Strategic Communication Under Perceived Misalignment of Interests

Alexandra Ballyk*

October 31, 2025

Latest version available here.

Abstract

I experimentally study why people fail to communicate strategically when they know their incentives align with a decision-maker's, but the decision-maker is wary of this alignment. I propose a game that characterizes a Sender's optimal communication strategy in this setting. I find that most Senders communicate sub-optimally. Many use a strategy that would only be optimal if Receivers were not wary. Senders' mistake stems primarily from a failure of contingent thinking, rather than incorrect beliefs or lying costs. Prompting Senders to consider how Receivers respond mitigates the failure of contingent thinking, which sheds light on the source of such failures.

JEL Classification Numbers: C90, D82, D83, D91.
Keywords: Strategic Communication, Failure of Contingent Thinking, Laboratory Experiment.

*Department of Economics, New York University (email: alexandra.ballyk@nyu.edu). I am thankful to my supervisor Yoram Halevy for his support and guidance throughout all stages of this project. I also wish to recognize my committee members Tanjim Hossain and Anne-Katrin Roesler, who provided invaluable feedback on the experimental design and theoretical framework. This paper has benefited from helpful suggestions from Billur Görgülü, En Hua Hu, Stanton Hudja, Manu Vespa, Ryan Webb, Georg Weizsäcker and Sevgi Yuksel, as well as seminar participants at the following universities, conferences and seminar series: University of Toronto, McMaster University, CIREQ Ph.D. Students' Conference, ESA North American Meetings, SITE Experimental Economics and the VIEE Seminar Series. Xiaoya Gao and Shanshan Liu provided outstanding research assistance. Financial support from the University of Toronto's economics department is gratefully acknowledged. The experiment was approved by the University of Toronto's Research Ethics Board (Protocol #45844), and pre-registered on AsPredicted.org (Main treatment: #164964, Beliefs treatment: #171009). An earlier version of this paper was circulated under the title "Paternalistic Persuasion".

1 Introduction

Advisors often seek to help decision-makers achieve their goals. For instance, financial advisors aim to help clients save for large purchases, and physicians strive to help patients adopt healthier lifestyles. While these advisors know their interests align with decision-makers', decision-makers are often unsure of this. Financial advisors, in particular, are widely viewed as untrustworthy (Egan et al., 2024), and decision-makers in both financial and medical settings frequently reject advisors' recommendations, especially when they do not fully trust the advisor (Bhattacharya et al., 2012; Cole et al., 2013; Alsan and Eichmeyer, 2024).

To persuade a decision-maker to change their behavior, advisors must therefore communicate strategically: they must offer recommendations the decision-maker trusts, and is therefore willing to follow. There is substantial evidence that people fail to communicate strategically to decision-makers whose preferences differ from their own (Blume et al., 2020; Lafky et al., 2022; Weizsäcker, 2023). This paper explores the possibility that people make similar mistakes when they know their interests align with decision-makers', but decision-makers do not. In doing so, it seeks to provide theoretical insights on the source of these failures, and practical insights on how to help advisors offer more effective recommendations.

To better understand communication failures that arise in this setting, I propose a model of a *recommendation game*. The game features an interaction between an aligned-type Sender and a Receiver: while this Sender's incentives align with the Receiver's, the Receiver is unsure of this alignment. I show that the aligned-type Sender's *optimal recommendations* maximize the Receiver's utility, subject to the Receiver being willing to follow them. I then use a laboratory experiment to evaluate why people acting as aligned-type Senders may fail to send optimal recommendations. This controlled environment allows me to cleanly identify the source of communication failures, while ruling out confounds related to uncertainty about Receivers' incentives, available actions and information.

The recommendation game is a variant of the canonical strategic communication model (Crawford and Sobel, 1982), capturing the idea that decision-makers may doubt an advisor's expertise or suspect ulterior motives. The game features three equidistant states, an action corresponding to each state, and two players: a Sender and a Receiver. Action 1, which the Receiver thinks is likely to correspond to the true state, is the Receiver's *status-quo*. The Receiver aims minimize the distance between the action they take and the action corresponding to the true state. One of two Sender types, who differ in the degree to which their preferences align with the Receiver's, communicates with the Receiver. The *aligned type* knows the state and has the same preferences as the Receiver. The *commitment type* always recommends action 3, which the Receiver believes is unlikely to correspond to the true state. To give a practical example of these Sender types, the aligned type may capture a highly-trained financial advisor with a fiduciary duty to their clients, while the commitment type may capture a financial advisor with commission incentives, or one who does not tailor their advice to each client. Upon receiving a recommendation that differs from their status-quo, the Receiver must decide between accepting that recommendation or sticking with their status-quo.

The recommendation game has up to two equilibria. If the Receiver's expected payoff from action 3 is high enough, the game has a *trusting equilibrium* where the Receiver accepts all recommendations (the *accept all* strategy), and the aligned-type Sender always recommends the action corresponding to the true state (the *full truth-telling* strategy). Otherwise, the game only has a *wary equilibrium*, where the Receiver accepts a recommendation to take action 2, but rejects a recommendation to take action 3 (the *accept-reject* strategy). The aligned type's best response is to recommend action 2 when action 3 corresponds to the true state, and otherwise to recommend the action corresponding to the true state (the *partial truth-telling strategy*).

In the experiment, participants play the recommendation game as either a Receiver or an aligned-type Sender. The commitment type is played by a "computer" that always recommends action 3. When presenting the experiment and results, I refer to participants in the aligned-type Sender role as "Senders". Since the experiment's main purpose is to evaluate whether Senders account for Receivers' reluctance to accept recommendations they do not trust, I parameterize the recommendation game such that only a wary equilibrium exists, ensuring Receivers are expected to show such reluctance. As a result, Senders' optimal recommendations in the experiment are captured by their partial truth-telling strategy.

In the *main treatment*, participants play twenty rounds of the recommendation game, and receive feedback about their co-player's strategy after each round. Importantly, this feedback reveals to Senders how a Receiver would have responded to each recommendation that involves switching from their status-quo. I begin the analysis of the main treatment's results by focusing on Receivers' strategies. In each round, most Receivers play their wary equilibrium strategy, and the distribution of Receivers' strategies ensures that Senders' partial truth-telling strategy yields a higher expected payoff than their full truth-telling strategy. Having confirmed that partial truth-telling is empirically optimal, and since my research objectives relate to Senders' behavior, I focus primarily on Senders' strategies in the remaining analyses.

Unlike Receivers, many Senders do not play their wary equilibrium strategy. In the first round, only 21% of Senders send optimal recommendations. Instead, nearly two-thirds (63%) play full truth-telling. That is, most Senders send recommendations that would be optimal if Receivers accepted all recommendations. One possible explanation may be that, while Senders are aiming to optimize their recommendations given the Receiver's strategy, they incorrectly believe that Receivers will accept all recommendations. This would be consistent with level-k thinking (see, e.g., Nagel (1995)): Senders may be level-1 thinkers, best-responding to the belief that Receivers are naïve, level-0 thinkers who accept all recommendations. If so, the feedback about the Receiver's strategy should enable Senders to update their beliefs about Receivers' level of sophistication, and respond accordingly. However, feedback has only a modest effect on Senders' strategies: by the twentieth and final round, only 45% of Senders send optimal recommendations. The modal Sender never plays partial truth-telling, a pattern that persists in later rounds and among Senders receiving more feedback that Receivers play accept-reject.

Alternatively, Senders may exhibit a failure of contingent thinking, which captures a failure to focus on relevant contingencies (e.g., states, strategies of other players) when making a decision (Niederle and Vespa, 2023). In the recommendation game, the Receiver strategy that is relevant to the Sender is their accept-reject strategy. Senders may fail to think contingently by failing to identify their best response to that Receiver strategy contingency. To test for this failure, after the final round, Senders play an additional round where they can condition their recommendations on the Receiver's strategy. When responding to Receivers' accept-reject strategy, Senders are placed in the relevant contingency before they send their recommendations.[1,2] I find that Senders are much more likely to send optimal recommendations when the game is presented in this way. While 45% play partial truth-telling in the final round, 78% do so in response to the Receiver's accept-reject strategy in the additional round. This rules out the possibility that Senders' initial failure was due to an inability to identify the best response to Receiver's accept-reject strategy, or driven by a preference for truth-telling. Moreover, it provides support for the failure of contingent thinking mechanism.

I run a second treatment, the *beliefs treatment*, as a second test of the failure of contingent thinking mechanism. While the additional round of the main treatment placed Senders in the relevant contingency, the beliefs treatment implements a weaker intervention: prompting Senders to think about the relevant contingency. I do so by eliciting Senders' beliefs about the modal Receiver strategy in the upcoming round.[3] If the failure to send optimal recommendations is driven by a failure of contingent thinking, Senders with correct beliefs about the modal Receiver strategy will focus on the relevant contingency: Receivers' accept-reject strategy. This in turn should make them more likely to send optimal recommendations.

The findings from the beliefs treatment provide further support for the failure of contingent thinking mechanism. Most Senders expect Receivers to play their accept-reject strategy. This confirms that the belief elicitation task draws Senders' attention to the relevant contingency, and that Senders' error in the main treatment was not due to a misprediction of Receiver's strategy. Consistent with a failure of contingent thinking, Senders are much more likely to send optimal recommendations in the beliefs treatment. In each round, the proportion of Senders sending

[1] Esponda and Vespa (2014, 2024) show that, when a problem is framed in such a way that places decision-makers in a relevant contingency, they are much less likely to display a failure of contingent thinking.

[2] Notice that, while the feedback about the Receiver's strategy is provided *ex-post* (i.e., after they send their recommendations), in the additional round, Senders are placed in the relevant contingency *ex-ante* (i.e., before they send their recommendations).

[3] The effect of eliciting players' beliefs about their opponents' strategies is unclear. While some studies find that belief elicitation changes players' strategies (Erev et al., 1993; Croson, 1999, 2000; Rutström and Wilcox, 2009; Gächter and Renner, 2010), others find that it has no effect (Nyarko and Schotter, 2002; Costa-Gomes and Weizsäcker, 2008).

optimal recommendations is 36%-84% higher than in the corresponding round of the main treatment. I also find that Senders learn to send optimal recommendations more quickly, and that Senders with correct beliefs about the modal Receiver strategy - who thus focus on the relevant contingency - are more likely to send optimal recommendations.

In summary, Senders play a naïve strategy when they do not focus on Receivers' behavior, and the optimal strategy when they do. These findings have implications for the cognitive underpinnings of failures of contingent thinking and strategic reasoning errors more broadly. Prior work has shown that reframing problems to emphasize the states where actions matter reduces failures of contingent thinking (Esponda and Vespa, 2014; Niederle and Vespa, 2023; Esponda and Vespa, 2024). I find that an even weaker intervention, simply prompting decision-makers to think about the relevant contingencies, can have a similar effect, suggesting that fully reframing the decision environment may not be necessary. This result implies that failures of contingent thinking may stem from a form of salience bias (Bordalo et al., 2012, 2013, 2020, 2022), in which the relevant contingency is not salient to the decision-maker. Moreover, Senders' failure does not stem from either of two well-studied types of strategic reasoning errors (Eyster, 2019): misprediction or a failure to best respond. Instead, the error appears to be driven by the lack of salience of a relevant strategic contingency. Future work may examine whether this error accounts for deviations from theoretical predictions in other games.

This paper builds upon well-established literature on strategic communication and failures of contingent thinking, and contributes to a growing body of work on paternalistic behavior. Research on strategic communication has yielded important insights into how people should communicate strategically, and whether they actually do so (see reviews by Blume et al. (2020) and Jin et al. (2022)). My recommendation game is similar to a "noisy" cheap talk model.[4] In this type of strategic communication game, the Receiver receives a noisy signal of the Sender's message, creating uncertainty about the content of the original message (Myerson, 1991; Blume

[4]These models are part of a class of mediated communication mechanisms, in which players send messages to a communication device (often described as a neutral mediator) that then uses those messages to send signals to each player. These mechanisms can improve information transmission (Forges, 1985, 1986; Myerson, 1986; Goltsman et al., 2009), and, in some cases, can be implemented without a communication device (Krishna, 2007).

et al., 2007; Lightle, 2014; Blume et al., 2023). In both noisy cheap talk and the recommendation game, the Sender's optimal strategy accounts for the Receiver's uncertainty. In the recommendation game, however, this uncertainty stems from the two possible Sender types rather than noisy communication by a single type of Sender.[5] This setup aligns the theory more closely with the empirical context that motivates this study, and facilitates extensions to settings where Senders whose preferences are misaligned with Receivers' are strategic (see Section 5).

Experimental studies of strategic communication typically examine interactions between an uninformed Receiver and a single type of informed Sender, whose incentives are imperfectly aligned with the Receiver's. Senders in these games tend to "over-communicate" by revealing more information than theory predicts (Cai and Wang, 2006; Wang et al., 2010; Lafky et al., 2022). Prior work has compared information transmission in settings with different degrees of preference alignment, where the degree of alignment is common knowledge (Dickhaut et al., 1995; Blume et al., 1998, 2001; Kawagoe and Takizawa, 2009; Altmann et al., 2025). To my knowledge, this paper provides the first analysis of a game where preferences are fully aligned, but strategic tension arises because one player is unsure of this alignment. I show that Senders over-communicate even in this setting, and identify the source of this error. In doing so, I also contribute to the literature on failures of contingent thinking (see Niederle and Vespa (2023) for a review). Such failures have been documented in other strategic environments, including auctions (Charness and Levin, 2009; Martínez-Marquina et al., 2019), voting (Esponda and Vespa, 2014, 2024; Ali et al., 2021), public-good provision (Calford and Cason, 2024) and markets (Ngangoué and Weizsäcker, 2021). I extend this work by identifying a failure of contingent thinking in a communication game, and shedding light on the mechanism underlying this failure. Future work may explore whether such failures explain deviations from theoretical predictions in other forms of communication, such as delegation (Holmström, 1978; Amador and Bagwell, 2013), disclosure (Milgrom, 1981) and Bayesian persuasion (Kamenica and Gentzkow, 2011).

[5] Li and Madarász (2008) also propose a cheap talk model with two possible Sender types. In their model, however, both types are biased: each has preferences that differ from the Receiver's. The authors study the effect of disclosing the Sender's bias, and find that nondisclosure often results in higher welfare for both parties.

Finally, this study complements research on paternalistic behavior, which examines when and how people intervene to improve others' choices. Much of this work focuses on the use of "hard" interventions that restrict decision-makers' choice sets. Like those studies, I study a situation where Senders attempt to prevent Receivers from taking a suboptimal action: the Receiver's status-quo. However, Senders are limited to a "soft" intervention that preserves Receivers' freedom of choice: a recommendation.[6] Prior work has shown that, while soft interventions are generally preferred (Bartling et al., 2025), hard interventions are used to align decision-makers' choices with one's own aspirations (Ambuehl et al., 2021). Unlike a hard intervention, a soft intervention is only effective if the paternalist strategically anticipates the decision-maker's response to it. I show that people acting paternalistically struggle to do so, determine the source of this error, and identify an intervention that mitigates it.

The rest of the paper proceeds as follows. In Section 2, I develop a model that characterizes players' optimal strategies in the recommendation game. Sections 3 and 4 present the design and results of the main and beliefs treatments, respectively. Section 5 concludes with a discussion of the experimental and theoretical results.

2 Theoretical Framework

To formalize the idea that well-meaning advisors should send recommendations strategically, I develop a model of a recommendation game. The model is prescriptive: its purpose is to outline the strategies that players should play, which may not be those that they actually play in the experiment. The recommendation game is a variant of the classic strategic communication model (Crawford and Sobel, 1982), which features a single type of Sender whose preferences are known to the Receiver. However, decision-makers are often unsure whether an advisor's interests are aligned with theirs. The recommendation game incorporates this uncertainty by introducing two types of Senders, where a Sender's type captures the alignment of

[6] Soft interventions are often the only type of intervention that a paternalist has at their disposal. For instance, an employer may be able to set a default or recommended retirement savings contribution rate, but may not be able to enforce that rate. Furthermore, prior work on the value of decision rights suggests that decision-makers prefer soft interventions to hard ones (Fehr et al., 2013; Bartling et al., 2014).

their preferences with the Receiver's. Furthermore, advisors often seek to persuade decision-makers to change their behavior. As a result, they often search for, or provide an alternative to, an action the decision-maker is already taking. Thus, while Receivers in the classic model are not biased towards any action, Receivers in the recommendation game must decide whether to retain a status-quo action.

2.1 The Recommendation Game

There are three states, $\theta \in \{1,2,3\}$, each with a corresponding action $a \in \{1,2,3\}$. Let $\hat{a}(\theta)$ denote the action corresponding to the true state. There are two players, a *Sender* and a *Receiver*, who have different information about the true state. The Receiver's prior belief assigns probability $q_\theta > 0$ to state θ, where $q_1 \geq q_3$ and $q_1 + q_2 + q_3 = 1$.

The Sender may be one of two types. The information the Sender has, as well as their preferences, depends on their type. With probability p, the Sender is an *aligned type*: they know the true state, and their utility function is the same as the Receiver's. With probability $1-p$, the Sender is a non-strategic *commitment type*: they always recommend action 3. p and q are independent, and are known to all players. The aligned type represents an expert advisor who has a decision-maker's best interest at heart, such as highly-trained financial advisors who have a fiduciary duty to their clients. The commitment type's behavior has two possible interpretations. First, it may represent a competent advisor whose incentives are not fully aligned with the decision-maker's. An example may be a financial advisor employed by a commercial bank, who is pressured to sell certain financial products to meet the bank's sales goals.[7] Second, the behavior may represent an advisor whose interests are aligned with the decision-maker's, but assigns the highest likelihood to a particular state (here, state 3) before the beginning of the game, and must exert effort to learn the true state. If such an advisor has not exerted the effort required to learn the true state, they will recommend the action corresponding to the most-likely state. For instance, a well-meaning financial advisor may always recommend an

[7]Of the two possible interpretations of the commitment type's behavior, this interpretation is most closely aligned with the version of the model that is implemented in the experiment.

investment strategy that works well for their average client, instead of determining a strategy that is well-suited to each client's goals.

The Receiver's goal is to minimize the distance between the action they take and the action corresponding to the true state. Their preferences are described by the utility function $u(\alpha|a_R(r) - \hat{a}(\theta)|)$, where $\alpha > 0$ and $u'(\cdot) < 0$. An aligned-type Sender has the same preferences as the Receiver. When an aligned-type Sender is indifferent between recommendations, I assume that they recommend the action closest to $\hat{a}(\theta)$. This assumption effectively states that aligned-type Senders prefer to tell the truth when indifferent between recommendations, which is in line with evidence on the psychological cost of lying (Abeler et al., 2019).

The game proceeds in three stages. First, Nature determines the state and the Sender's type. Second, the Sender recommends an action $r(\theta)$ to the Receiver. If the Sender is a commitment type, they recommend $r(\theta) = 3$ for all θ. If the Sender is an aligned type, they observe the realized state and can recommend any action; that is, $r(\theta) \in \{1, 2, 3\}$. Third, the Receiver observes the recommendation r, and either accepts it or rejects it. Accepting entails taking the recommended action, while rejecting entails taking action 1; that is, $a_R(r) \in \{1, r\}$.[8] I call action 1 the Receiver's *status-quo*,[9] which can be interpreted as an action the Receiver is already taking, or one that they have independently determined to be most likely to correspond to the true state.[10] The Receiver's action space captures situations where the Sender makes actions available to the Receiver, either by providing the opportunity to take the action or by alleviating search costs. For instance, a client may hire a financial advisor to alleviate the costs of determining an ideal investment portfolio, and the advisor may allow their client to purchase specialized or customized financial products.

[8] If the Receiver is recommended to take action 1, their action set is the singleton set $\{1\}$. It is therefore without loss to restrict attention to the Receiver's responses to recommendations of 2 and 3.

[9] The fact that action 1 is the Receiver's status-quo is a behavioral assumption. That is, under the assumed structure of the Receiver's prior ($q_1 \geq q_3$), action 1 is not necessarily the Receiver's optimal action based on their prior. In Lemma 1 (see Online Appendix A.2), I show that action 1 is the Receiver's ex ante optimal action for priors satisfying the slightly stronger condition $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$.

[10] The structure of the Receiver's prior captures a form of motivated reasoning (Kunda, 1990; Epley and Gilovich, 2016) where the Receiver believes that actions far from their status-quo are weakly less likely to be in their best interest.

2.2 Equilibria

Depending on the values of certain parameters, the recommendation game has up to two pure-strategy[11] weak perfect Bayesian equilibria. Propositions 1 and 2 describe the equilibria, and Proposition 3 shows that no other equilibria exist. The proofs of these propositions are provided in Online Appendix A.1. In Section 5, I discuss variations of the model that relax assumptions on the Receiver's action space, distance between states, and the non-strategic nature of the commitment type. In Online Appendix A.2, I show that my main theoretical results are robust to those variations, sometimes under mild conditions.

It is natural to think that the recommendation game may have an equilibrium where the aligned-type Sender always recommends the action corresponding to the true state. Proposition 1 shows that such an equilibrium can exist, but only for certain parameter values.

PROPOSITION 1: *When $q_3 \geq q_1(1-p)$, the recommendation game has a "trusting" equilibrium where...*

- *The aligned-type Sender plays the "full truth-telling" strategy $r(\theta) = \theta$ for all $\theta \in \{1, 2, 3\}$. That is, they always recommend the action corresponding to the true state.*
- *The Receiver plays the "accept all" strategy $a_R(r) = r$ for all $r \in \{2, 3\}$. That is, they accept all recommendations.*

Intuitively, when the Receiver accepts all recommendations, it is optimal for the aligned-type Sender to recommend the action corresponding to the true state. However, the Receiver will only want to accept a recommendation of 3 if their expected payoff from doing so is sufficiently high. This occurs when the probability of state 3 occurring is at least as high as the Receiver's likelihood of encountering the commitment-type Sender in state 1.

[11] For the purpose of specifying testable hypotheses for the experiments, I restrict my attention to pure-strategy equilibria. Experimental participants do not often use mixed strategies. For instance, in Prisoner's Dilemma experiments, most participants play one of several pure strategies (Dal Bó and Fréchette, 2018), and many participants play pure strategies even when they have the opportunity to explicitly randomize between actions (Romero and Rosokha, 2023). In Online Appendix C.2, I provide evidence that very few aligned-type Senders mix between strategies in the experiment.

It may be surprising that a trusting equilibrium does not always exist. That said, the recommendation game does have an equilibrium that exists for all parameter values, which is described in Proposition 2.

PROPOSITION 2: *For all parameter values, the recommendation game has a "wary" equilibrium where...*

- *The aligned-type Sender plays the "partial truth-telling" strategy* $r(1) = 1,\ r(2) = r(3) = 2$*. That is, they recommend 2 when the state is 3, and otherwise recommend the action corresponding to the true state.*
- *The Receiver plays the "accept-reject" strategy* $a_R(2) = 2,\ a_R(3) = 1$*. That is, they accept a recommendation of 2 and reject a recommendation of 3.*

Proposition 2 outlines the aligned-type Sender's optimal strategy when Receivers reject a recommendation that they are wary of. Intuitively, since the Receiver accepts a recommendation of 2 but rejects a recommendation of 3, it is optimal for the aligned-type Sender to recommend action 2 when the state is 3. Proposition 3 shows that the equilibria described in Propositions 1 and 2 are the unique[12] pure strategy equilibria of the recommendation game.

PROPOSITION 3: *The trusting equilibrium and the wary equilibrium are the unique pure-strategy weak perfect Bayesian equilibria of the recommendation game.*

3 Main Treatment

My main objective is to determine whether, in situations where decision-makers are wary of advisors' incentives, well-meaning advisors account for this wariness when issuing recom-

[12]Unlike the classic strategic communication model (Crawford and Sobel, 1982), the recommendation game does not have a babbling equilibrium: an equilibrium where the Sender's messages are completely uninformative, and the Receiver always takes the action that is optimal given their prior beliefs. As shown in the proof of Proposition 3, the lack of a babbling equilibrium is due to the assumption that the aligned-type Sender recommends the action closest to $\hat{a}(\theta)$ when indifferent between recommending two or more actions.

mendations. The main experiment thus features a version of the recommendation game where aligned-type Senders are required to account for Receivers' wariness: one where only a wary equilibrium exists. To make the recommendation game easier to understand, I frame it as a "Plant Pot" Game: Receivers' goal is to produce the largest possible plant by turning on a sprinkler (i.e., taking an action) that is as close as possible to a pot with a plant seed in it (i.e., the true state). They do so by deciding whether to accept a recommendation regarding which sprinkler to turn on. The recommendation to turn on sprinkler 3 may either be from the computer (the commitment-type Sender) or a Sender (the aligned-type Sender).[13]

My secondary research objective is to shed light on why advisors may fail to account for decision-makers' wariness. The main treatment assesses the explanatory power of two key mechanisms. To investigate whether Senders must update incorrect beliefs about Receivers' strategies, participants play multiple rounds of the Plant Pot Game, and receive feedback about their co-player's strategy at the end of each round. To investigate a possible failure of contingent thinking, Senders play a round of the Plant Pot Game where they can condition their strategy on the Receiver's. This intervention, which effectively reverses the sequence of the game, is akin to an intervention that has been shown to mitigate failures of contingent thinking in committee voting problems (Esponda and Vespa, 2014).

Section 3.1 describes the design of the main treatment. Its results are presented and discussed in Section 3.2.

3.1 Experimental Design

The Plant Pot Game

The Plant Pot Game is the core task in the main treatment. The game features three pots arranged in a line. One of the pots - the "plant pot" - has a plant seed in it, whereas the other two do not. Each pot has a sprinkler above it. When a sprinkler is turned on, it sprays water directly below it, as well as below to the right and left of it. The pot closest to it receives the

[13] In the experiment, Receivers were called "Choosers", and Senders were called "Advisors".

most water, while the pot farthest from it receives the least. As illustrated in Figure 1, a seed will grow into a larger plant if it receives more water.

[Figure 1]

Participants playing the role of Receiver are incentivized to produce the largest possible plant: they receive \$16 if they produce a large plant, \$13 if they produce a small plant, and \$10 if they do not produce a plant. Receivers know that pot 1 has a 9-in-20 chance of being the plant pot, pot 2 has a 7-in-20 chance and pot 3 has a 4-in-20 chance.

While sprinkler 1 will be turned on by default, Receivers may receive a recommendation to switch to one of the other sprinklers. If recommended to switch from the default, a Receiver must decide between accepting or rejecting the proposed switch, where rejecting entails sticking with the default sprinkler. If they are recommended to keep the default, they automatically do so.

Receivers know that they have a 1-in-3 chance of receiving a Sender's recommendation, and a 2-in-3 chance of receiving the computer's recommendation. Senders are participants who have the same incentives as Receivers[14] and know which pot is the plant pot. Senders are also aware of all the information Receivers have when deciding whether to accept a recommendation. The computer recommends for the Receiver to switch to sprinkler 3 regardless of which pot is the plant pot.

In summary, the Plant Pot Game's structure and incentives are identical to those of the recommendation game, with a linear payoff function for the Receiver and Sender. It also features parameters for which only a wary equilibrium exists. As a result, it is optimal for Receivers to accept a recommendation to switch to sprinkler 2 and reject a recommendation to switch to sprinkler 3. Senders should thus recommend the sprinkler corresponding to the plant pot when pots 1 and 2 are the plant pot, but recommend sprinkler 2 when pot 3 is the plant pot.

[14] If a Receiver received the computer's recommendation, the Sender they were paired with received \$10 regardless of the size of the plant the Receiver produced.

Structure of the Main Treatment

Experimental sessions were conducted with even numbers of participants. At the beginning of a session, half of the participants were randomly assigned to each role (Sender or Receiver). Participants retained their role for the entire experiment.

[Figure 2]

Figure 2 summarizes the structure of the main treatment. After reading the instructions, participants completed a comprehension quiz. They were required to complete all quiz questions correctly before proceeding to the first round. As an incentive to read the instructions carefully, participants received a $1 bonus payment if they answered all questions correctly on their first attempt.

Next, participants played 22 rounds of the Plant Pot Game. The first round was an unpaid practice round, and the rest were "main" rounds that were eligible to count for payment. At the beginning of each round, Senders and Receivers were randomly re-matched, and the plant pot and source of the recommendation shown to the Receiver were determined according to a draw from the relevant probability distribution. In all rounds, participants' strategies were elicited using the strategy method. Senders specified the sprinkler they would recommend for each possible plant pot (i.e., pot 1, 2 and 3), and Receivers specified whether they would accept or reject each recommendation that involved switching from the default sprinkler (i.e., recommendations to switch to sprinkler 2 and to switch to sprinkler 3). At the end of each round, participants received feedback on the outcome of the round and their co-player's strategy. Thus, a Sender learned how the Receiver they were paired with would have responded to all recommendations that involved switching sprinklers, regardless of which recommendation that Receiver actually received.[15]

[15] I use the strategy method to elicit participants' strategies because, unlike the direct response method, it allows me to give Senders feedback about Receivers' full strategies. This feedback is important for testing the learning mechanism: it ensures that, at the end of each round, all Senders have the opportunity to learn how Receivers respond to a recommendation to switch to sprinkler 3. Studies that have compared the strategy and direct response methods have shown that participants' responses do not tend to differ based on the method used (Brandts and Charness, 2011).

In the practice round and the first 20 main rounds, the game was played out in the sequence described in Section 2. That is, Senders sent their recommendations, then Receivers responded. In the final main round, the sequence was reversed: Receivers committed to a response to each possible recommendation, then Senders sent their recommendations. To implement this reversed sequence using the strategy method, I allowed Senders to specify the sprinkler they would recommend for each possible plant pot and each possible strategy the Receiver could play (i.e., accept all, accept-reject, reject-accept and reject all). I refer to the first 20 main rounds as Sender-first rounds, and the final main round as the Receiver-first round.

After completing all rounds of the Plant Pot Game, participants completed an end-of-experiment survey, which included questions about the rationale behind their strategies and a demographic questionnaire.

Procedures

Online Appendix B contains the main treatment's instructions. The treatment was pre-registered on AsPredicted.org, programmed and deployed in oTree (Chen et al., 2016), and conducted at the Toronto Experimental Economics Laboratory (TEEL). Participants were recruited from the University of Toronto student body using the ORSEE online recruitment system (Greiner, 2015). I conducted seven sessions, each with 16 to 20 participants, for a total of 136 participants (68 Senders, 68 Receivers).

Participants' total payment included a $5 show-up fee, their earnings from the comprehension quiz, and their earnings from one round of the game. All rounds except the practice round were eligible to count for payment. The round that counted for each participant was randomly selected at the beginning of the experiment and revealed to them at the end of the experiment. The average participant earned $18.69 and took 40 minutes to complete the experiment.

3.2 Results

I begin by establishing whether, given the empirical distribution of Receivers' strategies, it is indeed optimal for Senders to play their wary equilibrium strategy. Next, I investigate whether

Senders play their empirically-optimal strategy. Unless otherwise specified, all analyses focus exclusively on the 20 Sender-first rounds. I exclude one participant who withdrew from the experiment,[16] which yields a final sample of 135 participants (67 Senders, 68 Receivers).

Which Recommendations Are Optimal?

Given the parameters used in the experiment, it is optimal for Senders to play their partial truth-telling strategy in equilibrium; that is, to recommend the action corresponding to the state when the state is 1 or 2, and to recommend action 2 when the state is 3. This prediction relies on the assumption that a majority of Receivers play their wary equilibrium strategy, which is to accept a recommendation of 2 and reject a recommendation of 3. Before analyzing Senders' strategies, I therefore determine whether partial truth-telling is indeed a best-response to the empirical distribution of Receivers' strategies, and that Senders have the opportunity to learn this through the feedback they receive.

[Figure 3]

Consistent with the theoretical prediction, a clear majority of Receivers play their wary equilibrium strategy. Figure 3 displays the distribution of Receivers' strategies in each round. In the average round, 70% of Receivers play accept-reject, while at most 27% of Receivers play any other strategy in any given round. As a result, in every round, the modal Receiver plays accept-reject.

[Figure 4]

Partial truth-telling is indeed the best response to this empirical distribution of Receivers' strategies. Since Senders' equilibrium strategies differ only in the action recommended in state 3, Figure 4 compares the expected state-3 payoff of these strategies,[17] given the empirical distribution of Receivers' strategies. In all rounds, partial truth-telling yields a higher expected

[16]This participant was playing the role of Sender, and withdrew during the survey.

[17]Let $P_{a_R(2)a_R(3)}$ be the percent of Receivers who play strategy $(a_R(2), a_R(3))$. The expected state-3 payoff to full truth-telling is $(P_{21} + P_{11}) \cdot (M - 2d) + (P_{23} + P_{13}) \cdot M$. The expected state-3 payoff to partial truth-telling is $(P_{21} + P_{23}) \cdot (M - d) + (P_{11} + P_{13}) \cdot (M - 2d)$.

payoff than full truth-telling: the index $\frac{\text{Exp. payoff of full truth-telling}}{\text{Exp. payoff of partial truth-telling}}$ is always below 1. Across all rounds, the expected payoff of partial truth-telling is up to 15% higher, with an average payoff difference of 11%.[18]

RESULT 1: *Partial truth-telling is Senders' empirical best response.*

[Figure 5]

Furthermore, the fact that Receivers consistently play accept-reject is reflected in the feedback Senders receive. Figure 5 displays, for rounds 2 through 20, the percent of previous rounds in which the average Sender received the feedback that they were paired with a Receiver who played accept-reject. The average Sender has received this feedback in 66% to 76% of prior rounds. The average frequency across all rounds is 69%, which is only one percentage point lower than the true average frequency with which Receivers play accept-reject.

RESULT 2: *Senders receive feedback that reflects that Receivers consistently play accept-reject.*

Result 1 confirms that the optimal recommendations are those specified by Senders' wary equilibrium strategy, and Result 2 indicates that Senders have the opportunity to learn this. Next, I investigate whether Senders send their optimal recommendations.

Do Senders Send Optimal Recommendations?

[Figure 6]

Figure 6 displays the frequency with which participants in each role play their wary equilibrium strategies. Figure 6a defines frequency at the round-level, while Figure 6b defines

[18]The maximum and average percent payoff differences correspond to \$1.90 and \$1.41 absolute payoff differences, respectively. To put these absolute differences into perspective, recall that participants received a \$1 bonus payment for answering all comprehension questions correctly on their first attempt. Bonus payments of this magnitude are common in lab experiments, and tend to successfully encourage participants to read instructions carefully. This suggests that payoff differences that were up to 90% larger than the bonus payment would have been meaningful to participants.

frequency at the player-level. Regardless of how frequency is defined, it is clear that only a minority of Senders send optimal recommendations.

As shown in Figure 6a, in any given round, less than half of Senders play their partial truth-telling strategy. The proportion doing so begins at 21% in the first round, reaches a maximum of 46% in round 18, and settles at 45% in the final round. As a result, the proportions of Senders and Receivers playing their wary equilibrium strategies are significantly different in all rounds (proportions test, $p < 0.01$ in all rounds).

Similar results emerge when comparing the distributions of wary equilibrium strategy play at the player level. These distributions differ significantly by role (Wilcoxon rank-sum test, $p < 0.01$), with Receivers tending to play optimally in a larger number of rounds. As shown in Figure 6b, the median Receiver plays accept-reject in 15 rounds, and the modal Receivers - who make up 25% of all Receivers - do so in all rounds. In contrast, the median Sender plays partial truth-telling in 2 rounds, and the modal Senders - who make up 37% of all Senders - do so in none of the rounds. Finally, whereas 63% of Senders play optimally in fewer than half of all rounds, only 23% of Senders do so. Online Appendix Figures C.1 and C.2 repeat the analysis from Figure 6b, but focus on Senders' behavior in the second half of the rounds (in Figure C.1), and Senders who are paired with an above-median number of Receivers who played accept-reject (in Figure C.2). The Figures show that, even with extensive experience and feedback, the modal Sender fails to play optimally in all rounds.

[Figure 7]

Having established that Senders do not often send optimal recommendations, I now examine which recommendations they do send. Figure 7, which displays the distribution of Senders' strategies in each round, shows that full truth-telling is consistently the most popular strategy. It is the modal Sender strategy in 19 out of the 20 rounds, and 40% to 63% of Senders play it in any given round.

RESULT 3: *A minority of Senders play partial truth-telling in any given round. Full truth-telling is* *the modal Sender strategy in almost all rounds.*

While full truth-telling is almost always the modal Sender strategy, partial truth-telling becomes more common as rounds progress. The proportion of Senders playing partial truth-telling increases significantly between the first and last rounds (proportions test, $p < 0.01$), while the proportion playing full truth-telling significantly decreases (proportions test, $p = 0.06$).

[Table 1]

Table 1 reports the results of regressions that further explore the evolution of Senders' strategies. As seen in column 1, playing an additional round of the game increases a Sender's likelihood of playing partial truth-telling by 0.9 percentage points. The results from columns 2 and 3 reveal that this effect is larger in early rounds. In rounds 1 through 10 (column 2), playing an additional round increases a Sender's likelihood of playing partial truth-telling by 1.3 percentage points. However, in rounds 11 through 20 (column 3), an additional round only has a 0.9 percentage point effect.

RESULT 4: *Senders become slightly more likely to play partial truth-telling as rounds progress, especially in early rounds.*

Result 4 suggests that receiving feedback about Receivers' strategies helps Senders learn to play partial truth-telling. However, the learning effect is weak: by the final round, more than half of Senders still fail to send optimal recommendations. The Receiver-first round is designed to help Advisors do so by allowing them to directly respond to Receivers' wary equilibrium strategy. Next, I examine how this reversal of the sequence of the game impacts Senders' likelihood of sending optimal recommendations.

[Table 2]

Table 2 compares Senders' strategies in the final Sender-first round to their responses to accept-reject in the Receiver-first round. Recall that the Receiver-first round occurs immediately after the final Sender-first round, so Senders do not gain more experience with the game between these two rounds. Nevertheless, Senders are much more likely to send optimal recommendations when directly responding to Receivers' wary equilibrium strategy. As shown in the final row of Table 2, nearly 78% of Senders play partial truth-telling in the latter scenario. This proportion is significantly larger than the 45% of Senders who send optimal recommendations in the final Sender-first round (proportions test, $p < 0.01$). Furthermore, this change is largely driven by the Senders who do not play optimally in the final Sender-first round. Over 70% of these Senders switch to playing optimally when responding to accept-reject. In contrast, almost all Senders who sent optimal recommendations in the final Sender-first round continue to do so when responding to accept-reject.

RESULT 5: *Senders are much more likely to play their partial truth-telling strategy when directly responding to accept-reject.*

Discussion: Main Treatment Results

In all rounds, only a minority of Senders send optimal recommendations. Instead, they tend to send recommendations that would be optimal if Receivers accepted all recommendations, which Receivers do not. While Senders learn their optimal recommendations as rounds progress, this learning effect is much smaller than the effect of reversing the sequence of the game, which allows Senders to respond to Receivers' wary equilibrium strategy.

Why might Senders fail to send optimal recommendations? As summarized by Result 4, the learning effect is relatively weak. This suggests that a lack of experience with the game, and/or uncertainty about Receivers' strategy, can only partially explain Senders' behavior. Furthermore, the fact that Senders best-respond to Receiver's accept-reject strategy in the Receiver-first round rules out an inability to identify the best response to that strategy, as well as a preference for truth-telling. In Online Appendix C.2, I summarize additional analyses that explore a preference for truth-telling, as well as mixing between strategies, as potential mechanisms for this failure. I find further evidence against both.

Instead, Senders' behavior appears to be driven by a failure of contingent thinking (FCT). Such failures occur when decision-makers fail to focus on contingencies (e.g., states, strategies of other players) when making a decision (Niederle and Vespa, 2023). In the recommendation game, the Receiver strategy that is relevant to the Sender is their accept-reject strategy. Senders may fail to think contingently by failing to focus on how to respond to that Receiver strategy contingency.

Decision-makers who fail to think contingently tend to exhibit several behavioral hallmarks (Niederle and Vespa, 2023). First, they tend to play a "naïve" strategy instead of the optimal one. As summarized by Result 3, many Senders did exactly that: instead of playing a strategy that accounts for the fact that Receivers reject a specific recommendation, they play a strategy that would be optimal if Receivers accepted all recommendations. Second, repetitions and feedback do little to fix the FCT. The weak learning effect summarized in Result 4 is consistent with this trend. Third, many decision-makers overcome the FCT if placed in the relevant contingency or encouraged to focus on it. Changing the timing of a strategic game is an effective way of placing decision-makers in the relevant contingency (Esponda and Vespa, 2014). Result 5 indicates that Senders react quite strongly to such an intervention, becoming much more likely to send optimal recommendations.

The results of the main treatment thus suggest that Senders fail to send optimal recommendations because they fail to think contingently. If so, simply prompting Senders to think about the relevant contingency - the Receiver's accept-reject strategy - should make them more likely to send optimal recommendations. As an additional test of the mechanism, I therefore conduct a second treatment that allows me to study the effect of such an intervention.

4 Beliefs Treatment

In the beliefs treatment, I prompt Senders to predict the Receiver's strategy before submitting their own. This intervention is weaker than the one implemented in the additional round of the main treatment: it does not place Senders in the relevant contingency, but rather encourages them to think about it. If Senders' strategic mistake in the main treatment is driven by a failure of contingent thinking, Senders who correctly predict that Receivers play accept-reject - and thus focus on the relevant contingency - should be more likely to send optimal recommendations. Section 4.1 describes the design of the beliefs treatment. Its results are presented and discussed in Section 4.2.

4.1 Experimental Design

Structure of the Beliefs Treatment

The basic structure of the beliefs treatment was identical to that of the main treatment (see Figure 2). The only difference was that, before certain rounds, I elicited participants' beliefs about their co-player's strategy in the upcoming round. More precisely, participants indicated their belief about the modal strategy played by participants in the opposite role in previous sessions of the experiment. Receivers' belief elicitation task entailed specifying their beliefs about Senders' strategies and re-stating the recommendation the computer always sent.

Beliefs were elicited before every sixth Sender-first round, starting with the first one. There were thus four belief elicitation tasks, which occurred before the first, seventh, thirteenth and nineteenth Sender-first rounds. These tasks were incentivized using a simple scheme that paid $5 if the stated beliefs were correct,[19] and nothing otherwise.[20]

Procedures

Online Appendix B contains the beliefs treatment's instructions. Similar to the main treatment, the beliefs treatment was pre-registered on AsPredicted.org, programmed and deployed

[19]For payment purposes, the correct modal strategies were the modal strategies in the main treatment. As a reminder, in all rounds of the main treatment, the modal Receiver played accept-reject, and the modal Sender played their full truth-telling strategy. In addition to correctly stating Sender's strategy, Receivers were additionally required to correctly re-state the computer's recommendation in order to receive the $5 payment.

[20]I chose this simple incentive scheme over a more complex one (e.g., a scoring rule) for two reasons. First, recent evidence suggests that more complex incentive schemes may confuse participants (Danz et al., 2022). Second, the primary goal of this treatment was to prompt Senders to focus on the relevant Receiver strategy contingency. Using a scoring rule would have involved eliciting Senders' beliefs about the probability of Receivers playing each of their strategies, which would have made Senders less likely to focus on the relevant Receiver strategy.

in oTree (Chen et al., 2016), and conducted at the TEEL with participants recruited via ORSEE (Greiner, 2015). I conducted seven sessions, each with 16 to 22 participants, for a total of 130 participants (65 Senders, 65 Receivers).

Participants' total payment included a $5 show-up fee, their earnings from the comprehension quiz, and their earnings from one main round of the game and one belief elicitation task. The round and belief elicitation task that counted for each participant was randomly selected at the beginning of the experiment and revealed to them at the end of the experiment. The average participant earned $21.71 and took 41 minutes to complete the experiment. Online Appendix Table C.1 shows that participants' characteristics did not differ significantly between the two experiments.

4.2 Results

I first establish whether Senders focused on the relevant contingency. That is, did most Senders believe the modal Receiver played accept-reject? Next, I investigate whether participants' strategies differ by treatment. Recall that if Senders fail to send optimal recommendations because they fail to think contingently, Senders who focus on Receivers' modal strategy should be more likely to send optimal recommendations. As with the main treatment, I focus exclusively on the 20 Sender-first rounds and drop participants who withdrew from the experiment.[21] The final sample includes 129 participants (65 Senders, 64 Receivers). All reported p-values are from two-sided tests.

Do Senders Focus on the Relevant Contingency?

[Figure 8]

Figure 8 displays the percent of Senders who believe accept-reject is the modal Receiver

[21] One participant withdrew from the beliefs treatment. They were playing the role of Receiver, and withdrew after round 11. Given that one-to-one matching of Receivers to Senders requires an even number of participants to be present in all rounds, to avoid having to stop the session, a research assistant played at this participant's terminal for the remainder of the rounds. The research assistant played the modal Receiver strategy in all rounds of all prior experimental sessions, which was accept-reject.

strategy in a given round. In every round that beliefs are elicited, most Senders believe accept-reject is the most-commonly played Receiver strategy. Moreover, more Senders believe this as rounds progress, presumably due to the feedback they receive. Between rounds 1 and 19, the percentage increases from 60% to 88%. This increase, as well as the increase from rounds 1 to 7, are significant (proportions tests, $p \leq 0.01$ for both).[22]

RESULT 6: *In all rounds where beliefs are elicited, most Senders believe that accept-reject is the modal Receiver strategy.*

Result 6 indicates that eliciting Senders' beliefs prompted most of them to focus on the relevant contingency. Next, I investigate whether this intervention increases Senders' likelihood of sending optimal recommendations, as predicted by a failure of contingent thinking.

How Does Eliciting Senders' Beliefs Affect Their Strategies?

[Figure 9]

Figure 9 compares the frequency of wary equilibrium strategy play across the two treatments. Figure 9a shows that Receivers' strategies do not change much across the two experiments. In almost all rounds, the percent of Receivers playing accept-reject in the beliefs treatment is not significantly different from the main treatment (proportions tests, $p > 0.13$).[23]

Senders, on the other hand, are more likely to send optimal recommendations in the beliefs treatment. As shown in Figure 9b, the aggregate trend in Senders' strategies is the same across treatments: the percent of Senders playing partial truth-telling increases as rounds progress.

[22] Online Appendix Table C.2 documents how Receivers' beliefs about Senders' strategies evolve as rounds progress. The proportion of Receivers believing that the modal Sender plays partial truth-telling increases as rounds progress, while the proportion believing full truth-telling is the mode decreases as rounds progress. The changes in Receivers' beliefs thus match the empirical trends in Senders' strategies.

[23] Proportions tests indicate that significantly more Receivers in the beliefs treatment play accept-reject in rounds 4 ($p < 0.01$), 8 ($p < 0.05$), 7 and 14 ($p < 0.10$ for both). This means that, starting in round 5, Senders in the beliefs treatment may have been matched with more Receivers who played accept-reject than were Senders in the main treatment. However, beliefs treatment Senders are more likely to play partial truth-telling starting in round 1. Given that the change in Senders' strategies begins before the change in Receivers' strategies, it is unlikely that it was driven by an increased number of pairings with Receivers playing accept-reject.

However, the percent doing so in the beliefs treatment is consistently higher than in the main treatment, and significantly so in all rounds after and including round 4 (proportions tests, $p \leq 0.05$).[24] Even in round 1 - when Senders have not received any feedback about Receivers' strategies - the percent of Senders playing partial truth-telling is 55% higher in the beliefs treatment. These treatment-level differences are robust to using a stricter definition of strategic behavior. In Online Appendix Figure C.3, I define a Sender as "strategic" in round x if they send optimal recommendations in round x and all subsequent rounds.[25] As shown in the Figure, the distribution of strategic Senders in the beliefs treatment first-order stochastically dominates the distribution of strategic Senders in the main treatment (Wilcoxon rank-sum test, $p < 0.01$).

RESULT 7: *Senders are more likely to play partial truth-telling in the beliefs treatment than in the main treatment.*

[Table 3]

In addition to being more likely to send optimal recommendations, Senders learn to send these recommendations more quickly. Across all rounds (column 1 of Table 3), playing an additional round increases a Sender's likelihood of playing partial truth-telling by 1.3 percentage points in the beliefs treatment. This aggregate effect is larger than that in the main treatment (see column 1 of Table 1). Furthermore, column 2 of Table 3 reveals that Senders learn their optimal strategy more quickly in the beliefs treatment. In early rounds of the beliefs treatment, playing an additional round increases a Sender's likelihood of sending optimal recommendations by 3.3 percentage points. This effect is more than twice as large as the early-round effect

[24]In rounds 1 through 3, two-sided proportions tests are not significant at conventional levels ($p \in [0.14, 0.16]$), but one-sided tests are ($p \in [0.07, 0.08]$). The fact that the difference in the proportions becomes larger as rounds progress is consistent with the fact that more Senders hold correct beliefs, and thus focus on the relevant contingency, in later rounds.

[25]Esponda and Vespa (2014) use a similar definition to classify strategic subjects in a committee voting problem.

in the main treatment (see column 2 of Table 1).[26] Online Appendix Table C.3 provides additional support for the increase in the early-round learning effect, using a regression framework that pools the data from both treatments.

RESULT 8: *Senders learn their optimal strategy more quickly in the beliefs treatment than in the main treatment.*

Results 7 and 8 support the claim that Senders' failure to send optimal recommendations is driven by a failure of contingent thinking. Next, I provide an additional test of this mechanism. Recall that, in every round, some Senders hold incorrect beliefs about the modal Receiver strategy. In other words, eliciting Senders' beliefs does not prompt all of them to focus on the relevant contingency. I therefore next examine whether Senders who hold accurate beliefs are more likely to send optimal recommendations.

[Table 4]

When pooling the data from all rounds, I find that correctly predicting the modal Receiver strategy is positively correlated with sending optimal recommendations ($\rho = 0.43$, $p < 0.01$). Next, I run a series of regressions to confirm that this finding persists when controlling for the round number. Table 4 reports the results of regressions that identify the impact of Senders' beliefs about the modal Receiver strategy in a given round on the strategy they play in that round. As shown in column 1, Senders who hold correct beliefs about the modal Receiver strategy in the upcoming round are 44.6 percentage points more likely to send optimal recommendations in that round. Columns 2 and 3 indicate that the effect is similar in early and late rounds. These results confirm that Senders who focus on the relevant contingency - that is, who correctly predict the recommendations Receivers accept - are more likely to send optimal recommendations.

[26]The early-round learning effects can also be compared as fractions of the percent of Senders sending optimal recommendations in early rounds of each treatment. In the main treatment, the 1.3 percentage point effect represents $\frac{0.013}{0.32} = 4\%$. In the beliefs treatment, the effect is larger: $\frac{0.033}{0.522} = 6\%$.

RESULT 9: *Senders who hold correct beliefs about the modal Receiver strategy, and thus focus on the relevant contingency, are more likely to send optimal recommendations.*

Discussion: Beliefs Treatment Results

In the beliefs treatment, Senders are prompted to predict the Receiver's strategy before sending their recommendations. If the failure to send optimal recommendations were driven by incorrect beliefs about the Receiver's strategy, lying costs, or a failure to identify their best response, this task should have no effect on Senders' strategies. Instead, the task makes Senders much more likely to send optimal recommendations. Results 6 and 9 point to the reason for this effect: the belief elicitation task prompts most Senders to focus on Receivers' modal strategy, and those who do are much more likely to send optimal recommendations. Thus, across both treatments, directly or indirectly placing Senders in the relevant contingency makes them less likely to send naïve recommendations, and more likely to send optimal ones. This effect is consistent with a failure of contingent thinking.

5 Discussion and Conclusion

When attempting to persuade decision-makers to change their behavior, well-meaning advisors should account for how decision-makers may respond to the recommended change. Using a Sender-Receiver game in which Receivers are wary of well-meaning Senders' incentives, I show that many Senders fail to account for Receivers' wariness. Despite repeated interactions with Receivers and feedback about their responses to recommendations, most Senders send recommendations that would only be optimal if Receivers were required to follow them. When it comes to correcting this mistake, the most effective interventions are those that make the recommendations Receivers accept more salient to Senders.

In reality, the settings in which advisors issue recommendations can be more complex than the setting considered in this paper. First, decision-makers may consider alternatives other than their status-quo and the recommended alternative. Second, advisors whose incentives are misaligned with decision-makers' may also be strategic, which would require well-meaning advisors to account for their strategies. Third, states and their corresponding actions may not be equidistant from each other. In Online Appendix A.2, I show that the optimality of the aligned-type Sender's partial truth-telling strategy is robust to these variations of the recommendation game. Models of bad reputation (Morris, 2001; Ely et al., 2008) extend the insights of the recommendation game to dynamic settings by showing how a "good" advisor's reputational incentives may lead them to distort their information. It would conceivably be even more challenging for advisors to issue optimal recommendations in these more complex environments. Determining exactly how much more challenging may be a promising avenue for future research.

The optimality of a partial truth-telling strategy requires certain restrictions on the Receiver's beliefs and action set. In order for the recommendation game to have a wary equilibrium, the Receiver must be able to take an action that is in-between their status-quo and the action they are wary of taking. Put differently, a partial truth-telling strategy can only be optimal situations where the Receiver benefits from making a large behavioral change. Such situations are very common empirically. For instance, whereas the Canadian government recommends drinking no more than two alcoholic beverages per week (Paradis et al., 2023), the average Canadian adult drinks nearly five times that amount (Statistics Canada, 2024). Similarly, whereas the World Health Organization recommends that adults perform at least 150 minutes of moderate exercise per week (World Health Organization, 2020), up to one-third of adults worldwide do not exercise at all (Ipsos Global Advisor, 2021). In these scenarios, as well as many others, it is thus possible to recommend an action that is in-between many decision-makers' status-quo and their optimal action. The optimality of a partial truth-telling strategy is also limited to situations where the Receiver believes that a large behavioral change is unlikely to be in their best interest, as modeled by the assumption $q_1 \geq q_3$. This assumption captures a common form of motivated reasoning, where a decision-maker convinces themselves that they don't need to change their behavior. For instance, decision-makers often persuade themselves that they do not need to drink less alcohol, exercise more or save more, even when presented evidence that such changes are in their best interest.

My findings point to several interesting avenues for future work. For one, they indicate that simply prompting decision-makers to think about a relevant contingency may help correct a failure of contingent thinking. This intervention is weaker than those studied in prior work, which involve re-framing the decision problem (Esponda and Vespa, 2014; Niederle and Vespa, 2023; Esponda and Vespa, 2024). Studying the effectiveness of this intervention in other strategic settings may be a promising avenue for future work. More broadly, though, my findings raise questions about the practice of eliciting players' beliefs about other players' strategies. Obtaining data on these beliefs can provide valuable insight into a player's level of strategic sophistication (see, e.g., Aoyagi et al. (2024)). My results suggest, however, that eliciting those beliefs may change a player's choice of strategy. Future research may study the impact of belief elicitation in a broader class of games, to better understand when this practice affects players' strategies.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica 87*(4), 1115–1153.

Ali, S. N., M. Mihm, L. Siga, and C. Tergiman (2021). Adverse and Advantageous Selection in the Laboratory. *American Economic Review 111*(7), 2152–78.

Alsan, M. and S. Eichmeyer (2024). Experimental Evidence on the Effectiveness of Nonexperts for Improving Vaccine Demand. *American Economic Journal: Economic Policy 16*(1), 394–414.

Altmann, S., A. Falk, and A. Grunewald (2025). Communicating through Defaults. *Review of Economics and Statistics 107*(1), 256–268.

Amador, M. and K. Bagwell (2013). The Theory of Optimal Delegation with an Application to Tariff Caps. *Econometrica 81*(4), 1541–1599.

Ambuehl, S., B. D. Bernheim, and A. Ockenfels (2021). What Motivates Paternalism? An Experimental Study. *American Economic Review 111*(3), 787–830.

Aoyagi, M., G. R. Fréchette, and S. Yuksel (2024). Beliefs in Repeated Games: An Experiment. *American Economic Review 114*(12), 3944–3975.

Bartling, B., A. W. Cappelen, H. Hermes, M. Skivenes, and B. Tungodden (2025). Free to Fail? Paternalistic Preferences in the United States. *DICE Discussion Paper No. 400.*

Bartling, B., E. Fehr, and H. Herz (2014). The Intrinsic Value of Decision Rights. *Econometrica 82*(6), 2005–2039.

Bhattacharya, U., A. Hackethal, S. Kaesler, B. Loos, and S. Meyer (2012). Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study. *The Review of Financial Studies 25*(4), 975–1032.

Blume, A., O. J. Board, and K. Kawamura (2007). Noisy Talk. *Theoretical Economics 2*(4), 395–440.

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (1998). Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games. *American Economic Review 88*(5), 1323–1340.

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (2001). Evolution of Communication with Partial Common Interest. *Games and Economic Behavior 37*(1), 79–120.

Blume, A., E. K. Lai, and W. Lim (2020). Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations. In C. M. Capra, R. T. Croson, M. L. Rigdon, and T. S. Rosenblat (Eds.), *Handbook of Experimental Game Theory*, pp. 311–347. Edward Elgar Publishing.

Blume, A., E. K. Lai, and W. Lim (2023). Mediated Talk: An Experiment. *Journal of Economic Theory 208*, 105593.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice Under Risk. *The Quarterly Journal of Economics 127*(3), 1243–1285.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020). Memory, attention, and choice. *The Quarterly Journal of Economics 135*(3), 1399–1442.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022). Salience. *Annual Review of Economics 14*(1), 521–544.

Brandts, J. and G. Charness (2011). The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons. *Experimental Economics 14*, 375–398.

Cai, H. and J. T.-Y. Wang (2006). Overcommunication in Strategic Information Transmission Games. *Games and Economic Behavior 56*(1), 7–36.

Calford, E. M. and T. N. Cason (2024). Contingent Reasoning and Dynamic Public Goods Provision. *American Economic Journal: Microeconomics 16*(2), 236–266.

Charness, G. and D. Levin (2009). The Origin of the Winner's Curse: A Laboratory Study. *American Economic Journal: Microeconomics 1*(1), 207–36.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery (2013). Barriers to Household Risk Management: Evidence from India. *American Economic Journal: Applied Economics 5*(1), 104–135.

Costa-Gomes, M. A. and G. Weizsäcker (2008). Stated Beliefs and Play in Normal-Form Games. *The Review of Economic Studies 75*(3), 729–762.

Crawford, V. P. and J. Sobel (1982). Strategic Information Transmission. *Econometrica 50*(6), 1431–1451.

Croson, R. T. (1999). The Disjunction Effect and Reason-Based Choice in Games. *Organizational Behavior and Human Decision Processes 80*(2), 118–133.

Croson, R. T. (2000). Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play. *Journal of Economic Behavior & Organization 41*(3), 299–314.

Dal Bó, P. and G. R. Fréchette (2018). On the Determinants of Cooperation in Infinitely Repeated Games: A Survey. *Journal of Economic Literature 56*(1), 60–114.

Danz, D., L. Vesterlund, and A. J. Wilson (2022). Belief Elicitation and Behavioral Incentive Compatibility. *American Economic Review 112*(9), 2851–2883.

Dickhaut, J. W., K. A. McCabe, and A. Mukherji (1995). An Experimental Study of Strategic Information Transmission. *Economic Theory 6*, 389–403.

Egan, M., G. Matvos, and A. Seru (2024). The Problem of Good Conduct among Financial Advisers. *Journal of Economic Perspectives 38*(4), 193–210.

Ely, J., D. Fudenberg, and D. K. Levine (2008). When is Reputation Bad? *Games and Economic Behavior 63*(2), 498–526.

Epley, N. and T. Gilovich (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives 30*(3), 133–140.

Erev, I., G. Bornstein, and T. S. Wallsten (1993). The Negative Effect of Probability Assessments on Decision Quality. *Organizational Behavior and Human Decision Processes 55*(1), 78–94.

Esponda, I. and E. Vespa (2014). Hypothetical Thinking and Information Extraction in the Laboratory. *American Economic Journal: Microeconomics 6*(4), 180–202.

Esponda, I. and E. Vespa (2024). Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory. *Review of Economic Studies 91*(5), 2806–2831.

Eyster, E. (2019). Errors in Strategic Reasoning. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 187–259. Elsevier.

Fehr, E., H. Herz, and T. Wilkening (2013). The Lure of Authority: Motivation and Incentive Effects of Power. *American Economic Review 103*(4), 1325–1359.

Forges, F. (1985). Correlated Equilibria in a Class of Repeated Games with Incomplete Information. *International Journal of Game Theory 14*, 129–149.

Forges, F. (1986). An Approach to Communication Equilibria. *Econometrica 54*(6), 1375–1385.

Gächter, S. and E. Renner (2010). The Effects of (Incentivized) Belief Elicitation in Public Goods Experiments. *Experimental Economics 13*, 364–377.

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

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Holmström, B. R. (1978). *On Incentives and Control in Organizations.* Stanford University.

Ipsos Global Advisor (2021). Global views on exercise and team sports.

Jin, G. Z., M. Luca, and D. Martin (2022). Complex Disclosure. *Management Science 68*(5), 3236–3261.

Kamenica, E. and M. Gentzkow (2011). Bayesian Persuasion. *American Economic Review 101*(6), 2590–2615.

Kawagoe, T. and H. Takizawa (2009). Equilibrium Refinement vs. Level-k Analysis: An Experimental Study of Cheap-Talk Games with Private Information. *Games and Economic Behavior 66*(1), 238–255.

Krishna, R. V. (2007). Communication in Games of Incomplete Information: Two Players. *Journal of Economic Theory 132*(1), 584–592.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin 108*(3), 480.

Lafky, J., E. K. Lai, and W. Lim (2022). Preferences vs. Strategic Thinking: An Investigation of the Causes of Overcommunication. *Games and Economic Behavior 136*, 92–116.

Li, M. and K. Madarász (2008). When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests. *Journal of Economic Theory 139*(1), 47–74.

Lightle, J. (2014). The Paternalistic Bias of Expert Advice. *Journal of Economics & Management Strategy 23*(4), 876–898.

Martínez-Marquina, A., M. Niederle, and E. Vespa (2019). Failures in Contingent Reasoning: The Role of Uncertainty. *American Economic Review 109*(10), 3437–3474.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 380–391.

Morris, S. (2001). Political Correctness. *Journal of Political Economy 109*(2), 231–265.

Myerson, R. (1991). Games with Communication: Sender-Receiver Games. In *Game Theory: Analysis of Conflict*, pp. 283–288. Harvard University Press.

Myerson, R. B. (1986). Multistage Games with Communication. *Econometrica 54*(2), 323–358.

Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *American Economic Review 85*(5), 1313–1326.

Ngangoué, M. K. and G. Weizsäcker (2021). Learning from Unrealized versus Realized Prices. *American Economic Journal: Microeconomics 13*(2), 174–201.

Niederle, M. and E. Vespa (2023). Cognitive Limitations: Failures of Contingent Thinking. *Annual Review of Economics 15*, 307–328.

Nyarko, Y. and A. Schotter (2002). An Experimental Study of Belief Learning Using Elicited Beliefs. *Econometrica 70*(3), 971–1005.

Paradis, C., P. Butt, K. Shield, N. Poole, S. Wells, T. Naimi, A. Sherk, and the Low-Risk Alcohol Drinking Guidelines Scientific Expert Panels (2023). Canada's Guidance on Alcohol and Health: Final Report. Ottawa: Canadian Centre on Substance Use and Addiction.

Romero, J. and Y. Rosokha (2023). Mixed Strategies in the Indefinitely Repeated Prisoner's Dilemma. *Econometrica 91*(6), 2295–2331.

Rutström, E. E. and N. T. Wilcox (2009). Stated Beliefs Versus Inferred Beliefs: A Methodological Inquiry and Experimental Test. *Games and Economic Behavior 67*(2), 616–632.

Statistics Canada (2024). Control and sale of alcoholic beverages and cannabis, April 1, 2022, to March 31, 2023.

Wang, J. T.-Y., M. Spezio, and C. 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*(3), 984–1007.

Weizsäcker, G. (2023). *Misunderstandings: False Beliefs in Communication*. Open Book Publishers.

World Health Organization (2020). Who guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.

Tables

Table 1
Evolution of Senders' Likelihood of Playing Partial Truth-Telling (Main Treatment)

Rounds	All (1)	1-10 (2)	11-20 (3)
Round	0.009***	0.013**	0.009**
	(0.003)	(0.006)	(0.004)
Constant	0.226**	0.157	0.280*
	(0.115)	(0.111)	(0.144)
Observations	1340	670	670
Mean of DV	0.37	0.32	0.41

Notes. Coefficients from OLS regressions conducted among Senders in the main treatment, with standard errors in parentheses. The dependent variable is an indicator for playing partial truth-telling (Senders' wary equilibrium strategy) in a given round, and the independent variable is the round number. All regressions include controls for the session and standard errors clustered at the participant-level. * denotes $p < 0.10$, ** denotes $p < 0.05$, and *** denotes $p < 0.01$.

Table 2

Senders' Strategies: Final Sender-First Round vs. Receiver-First Round

	Response to accept-reject		
	Partial truth-telling	Other strategy	*Total*
Final Sender-first round			
Partial truth-telling	86.7%	13.3%	100.0%
Other strategy	70.3%	29.7%	100.0%
Total	*77.6%*	*22.4%*	*100.0%*

Notes. This table displays main treatment Senders' strategies in the Receiver-first round, as a function of the strategies they play in the final Sender-first round. The cells display the proportion of Senders who play a strategy shown in a given row in the final Sender-first round, who then play a strategy shown in a given column in response to accept-reject in the Receiver-first round.

Table 3
Evolution of Senders' Likelihood of Playing Partial Truth-Telling (Beliefs Treatment)

Rounds	All	1-10	11-20
	(1)	(2)	(3)
Round	0.013***	0.033***	0.005
	(0.003)	(0.008)	(0.003)
Constant	0.258*	0.201	0.338**
	(0.150)	(0.148)	(0.161)
Observations	1300	650	650
Mean of DV	0.578	0.522	0.634

Notes. Coefficients from OLS regressions conducted among Senders in the beliefs treatment, with standard errors in parentheses. The dependent variable is an indicator for playing partial truth-telling (Senders' wary equilibrium strategy) in a given round, and the independent variable is the round number. All regressions include controls for the session and standard errors clustered at the participant-level. * denotes $p < 0.1$, ** denotes $p < 0.05$, and *** denotes $p < 0.01$.

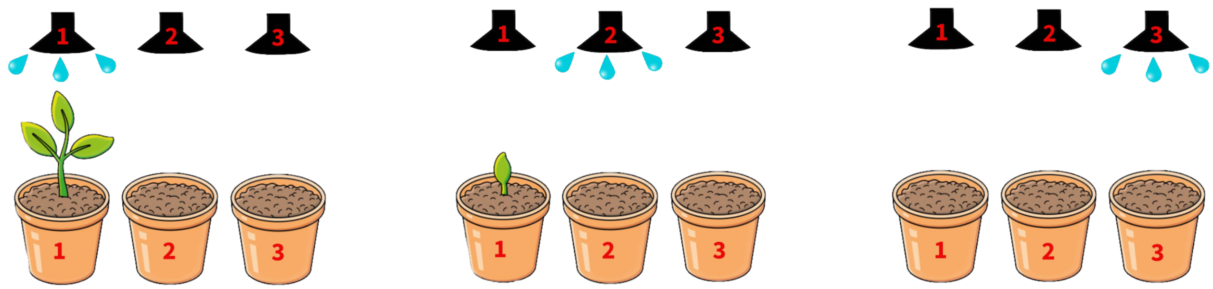
Table 4
Effect of Beliefs on Senders' Likelihood of Playing Optimal Strategy

Rounds	All	Rounds 1 & 7	Rounds 13 & 19
	(1)	(2)	(3)
Believes accept-reject is mode	0.446^{***}	0.377^{***}	0.541^{***}
	(0.075)	(0.096)	(0.096)
Constant	0.017	0.017	-0.068
	(0.881)	(0.897)	(0.666)
Observations	260	130	130
Mean of DV	0.546	0.469	0.623

Notes. Coefficients from OLS regressions conducted among Senders in the beliefs treatment, with standard errors in parentheses. The dependent variable is an indicator for playing partial truth-telling (Senders' wary equilibrium strategy) in a given round, and the independent variable is an indicator for believing accept-reject (Receivers' wary equilibrium strategy) is the modal Receiver strategy in a given round. All regressions include controls for the round and session. Standard errors clustered at the participant-level. * denotes $p < 0.10$, ** denotes $p < 0.05$, and *** denotes $p < 0.01$.

Figures

Figure 1
Illustration of the Plant Pot Game



(a) Pot 1 is the plant pot; sprinkler 1 is turned on

(b) Pot 1 is the plant pot; sprinkler 2 is turned on

(c) Pot 1 is the plant pot; sprinkler 3 is turned on

Figure 2
Structure of the Main Treatment

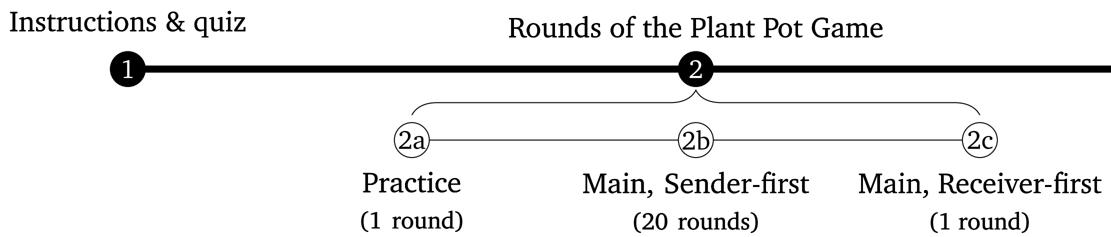
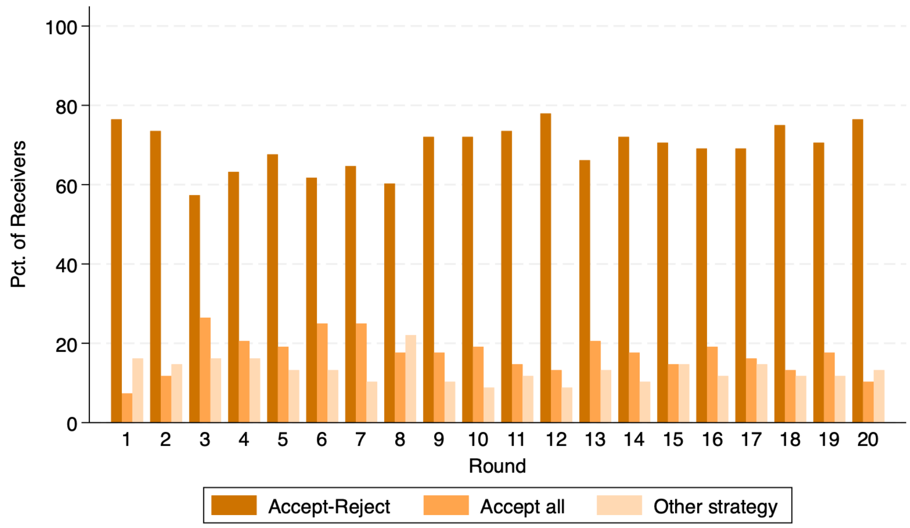
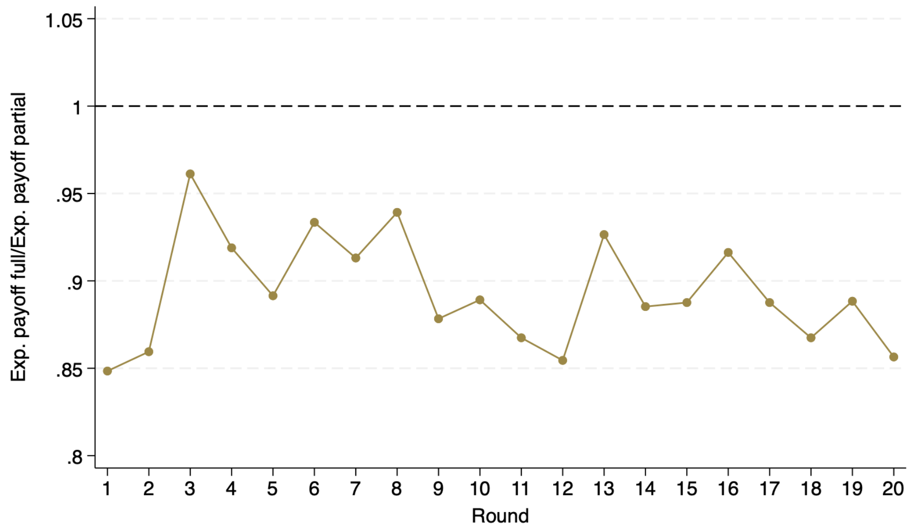


Figure 3
Distribution of Receivers' Strategies



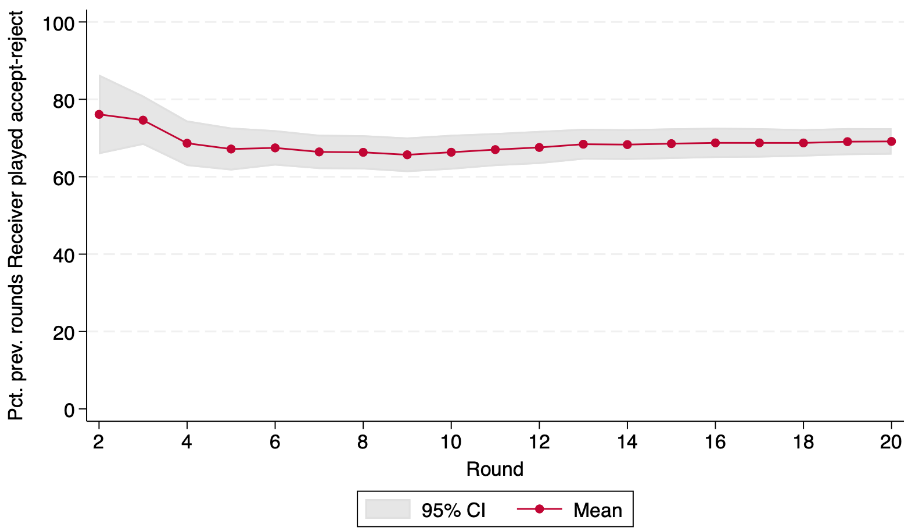
Notes. This figure displays the percent of Receivers in the main treatment playing accept-reject (their wary equilibrium strategy), accept-all (their trusting equilibrium strategy), and other strategies in a given round.

Figure 4
Expected State-3 Payoff of Senders' Equilibrium Strategies



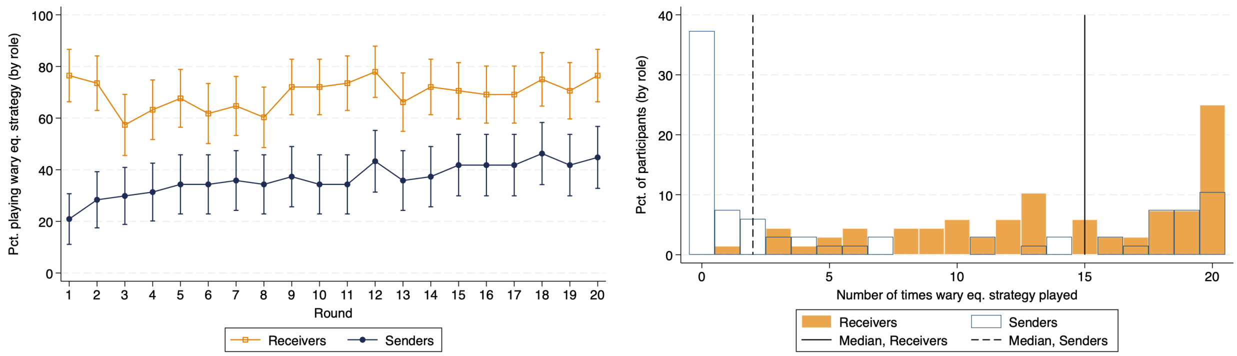
Notes. This figure displays the expected state-3 payoff of Senders' full truth-telling strategy as a fraction of the expected state-3 payoff of their partial truth-telling strategy. Expected payoffs are calculated using the empirical distribution of Receivers' strategies in the main treatment.

Figure 5
Feedback Received by Senders



Notes. This figure summarizes the feedback Senders in the main treatment have received (on the y-axis) when choosing their strategy in round x (on the x-axis). The figure displays the percent of previous rounds in which Senders received the feedback that the Receiver they were paired with played accept-reject. The pink line indicates the mean (across Senders), and the grey area indicates the 95% confidence interval of that mean.

Figure 6
Frequency of Optimal Strategy Play

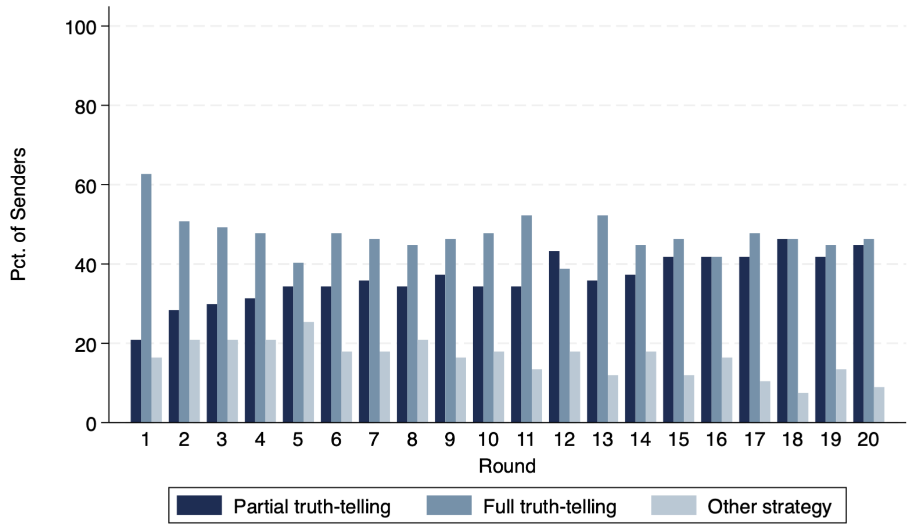


(a) Round-level

(b) Player-level

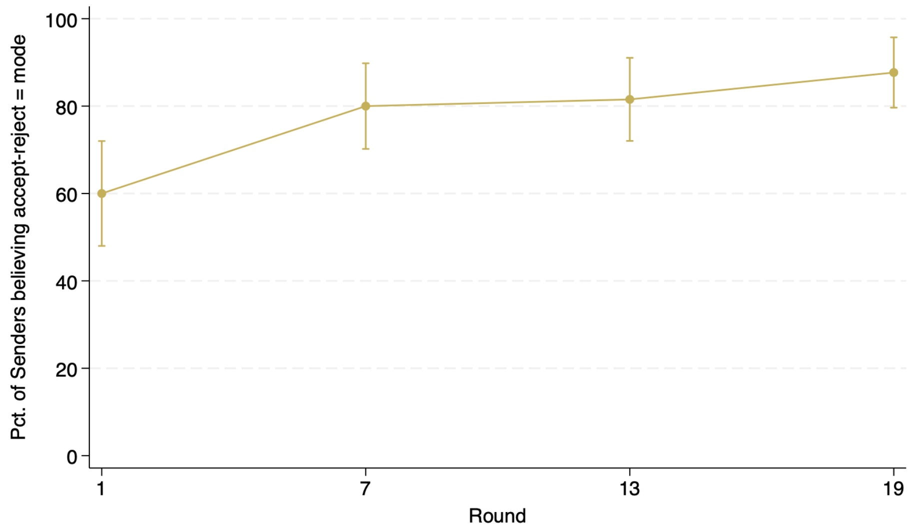
Notes. This figure displays the frequency with which main treatment participants in each role play their wary equilibrium strategies. In Figure 6a, frequency is defined at the round-level: the Figure shows the percent of participants in each role playing their wary equilibrium strategy in each round. Bars indicate the 95% confidence interval. In Figure 6b, frequency is defined at the player-level: the Figure shows the percent of participants in each role playing their wary equilibrium strategy in x of the 20 rounds.

Figure 7
Distribution of Senders' Strategies



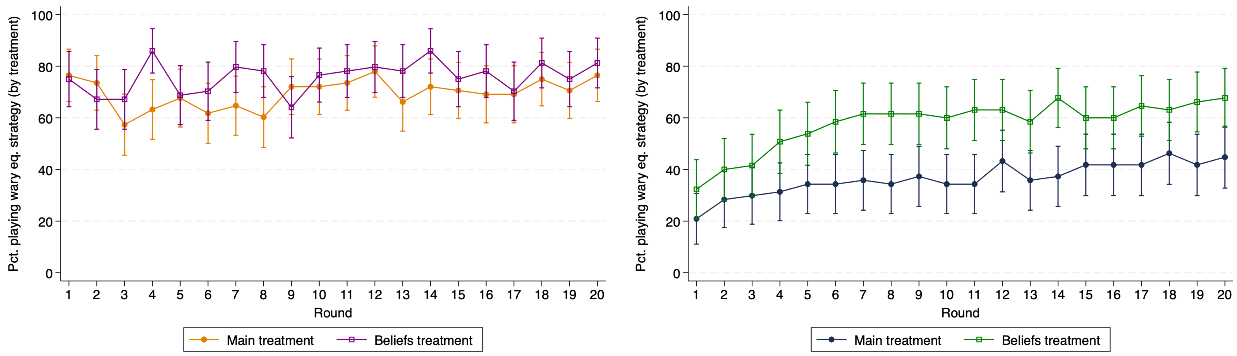
Notes. This figure displays the percent of Senders in the main treatment playing partial truth-telling (their wary equilibrium strategy), full truth-telling (their trusting equilibrium strategy), and other strategies in a given round.

Figure 8
Senders' Beliefs about Receivers' Strategies



Notes. This figure displays the percent of Senders in the beliefs treatment who believe that, in prior experimental sessions, the modal Receiver strategy in round x was accept-reject. Bars indicate the 95% confidence interval.

Figure 9
Frequency of Optimal Strategy Play (Round-Level), by Treatment



(a) Receivers **(b)** Senders

Notes. This figure displays the percent of participants in each treatment playing their optimal strategy in each round. Bars indicate the 95% confidence interval. Senders and Receivers are plotted in separate sub-figures.

Online Appendix for "Strategic Communication under Perceived Misalignment of Interests"

A Proofs

In all proofs, I use $t \in \{A, C\}$ to denote the Sender's type, where A indicates the aligned type and C indicates the commitment type. I use $\mu(\theta, t|r)$ to denote the Receiver's belief about the probability of the history (θ, t) having occurred, conditional on receiving recommendation r.

A.1 Proofs of Propositions

Note that if the Receiver is recommended to take action 1 (i.e., to stick with their status-quo), their action set is the singleton set $\{1\}$. It is therefore without loss to restrict one's attention to the Receiver's responses to recommendations of 2 and 3, which is the approach I take in all proofs in Section A.1. Furthermore, given that the Sender's commitment type is non-strategic, the Receiver and the Sender's aligned type are the only players in the recommendation game. When describing each equilibrium, I therefore include players' strategies, as well as a reminder that the commitment type plays $r(\theta) = 3$ for all θ.

Proof of Proposition 1

Suppose the Receiver plays $a_R(r) = r$ for all $r \in \{2, 3\}$; that is, they accept all recommendations to take an action other than their status-quo. The aligned-type Sender's unique best response is $r(\theta) = \hat{a}(\theta)$ for all θ, given that they want the Receiver's action to be as close as possible to $\hat{a}(\theta)$. If the aligned-type Sender uses this strategy, the Receiver forms the following beliefs:

$$\mu(1, C|3) = \frac{q_1(1-p)}{q_1(1-p) + q_2(1-p) + q_3} \tag{1}$$

$$\mu(2, C|3) = \frac{q_2(1-p)}{q_1(1-p) + q_2(1-p) + q_3} \tag{2}$$

$$\mu(3, C|3) = \frac{q_3(1-p)}{q_1(1-p) + q_2(1-p) + q_3} \tag{3}$$

$$\mu(3, A|3) = \frac{q_3 p}{q_1(1-p) + q_2(1-p) + q_3} \tag{4}$$

$$\mu(1, A|1) = 1 \tag{5}$$

$$\mu(2, A|2) = 1 \tag{6}$$

$$\mu(\theta, A|r) = 0 \text{ for all } r \neq \theta \tag{7}$$

Given these beliefs, it is always better for the Receiver to accept a recommendation of 2 than to stick with their status-quo. Accepting a recommendation of 3 is better than sticking with their status-quo if

$$\frac{1}{q_1(1-p)+q_2(1-p)+q_3}[q_1(1-p)u(2\alpha)+q_2(1-p)u(\alpha)+q_3u(0)] \geq \frac{1}{q_1(1-p)+q_2(1-p)+q_3}[q_1(1-p)u(0)+q_2(1-p)u(\alpha)+q_3u(2\alpha)]$$

which simplifies to

$$q_3 \geq q_1(1-p) \tag{8}$$

Thus, if Equation 8 holds, there exists a "trusting" weak perfect Bayesian equilibrium where the aligned-type Sender plays $r(\theta)=\hat{a}(\theta)$ for all θ, the Receiver plays $a_R(r)=r$ for all r, and the Receiver holds the beliefs described in Equations 1 through 7. Additionally, by assumption, the commitment-type Sender recommends $r(\theta)=3$ for all θ. ■

Proof of Proposition 2

Suppose the Receiver plays $a_R(2)=2$ and $a_C(3)=1$; that is, they accept a recommendation of 2 and reject a recommendation of 3. The aligned-type Sender's unique best response is $r(1)=1$ and $r(2)=r(3)=2$, given that they want the Receiver's action to be as close as possible to $\hat{a}(\theta)$. The uniqueness of the best response for $\theta=1$ arises from the assumption that the aligned-type Sender recommends the action closest to $\hat{a}(\theta)$ when indifferent between recommendations. If the aligned-type Sender uses this strategy, the Receiver forms the following beliefs:

$$\mu(\theta,C|3)=q_\theta \text{ for all } \theta \tag{9}$$

$$\mu(\theta,A|3)=0 \text{ for all } \theta \tag{10}$$

$$\mu(1,A|2)=0 \tag{11}$$

$$\mu(2,A|2)=\frac{q_2}{q_2+q_3} \tag{12}$$

$$\mu(3,A|2)=\frac{q_3}{q_2+q_3} \tag{13}$$

$$\mu(1,A|1)=1 \tag{14}$$

$$\mu(\theta,A|1)=0 \text{ for all } \theta \neq 1 \tag{15}$$

According to these beliefs, accepting a recommendation of 2 will always bring the Receiver closer to the action corresponding to the true state than will rejecting it (i.e., sticking with their status-quo). Thus, it is always best for the Receiver to accept a recommendation of 2. Rejecting a recommendation of 3 is better than accepting it if

$$q_1 u(2\alpha) + q_2 u(\alpha) + q_3 u(0) \leq q_1 u(0) + q_2 u(\alpha) + q_3 u(2\alpha)$$

which simplifies to

$$q_3 \leq q_1 \tag{16}$$

Equation 16 holds by assumption. Thus, for all parameter values, there exists a "wary" weak perfect Bayesian equilibrium where the aligned-type Sender plays $r(3) = 2$ and $r(\theta) = \hat{a}(\theta)$ for all $\theta \neq 3$, the Receiver plays $a_R(2) = 2$ and $a_R(3) = 1$, and the Receiver holds the beliefs described in Equations 9 through 15. Additionally, by assumption, the commitment-type Sender recommends $r(\theta) = 3$ for all θ. ■

Proof of Proposition 3

In Propositions 1 and 2, I showed that the aligned-type Sender has a unique best response to each of the Receiver's strategies that involve accepting a recommendation of 2.[1] To show that the trusting and wary equilibria are the unique equilibria of the recommendation game, I therefore prove that there can be no equilibrium where the Receiver rejects a recommendation of 2.

Suppose the Receiver plays a strategy that involves rejecting a recommendation of 2. The aligned-type Sender's unique best response is $r(\theta) = \hat{a}(\theta)$ for all θ, given that they want the Receiver's action to be as close as possible to $\hat{a}(\theta)$ and recommend the action closest to $\hat{a}(\theta)$ when indifferent[2] between recommendations. If the aligned-type Sender uses this strategy, the Receiver would hold the belief $\mu(2, A|2) = 1$, $\mu(\theta, A|2) = 0$ for all $\theta \neq 2$. Stated otherwise, receiving a recommendation of 2 indicates that the state is 2. The Receiver could thus profitably deviate to accepting a recommendation of 2. Thus, there is no weak perfect Bayesian equilibrium where the Receiver rejects a recommendation of 2. As a result, the trusting and wary equilibria are the unique equilibria of the recommendation game. ■

[1] Recall that the Receiver has two strategies that involve accepting a recommendation of 2: accept all (i.e., accept 2, accept 3) and accept-reject (i.e., accept 2, reject 3).

[2] When the Receiver rejects a recommendation of 2, in state 2, the aligned-type Sender is indifferent between all possible recommendations. When the Receiver rejects all recommendations, the aligned-type Sender is additionally indifferent between all possible recommendations in states 1 and 3.

A.2 Proofs of Model's Robustness to Alternative Assumptions

In the model presented in Section 2, which is the version of the model used in the experiment, I impose various assumptions that may seem stark. I chose to implement that version of the model to keep strategies simple, and the game and equilibrium easy to understand for participants. In this section, I show that the wary equilibrium that I focus on is robust to many of these assumptions, sometimes under mild conditions. In particular, I show that the wary equilibrium is robust to:

1. Allowing the Receiver to choose any action, instead of just selecting between the recommendation and their status-quo
2. Allowing states to be non-equidistant
3. Making the commitment-type Sender a strategic, self-interested player

1. Receiver can take any action

In this variation of the model, the Receiver's action space is $a_R(r) \in \{1,2,3\}$ instead of $a_R(r) \in \{1,r\}$.

Proposition A.2.1: Suppose $a_R(r) \in \{1,2,3\}$, $q_2 \geq q_3$ and $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$. Then, the recommendation game has a wary equilibrium.

First, I define and prove a result that will help to interpret the proof of Proposition A.2.1.

Lemma 1: When $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$, action 1 is the Receiver's ex-ante optimal action; that is, the optimal action based on their prior.

Proof of Lemma 1: In order for action 1 to be the Receiver's ex-ante optimal action, its ex-ante expected payoff must be weakly greater than that of actions 2 and 3. That is, the following conditions must hold:

$$q_1 u(0) + q_2 u(\alpha) + q_3 u(2\alpha) \geq (q_1 + q_3)u(\alpha) + q_2 u(0) \tag{17}$$

$$q_1 u(0) + q_2 u(\alpha) + q_3 u(2\alpha) \geq q_3 u(0) + q_2 u(\alpha) + q_1 u(2\alpha) \tag{18}$$

Equations 17 and 18 simplify to the following conditions, respectively:

$$q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha) - u(2\alpha)}{u(0) - u(\alpha)} \tag{19}$$

$$q_1 \geq q_3 \tag{20}$$

Note that Equation 20 is true by assumption. Thus, when $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$, action 1 is the Receiver's ex-ante optimal action. ■

Proof of Proposition A.2.1: Suppose the Receiver plays $a_R(1) = 1$, $a_R(2) = 2$ and $a_R(3) = 1$. This strategy is a natural extension of the Receiver's accept-reject strategy in a setting where they can choose any action, and action 1 is optimal under their prior. The aligned-type Sender's unique[3] best response is their partial truth-telling strategy; that is, $r(1) = 1$ and $r(2) = r(3) = 2$.

Now, suppose the aligned-type Sender plays their partial truth-telling strategy. The Receiver then forms the beliefs outlined in Equations 9 through 15. I now show that, under the conditions of the Proposition, it is optimal for the Receiver to play $a_R(1) = 1$, $a_R(2) = 2$ and $a_R(3) = 1$, given their beliefs. Since the Receiver knows the state is 1 if they are recommended 1, it is always optimal for them to play $a_R(1) = 1$. If the Receiver is recommended 2, they know the state is either 2 or 3. From the beliefs outlined in Equations 12 and 13, it is easy to see that $a_R(2) = 2$ is optimal if $q_2 \geq q_3$. Finally, a recommendation of 3 reveals no new information about the true state. In order for $a_R(3) = 1$ to be optimal, it must yield a greater expected payoff than $a_R(3) = 3$ and $a_R(3) = 2$, which is true under the following conditions:

$$(1-p)[q_1 u(0) + q_2 u(\alpha) + q_3 u(2\alpha)] \geq (1-p)[q_1 u(2\alpha) + q_2 u(\alpha) + q_3 u(0)] \tag{21}$$

$$(1-p)[q_1 u(0) + q_2 u(\alpha) + q_3 u(2\alpha)] \geq (1-p)[q_1 u(\alpha) + q_2 u(0) + q_3 u(\alpha)] \tag{22}$$

Equation 21 simplifies to $q_1 \geq q_3$, which is true by assumption. Equation 22 simplifies to $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$. Per Lemma 1, this condition implies that action 1 is ex-ante optimal for the Receiver. Thus, when the Receiver can take any action, the aligned-type Sender can play a partial truth-telling strategy in equilibrium if $q_2 \geq q_3$ and $q_1 \geq q_2 + q_3 \cdot \frac{u(\alpha)-u(2\alpha)}{u(0)-u(\alpha)}$. ■

2. Non-equidistant states

In this variation of the model, instead of there being distance 1 between adjacent states, there is distance d_{12} between states 1 and 2 and d_{23} between states 2 and 3, where $d_{12},\ d_{23} \geq 0$.

PROPOSITION A.2.2: Suppose there is distance d_{12} between states 1 and 2 and d_{23} between states 2 and 3, where $d_{12},\ d_{23} \geq 0$.

(i) Suppose further that $d_{23} \geq d_{12}$. Then, the recommendation game has a wary equilibrium.

(ii) Suppose further that $d_{23} < d_{12}$ and $q_2 \frac{u(\alpha d_{23})-u(\alpha d_{12})}{u(0)-u(\alpha(d_{12}+d_{23}))} + q_3$. Then, the recommendation game has a wary equilibrium.

[3] The uniqueness of the best response arises from the assumption that the aligned-type Sender recommends the action closest to $\hat{a}(\theta)$ when indifferent between recommendations.

Proof of Proposition A.2.2: Suppose the Receiver plays their accept-reject strategy; that is, $a_R(2) = 2$ and $a_R(3) = 1$. The aligned-type Sender's unique best response is their partial truth-telling strategy; that is, $r(1) = 1$ and $r(2) = r(3) = 2$.

Now, suppose the aligned-type Sender plays their partial truth-telling strategy. The Receiver then forms the beliefs outlined in Equations 9 through 15. I now show that, under the conditions of the Proposition, it is optimal for the Receiver to play accept-reject, given their beliefs. According to these beliefs, accepting a recommendation of 2 will bring the Receiver closer to the action corresponding to the true state than will rejecting it. Thus, it is always optimal for the Receiver to accept a recommendation of 2. It is optimal for the Receiver to reject a recommendation of 3 if

$$\begin{aligned}(1-p)[q_1u(0)+q_2u(\alpha d_{12})+q_3u(\alpha(d_{12}+d_{23}))] \geq \\ (1-p)[q_1u(\alpha(d_{12}+d_{23}))+q_2u(\alpha d_{23})+q_3u(0)]\end{aligned}$$

which simplifies to

$$q_1 \geq q_2\frac{u(\alpha d_{23})-u(\alpha d_{12})}{u(0)-u(\alpha(d_{12}+d_{23}))}+q_3 \tag{23}$$

When $d_{23} \geq d_{12}$, $u(\alpha d_{23}) \leq u(\alpha d_{12})$ (recall that $u'(\cdot) < 0$). For those relative distances, the left-hand side of Equation 23 is thus strictly smaller than q_3, which means Equation 23 is always satisfied (recall that $q_1 \geq q_3$ by assumption). Thus, when $d_{23} \geq d_{12}$, the recommendation game always has a wary equilibrium. When $d_{23} < d_{12}$, the recommendation game has a wary equilibrium if Equation 23 is satisfied. ■

3. Strategic self-interested Sender

In this variation of the model, the non-strategic commitment-type Sender becomes a strategic, self-interested player. This player, who I will call the selfish-type Sender, either (1) receives a bonus if the Receiver takes action 3, or (2) places some weight on each of two outcomes: maximizing the Receiver's wellbeing and recommending action 3. More precisely, the selfish-type Sender's preferences are described by one of the following payoff functions:

$$U_1(a_R(r)) = \begin{cases}\pi & \text{if } a_R(r) \neq 3 \\ \pi+B & \text{if } a_R(r)=3\text{, where } B>0\end{cases}$$

$$U_2(a_R(r)) = \gamma\cdot v(\alpha|a_R(r)-\hat{a}(\theta)|)+(1-\gamma)\cdot B\cdot \mathbb{1}\{r=3\}\text{, where } B>0 \text{ and } \gamma\in[0,1]$$

$U_1(\cdot)$ captures the preferences of a Sender who stands to benefit significantly more when the

Receiver takes one particular action. Physicians, for instance, may receive a much larger compensation from performing a particular procedure. $U_2(\cdot)$ captures the preferences of a Sender who is pressured to recommend certain products, which is common in the financial industry.[4]

PROPOSITION A.2.3:

(i) Suppose the selfish-type Sender's preferences are described by $U_1(\cdot)$. Then, the recommendation game has a wary equilibrium.

(ii) Suppose the selfish-type Sender's preferences are described by $U_2(\cdot)$, and $\frac{B}{u(0)-u(\alpha)+B} \geq \gamma$. Then, the recommendation game has a wary equilibrium.

Proof of Proposition A.2.3: I wish to show that, for each of the selfish-type Sender's possible payoff functions, the recommendation game has an equilibrium where the Receiver plays accept-reject, the aligned-type Sender plays their partial truth-telling strategy, and the selfish-type Sender always recommends action 3. If each Sender type uses these strategies, the Receiver forms the beliefs outlined in Equations 9 through 15. As shown in the proof of Proposition 2, given these beliefs, it is optimal for the Receiver to accept a recommendation of 2 and reject a recommendation of 3.

I next check that the strategy of each of the Sender's types is a best response to the Receiver's strategy. As shown in the proof of Proposition 2, the aligned-type Sender's unique best response to the Receiver's accept-reject strategy is their partial truth-telling strategy. When the selfish-type Sender has the preferences described by $U_1(\cdot)$, they are indifferent between the Receiver accepting a recommendation of 2 and rejecting any given recommendation. Thus, given the Receiver's strategy, a selfish-type Sender with these preferences cannot profitably deviate from recommending 3 in all states. As a result, when the selfish-type Sender has the preferences described by $U_1(\cdot)$, the recommendation game always has a wary equilibrium.

When the selfish-type Sender has the preferences described by $U_2(\cdot)$, they may have a profitable deviation if recommending an action other than 3 leads the Receiver to take an action that is closer to the true state. This can only be the case in state 2, where recommending 2 leads the Receiver to take the action corresponding to the true state. The selfish-type Sender prefers recommending 3 in state 2 if

$$\gamma \cdot v(\alpha) + (1-\gamma)B \geq \gamma \cdot v(0)$$

[4]For instance, a major Canadian news outlet has documented how financial advisors at large Canadian banks are pressured - and sometimes even "coached" - to recommend certain financial products to their clients (Johnson, 2017; Johnson et al., 2024).

which simplifies to

$$\frac{B}{u(0)-u(\alpha)+B} \geq \gamma \tag{24}$$

Intuitively, the selfish-type Sender must place a sufficiently high weight on their bonus, relative to the weight they place on the Receiver's wellbeing. Thus, when the selfish-type Sender has the preferences described by $U_2(\cdot)$, the recommendation game has a wary equilibrium if Equation 24 is satisfied. ■

B Experiment Instructions

Below is a transcription of the instructions and tasks in both experiments. Recall that, in the experiment, Receivers were called "Choosers" and aligned-type Senders were called "Advisors". Any instructions or tasks that are specific to a particular role (Chooser or Advisor) or a particular treatment (Main or Beliefs) are indicated by text boxes.

Welcome to the experiment!

General instructions. This is an experiment designed to study decision-making. If you pay close attention to the instructions, you can earn a significant amount of money. Please ensure all of your electronics (cell phones, smart watches, etc.) are put away, and do not talk with others during the experiment. If you have a question, please raise your hand and an experimenter will come answer it in private.

MAIN TREATMENT ONLY

> ***Structure of the experiment.*** The experiment consists of multiple rounds of a game, followed by a short survey. You will receive detailed instructions prior to each part of the experiment.

BELIEFS TREATMENT ONLY

> ***Structure of the experiment.*** The experiment consists of multiple rounds of a game. Before several of those rounds, you will complete a guessing task. Finally, you will complete a short survey. You will receive detailed instructions prior to each part of the experiment.

Payment. Your total earnings from the experiment will consist of several components.

- *Show-up fee.* You will receive a $5.00 payment for completing the experiment.
- *Quiz.* After reading the experiment's instructions but before proceeding to the main task, you will complete a quiz that will test your understanding of the instructions. You will receive $1.00 if you answer all quiz questions correctly on your first try.

MAIN TREATMENT ONLY

> - *Randomly-selected round of the game.* One round of the game has been randomly selected to count for your payment.

BELIEFS TREATMENT ONLY

- *Randomly-selected round and guessing task.* One round of the game, as well as one guessing task, have each been randomly selected to count for your payment.

The Plant Pot Game - Instructions

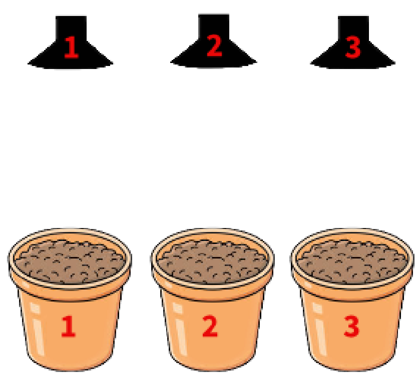
There are three pots, each with a sprinkler above it. The "plant pot" has a plant seed in it; the other two do not.

CHOOSERS ONLY

You are playing the role of **Chooser**. You will not know for sure which pot is the plant pot, but you will know each pot's chance of being the plant pot. Specifically, pot 1 has a 9-in-20 chance, pot 2 has a 7-in-20 chance and pot 3 has a 4-in-20 chance.

ADVISORS ONLY

Participants playing the role of Chooser know that pot 1 has a 9-in-20 chance of being the plant pot, pot 2 has a 7-in-20 chance and pot 3 has a 4-in-20 chance. You, however, are playing the role of **Advisor**. As we'll explain later, Advisors have more information about the plant pot than Choosers do.



9-in-20 7-in-20 4-in-20

CHOOSERS ONLY

The pots, the sprinklers and each pot's chance of being the plant pot.

ADVISORS ONLY

The pots, the sprinklers and the information Choosers have (each pot's chance of being the plant pot).

When a sprinkler is turned on, it sprays water directly below it, as well as below to the right and left of it. The pot closest to it receives the most water; the pot farthest from it receives the least. A seed will grow into a larger plant if it receives more water.

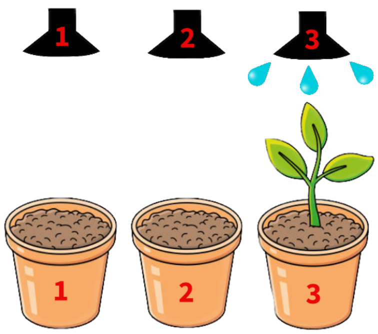


Figure 1

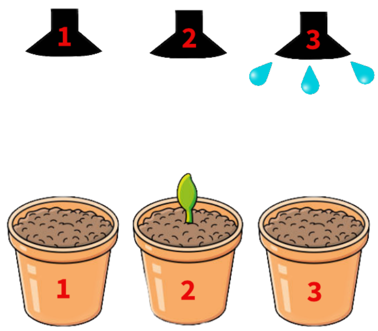


Figure 2

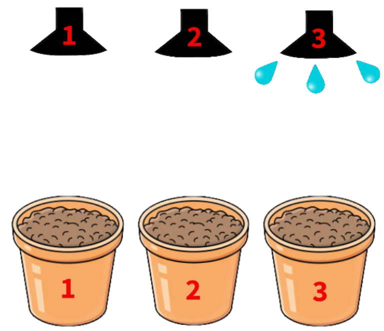


Figure 3

Example: Turning on sprinkler 3 will produce a large plant if pot 3 is the plant pot (as in Figure 1), a small plant if pot 2 is the plant pot (as in Figure 2), and no plant if pot 1 is the plant pot (as in Figure 3).

Your goal

CHOOSERS ONLY

As a Chooser, your goal is to turn on the sprinkler that's as close as possible to the plant pot, and thus to produce the largest possible plant.

ADVISORS ONLY

As an Advisor, you will be paired with a participant playing the role of Chooser. Choosers and Advisors have the same goal: for the Chooser to turn on the sprinkler that's as close as possible to the plant pot, and thus to produce the largest possible plant.

Your task

CHOOSERS ONLY

While sprinkler 1 will be turned on by default, you may receive a recommendation to switch to one of the other sprinklers. Your task requires you to consider two scenarios: the scenario where you receive a recommendation to switch to sprinkler 2, and the one where you receive a recommendation to switch to sprinkler 3. You must decide whether you will accept the proposed switch or stick with the default in each of these scenarios. If you are recommended to stick with the default, you will automatically do so.

ADVISORS ONLY

Your task requires you to consider three scenarios: the scenario where pot 1 is the plant pot, the one where pot 2 is the plant pot, and the one where pot 3 is the plant pot. You must choose a sprinkler to recommend to the Chooser in each scenario. Think of this task as if you know a specific pot (pot 1, 2 or 3) is the plant pot, and with this knowledge, must recommend a sprinkler to the Chooser.

Sprinkler 1 will be turned on by default; you may recommend that the Chooser either stick with this default or switch to one of the other sprinklers. While you are completing this task, the Chooser will decide whether, if recommended to switch to a given sprinkler, they will accept the proposed switch or stick with the default. If they are recommended to stick with the default, they will automatically do so.

Sources of recommendations

CHOOSERS ONLY

The recommendation you receive will be from one of two sources: an Advisor or the computer. An Advisor is another participant who has the same goal as you, and can recommend sticking with the default or switching to one of the other sprinklers. While you are completing your task, they will choose a sprinkler to recommend for each possible plant pot. You can think of the Advisor's task as if they know a specific pot (pot 1, 2 or 3) is the plant pot, and with this knowledge, must recommend you a sprinkler. The computer recommends to switch to sprinkler 3 regardless of which pot is the plant pot. There is a 1-in-3 chance that you will receive an Advisor's recommendation, and a 2-in-3 chance that you will receive the computer's recommendation.

ADVISORS ONLY

The recommendation the Chooser receives will be from one of two sources: you or the computer. The computer recommends to switch to sprinkler 3 regardless of which pot is the plant pot. There is a 1-in-3 chance that the Chooser will receive your recommendation, and a 2-in-3 chance that they will receive the computer's recommendation. When making their decisions, Choosers know their chance of receiving a recommendation from each potential source. They also know that Advisors have the same goal as them, and must choose a sprinkler to recommend for each possible plant pot.

Rounds

CHOOSERS ONLY

You will play 22 rounds of this game. At the beginning of each round, we will randomly re-match each Chooser with an Advisor. We will also determine the plant pot and source of the recommendation you will receive according to the chances mentioned earlier. After you and the Advisor you are paired with have made your decisions, the outcome of the game will be determined. At the end of each round, you will receive feedback about that outcome, as well as the decisions of the Advisor you were paired with.

ADVISORS ONLY

You will play 22 rounds of this game. At the beginning of each round, we will randomly re-match each Chooser with an Advisor. We will also determine the plant pot and source of the recommendation the Chooser will receive according to the chances mentioned earlier. After you and the Chooser you are paired with have made your decisions, the outcome of the game will be determined. At the end of each round, you will receive feedback about that outcome, as well as the decisions of the Chooser you were paired with.

Round 1 is a "practice round". It is intended to familiarize you with your task, and is not eligible to determine your payment from the experiment. At the end of round 1, you will see an outline of the feedback you will receive at the end of the rest of the rounds.

BELIEFS TREATMENT ONLY

Guessing tasks

CHOOSERS ONLY

Before rounds 2, 8, 14 and 20, you will (1) make guesses about the most-common decisions Advisors have made in that round in previous sessions of the experiment, and (2) state which recommendation the computer sends when any given pot is the plant pot. Part (2) does not involve guessing; its purpose is to ensure you remember which recommendation the computer always sends.

ADVISORS ONLY

Before rounds 2, 8, 14 and 20, you will make guesses about the most-common decisions Choosers have made in that round in previous sessions of the experiment.

Your payment

CHOOSERS ONLY

One round of the game has been randomly selected to count for your payment. You will receive $16.00 if you produce a large plant in that round, $13.00 if you produce a small plant, and $10.00 if you produce no plant.

BELIEFS TREATMENT ONLY

One of the guessing tasks has also been randomly selected to count for your payment. You will receive $5.00 if your guesses about Advisors' recommendations, as well as your statements about the recommendation the computer sends, are correct.

ADVISORS ONLY

One round of the game has been randomly selected to count for your payment. If the Chooser you are paired with receives your recommendation in that round, you will receive $16.00 if they produce a large plant, $13.00 if they produce a small plant, and $10.00 if they produce no plant. If the Chooser you are paired with receives the computer's recommendation in that round, you will receive $10.00 regardless of the size of the plant they produce.

BELIEFS TREATMENT ONLY

One of the guessing tasks has also been randomly selected to count for your payment. You will receive $5.00 if your guesses are correct.

Quiz

On the next page, you will complete a quiz that tests your understanding of these instructions. You must answer all quiz questions correctly before proceeding to the first round. You will receive $1.00 if you answer all quiz questions correctly on your first try. You will not be able to return to the instructions while completing the quiz, so please ensure you have read them carefully.

Quiz

Question 1

When a sprinkler is turned on...

- ☐ only the pot closest to it receives any water.
- ☐ the pot closest to it receives the most water; the pot farthest from it receives the least.

Question 2

The computer always recommends to...

- ☐ stick with sprinkler 1.
- ☐ switch to sprinkler 2.
- ☐ switch to sprinkler 3.

CHOOSERS ONLY

Question 3

Which of these statements is true?

- ☐ Advisors have the same goal as you.
- ☐ Advisors choose a sprinkler to recommend for each possible plant pot.
- ☐ Advisors can recommend sticking with the default or switching to one of the other sprinklers.

Question 4

In each round, you will be...

- ☐ randomly matched to an Advisor.
- ☐ matched with the same Advisor.

ADVISORS ONLY

Question 3

After receiving a recommendation to switch sprinklers, a Chooser can turn on...

- ☐ the default sprinkler or the recommended sprinkler.
- ☐ any sprinkler.

Question 4

Which of these statements is true?

- ☐ When making their decisions, Choosers know their chance of receiving a recommendation from each potential source.
- ☐ Choosers know that Advisors have the same goal as them.
- ☐ Choosers know that Advisors choose a sprinkler to recommend for each possible plant pot.
- ☐ All of the above statements are true.

Question 5

In each round, you will be...

- ☐ randomly matched to a Chooser.
- ☐ matched with the same Chooser.

Question 6

Your task consists of three choices: you must choose a sprinkler to recommend to the Chooser for each possible plant pot (pot 1, 2 or 3). You should think of this task as if…

- ☐ you know a specific pot (pot 1, 2 or 3) is the plant pot, and with this knowledge, must recommend a sprinkler.
- ☐ you are sending recommendations without knowing which pot is the plant pot.

Quiz complete

Congratulations! You have answered all of the quiz questions correctly. Once all participants have done so, we will proceed to the practice round.

Round 1 (practice round)

CHOOSERS ONLY

A new round has begun. You have been randomly matched to an Advisor, and the plant pot and source of the recommendation shown to you have been determined.

ADVISORS ONLY

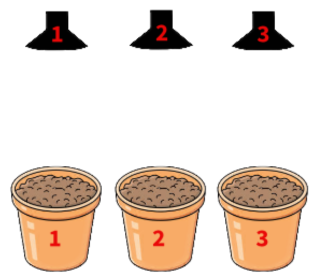
A new round has begun. You have been randomly matched to a Chooser, and the plant pot and source of the recommendation shown to that Chooser have been determined.

Round 1 (practice round)

CHOOSERS ONLY

Summary of instructions

- A seed will grow into a larger plant if it receives more water.
- Your goal (same as the Advisor's): Produce the largest possible plant.
- Your task: Decide whether, if recommended to switch to a given sprinkler, you will accept the proposed switch or stick with the default (sprinkler 1).
- There is a 1-in-3 chance you will receive the Advisor's recommendation, and a 2-in-3 chance you will receive the computer's recommendation. The Advisor chooses a sprinkler to recommend for each possible plant pot. The computer always recommends to switch to sprinkler 3.



9-in-20 7-in-20 4-in-20

The pots, the sprinklers and each pot's chance of being the plant pot.

Your task

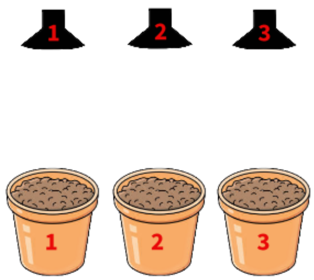
Please decide whether you will accept or reject the recommendation in each of the scenarios below.

- You receive a recommendation to switch to SPRINKLER 2. Will you accept or reject this recommendation?
  - ☐ Accept (switch to sprinkler 2)
  - ☐ Reject (stick with sprinkler 1)
- You receive a recommendation to switch to SPRINKLER 3. Will you accept or reject this recommendation?
  - ☐ Accept (switch to sprinkler 3)
  - ☐ Reject (stick with sprinkler 1)

Advisors only

Summary of instructions

- A seed will grow into a larger plant if it receives more water.
- Your goal (same as the Chooser's): Produce the largest possible plant.
- Your task: Choose a sprinkler to recommend to the Chooser for each possible plant pot. The Chooser decides whether, if recommended to switch to a given sprinkler, they will accept the proposed switch or stick with the default (sprinkler 1).
- There is a 1-in-3 chance the Chooser will receive your recommendation, and a 2-in-3 chance they will receive the computer's recommendation. The computer always recommends to switch to sprinkler 3.



9-in-20 7-in-20 4-in-20

The pots, the sprinklers and the information Choosers have (each pot's chance of being the plant pot)

Your task

Please choose a recommendation to send the Chooser in each of the scenarios below.

- You know that POT 1 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- You know that POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- You know that POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

Round 1 (practice round)

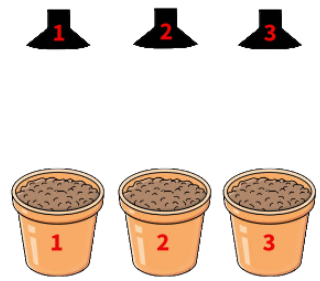
Feedback

Here is an outline of the feedback you will receive at the end of a round.

Choosers only

Advisor’s decisions: The Advisor you were paired with made the following decisions.

- If Pot 1 is the plant pot, recommend to *[stick with sprinkler 1/switch to sprinkler 2/switch to sprinkler 3]*
- If Pot 2 is the plant pot, recommend to *[stick with sprinkler 1/switch to sprinkler 2/switch to sprinkler 3]*
- If Pot 3 is the plant pot, recommend to *[stick with sprinkler 1/switch to sprinkler 2/switch to sprinkler 3]*



This image will additionally show the sprinkler that was turned on and the plant that was produced.

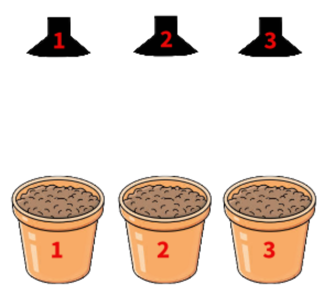
Outcome of this round: *Here, you will receive feedback on...*

- *which pot was the plant pot*
- *the source of the recommendation you received*
- *the sprinkler that was turned on*
- *the size of the plant produced*

Advisors only

Chooser’s decisions: The Chooser you were paired with made the following decisions.

- If recommended to switch to sprinkler 2, *[accept/reject]* this recommendation
- If recommended to switch to sprinkler 3, *[accept/reject]* this recommendation



This image will additionally show the sprinkler that was turned on and the plant that was produced.

Outcome of this round: *Here, you will receive feedback on...*

- *which pot was the plant pot*
- *the source of the recommendation the Chooser received*
- *the sprinkler that was turned on*
- *the size of the plant produced*

Practice round complete

Congratulations! You have completed the practice round. Once all participants have done so, we will proceed to the rest of the experiment.

MAIN TREATMENT ONLY

Recall that one of the rounds that follow has been randomly selected to count for your payment. You should therefore treat each round as if it is the one that counts.

BELIEFS TREATMENT ONLY

Recall that one of the rounds that follow, as well as one of the guessing tasks, have been randomly selected to count for your payment. You should therefore treat each round and guessing task as if it is the one that counts.

AUTHOR'S NOTE: Participants then played 21 more rounds of the Plant Pot Game. In rounds 2 through 21, the introduction and decision pages were the same as in the practice round. The structure of the feedback page was the same in rounds 2 through 22, but its content depended on the outcome of the round. Below I show an example of the feedback that players in each role may receive.

Round *n*

Feedback

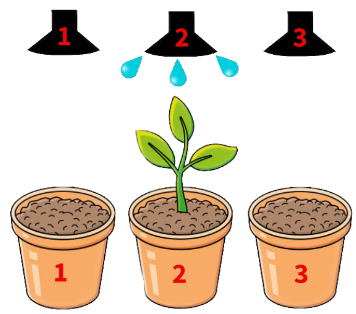
CHOOSERS ONLY

Advisor's decisions: The Advisor you were paired with made the following decisions.

- If Pot 1 is the plant pot, recommend to stick with sprinkler 1
- If Pot 2 is the plant pot, recommend to switch to sprinkler 2
- If Pot 3 is the plant pot, recommend to switch to sprinkler 3

Outcome of this round:

- Pot 2 was the plant pot.
- You received the Advisor's recommendation, which, as described above, was to switch to sprinkler 2.
- You accepted this recommendation, thereby switching to sprinkler 2.
- As a result, you produced a large plant.



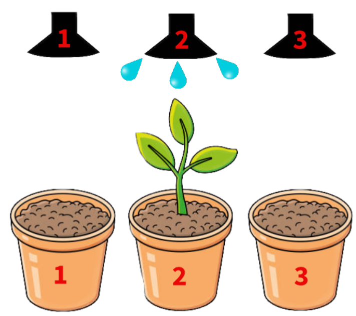
ADVISORS ONLY

Chooser's decisions: The Chooser you were paired with made the following decisions.

- If recommended to switch to sprinkler 2, accept this recommendation
- If recommended to switch to sprinkler 3, reject this recommendation

Outcome of this round:

- Pot 2 was the plant pot.
- The Chooser received your recommendation, which was to switch to sprinkler 2.
- As described above, the Chooser accepted this recommendation, thereby switching to sprinkler 2.
- As a result, the Chooser produced a large plant.



AUTHOR'S NOTE: In the beliefs treatment, participants viewed and completed the page below before the introduction page in rounds 2, 8, 14 and 20.

Guessing task for round m

CHOOSERS ONLY

Instructions

- **Your task:** We want you to think about the recommendations that tend to be sent in the round you are about to play, depending on which pot is the plant pot. You must (1) guess which recommendation has most-frequently been sent by Advisors when they knew a given pot was the plant pot in round m of previous sessions of this experiment, and (2) state which recommendation the computer sends when any given pot is the plant pot. Part (2) does not involve guessing; its purpose is to ensure you remember which recommendation the computer always sends.
- **Payment:** If this guessing task counts for your payment, you will be paid $5.00 if all of your guesses about Advisors' recommendations, as well as all of your statements about the recommendation the computer sends, are correct.

Your task

(1) In ROUND m, which recommendation is most-frequently sent by **Advisors** when they know...

- ... POT 1 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... POT 2 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... POT 3 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

(2) In ALL ROUNDS, which recommendation does **the computer** send when...

- ... POT 1 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... POT 2 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... POT 3 is the plant pot?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

ADVISORS ONLY

Instructions

- **Your task:** We want you to think about the decisions Choosers tend to make in the round you are about to play. You must guess Choosers' most-common response to a recommendation to switch to a given sprinkler in round m of previous sessions of this experiment.

- **Payment:** If this guessing task counts for your payment, you will be paid $5.00 if both of your guesses are correct.

Your task

In ROUND m, what is Choosers' most-common response to a recommendation to...

- ... switch to SPRINKLER 2?
  - ☐ Accept (switch to sprinkler 2)
  - ☐ Reject (stick with sprinkler 1)
- ... switch to SPRINKLER 3?
  - ☐ Accept (switch to sprinkler 3)
  - ☐ Reject (stick with sprinkler 1)

AUTHOR'S NOTE: In round 22, Choosers' introduction and decision pages were the same as in all previous rounds. These pages were different for Advisors (see below).

Round 22

AUTHOR'S NOTE: The two introduction sentences were the same as in all previous rounds.

*In this final round, you will be able to make different recommendations depending on (1) which pot is the plant pot **AND** (2) the Chooser's decision to accept or reject recommendations to switch sprinklers.*

Round 22

AUTHOR'S NOTE: The summary of the instructions was the same as in all previous rounds.

*In this final round, you will be able to make different recommendations depending on (1) which pot is the plant pot **AND** (2) the Chooser's decision to accept or reject recommendations to switch sprinklers.*

Your task

Please choose a recommendation to send the Chooser in each of the scenarios below.

You know the Chooser would ACCEPT to switch to SPRINKLER 2 and ACCEPT to switch to SPRINKLER 3, and...

- ... you know POT 1 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 3 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

You know the Chooser would ACCEPT to switch to SPRINKLER 2 and REJECT to switch to SPRINKLER 3, and...

- ... you know POT 1 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 3 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

You know the Chooser would REJECT to switch to SPRINKLER 2 and ACCEPT to switch to SPRINKLER 3, and...

- ... you know POT 1 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2

- ☐ Switch to sprinkler 3
- ... you know POT 3 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

You know the Chooser would REJECT to switch to SPRINKLER 2 and REJECT to switch to SPRINKLER 3, and...

- ... you know POT 1 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 2 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3
- ... you know POT 3 is the plant pot. What do you recommend to the Chooser?
  - ☐ Stick with sprinkler 1
  - ☐ Switch to sprinkler 2
  - ☐ Switch to sprinkler 3

MAIN TREATMENT ONLY

Rounds complete

Congratulations! You have now completed all rounds of the game. Click the "Next" button to proceed to the survey.

BELIEFS TREATMENT ONLY

Rounds and guessing tasks complete

Congratulations! You have now completed all rounds of the game and all of the guessing tasks. Click the "Next" button to proceed to the survey.

Survey (Page 1 of 2)

CHOOSERS ONLY

We're interested in knowing more about how you decided whether to accept a recommendation.

Advisors only

We're interested in knowing more about how you decided which recommendations to send.

Did your approach to this decision remain constant, or did it change as the rounds progressed?

- ☐ My approach remained constant.
- ☐ My approach changed as the rounds progressed.

Please explain why your approach did or did not change as the rounds progressed. [Text field where participant types answer]

Do you have any other thoughts or comments on the experiment? If not, leave this field blank. [Text field where participant types answer]

Survey (Page 2 of 2)

We want to know more about you! Please take a moment to answer these demographic questions.

What degree are you currently pursuing?

- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ Doctoral degree
- ☐ Other degree (not listed)

In what year did you begin the degree you are currently pursuing?

- ☐ 2024
- ☐ 2023
- ☐ 2022
- ☐ 2021
- ☐ 2020
- ☐ 2019
- ☐ 2018 or earlier

What is your primary field of study (e.g., Anthropology, Chemistry, Mechanical Engineering, etc.)? [Text field where participant types answer]

How old are you? [Text field where participant types answer]

What is your gender?

- ☐ Female
- ☐ Male
- ☐ Non-binary
- ☐ Other (not listed)
- ☐ Prefer not to answer

AUTHOR'S NOTE: Below I show an example of a participant's earnings page.

Your earnings

MAIN TREATMENT ONLY

Round 10 was randomly selected to count for your payment. As summarized in the table below, your total earnings are **$22.00**.

Earnings breakdown	
Show-up fee	$5.00
Earnings from quiz	$1.00
Earnings from round 10	$16.00
Total earnings	*$22.00*

BELIEFS TREATMENT ONLY

Round 10 and the guessing task for round 2 were was randomly selected to count for your payment. As summarized in the table below, your total earnings are **$27.00**.

Earnings breakdown	
Show-up fee	$5.00
Earnings from quiz	$1.00
Earnings from round 10	$16.00
Earnings from guessing task for round 2	$5.00
Total earnings	*$27.00*

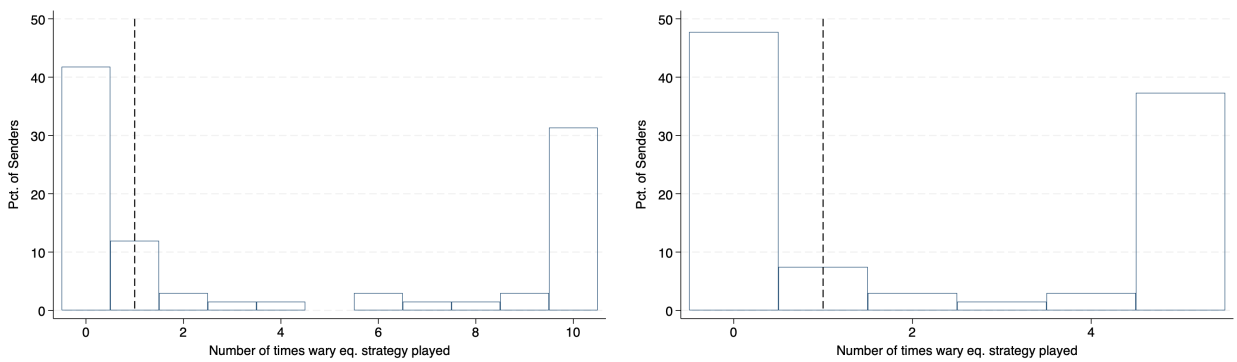
An experimenter will come by your desk shortly and pay you this amount in cash. Thanks for participating!

C Supplementary Analyses

C.1 Supplementary Tables and Figures

Figure C.1

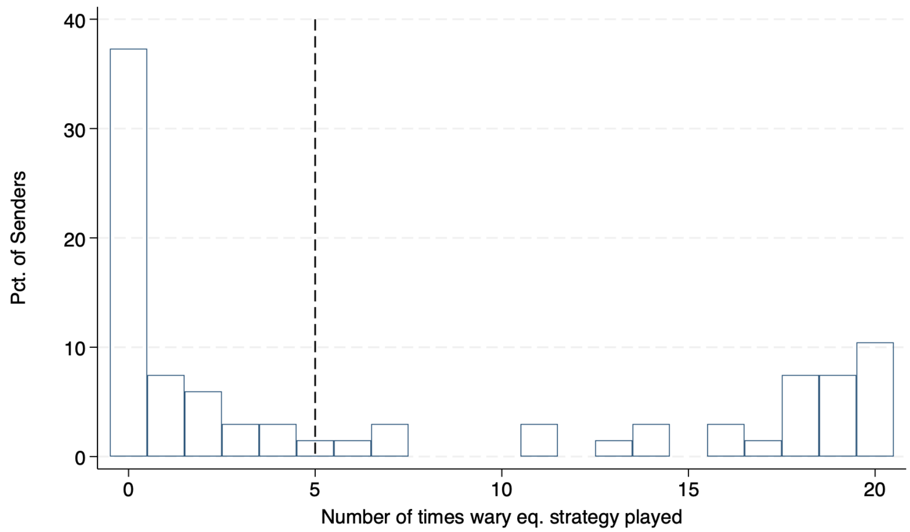
Senders' Frequency of Optimal Strategy Play: Main Treatment, Player-Level, Final Rounds



(a) Rounds 11 through 20 **(b)** Rounds 15 through 20

Notes. This figure displays the percent of Senders in the main treatment playing their wary equilibrium strategy in a given number of the final main rounds. Figure C.1a restricts the data to the final 10 rounds, and Figure C.1b restricts the data to the final 5 rounds.

Figure C.2
Senders' Frequency of Optimal Strategy Play: Main Treatment, Player-Level, Above-Median Feedback



Notes. The sample in this figure is restricted to main-treatment Senders who, by round 20, were paired with more Receivers that played accept-reject than the median Sender was. The median Sender was paired with such Receivers in 70% of the 20 rounds. The figure shows the percent of such main-treatment Senders who played their wary equilibrium strategy in a given number of rounds.

Table C.1
Participant Characteristics

	Main	Beliefs	p-value
Age	21.1	21.3	0.95
% women	60.0	67.4	0.21
% bachelor's degree program	85.9	80.6	0.25
Comprehension quiz mistakes	0.99	1.15	0.21

Notes. Mean participant characteristics by treatment. Excludes the data of one participant from each treatment who withdrew from their experimental session. The final column reports the p-value from a Wilcoxon-rank-sum test.

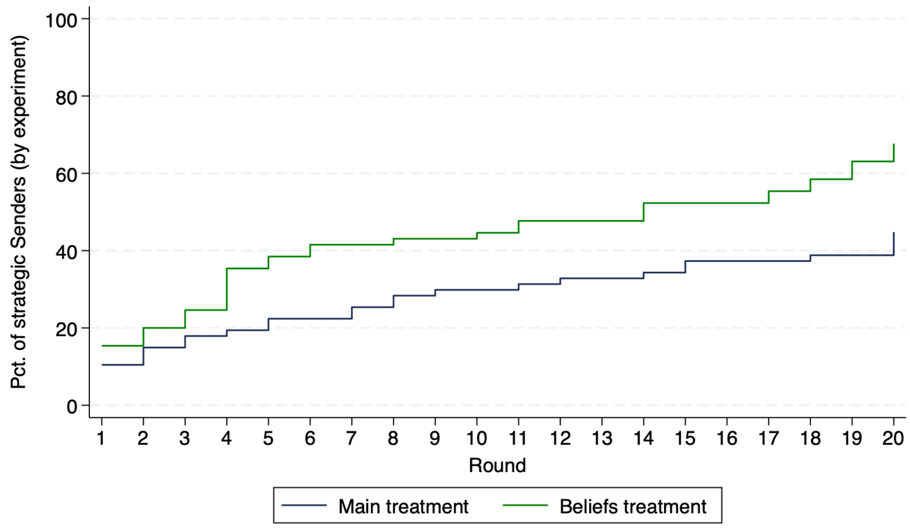
Table C.2
Receivers' Beliefs about Senders' Strategies (Beliefs Treatment)

	Round 1	Round 7	Round 13	Round 19
Partial truth-telling	14.0	26.6	45.3	50.0
Full truth-telling	75.0	65.6	48.4	45.3
Other strategy	10.9	7.8	6.3	4.7

Notes. Percent of Receivers in the beliefs treatment who believe that a given strategy is the modal Sender strategy in a given round.

Figure C.3

Percent of Strategic Senders (By Round and Treatment)



Notes. This figure displays the percent of strategic Senders in each round of each treatment. A Sender is categorized as "strategic" in round x if they play their optimal strategy (partial truth-telling) in round x and all subsequent rounds.

Table C.3
Evolution of Senders' Likelihood of Playing Partial Truth-Telling (Both Treatments)

Rounds	All	1-10	11-20
	(1)	(2)	(3)
Round	0.009***	0.013**	0.009**
	(0.003)	(0.006)	(0.004)
Beliefs treatment	0.164	0.128	0.239
	(0.145)	(0.145)	(0.197)
Beliefs treatment × round	0.004	0.019**	-0.004
	(0.341)	(0.010)	(0.005)
Constant	0.226**	0.157	0.280*
	(0.114)	(0.111)	(0.143)
Observations	2640	1320	1320
Mean of DV	0.470	0.420	0.520

Notes. Coefficients from OLS regressions conducted among Senders in both treatments, with standard errors in parentheses. The dependent variable is an indicator for playing partial truth-telling (Senders' wary equilibrium strategy) in a given round. Independent variables include the round number, an indicator for participating in the beliefs treatment, and the interaction of these variables. All regressions include controls for the session and standard errors clustered at the participant-level. * denotes $p < 0.10$, ** denotes $p < 0.05$, and *** denotes $p < 0.01$.

C.2 Discussion of Alternative Mechanisms

The analysis of the main treatment's results presented in Section 3.2 focuses on two main mechanisms that may have driven Senders' failure to send optimal recommendations: incorrect beliefs about Receivers' strategies and a failure of contingent thinking. In the current Section, I present additional analyses of the main treatment's results that explore two alternative mechanisms: lying costs and mixing between strategies. The analyses provide evidence against each of these mechanisms.

1. Lying costs

Prior work has shown that lying can be psychologically costly (see, e.g., Abeler et al. (2019)). If Senders had a high psychological lying cost, this cost might prevent them from playing a strategy that involves recommending an action that does not correspond to the true state, such as the partial truth-telling strategy.

I use data from the Receiver-first round of the main treatment to test for high psychological lying costs. If it were costly for Senders to play non-truth-telling strategies, Senders should play their full truth-telling strategy in response to each of the Receiver's strategies. However, only 6% of Senders do so. It therefore seems unlikely that lying costs were responsible for most Senders' failure to send optimal recommendations.

2. Mixing between strategies

Suppose Senders were mixing between their partial and full truth-telling strategies. For instance, this may have occurred if Senders believed that Receivers were mixing between their accept-reject and accept all strategies.

If Senders were mixing in this way, they should then play each of their equilibrium strategies (partial truth-telling and full truth-telling) in at least one of the 20 rounds. However, only 46% of Senders do so. We would also expect Senders to switch multiple times between their equilibrium strategies.[5] That said, among Senders who play each equilibrium strategy at least once, nearly half (45%) switch only once, 29% switch twice, and only 26% switch three or more times. Therefore, Senders do not appear to be playing full truth-telling more frequently than theoretically predicted because they are mixing between their equilibrium strategies.

[5]Switching only once would be consistent with a realization that a given strategy is the unique best response to the empirical distribution of Receivers' strategies.

References for Online Appendix

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica 87*(4), 1115–1153.

Johnson, E. (2017, March). 'I will do anything I can to make my goal': TD teller says customers pay price for 'unrealistic' sales targets. *CBC News*. Available at: https://www.cbc.ca/news/canada/british-columbia/td-tellers-desperate-to-meet-increasing-sales-goals-1.4006743.

Johnson, E., J. McDonald, M. McNair, K. Ivany, and M. McCann (2024, March). Hidden cameras capture bank employees misleading customers, pushing products that help sales targets. *CBC News*. Available at: https://www.cbc.ca/news/business/marketplace-hidden-camera-banks-1.7142427.