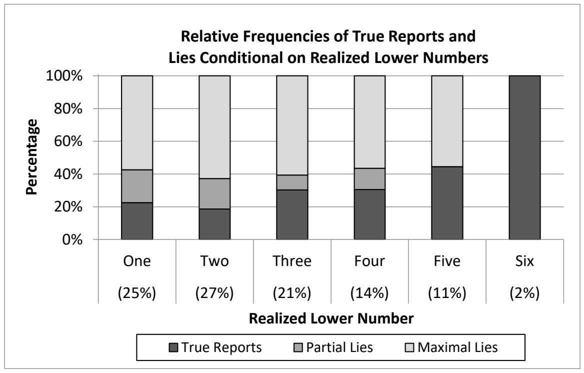


Figure 8: Reporting Behavior in Task III

which indicates that some intensity of truthful preferences are at work. Excluding the cases of a realized six in which there are 100% of truthfully reported sixes, the relative frequencies of true reports range from 19% to 44%. Relative to telling maximal lies, the opportunity cost of being truthful in monetary terms is decreasing in the realized number. This is reflected in our observations: with the exception of a transient drop at two, the relative frequencies of true reports are monotone in the realized number.

Our main interest in Task III is to examine how truthful behavior in non-strategic communication relates to that in strategic communication. As a starting point, Figure 9 presents the relative frequencies of truthful behavior in Tasks I and II conditional on truthful and lying behavior in Task III. With only one kind of true messages, the comparisons for Task II are straightforward. Among the cases where senders are truthful in Task III, 65% are also truthful in Task II. The frequency falls to 58% among cases where senders are not truthful in Task III. The difference is, however, not statistically significant ($p = 0.44$).

Comparisons of Task I with Task III are more nuanced due to the two classes of true messages in Task I. The relative frequency of completely truthful behavior in Task I is 60% conditional on being truthful in Task III, which is higher, but not significantly so, than the 53% conditional on lying in Task III ($p = 0.25$). For incompletely truthful behavior, the relative frequency is 29% conditional on being truthful in Task III, which is significantly higher at the 5% level than the 22% conditional on lying in Task III ($p = 0.05$). Counting both classes of true messages as truthful behavior in Task I, the two sets of frequencies stack up to be significantly different at the 1% level (89% vs. 75%, $p < 0.01$).

The comparisons above show a positive relationship between truthful behavior across Tasks I and III, although the strength of that relationship is sensitive to how one defines truthful-

defined when the realized lower number is five because any lie in this case must involve a report of six, which is defined as a maximal lie.

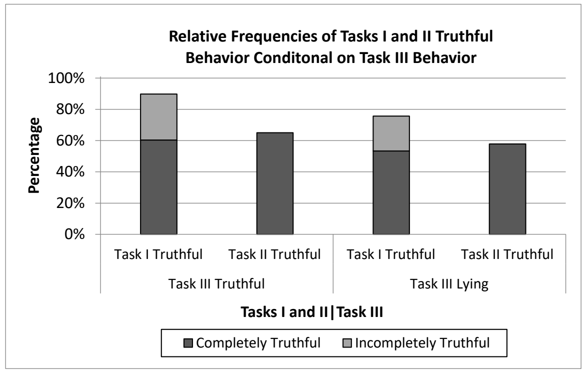


Figure 9: Truthful Behavior in Tasks I and II
Conditional on Behavior in Task III

ness.[39] In light of the discussion in Section 2.4 about the two subtypes of truthful preferences, the significant relationship for "Either" but not for the completely true "Blue" or "Green" is more in line with truthful behavior being driven by aversion to overt lying rather than a desire to be fully accurate and descriptive. These results, based on data aggregated at the matching-cohort level, suggest that there is some portability of lying aversion, but not truth affinity, from non-strategic to strategic communication.

While our analysis thus far is based on aggregated data to ensure complete independence of observations, some information is necessarily lost in the process of aggregation. We explore more thoroughly the relationship between strategic and non-strategic truthful behavior with regression analysis using subject-level data.

We estimate two sets of regressions, separately examining the correlations between Tasks III and I and between Tasks III and II. Treating Task III behavior as the dependent variable, we use two different measures of Task III truthfulness: a dummy variable, $True_i^{\mathrm{III}}$, which takes the value of one if sender i is truthful in Task III and zero otherwise, and a distance variable, $DLie_i^{\mathrm{III}} = |m_i^{\mathrm{III}} - \theta_i^{\mathrm{III}}|$, which measures the magnitude of sender i's lying in Task III. The independent variables, $True_i^{\mathrm{I}}$ and $True_i^{\mathrm{II}}$, measure sender i's truthfulness in Tasks I and II respectively.[40] For Task I, $True_i^{\mathrm{I}}$ counts both completely and incompletely true messages. We also include an alternative Task I specification in which we replace $True_i^{\mathrm{I}}$ with two mutually exclusive independent variables, $CTrue_i^{\mathrm{I}}$, which captures completely truthful behavior, and $ITrue_i^{\mathrm{I}}$, which captures incompletely truthful behavior.

[39] The findings are similar if we restrict the sample to state Green in Task I where interests are not aligned. We still see significantly more combined truthful cases conditional on being truthful in Task III, but the difference becomes insignificant in the breakdown for each class of true messages.

[40] Since there are two rounds in each of the strategic tasks, we measure "average truthfulness" across the two rounds, where $True_i^{\mathrm{I}}$ and $True_i^{\mathrm{II}}$ each take the value of one, one half, and zero if sender i is truthful in both rounds, only one round, and no round respectively in the corresponding task.

Table 13 reports our first set of regression results comparing behavior between Tasks III and I. Columns (1) and (3) in panel A report the respective estimates from regressing our binary ($True_i^{\text{III}}$) and lying-distance ($DLie_i^{\text{III}}$) measures of Task III truthfulness on the broad measure of Task I truthfulness ($True_i^{\text{I}}$). Columns (2) and (4) report instead the estimates from the alternative specifications breaking down Task I truthfulness into complete truthfulness ($CTrue_i^{\text{I}}$) and incomplete truthfulness ($ITrue_i^{\text{I}}$). Across all four of our specifications, we find no evidence that being truthful in Task I is associated with being truthful in Task III.

In the preceding analysis of aggregate behavior, we separately examine sender behavior in state Green, where ideal actions of the sender and receiver are misaligned (Section 4.3), and explore the differences between opponent payment and non-payment rounds to gauge the potential effects of other-regarding preferences (Section 4.5). We conduct parallel subject-level analyses by re-estimating the regressions in panel A of Table 13 using four different restricted samples based on state and type of opponent payment round. The results are reported in panel B. With one exception, it remains the case that there is no significant relationship between truthful behavior in Task III and that in Task I in each of the four restricted samples.

The one exception occurs when we restrict the sample to opponent non-payment rounds and regress $True_i^{\text{III}}$ on $True_i^{\text{I}}$. The relationship is found to be significant at the 5% level. Opponent non-payment rounds provide the most generous test of the presence of truthful preferences in strategic communication, since senders' choice of messages do not influence the experimental rewards of receivers. Stripping away any concern for the payoff effects their messages have on receivers, strategic communication here is closer to a strictly self-interested decision problem. The truthful motives that operate in the non-strategic Task III may therefore be most portable to the strategic Task I in the opponent non-payment rounds, and this may explain why we observe a significant relationship in this particular case but not in opponent payment rounds.

Despite the one significant correlation observed in opponent non-payment rounds, comparing the estimates in columns (1)–(4) with those in the corresponding columns (5)–(8) under the two payment-round types, Chow tests indicate that there are no significant differences between the two sets of estimates ($p \geq 0.54$). The test results corroborate the aggregate finding in Section 4.5 that other-regarding preferences do not play a significant role in Task I.

A. Both States and Types of Opponent Payment Round

	(1)	(2)	(3)	(4)
	$True_i^{\text{III}}$		$DLie_i^{\text{III}}$	
$True_i^{\text{I}}$	0.25	–	−0.51	–
	(0.14)	–	(0.56)	–
$CTrue_i^{\text{I}}$	–	0.24	–	−0.80
	–	(0.13)	–	(0.58)
$ITrue_i^{\text{I}}$	–	0.27	–	−0.01
	–	(0.18)	–	(0.64)
Constant	0.13	0.13	2.45***	2.48***
	(0.13)	(0.13)	(0.55)	(0.54)
Observations	78	78	78	78

B. By State and Type of Opponent Payment Round

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$True_i^{\text{III}}$		$DLie_i^{\text{III}}$		$True_i^{\text{III}}$		$DLie_i^{\text{III}}$	
	State Green				State Blue			
$True_i^{\text{I}}$	0.21	–	−0.11	–	0.14	–	−0.50	–
	(0.15)	–	(0.54)	–	(0.14)	–	(0.42)	–
$CTrue_i^{\text{I}}$	–	0.43	–	−1.12	–	0.13	–	−0.49
	–	(0.39)	–	(1.26)	–	(0.14)	–	(0.42)
$ITrue_i^{\text{I}}$	–	0.19	–	−0.01	–	−0.06	–	−0.26
	–	(0.15)	–	(0.55)	–	(0.20)	–	(0.81)
Constant	0.20	0.20	2.22***	2.21***	0.19	0.21	2.45***	2.43***
	(0.12)	(0.12)	(0.36)	(0.35)	(0.12)	(0.12)	(0.50)	(0.50)
Observations	53	53	53	53	65	65	65	65
	Opponent Payment Round				Opponent Non-Payment Round			
$True_i^{\text{I}}$	0.06	–	−0.13	–	0.20*	–	−0.40	–
	(0.13)	–	(0.39)	–	(0.09)	–	(0.43)	–
$CTrue_i^{\text{I}}$	–	0.05	–	−0.25	–	0.19	–	−0.54
	–	(0.13)	–	(0.44)	–	(0.10)	–	(0.47)
$ITrue_i^{\text{I}}$	–	0.07	–	0.21	–	0.21	–	−0.18
	–	(0.15)	–	(0.46)	–	(0.12)	–	(0.53)
Constant	0.29*	0.29*	2.14***	2.14***	0.18*	0.18*	2.35***	2.35***
	(0.13)	(0.13)	(0.41)	(0.41)	(0.08)	(0.08)	(0.46)	(0.46)
Observations	78	78	78	78	78	78	78	78

Note: Panel A reports OLS estimates including both states and using sender i's Task I average truthfulness across the two rounds. Panel B reports OLS estimates using sender i's Task I average truthfulness calculated over the round or rounds that satisfy the condition of the restricted sample. For state Green, e.g., if sender i encounters Green in both rounds, the average over both rounds is used; if sender i encounters Green in only one round, data from only that round is used. Two senders who observed two rolls of six in Task III were dropped as they had no incentive to lie. Standard errors clustered at the matching-cohort level are in parentheses. *** indicates significance level at 0.1%, ** at 1%, and * at 5%.

Table 13: Correlations Between Truthful Behavior in Task III and Task I

Table 14 reports the regression results for Tasks III and II. Similar to Table 13, estimates using all data are in panel A, while those based on restricted samples from opponent payment and non-payment rounds are in panel B. In all cases, none of the correlations are significantly different from zero; being truthful in Task II is not associated with being truthful in Task III. In line with our finding from Section 4.5 that other-regarding preferences do not influence behavior in Task II, Chow tests reveal that there are no significant differences between the two sets of estimates in Panel B ($p \geq 0.84$).

A. Both Types of Opponent Payment Round

	(1)	(2)
	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$
$True_i^{\text{I}}$	0.11	0.02
	(0.17)	(0.44)
Constant	0.27*	2.03***
	(0.12)	(0.28)
Observations	78	78

B. By Type of Opponent Payment Round

	(1)	(2)	(3)	(4)
	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$
	Opponent Payment Round		Opponent Non-Payment Round	
$True_i^{\text{I}}$	0.06	−0.05	0.07	0.07
	(0.15)	(0.47)	(0.10)	(0.29)
Constant	0.30*	2.07***	0.30**	2.00***
	(0.11)	(0.34)	(0.08)	(0.21)
Observations	78	78	78	78

Note: Panel A reports OLS estimates using sender i's Task II average truthfulness across the two rounds. Panel B reports OLS estimates using sender i's Task I average truthfulness calculated over the round or rounds that satisfy the condition of the restricted sample. Two senders who observed two rolls of six in Task III were dropped as they had no incentive to lie. Standard errors clustered at the matching-cohort level are in parentheses. *** indicates significance level at 0.1%, ** at 1%, and * at 5%.

Table 14: Correlations Between Truthful Behavior in Task III and Task II

We summarize our matching-cohort-level and subject-level findings about the relationship between truthful behavior in the strategic and non-strategic tasks:

Finding 12. *At the aggregate level, there is a positive relationship between truthful behavior in Task III and truthful behavior in Task I. At the individual level, with one exception observed in opponent non-payment rounds, truthful behavior in Task III is not correlated with truthful behavior in Task I. Truthful behavior in Tasks III and II is not correlated at the aggregate or individual level.*

Finding 12 suggests that truthful behavior in Task III has limited ability to predict behavior in the two strategic tasks. While we do observe some relationship between the truthful behavior in Tasks III and I, our overall findings suggest caution in attempting to use truthful preferences documented in decision-theoretic environments as an explanation for strategic overcommunication. Another insight from our analysis of Tasks III and I concerns the nature of truthful preferences and the implications for interpreting sender behavior. Our findings indicate that truth affinity and lying aversion as distinct subtypes of truthful preferences may operate differently with different portability between non-strategic and strategic communication. In both our aggregate and individual analysis, completely true messages in Task I are never significantly correlated with truthful reporting in Task III, suggesting that the limited overall relationship we observe between strategic and non-strategic communication is derived from an aversion to lying, rather than an affinity for the truth.

4.7 A Summary of Findings from the Robustness SOLO Treatment

We conclude our experimental findings by summarizing the key observations from the SOLO treatment. SOLO serves as a control to gauge if our findings from TEAM might have been confounded by the concurrent variations from strategic to non-strategic tasks and from team to individual decisions. The detailed analysis of the data from SOLO can be found in Appendix C, in which we repeat the analysis in Sections 4.2, 4.5, and 4.6. Below is a summary of Findings C.1 – C.8 reported in the appendix:

Finding 13. *The findings from SOLO can be summarized as:*

(a) *Task I: The principle difference from TEAM concerns sender truthful behavior, in which there are more completely truthful instances of message "Green." The message nevertheless remains least frequently used, and thus the overall impacts on communication outcomes are limited with a moderate increase in average receiver payoffs.*

(b) *Task II: The principle difference from TEAM is that the pre-play communication is somewhat informative, and average receiver payoffs are higher and pass the statistically significant threshold to indicate that there is overcommunication.*

(c) Task III: There is no substantive difference from the reporting behavior in TEAM.

(d) Cross-task comparisons: The non-strategic truthful behavior in Task III is not positively correlated with strategic truthful behavior in Task I or II.

Finding 13(d) is the key finding from SOLO, as it suggests that our results from TEAM are not a by-product of switching between the team and individual settings.[41] The robustness across the team and individual designs provides further support to the idea that one should be cautious in extrapolating truthful preferences documented in decision-theoretic communication to account for strategic truthful behavior.

5 Conclusion and Discussion

The well-documented phenomenon of overcommunication in experimental games of strategic communication has been rationalized via two distinct approaches: as a manifestation of homegrown preferences, in particular truthful preferences, or as a consequence of heterogeneity in strategic thinking operationalized as level-k reasoning. Utilizing an experimental design inspired by prior studies, we have provided to our knowledge the first investigation of the relative explanatory power of these competing explanations for overcommunication. We leverage within-subject variation of strategic and non-strategic communication tasks, asymmetric knowledge of payment rounds, and, most importantly, elicitation of subject reasoning through team choices to dissect the causes of overcommunication.

We find evidence of strategic overcommunication. Non-strategic reporting behavior indicates the presence of truthful preferences. Elicited beliefs and motives show that subjects in strategic communication have heterogeneous beliefs about their opponents, holding equilibrium beliefs only in a minority of cases. The majority of subjects best respond to their beliefs, which suggests that some truthful behavior observed in the strategic environments may in fact be best responses to non-equilibrium beliefs. Alongside this finding, we further observe that truthful behavior in the non-strategic communication task has limited ability to predict behavior in the strategic tasks. Taken together, our results point to strategic thinking—optimizing

[41] While our main interests in comparing TEAM and SOLO are not the effects of teams on communication *per se* but the effects on the *relationship between* strategic and non-strategic communication, it is worth noting that Findings 13(a)–(b) echo the team effects documented in some of the prior studies. To the extent that more informative communication corresponds to being less strategic in our environments, our findings echo those in Cooper and Kagel (2005), who find that teams behave more strategically in signaling games. In a more recent study, Conrads et al. (2013) find that lying is more pronounced in teams, which is in line with the more truthful behavior observed in SOLO than in TEAM, although the differences in our case are not substantial.

given beliefs that need not be consistent with equilibrium—as the primary explanation for overcommunication in strategic environments.

Our findings suggest caution in attempting to use truthful preferences observed in a decision-theoretic environment to rationalize strategic communication behavior. The discrepancy we find between strategic and non-strategic truthful behavior is also sensitive to how one defines truthful behavior, specifically whether truthfulness is simply the absence of overt lying, or adherence to fully accurate description. These two types of truthful preferences may operate differently across strategic and non-strategic environments.

We propose several directions for future research. Sobel (2020) defines deception as a "deliberate attempt by the sender to induce incorrect beliefs." Deception so defined can be truthful in a strategic setting. "Truthful deception" is a useful theoretical construct to elucidate the important difference between exogenous and endogenous meanings in strategic communication. Yet to our knowledge no study has been devoted to investigate whether and under what circumstances truthful deception may systematically arise as an observed strategic phenomenon. For instance, does the manner of deception depend upon the complexity of the communication environment, the need for plausible deniability, or the natures of incentive conflicts? In situations where lies are difficult to understand, telling the truth may be a more effective way to deceive. Being truthful may also serve the purpose of plausible deniability, providing a cover to hide an ulterior deception motive. In our strategic tasks, whenever conflicts arise they involve unfair payoff allocations between the sender and the receiver. It is worthwhile to explore how deception attempts may manifest as truth telling or lying in other payoff settings, e.g., the trust-game environment studied in Charness and Dufwenberg (2006).

There are also important unanswered questions with regard to the distinction between lying aversion and truth affinity. Controlling for the consequences of deception by, for example, inducing an expectation that the receiver would ignore the sender's message, to what extent are senders motivated by an aversion to lying versus an affinity for the truth? Does the mere context of a strategic environment, which may suggest that the goal is to "outsmart" others, play a role in the relative importance of these two subtypes of truthful preferences? With overcommunication rather than truthful behavior *per se* being the primary subject matter of our study, we touch on some of these issues but by no means provide direct answers to them. In our view, designing experiments to answer these and other related questions to unravel the motives of truth telling, lying, and deception in different communication environments present important and challenging opportunities for future research.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019), "Preferences for truth-telling." *Econometrica*, 87, 1115–1153.

Blume, Andreas, Douglas V. Dejong, Yong-Gwan Kim, and Geoffrey B. Sprinkle (2001), "Evolution of communication with partial common interest." *Games and Economic Behavior*, 37, 79–120.

Blume, Andreas, Ernest K. Lai, and Wooyoung Lim (2020), "Strategic information transmission: A survey of experiments and theoretical foundations." In *Handbook of Experimental Game Theory* (C. Monica Capra, Rachel Croson, Mary Rigdon, and Tanya Rosenblat, eds.), Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 90, 166–193.

Burchardi, Konrad B. and Stefan P. Penczynski (2014), "Out of your mind: Eliciting individual reasoning in one shot games." *Games and Economic Behavior*, 84, 39–57.

Cai, Hongbin and Joseph Tao-Yi Wang (2006), "Overcommunication in strategic information transmission games." *Games and Economic Behavior*, 56, 7–36.

Charness, Gary and Martin Dufwenberg (2006), "Promises and partnership." *Econometrica*, 74, 1579–1601.

Chen, Ying (2011), "Perturbed communication games with honest senders and naive receivers." *Journal of Economic Theory*, 146, 401–24.

Chen, Ying, Navin Kartik, and Joel Sobel (2008), "Selecting cheap-talk equilibria." *Econometrica*, 76, 117–136.

Conrads, Julian, Bernd. Irlenbusch, Rainer M. Rilke, and Gari Walkowitz (2013), "Lying and team incentives." *Journal of Economic Psychology*, 34, 1–7.

Cooper, David J. and John H. Kagel (2005), "Are two heads better than one? Team versus individual play in signaling games." *American Economic Review*, 95, 477–509.

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

Crawford, Vincent P., Miguel A Costa-Gomes, and Nagore Iriberri (2013), "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications." *Journal of Economic Literature*, 51, 5–62.

Crawford, Vincent P and Joel Sobel (1982), "Strategic information transmission." *Econometrica*, 50, 1431–1451.

Dufwenberg, Martin and Martin A. Dufwenberg (2018), "Lies in disguise—A theoretical analysis of cheating." *Journal of Economic Theory*, 175, 248–264.

Fehr, Ernst and Klaus M. Schmidt (1999), "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, Urs (2007), "z-tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10, 171–178.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013), "Lies in disguise: An experimental study on cheating." *Journal of the European Economic Association*, 11, 525–547.

Forsythe, Rober, Russell Lundholm, and Thomas Rietz (1999), "Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence." *Review of Financial Studies*, 12, 481–518.

Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner (2013), "Preferences for truthfulness: Heterogeneity among and within individuals." *American Economic Review*, 103, 532–48.

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

Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018), "Lying aversion and the size of the lie." *American Economic Review*, 108, 419–53.

Gneezy, Uri, Bettina Rockenbachb, and Marta Serra-Garciac (2013), "Measuring lying aversion." *Journal of Economic Behavior & Organization*, 93, 293–300.

Hurkens, Sjaak and Navin Kartik (2009), "Would I lie to you? On social preferences and lying aversion." *Experimental Economics*, 12, 180–192.

Kajackaite, Agne and Uri Gneezy (2017), "Incentives and cheating." *Games and Economic Behavior*, 102, 433–444.

Kartik, Navin (2009), "Strategic communication with lying costs." *Review of Economic Studies*, 76, 1359–1395.

Kartik, Navin, Marco Ottaviani, and Francesco Squintani (2007), "Credulity, lies, and costly talk." *Journal of Economic theory*, 134, 93–116.

Kawagoe, Toshiji and Hirokazu Takizawa (2009), "Equilibrium refinement vs. level-k analysis: An experimental study of cheap-talk games with private information." *Games and Economic Behavior*, 66, 238–255.

Khalmetski, Kiryl, Bettina Rockenbach, and Peter Werner (2017), "Evasive lying in strategic communication." *Journal of Public Economics*, 156, 59–72.

Khalmetski, Kiryl and Dirk Sliwka (2019), "Disguising lies–Image concerns and partial lying in cheating games." *American Economic Journal: Microeconomics*, 11, 79–110.

Lafky, Jonathan (2014), "Why do people rate? Theory and evidence on online ratings." *Games and Economic Behavior*, 87, 554–570.

Lai, Ernest K. and Wooyoung Lim (2012), "Authority and communication in the laboratory." *Games and Economic Behavior*, 74, 541–560.

Lim, Wooyoung and Siyang Xiong (2021), "Does jump-bidding increase sellers' revenue? Theory and experiment." *Journal of Economic Behavior & Organization*, 189, 84–110.

Roth, Alvin E. (1987), "Laboratory experimentation in economics." In *Advances in Economic Theory, Fifth World Congress* (Truman Bewley, ed.), 269–299, Cambridge University Press.

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

Sánchez-Pagés, Santiago and Marc Vorsatz (2009), "Enjoy the silence: an experiment on truth-telling." *Experimental Economics*, 12, 220–241.

Sobel, Joel (2013), "Ten possible experiments on communication and deception." *Journal of Economic Behavior & Organization*, 93, 408–413.

Sobel, Joel (2020), "Lying and deception in games." *Journal of Political Economy*, 128, 907–947.

Sutter, Matthias (2009), "Deception through telling the truth?! Experimental evidence from individuals and teams." *Economic Journal*, 119, 47–60.

Vanberg, Christoph (2008), "Why do people keep their promises? An experimental test of two explanations." *Econometrica*, 76, 1467–1480.

Wang, Joseph Tao-yi, Michael Spezio, and Colin F. Camerer (2010), "Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games." *American Economic Review*, 100, 984–1007.

Wood, Daniel (2016), "Communication-enhancing vagueness." Working Paper.

Appendix A Proof for Task I

We argue that pooling is the unique equilibrium outcome of the game in Task I. To any message, there are seven possible sets of receiver responses: randomizing between all three or any two actions, which accounts for four possibilities, or taking one of the three actions with probability one, which accounts for three possibilities. Note, however, that any randomization with X and Y in the support can never be a best response: the receiver is indifferent between X and Y only if Blue and Green are believed to be equally likely, in which case the receiver strictly prefers Z. This rules out two possibilities as candidates for the receiver best responses.

Since the sender in state Blue has strict preference ordering $X \succ Z \succ Y$, there is a strict preference in any pairwise comparison of the remaining five possible sets of receiver responses. Thus, if there are at least two messages sent with positive probability and the receiver responds to them differently, the sender in state Blue will be willing to send only one of them. In order to have an equilibrium here, one of the messages must then be sent exclusively by the sender in state Green, in which case the receiver responds with Y. This is, however, the worst outcome for the sender in state Green, and thus the sender will not want to send an exclusive message. By the contrapositive, we conclude that in any equilibrium either the sender sends only one message or the receiver responds the same to all messages sent with positive probability.

Appendix B Receiver Motives in TEAM Treatment

Task I

Each receiver's answers to the multiple-choice questions in the opinion exchanges allow us to construct the receiver's beliefs. The most common beliefs are that the sender adopts a separating strategy, accounting for 65% of the cases; the most common type of anticipated separating strategy is the one that separates by sending "Blue" for Blue and "Either" for Green (40%). For the 35% of the cases where the receivers believe that the sender adopts a pooling strategy, the most frequently anticipated pooling message is "Blue" (23%).

Using these constructed beliefs, we next analyze individual best responses. In analyzing receiver aggregate behavior in Section 4.2, the aggregate frequencies allow us to speak of contingent choices of actions as if they were the strategy of a representative receiver. At the individual level, however, given that a receiver chooses an action rather than a strategy, notions such as naive responses are no longer applicable. Instead, we categorize the actions themselves and speak of X and Y as the *separating actions* and Z as the *pooling action*. In

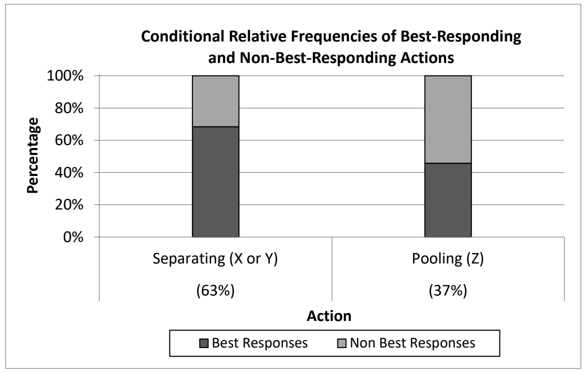


Figure B.1: Individual Receiver Best Responses in Task I

determining whether a separating action is a best response, we require that (a) the receiver believes that the sender adopts a separating strategy, and (b) the separating action is optimal given the receiver's belief about the current state. For the pooling action Z, while it is never an optimal action under either state and thus under the receiver's belief about the current state, it should plausibly be considered a best response when the receiver believes that the sender adopts a pooling strategy.

Based on these definitions of receiver best responses, Figure B.1 presents, for each class of action(s), separating (X or Y) or pooling (Z), the conditional relative frequency of best-responding actions among those cases in which the action(s) are chosen. We find that 68% of X or Y are chosen as best responses, while only 46% of Z are chosen as such. Multiplying these conditional frequencies with the respective frequencies of separating actions (63%) and pooling action (37%), we further categorize all cases of actions into four classes, which are presented in Table B.1. Separating best responses are the most frequent type of receiver actions (43%), while pooling best responses, which are consistent with equilibrium, are the least frequent (17%). We further find that best responding to messages as if they were true, taking actions X, Y, and Z in response to "Blue," "Green," and "Either" respectively, without regard to the sender's strategy accounts for half of non-best responses of receivers.

	Best-Responding (60%)	Non-Best-Responding (40%)
Separating (63%)	43%	20%
Pooling (37%)	17%	20%

Table B.1: Receiver Best-Responding Classes of Actions in Task I

Taste for Trust	Payoff Concern	Image Concern	Credulity	Confusion	None
5%	99%	1%	27%	8%	0%

Note: The five motive attributes are not mutually exclusive. “None” refers to the cases in which none of the five attributes are identified.

Table B.2: Receiver Motive Attributes in Task I

We further examine receiver motives using the written transcripts from the opinion exchanges. Table B.2 presents the relative frequencies with which each of the five attributes are identified, together with a “None” category. Payoff concern is virtually always present. Credulity is the second most frequent attribute, with 27% of the transcripts indicating that the receiver believes the message received to be literally true.[42]

Task II

We proceed to analyze receiver beliefs and best responses in Task II. For beliefs, in 49% of the cases the receivers believe that the sender adopts a separating strategy, which can be further divided into whether the separating strategy is completely truthful (30%) or lying (19%). In the remaining 51% of the cases, the receivers believe that the sender adopts a pooling strategy, which can again be further divided into whether the pooling is over “H” (31%) or “T” (20%).

Based on these constructed beliefs, we find that in 68% of the cases the receivers best respond to their beliefs about the sender’s choice of action. How frequent best responses are does not significantly vary across the two broad types of beliefs. Conditional on anticipated separating and pooling strategies, the relative frequencies of best responses are, respectively, 69% and 65% (two-sided $p = 0.94$).

Taste for Trust	Payoff Concern	Image Concern	Credulity	Confusion	None
3%	73%	0%	13%	2%	26%

Note: The five motive attributes are not mutually exclusive. “None” refers to the cases in which none of the five attributes are identified.

Table B.3: Receiver Motive Attributes in TEAM Task II

We further examine receiver motives using the written transcripts from the opinion ex-

[42]Here is a sample of representative transcripts for each type of motives (edited slightly for grammaticality): (a) taste for trust: “Blue, believe in humanities”; (b) payoff concern: “I don’t think they will cheat if the color is really blue as it does hurt both of us”; (c) image concern: “Like the last round, we should assume that they are being honest. A reward of 30 is much less than even the minimum of 50 in both colors. Perhaps they can even see our letter responses. I am going to choose X, and follow the strategy”; (d) credulity: “The message said is blue. Then assume it is really blue”; and (e) confusion: “Because we have to cooperate so as to maximize both payoff.”

changes, which are presented in Table B.3. Payoff concern is again the dominant attribute, with 73% of the transcripts indicating payoff-related motives. Credulity is the second most frequent, with 13% of the transcripts indicating the attribute.[43]

Appendix C SOLO Treatment Analysis

We repeat our analysis in Sections 4.2, 4.5, and 4.6 using the data from the robustness SOLO treatment (Sections 4.3 and 4.4, which examine sender and receiver motives elicited through the team design, are not applicable). SOLO serves as a control to explore whether our findings from the TEAM treatment are due to the concurrent variations in two dimensions of our design. Throughout this appendix, we reiterate some of the key figures from TEAM and point out any major difference. Our robustness inquiry culminates in the counterpart of Section 4.6 (Section C.3), in which we examine whether in SOLO non-strategic truthful behavior in Task III is correlated with strategic truthful behavior in Task I and II.

C.1 SOLO: Aggregate Communication Outcomes and Behavior

Task I

Table C.1 presents the joint frequencies over states and actions in Task I, alongside the joint distribution predicted by equilibrium. The unique pooling equilibrium outcome occur in 39% (37% in TEAM) of observations. The separating state-action pairs, (Blue, X) and (Green, Y), occur in 54% (47% in TEAM) of observations. "Miscommunication," represented by (Blue, Y) and (Green, X), occur in 7% of observations.

	X	Y	Z
Blue	**34%**	3%	13%
Green	4%	**20%**	26%

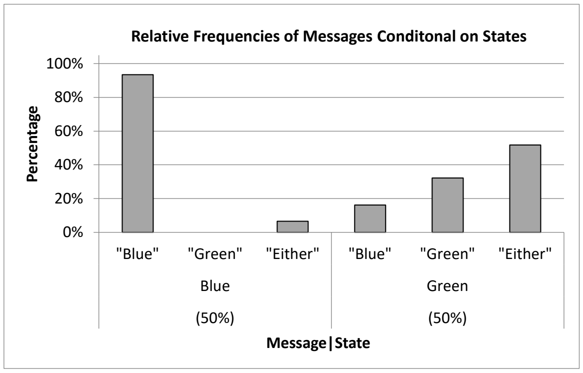
(a) Observed Outcome

	X	Y	Z
Blue	0%	0%	50%
Green	0%	0%	50%

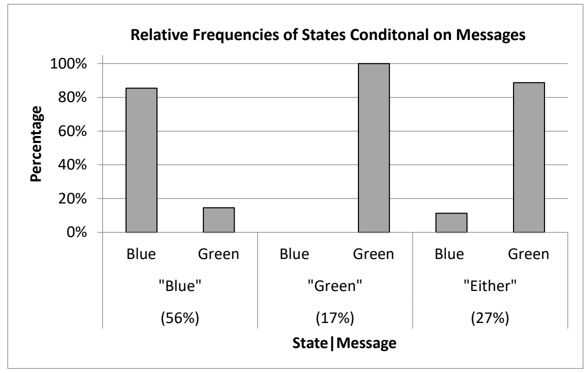
(b) Predicted Outcome

Table C.1: Observed and Predicted Outcomes in SOLO Task I

[43] Here is a sample of representative transcripts for each type of motives (edited slightly for grammaticality): (a) taste for trust: "I still believe they are honest"; (b) payoff concern: "They may lie to us and induce us to choose the one which could maximize their revenue. Their actual choose may be H"; (c) credulity: "Maybe they tell the truth"; (d) confusion: "If they tell us they will choose T, then we'll choose H in order to earn more money. As a result, they will earn more money"; (e) none: "Just make our decision as random and as unpredictable as possible. Pick any one you like."



(a) Message Behavior



(b) Implied Meanings of Messages

Figure C.1: Aggregate Message Behavior and Implied Meanings in SOLO Task I

Average receiver payoffs are 63.58 (57.56 in TEAM), which are significantly greater than the equilibrium benchmark of 50 of the pooling outcome ($p < 0.01$). We summarize this:

Finding C.1. *Average receiver payoffs in SOLO Task I are significantly greater than the equilibrium level, indicating the presence of overcommunication.*

For sender aggregate behavior, Figure C.1(a) presents the relative frequencies of the three messages conditional on the two randomly-determined states. Conditional on Blue, the relative frequency of the completely true "Blue" is 94%. Conditional on Green, the two most frequent messages are the incompletely "Either" (52%) and the completely true "Green" (32%). (The second most frequent message conditional on Green is "Blue" in TEAM.) Grouping both states, senders are completely truthful more often than not at 65% (54% in TEAM); including the incompletely true "Either," senders are truthful 92% of the time (80% in TEAM).

For empirically implied meanings, Figure C.1(b) presents the relative frequencies of the two states conditional on the three messages that are derived via Bayes' rule from the relative frequencies in Figure C.1(a). The modal meaning of "Blue" is Blue, with 85% (75% in TEAM) of all cases of "Blue" sent in state Blue. The modal meanings of "Green" and "Either" are both Green, with 100% (44% in TEAM) of all cases of "Green" and 89% (92% in TEAM) of all cases of "Either" sent in state Green. (The modal meaning of "Green" is Blue in TEAM.) "Green" is the least frequent message, sent in only 17% of all cases. An aggregate tendency to separate occurs primarily via messages "Blue" used to indicate Blue and "Either" used to indicate Green. Unlike the corresponding finding from TEAM, "Green" being used completely truthfully also plays a non-negligible role in the separation. We summarize the above:

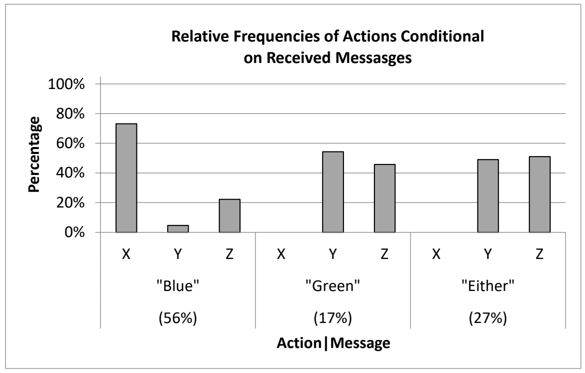


Figure C.2: Aggregate Action Behavior in SOLO Task I

Finding C.2. *Sender aggregate behavior in SOLO Task I indicates a tendency to separate. The separation occurs via messages "Blue," "Either," and, to a lesser extent, "Green," where the modal endogenous meaning of "Blue" is Blue and those of "Either"/"Green" are Green.*

For receivers, Figure C.2 presents the relative frequencies of receiver actions conditional on messages. Similar to the finding from TEAM, naive responses emerge as the modal behavior of receivers. The relative frequency of action X conditional on message "Blue" is 73% (62% in TEAM), that of Y conditional on the least frequently used "Green" is 54% (92% in TEAM), and that of Z conditional on "Either" is 51% (49% in TEAM). We summarize this:

Finding C.3. *The modal behavior of receivers in SOLO Task I is consistent with naive responses.*

The principal difference between Task I findings from SOLO and TEAM concerns how message "Green" is used. In both treatments, it is least frequently used among the three messages, although it is used more often in SOLO than in TEAM. Remarkably, "Green" in SOLO is always completely truthfully sent, never being sent in state Blue. This contributes to more informative communication (the moderately higher average receiver payoffs in SOLO than in TEAM), but the overall impact is limited because after all "Green" is infrequently sent, accounting for only 17% of all messages.

Task II

Table C.2 presents the relative frequencies of action profiles in Task II, alongside the equilibrium outcome. The sum of the observed frequencies of the two mis-matching action profiles, (H, T) and (T, H), is 63% (53% in TEAM), higher than the 50% prediction of the unique mixed-strategy equilibrium with or without pre-play communication. The fact that receivers' favorable action profiles occur more often than their unfavorable action profiles translates into

(a) Observed Outcome	Receiver: H	Receiver: T
Sender: H	16%	**42%**
Sender: T	**20%**	22%

(b) Predicted Outcome	Receiver: H	Receiver: T
Sender: H	25%	25%
Sender: T	25%	25%

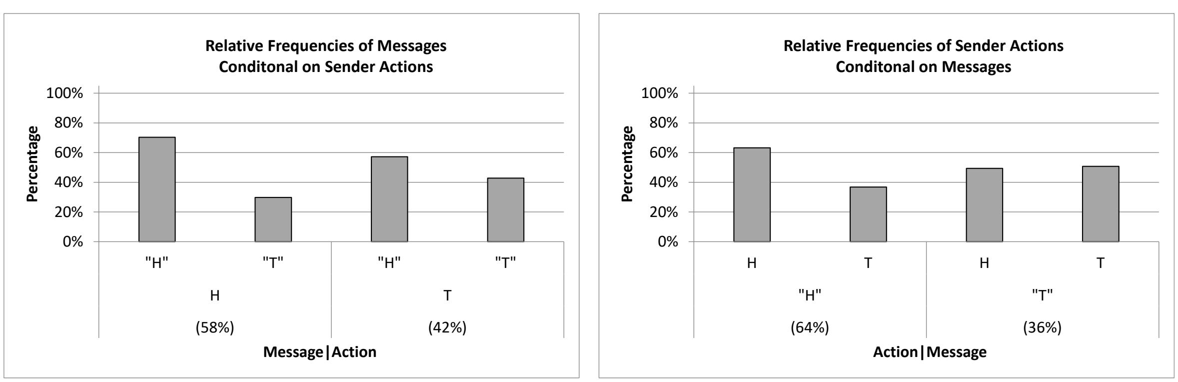
(a) Observed Outcome (b) Predicted Outcome

Table C.2: Observed and Predicted Outcome in SOLO Task II

average receiver payoffs of 32.28 (30.63 in TEAM), which are significantly greater than the equilibrium level of 30 ($p = 0.03$). Strictly following Definition 2 with regard to the presence of overcommunication in Task II, we draw a different conclusion from the counterpart in TEAM (in which we conclude no overcommunication is present given that average receiver payoffs therein are not significantly different from the equilibrium level):

Finding C.4. *Average receiver payoffs in SOLO Task II are significantly greater than the equilibrium level, indicating the presence of overcommunication.*

We further examine sender and receiver aggregate behavior to shed light on this somewhat different communication outcome. For senders, Figure C.3(a) presents the relative frequencies of their messages conditional on their actions. The frequency of message "H" conditional on action H is 70% (64% in TEAM), which is higher than the 57% (62% in TEAM) conditional on T ($p = 0.09$). Though not statistically significant, the frequencies vary across actions more so than in TEAM, suggesting that the messages are relatively more informative in SOLO.



(a) Message Behavior (b) Implied Meanings of Messages

Figure C.3: Aggregate Message Behavior and Implied Meanings in SOLO Task II

The empirically implied meanings provide clarification. Figure C.3(b) presents the relative frequencies of sender actions conditional on sender messages. The frequencies of action H conditional on messages “H” and “T” are, respectively, 63% and 49% (75% and 73% in TEAM). The aggregate pre-play communication provides some information in the sense that the “posteriors” vary more noticeably across messages, which stands in contrast to the equilibrium-conforming babbling observed in TEAM. Finally, in terms of truthful behavior, we note that there are more true messages when the action is H (70%) than when it is T (43%), with an overall relative frequency of truth messages at 63% (59% in TEAM).

The findings on sender actions are also different from those in TEAM. While as in TEAM senders in SOLO choose H more often, the relative frequency of H at 58% (74% in TEAM) is not significantly higher than the equilibrium prediction of 50% ($p = 0.11$). Sender actions in SOLO are thus better in line with the uniform equilibrium mixed strategy.

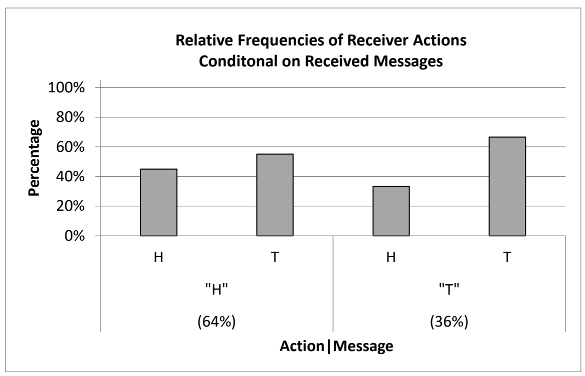


Figure C.4: Aggregate Action Behavior in SOLO Task II

Receiver aggregate behavior is closer to that in TEAM. The aggregate actions depart significantly from the uniform equilibrium mixed strategy with the same proclivity toward action T. The relative frequency of T is 64% (59% in TEAM), significantly higher than the equilibrium prediction of 50% ($p = 0.02$). In terms of communication, as in TEAM, receiver actions are not significantly influenced by sender messages; consistent with the predicted babbling, their proclivity toward T is largely message-independent. Figure C.4 presents the relative frequencies of receiver actions conditional on received messages. The frequency of action T conditional on message “T” is 67% (65% in TEAM), which is higher, but not significantly so, than the 55% (55% in TEAM as well) conditional on “H” ($p = 0.29$). We summarize the analysis of sender and receiver aggregate behavior in Task II with our next finding:

Finding C.5. *Aggregate behavior in SOLO Task II indicates that (a) the pre-play communication is partially consistent with equilibrium babbling, in which sender intended actions are limitedly revealed but receiver actions are not significantly influenced by the communication,* *and (b) the actions chosen by receivers, which gravitate toward T, are not consistent with the uniform-mixing uniquely predicted by equilibrium, while sender actions, with only a slight gravitation toward H, are closer to the equilibrium prediction.*

The relatively strong gravitation toward action T by receivers, combined with a slight gravitation toward action H by senders, generates frequencies of action profiles that are more favorable to receivers, resulting in average receiver payoffs that are higher than the mixed-strategy equilibrium level. While overcommunication is deemed to have occurred in light of payoffs, receiver responses to messages suggest that in aggregate they do not quite pick up the limited information revealed by senders in the pre-play communication. Yet the principal differences with the Task II findings in TEAM rest with the higher receiver payoffs and the somewhat more informative messages in SOLO.

Task III

Table C.3 presents the observed frequency of each reported number, together with the predicted distribution of reports and the associated payoffs. Only 58% (60% in TEAM) of observed reports are sixes. The profile of observed reports translates into average payoffs of 50.2, significantly lower than the predicted level of 60 ($p < 0.01$). We report the following finding:

Choice of Report	1	2	3	4	5	6
Payoff	10	20	30	40	50	60
Optimal Report	0%	0%	0%	0%	0%	100%
Observed Report	**5%**	**5%**	**6%**	**7%**	**19%**	**58%**

Table C.3: Observed Reports in SOLO Task III

Finding C.6. *Average payoffs in SOLO Task III are significantly less than the predicted level, indicating the presence of truthful behavior.*

C.2 SOLO: Opponent Payment Rounds

Table C.4 presents, for each of the subsamples from opponent payment and non-payment rounds, the relative frequencies of sender true messages and average receiver payoffs. Although senders tend to send more completely true messages in opponent payment rounds than in non-payment rounds in both tasks, none of the differences, including those in average receiver

	Opponent Payment Round	Opponent Non-Payment Round	Wilcoxon Signed Rank Test
Task I			
Senders			
True Messages	89%	95%	$p = 0.38$
Completely True	67%	63%	$p = 0.84$
Incompletely True	22%	32%	$p = 0.64$
Receivers			
Payoffs	63.94	63.22	$p = 0.98$
Task II			
Senders			
True Messages	71%	56%	$p = 0.16$
Receivers			
Payoffs	30.67	33.89	$p = 0.34$

Note: Opponent payment round is relative to the role listed for each set of rows, i.e., the "Senders" values are based on receiver payment and non-payment rounds, while the "Receivers" values are based on sender payment and non-payment rounds. Reported p-values are from two-sided tests. All differences between the two payment-round types remain insignificant at the 5% level using one-sided tests.

Table C.4: Comparisons between Opponent Payment and Non-Payment Rounds in SOLO

payoffs, are statistically significant. Regarding the impacts of other-regarding preferences, we reach the same conclusion for SOLO as we have done for TEAM:

Finding C.7. *Comparisons of aggregate behavior between opponent payment and non-payment rounds indicate that other-regarding preferences have minimal effects on the key strategic communication behavior in Tasks I and II.*

C.3 SOLO: Non-Strategic Reporting Behavior in Task III

Among all observed reports in Task III, 23% (28% in TEAM) are true, 22% (14% in TEAM) are partial lies, and 55% (58% in TEAM) are maximal lies.[44] Figure C.5 presents the relative frequencies of the three types of reports conditional on each realized lower number. The relative frequencies of maximal lies range from 42% to 64% (56% to 63% in TEAM). For true reports, excluding the cases of realized six in which we observe 100% of truthfully reported

[44]There is a small fraction (3%) of reported sixes that are true, which occur when both dice roll six. These 3% cases all occurred with subjects who were receivers in the strategic tasks.

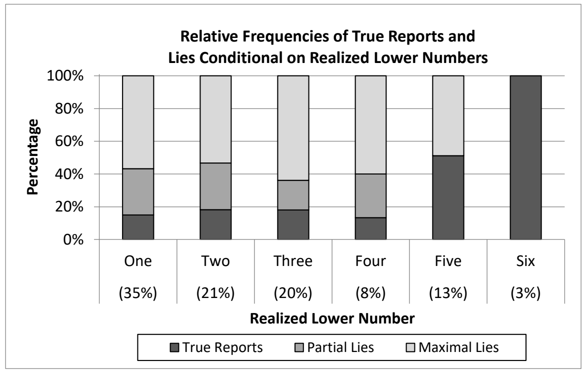


Figure C.5: Reporting Behavior in SOLO Task III

sixes, the relative frequencies range from 15% to 58% (19% to 44% in TEAM).

Turning to our principal inquiry, we present in Figure C.6 the relative frequencies of truthful behavior in Tasks I and II conditional on truthful and lying behavior in Task III. For Task II, the relative frequency of truthful behavior conditional on being truthful in Task III is 50%, which is lower, but not significantly so, than the 70% conditional on lying in Task III ($p = 0.23$). For Task I, the relative frequency of completely truthful behavior conditional on being truthful in Task III is 71%, which is higher, but not significantly so, than the 65% conditional on lying in Task III ($p = 0.37$). For incompletely truthful behavior, the relative frequency is 18% conditional on being truthful in Task III, which is lower, but not significantly so, than the 28% conditional on lying in Task III ($p = 0.41$). Counting both classes of true messages, the cases stack up to 89% conditional on being truthful in Task III, which is lower, but not significantly so, than the 93% conditional on lying in Task III ($p = 0.06$).

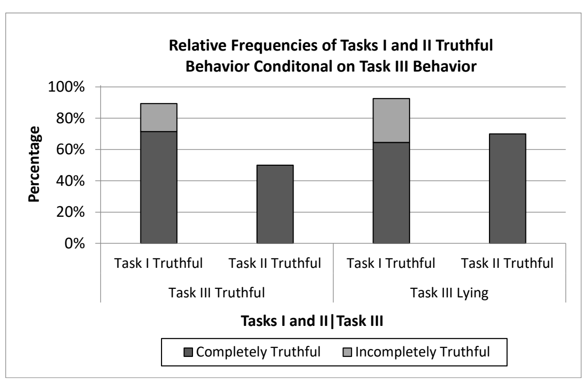


Figure C.6: Truthful Behavior in SOLO Tasks I and II
Conditional on Behavior in Task III

We further examine the relationship between non-strategic and strategic truthful behavior by running the same set of regressions as we have done for TEAM. Table C.5 reports the

A. Both States and Types of Opponent Payment Round

	(1)	(2)	(3)	(4)
	$True_i^{\text{III}}$		$DLie_i^{\text{III}}$	
$True_i^{\text{I}}$	−0.25	–	3.12*	–
	(0.38)	–	(1.13)	–
$CTrue_i^{\text{I}}$	–	−0.22	–	3.80*
	–	(0.34)	–	(1.44)
$ITrue_i^{\text{I}}$	–	−0.28	–	2.75
	–	(0.44)	–	(1.21)
Constant	0.46	0.44	−0.56	0.90
	(0.35)	(0.33)	(0.96)	(1.06)
Observations	40	40	40	40

B. By State and Type of Opponent Payment Round

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$True_i^{\text{III}}$		$DLie_i^{\text{III}}$		$True_i^{\text{III}}$		$DLie_i^{\text{III}}$	
	State Green				State Blue			
$True_i^{\text{I}}$	−0.23	–	1.60*	–	/	–	/	–
	(0.20)	–	(0.68)	–	/	–	/	–
$CTrue_i^{\text{I}}$	–	−0.19	–	2.12	–	0.25**	–	−1.07
	–	(0.22)	–	(1.05)	–	(0.07)	–	(0.61)
$ITrue_i^{\text{I}}$	–	−0.24	–	1.41	–	–	–	/
	–	(0.24)	–	(0.70)	–	–	–	/
Constant	0.42*	0.42*	0.84	0.79	0.23**	/	2.50***	3.50***
	(0.18)	(0.18)	(0.58)	(0.61)	(0.07)	/	(0.36)	(0.38)
Observations	30	30	30	30	30	30	30	30
	Opponent Payment Round				Opponent Non-Payment Round			
$True_i^{\text{I}}$	0.07	–	1.10	–	−0.37	–	2.33**	–
	(0.23)	–	(0.99)	–	(0.39)	–	(0.59)	–
$CTrue_i^{\text{I}}$	–	0.11	–	1.25	–	−0.43	–	2.65**
	–	(0.21)	–	(1.14)	–	(0.40)	–	(0.65)
$ITrue_i^{\text{I}}$	–	−0.01	–	0.74	–	−0.27	–	1.78
	–	(0.30)	–	(1.10)	–	(0.40)	–	(0.87)
Constant	0.15	0.15	1.32	1.29	0.50	0.50	0.50	0.50
	(0.21)	(0.20)	(0.87)	(0.91)	(0.38)	(0.39)	(0.38)	(0.39)
Observations	32	32	32	32	32	32	32	32

Note: Panel A reports OLS estimates including both states and using sender i's Task I average truthfulness across the two rounds. Panel B reports OLS estimates using sender i's Task I average truthfulness calculated over the round or rounds that satisfy the condition of the restricted sample. For state Green, e.g., if sender i encounters Green in both rounds, the average over both rounds is used; if sender i encounters Green in only one round, data from only that round is used. Standard errors clustered at the matching-cohort level are in parentheses. The "/" mark refers to the case where there is no variation in the variable in the state Blue sample, and thus an estimate is not obtained. *** indicates significance level at 0.1%, ** at 1%, and * at 5%.

Table C.5: Correlations Between Truthful Behavior in SOLO Task III and Task I

regression results for Tasks III and I. The estimates covering all data are presented in panel A, and those based on four restricted samples are in Panel B. None of the specifications using the binary truthfulness measure $True_i^{\text{III}}$ show any significant effects, while results are more mixed when truthfulness is measured via the magnitude of lies $DLie_i^{III}$. The overall *frequency* of Task III truthfulness is not correlated with Task I truthfulness; however, conditional upon lying in Task III, we see lies with larger magnitudes coming from those who are truthful in Task I. In other words, those who are truthful in Task I are no more likely to be truthful in Task III, but those who do choose to be dishonest tell larger lies. We further note that, similar to the finding from TEAM, Chow tests indicate no significant differences between the two sets of estimates in panel B under opponent payment and non-payment rounds ($p \geq 0.24$).

A. Both Types of Opponent Payment Round

	(1)	(2)
	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$
$True_i^{\text{I}}$	−0.21	1.56
	(0.26)	(0.89)
Constant	0.36	1.35
	(0.19)	(0.67)
Observations	40	40

B. By Type of Opponent Payment Round

	(1)	(2)	(3)	(4)
	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$	$True_i^{\text{III}}$	$DLie_i^{\text{III}}$
	Opponent Payment Round		Opponent Non-Payment Round	
$True_i^{\text{I}}$	−0.24	0.75	−0.05	1.14
	(0.23)	(0.89)	(0.17)	(0.59)
Constant	0.40	1.80*	0.29	1.53**
	(0.19)	(0.75)	(0.13)	(0.38)
Observations	34	34	34	34

Note: Panel A reports OLS estimates using sender i's Task II average truthfulness across the two rounds. Panel B reports OLS estimates using sender i's Task I average truthfulness in the round or rounds that satisfy the condition of the restricted sample. Standard errors clustered at the matching-cohort level are in parentheses. *** indicates significance level at 0.1%, ** at 1%, and * at 5%.

Table C.6: Correlations Between Truthful Behavior in SOLO Task III and Task II

Table C.6 reports the regression results for Tasks III and II. Regardless of choice of dependent variable or type of opponent payment round, none of the coefficients are significantly different from zero; being truthful in Task II is not correlated with being truthful in Task III. With regards to the role of other-regarding preferences, Chow tests once again show that there are no significant differences between the two sets of estimates in panel B ($p \geq 0.43$). We summarize our finding regarding the truthful behavior across tasks:

Finding C.8. *At both aggregate and individual levels, truthful behavior in Task III is not positively correlated with truthful behavior in Task I or II. In some cases, there is significant negative correlation between the magnitude of lies in Task III and truthful behavior in Task I.*

Finding C.8 provides evidence that one of our main findings from TEAM, that non-strategic truthful behavior may not explain non-equilibrium behavior in strategic settings, is unlikely to be a consequence of strategic communication being conducted in teams and non-strategic communication conducted individually.

Appendix D Experimental Instructions: TEAM Treatment

D.1 Instructions

INSTRUCTIONS

Welcome to the experiment. This experiment studies decision making in three different settings. Please read these instructions carefully; the cash payment you will receive at the end of the experiment will depend on the decisions you make.

Overview

You will participate in three decision tasks. Below are the instructions for Task I. You will receive the instructions for Tasks II and III later.

There are 16 participants divided into two "matching cohorts," each with 8 participants. The two cohorts are independent and never interact with each other at any point during the experiment.

Your earnings from the experiment will be determined by your decisions as well as the decisions of other participants. Upon finishing Task III, you will be told what your earnings were in each of the three tasks and will receive your cash payment. The final cash payment will be the sum of your earnings from the three tasks, translated into HKD at an exchange rate of 1 reward point = 0.7 HKD, plus a "show-up payment" of 30 HKD.

Your decisions as well as your monetary payment will be kept confidential. You will not be told the identity of the other participants during or after the experiment, nor will they ever be told your identity.

Task I

Teams and Decision Groups

Figure 1 illustrates the participation structure in a matching cohort. There are 4 teams in each cohort. You will be randomly placed into a *team* with another participant. The computer will randomly match your team with another team to form a *decision group*. In each decision group, one team will be randomly designated as *Team A* and the other as *Team B*.

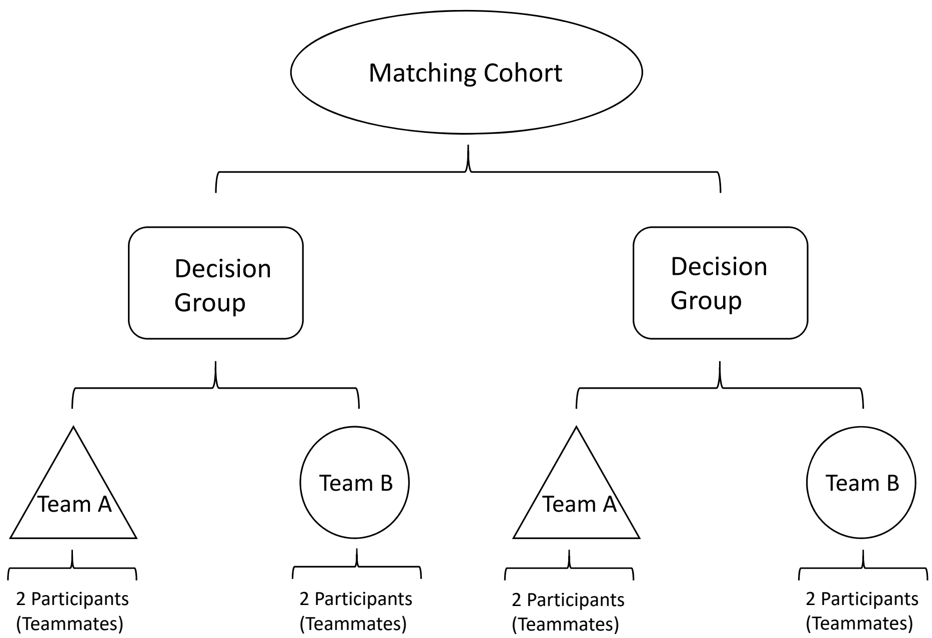


Figure 1: Matching Cohort, Decision Group, and Team

Rounds and Rewards

You will participate in *two rounds of decisions*. Your team designation (either Team A or Team B) will be the same in both rounds. After the first round, you will be randomly rematched with another participant to form a new team; your new team will then be randomly rematched with an opposite team to form a new decision group.

Table 1 contains the potential rewards that you and the other participants in your decision group may earn in a round.

One of the 6 cells, determined by the selected color and action, will be applicable to the current round. In each cell, the first number is the reward for *each* participant in Team A, and the second number is the reward for *each* participant in Team B. For example, if the selected color is Green and the selected action is Y (how they are selected will be explained below), each

		Action		
		X	Y	Z
Color	Blue	80, 80	10, 10	50, 50
	Green	30, 10	10, 80	50, 50

Table 1: Potential Rewards for Task I

Team-A member will earn 10 points while each Team-B participant will earn 80 points.

Color, Message, and Action

At the beginning of each round, the computer will randomly select with equal chance either Blue or Green. The selected color will be revealed *only* to the two members of Team A. After learning the selected color, the two Team-A members will engage in an "exchange of opinions" to be described below. After the exchange, each member will independently choose one of the following three messages to send to Team B:

- "The color is Blue,"
- "The color is Green," and
- "The color is either Blue or Green."

The message chosen by *one* Team-A member becomes the message for the team: the computer will *randomly* select a member and transmit the message chosen by that member to Team B.

Both members of Team B will learn the message transmitted from Team A, after which they exchange opinions. Each member of Team B will then independently choose an action, either X, Y, or Z. Similar to the determination of Team A's message, the action chosen by *one* randomly selected Team-B member will become the team's action.

Team B's final action, together with the selected color, will then be used to determine the rewards for the four participants in the decision group according to Table 1.

Exchange of Opinions between Teammates

After learning the color but before choosing a message, the two Team-A members will engage in an exchange of opinions. Each member will be prompted to answer several questions about what message should be sent to Team B. The answers will then be revealed between teammates, providing an opportunity for each Team-A member to suggest a choice of message to his/her teammate.

After viewing the message transmitted from Team A but before choosing an action, the two Team-B members will also exchange opinions. Each member will answer several questions about what action to choose. The answers will be revealed between teammates and provide an opportunity for suggestions of action within Team B.

Earnings from Task I

For each team, one of the two rounds will be *randomly* selected with equal chance for payment. The rewards you earn in this selected round will become your cash earnings for Task I, after the conversion to HKD as outlined above.

You will *not* be told which round is chosen to be your team's payment round for Task I until the end of the experiment (i.e., after the completion of Task III). As a result, it is in your best interest to take each round seriously.

Payment Round of the Opposing Team in Your Decision Group

The payment rounds of the two teams in a decision group are determined *independently*. While you will *not* know whether the current round is the payment round for your team, you *will* be able to see if it is the payment round for the opposing team, which may or may not be the same as your own team's payment round. *You can only influence the other team's earnings when it is their payment round.*

Summary of the Events in Each Round

Team A:

1. The computer randomly selects with equal chance either Blue or Green.
2. Members of Team A learn the color for the round. (Members of Team B do not.)
3. Members of Team A exchange opinions.
4. Each member of Team A chooses a message for Team B.
5. One member of Team A is randomly selected by the computer. That member's choice of message is transmitted to Team B.

Team B:

6. Both members of Team B view the message transmitted from Team A.
7. Members of Team B exchange opinions.

8. Each member of Team B chooses an action.

9. One member of Team B is randomly selected by the computer. That member's choice of action is Team B's action for the round.

10. Rewards for all four participants of the decision group are determined by Team B's choice of action and the color selected by the computer for the round.

Instructions for Tasks II and III

You will receive the instructions for the next task after you complete the current task.

If you have any questions, please raise your hand now.

Task II

Teams, Decision Groups, and Rounds

The teams and decision groups in Task I (Figure 1) remain the same in Task II. Your team designation from Task I (either Team A or Team B) continues to hold. There are also *two rounds of decisions* in Task II, with teammates and teams randomly matched for each round.

Rewards

Table 2 contain the potential rewards that you and the other participants in your decision group may earn in a round.

		Team B: H	Team B: T
Team A	H	40, 20	20, 40
	T	20, 40	40, 20

Table 2: Potential Rewards for Task II

One of the 4 cells, determined by the actions, H or T, chosen by Team A (row) and Team B (column), will be applicable to the current round. In each cell, the first number is the reward for *each* Team-A member, and the second number is the reward for *each* Team-B member.

Message and Action

Prior to choosing actions, the two Team-A members will engage in an exchange of opinions. After the exchange, each member will independently choose a message, either "We will choose

H" or "We will choose T," to send to Team B. Each member will also independently choose an action. The message and action chosen by *one* member become the choices for the team: the computer will randomly select a member and transmit the message chosen by that member to Team B; the action chosen by that member will also become the team's action.

Both members of Team B will learn the message transmitted from Team A, after which they will exchange opinions. Each member will then independently choose an action. The action chosen by *one* randomly selected member will become the team's action.

The actions chosen by the two selected members, one from Team A and one from Team B, will be used to determine the round rewards for the four participants in the decision group according to Table 2.

Exchange of Opinions between Teammates

Before choosing the message and the action, the two Team-A members will engage in an exchange of opinions. Each member will be prompted to answer several questions related to the message and action choices. *After viewing the message transmitted from Team A and before choosing the action*, the two Team-B members will also exchange opinions. Each member will answer several questions related to the received message and action choice.

Similar to the exchanges of opinions in Task I, answers to the questions will be revealed between teammates, providing an opportunity for each member to suggest a choice of message and action (Team A) or just a choice of action (Team B) to his/her teammate.

Earnings from Task II

One of the two rounds will be *randomly* selected with equal chance for payment, where the reward points will be converted to HKD as outlined above. You will *not* be told which round is chosen to be your team's payment round for Task II until the end of the experiment.

Payment Round of the Opposing Team in Your Decision Group

Similar to the payments rounds for Task I, while you will *not* know whether the current round is the payment round for your team, you *will* be able to see if it is the payment round for the opposing team, which may or may not be the same as your own team's payment round.

Summary of the Events in Each Round

Team A:

1. Members of Team A exchange opinions.

2. Each member of Team A chooses a message for Team B and an action.

3. One member of Team A is randomly selected by the computer. That member's choice of message and action is used for the team.

Team B:

6. Both members of Team B view the message transmitted from Team A.
7. Members of Team B exchange opinions.
8. Each member of Team B chooses an action.
9. One member from Team B is randomly selected by the computer. That member's choice of action becomes the team's action.
10. Rewards for all participants of the decision group are determined by the teams' actions.

Task III

Task III is your final task in the experiment. Unlike Tasks I and II, in this task you will make a decision without being matched or interacting with any other participant. Your reward will be determined solely by your own decision.

Dice Rolling and Report

The computer will randomly generate two numbers, each of which is between 1 and 6, where all six numbers are equally likely. It is as if the computer rolls two six-sided dice. You will be shown the two numbers, after which you will be asked to report the *lower* of the two.

Your reward will be determined based *only* on the number you report, as described in Table 3; the randomly generated numbers will *not* affect your reward.

Reported Number	1	2	3	4	5	6
Reward	10	20	30	40	50	60

Table 3: Potential Reward for Task III

The two randomly generated numbers will be recorded by the computer, but they will be *completely anonymous*. No one will be able to match your identity to the numbers you observed or the value you reported, either during the experiment or after it is completed.

Earnings from Task III

Your earning for Task III is simply one of the values given in Table 3, after the conversion to HKD as outlined above.

Completion of the Experiment

After finishing Task III, the experiment will be over. You will be instructed to receive your payment according to the payment procedure outlined above.

D.2 Survey Questions

The following survey questions were administered for Task I:

1. Team A
   (a) What action do you think Team B would take if your team says “the color is BLUE”? Possible answer: “X”; “Y”; “Z”
   (b) What action do you think Team B would take if your team says “The color is GREEN”? Possible answer: “X”; “Y”; “Z”
   (c) What action do you think Team B would take if your team says “The color is either BLUE or GREEN”? Possible answer: “X”; “Y”; “Z”
   (d) What color do you think your team should report to Team B? Possible answer: “Blue”; “Green”; “Blue or Green”
   (e) Do you have any other comment or opinion that you would like to share with your teammate? Possible answer: Written responses
2. Team B
   (a) Given the message sent by Team A, what do you think the true color is? Possible answer: “Blue”; “Green”
   (b) If the true color was different from what you think it is, what do you think Team A would report to you? Possible answer: “Blue”; “Green”; “Blue or Green”
   (c) Given the message sent by Team A, what action do you think your team should choose? Possible answer: “X”; “Y”; “Z”
   (d) Do you have any other comment or opinion that you would like to share with your teammate? Possible answer: Written responses

The following survey questions were administered for Task II:

1. Team A

   (a) What action do you think Team B would take if your team says "We will choose H"? Possible answer: "H"; "T"

   (b) What action do you think Team B would take if your team says "We will choose T"? Possible answer: "H"; "T"

   (c) What action do you think your team should choose? Possible answer: "H"; "T"

   (d) What message do you think your team should send to Team B? Possible answer: "We will choose H"; "We will choose T"

   (e) Do you have any other comment or opinion that you would like to share with your teammate? Possible answer: Written responses

2. Team B

   (a) Given the message sent by Team A, what action do you think they will choose? Possible answer: "H"; "T"

   (b) If Team A intended to choose a different action from what you think they will, what message do you think they would send to you? Possible answer: "We will choose H"; "We will choose T"

   (c) Given the message sent by Team A, what action do you think your team should choose? Possible answer: "X"; "Y"; "Z"

   (d) Do you have any other comment or opinion that you would like to share with your teammate? Possible answer: Written responses

Appendix E Classification Instructions for Research Assistants

CLASSIFICATION INSTRUCTIONS

Your task is to classify the motives exhibited by subjects in a communication game experiment. We would like you to classify different types of motives as implied by transcripts we collected from pairs of subjects as they discussed how to behave in the experiment.

If at any point you have questions about these instructions, please let us know. We will gladly clarify the details of the experiment and your task as a classifier; however, we cannot give you guidance as to how to classify any specific transcript. We want you to use your best judgment to assess each transcript, and do not want to influence your choices.

You will classify the transcripts from two distinct decision tasks in the experiment, Task I and Task II, each involving a different communication game. Please read through all of these instructions before you begin classifying.

<u>Task I</u>

Here we briefly summarize Task I, before asking you below to read through the experimental instructions in detail. Task I is a communication game of *private information*, in which subjects play the roles of senders and receivers. Each round of the game is played between one pair of senders and one pair of receivers. The senders are privately informed whether a randomly-determined color variable is Blue or Green. This information is not available to the receivers. After learning the color, the senders send a single message to the receivers regarding the observed color. The available messages for the senders to choose are "The color is Blue," "The color is Green," and "The color is Green or Blue." The senders are free to send any one of these three messages, regardless of the actual randomly-determined color. After receiving the message, the receivers takes one of three actions, X, Y, or Z, which determine the payoffs for both the senders and receivers.

The above is only a brief overview of Task I. To gain a full understanding of our experiment, we will ask you in a moment to put these classification instructions aside and carefully read through the experimental instructions that were provided to the subjects. Among other details, the instructions describe additional communication that took place *within* each team, from which we obtained the transcripts you are about to analyze.

For now, you will only need to read through the end of *page 4* of the experimental instructions, which cover Task I. You will later read through the experimental instructions for Task II. You do not need to read the instructions for Task III, as it produced no transcripts to classify.

When you have finished reading the experimental instructions for Task I, you should continue with these classification instructions. Please set these classifications instructions aside now and read the experimental instructions, returning here when you have finished.

Classification of Senders

Each sender in Task I answered five survey questions that were shared with their teammate prior to choosing what message to send to the receivers. You have been provided with a data file that contains the answers each sender provided to those five questions. Your first job is to identify for each transcript whether the sender has demonstrated any of four different motives. After you have classified the transcripts of senders, you will then move on to the transcripts of receivers, as will be described later.

The five questions answered by senders were:

1. What action do you think Team B would take if your team says "The color is BLUE"?
2. What action do you think Team B would take if your team says "The color is GREEN"?
3. What action do you think Team B would take if your team says "The color is either BLUE or GREEN"?
4. What color do you think your team should report to Team B?
5. Do you have any other comment or opinion that you would like to share with your teammate?

Below is a description of the motives you will look for in sender transcripts from Task I. There are four different motives: preference for honesty, payoff-concern, confusion, and image-concern. You should record each attribute as a binary variable, taking a value 1 if the transcript indicates the motive and 0 if it does not. Please refer back to these descriptions as necessary during your classification of the sender transcripts.

- The "preference for honesty" variable indicates whether a subject expressed a desire to be honest or to avoid lying *for its own sake*. It takes a value of 1 if subjects in any way indicated that truthfulness was desirable on its own merits. Transcripts should *not* be flagged as honest if truthful behavior seems to be motivated solely by payoff-seeking, such as if a sender said something along the lines of "They'll think we're lying, so tell the truth instead."
- The "payoff-concern" variable indicates any direct concern for payoffs, including when subjects expressed a desire for "fair" or "equal" payoffs between senders and receivers. It does not require a demonstration of self-interested payoff maximization, only that

the subject indicated that concern for *some* monetary payoffs motivated their behavior. Subjects are given a 1 for this variable if their transcript indicates that they were motivated by concern for payoffs, even if they did not explicitly say so. For example, subjects who indicated that a particular message should be sent to "get the other team to do what we want" would be classified as payoff-concern, even though payoffs were not explicitly mentioned. Likewise, a sender who wrote something along the lines of "Let's help them earn more" is also classified as demonstrating payoff-concern.

- The "confusion" variable takes the value 1 if a transcript indicates the subject fundamentally misunderstood some aspect of the game. For example, a sender would be given a 1 for confusion if they claimed that the action X always gives receivers a higher payoff, regardless of what the actual color is.

- The "image-concern" variable takes the value 1 if a transcript indicates any type of concern for what others might think of the subject as a result of the choices made. For example, a subject would demonstrate image-concern if they said something along the lines of "We have to tell the truth, otherwise they'll think we're liars" or "I don't want the people running this experiment to think we're dishonest."

The motives are *not* mutually exclusive, so, e.g., a message such as "Just tell the truth. Not their payment round anyways" receives a 1 for both honesty and payoff-concern. It is in principle possible for a transcript to be given all 1s, all 0s, or any combination. Please evaluate each motive independently of one another, and ask yourself "does this transcript provide evidence that the subject was influenced by this motive?"

After you have decided whether a motive is present or not, you will also indicate how confident you are in your assessment. For each motive, you will record an associated confidence rating. Each confidence rating should be an integer ranging from 0 to 3:

0 = little to no confidence—it is very difficult or impossible to judge what the correct classification should be.

1 = low confidence—you believe this classification to be most likely correct, but it is easy to argue for the opposite classification instead. (The opposite classification of the presence of a motive, i.e., a 1, is the absence of it, i.e., a 0, and vice versa).

2 = high confidence—you strongly believe this classification to be correct, but the opposite classification still seems plausible.

3 = certain or nearly certain—it is hard to imagine any argument for the opposite classification.

Please note that, if a transcript in no way suggests a particular motive, you should give that motive a classification of 0, and confidence of 3. This indicates that you are certain the subject did not indicate the motive, regardless of whether you are able to tell what *did* motivate them. By contrast, you would give a confidence of 0 if you are unable to tell *whether or not* that particular motive was expressed in their transcript.

Classification of Receivers

Recall that receivers did not observe the actual randomly-determined color, but they did receive a message about the color from the senders. After observing the message sent by the senders, receivers were asked to answer four survey questions. Each receiver's survey questions were shared with their teammate prior to the receivers making their final choice of action.

The four questions answered by receivers were:

1. Given the message sent by Team A, what do you think the true color is?
2. If the true color was different from what you think it is, what do you think Team A would report to you?
3. Given the message sent by Team A, what action do you think your team should choose?
4. Do you have any other comment or opinion that you would like to share with your teammate?

Your next job is to classify each receiver transcript for whether the receiver has demonstrated any of five different motives. Receiver transcripts are analyzed similarly to those of senders, though for slightly different motives, reflecting the different roles they took in the experiment. The five different motives are: taste-for-trust, payoff-concern, confusion, image-concern, and credulity. Below is a description of each of the motives. Please refer back to these as necessary during your classification of the transcripts.

- The "taste-for-trust" variable indicates that a receiver finds it desirable to trust the message to be true *for its own sake.* In other words, the receiver indicated that the act of trusting the sender to tell the truth is intrinsically desirable, e.g., "my mom told me to always trust people." Note that this motivation is about it being desirable to trust the sender, separate from actually believing the sender's message to be true, which would be about a different motive—credulity—to be discussed below.

- The "payoff-concern" and "confusion" variables are defined the same as they were for senders. If a receiver indicates concern for *any* subject's payoffs, the transcript receives a value of 1 for the payoff-concern variable. If a receiver indicates any misunderstanding of the rules of the game, the transcript receives a 1 for the confusion variable.
- The "image-concern" variable indicates that a receiver showed any type of concern for what others might think of the subject as a result of the choices made, e.g., "I don't want the senders to think we don't trust them," or "I like other people to know I'm trustworthy."
- The "credulity" variable indicates the receiver believed the message to be literally true, regardless of the sender's motives. This does not require the receiver to indicate that trust or trustworthiness is in any way desirable. They simply need to express that they believe the message to be true. For example, "I think they're telling the truth that it's Blue" or "They said Blue to make us think it's Green, but I bet it really is Blue" would both demonstrate that the receiver was credulous.

As before, after deciding whether a motive is present or not, please indicate how confident you are in your assessment. For each motive, you will record an associated confidence rating, using the same 0 to 3 scale as above.

Task II

For Task II, you are asked to conduct a similar classification exercise but for a different game. Task II is a version of the matching pennies game, augmented with one-sided communication of *intention*. Each round of the game is played between two teams of players, where each team chooses between two actions, H and T. Before they choose an action, one team (the senders) conveys a message to the other team (the receivers), either "We will choose T" or "We will choose H." The teams each choose their own action after the senders' message is viewed by the receivers. Note that the senders' conveyed message does not have to be the same as the action they actually choose.

The above is only a brief overview of Task II. To gain a full understanding of our experiment, we will ask you in a moment to put these classification instructions aside and carefully read through the experimental instructions that were provided to the subjects. Please read from *page 5 through 6*, which cover Task II. You do not need to read page 7, which covers Task III.

When you have finished reading the experimental instructions for Task II, you should continue with these classification instructions. Please set these classifications instructions aside now and read the experimental instructions, returning here when you have finished.

Classification of Senders

Each sender in Task II answered five survey questions that were shared with their teammate prior to choosing what message to send to the receivers. Your next job is to identify for each transcript whether the sender has demonstrated any of four different motives. After you have classified the transcripts of senders, you will then move on to the transcripts of receivers, similar to what you did in Task I.

The five questions answered by senders were:

1. What action do you think Team B would take if your team says "We will choose H"?
2. What action do you think Team B would take if your team says "We will choose T"?

3.a What action do you think your team should choose?

3.b What message do you think your team should send to Team B?

4. Do you have any other comment or opinion that you would like to share with your teammate?

As was the case in Task I, your goal is to classify different motives indicated in each transcript. You will look for the same set of four motives as you did for senders in Task I: preference for honesty, payoff-concern, confusion, and image-concern. For your convenience, the description of those four motives are repeated here. Please refer back to these descriptions as needed.

- The "preference for honesty" variable indicates whether a subject expressed a desire to be honest or to avoid lying *for its own sake.* It takes a value of 1 if subjects in any way indicated that truthfulness was desirable on its own merits. Transcripts should *not* be flagged as honest if truthful behavior seems to be motivated solely by payoff-seeking, such as if a sender said something along the lines of "They'll think we're lying, so tell the truth instead."

- The "payoff-concern" variable indicates any direct concern for payoffs, including when subjects expressed a desire for "fair" or "equal" payoffs between senders and receivers. It does not require a demonstration of self-interested payoff maximization, only that the subject indicated that concern for *some* monetary payoffs motivated their behavior. Subjects are given a 1 for this variable if their transcript indicates that they were motivated by concern for payoffs, even if they did not explicitly say so. For example, subjects who indicated that a particular message should be sent to "get the other team to do what we want" would be classified as payoff-concern, even though payoffs were not explicitly mentioned. Likewise, a sender who wrote something along the lines of "Let's help them earn more" is also classified as demonstrating payoff-concern.

- The "confusion" variable takes the value 1 if a transcript indicates the subject fundamentally misunderstood some aspect of the game. For example, a sender would be given a 1 for confusion if they claimed that the action X always gives receivers a higher payoff, regardless of what the actual color is.

- The "image-concern" variable takes the value 1 if a transcript indicates any type of concern for what others might think of the subject as a result of the choices made. For example, a subject would demonstrate image-concern if they said something along the lines of "We have to tell the truth, otherwise they'll think we're liars" or "I don't want the people running this experiment to think we're dishonest."

Once again, please indicate how confident you are in your assessment of each motive, using the same 0 to 3 scale as above.

Classification of Receivers

After observing the message sent by the senders, receivers were asked to answer four survey questions. Each receiver's survey questions were shared with their teammate prior to the receivers making their final choice of action.

The four questions answered by receivers were:

1. Given the message sent by Team A, what action do you think they will choose?

2. If Team A intended to choose a different action from what you think they will, what message do you think they would send to you?
3. Given the message sent by Team A, what action do you think your team should choose?
4. Do you have any other comment or opinion that you would like to share with your teammate?

When classifying Task II receiver transcripts, you will look for the same set of five motives as you did for receivers in Task I: taste-for-trust, payoff-concern, confusion, image-concern, and credulity.

- The "taste-for-trust" variable indicates that a receiver finds it desirable to trust the message to be true *for its own sake*. In other words, the receiver indicated that the act of trusting the sender to tell the truth is intrinsically desirable, e.g., "my mom told me to always trust people." Note that this motivation is about it being desirable to trust the sender, separate from actually believing the sender's message to be true, which would be about a different motive—credulity—to be discussed below.
- The "payoff-concern" and "confusion" variables are defined the same as they were for senders. If a receiver indicates concern for *any* subject's payoffs, the transcript receives a value of 1 for the payoff-concern variable. If a receiver indicates any misunderstanding of the rules of the game, the transcript receives a 1 for the confusion variable.
- The "image-concern" variable indicates that a receiver showed any type of concern for what others might think of the subject as a result of the choices made, e.g., "I don't want the senders to think we don't trust them," or "I like other people to know I'm trustworthy."
- The "credulity" variable indicates the receiver believed the message to be literally true, regardless of the sender's motives. This does not require the receiver to indicate that trust or trustworthiness is in any way desirable. They simply need to express that they believe the message to be true. For example, "I think they're telling the truth that it's Blue" or "They said Blue to make us think it's Green, but I bet it really is Blue" would both demonstrate that the receiver was credulous.

Once again, please indicate how confident you are in your assessment of each motive, using the same 0 to 3 scale as above.

Clarifying questions

As a reminder, please let us know if you have questions about the details of the experiment or your task as a classifier. We will gladly answer any clarifying questions; however, we cannot give you guidance as to how to classify any specific transcript. Barring any questions you may have, you are now ready to begin classifying the transcripts.