Working Paper Series
No. 2026/01
May 2026

CEEL Working Paper Series

The Effects of Non-Binding Promises on Sequential Cooperation

Experimental Evidence from a Multiplayer Centipede Game

Sandro Casal[1], Luigi Mittone[1], Matteo Ploner[1,*]

[1] Department of Economics and Management, University of Trento

[*] Corresponding author: Email: matteo.ploner@unitn.it

Abstract

We study whether non-binding promises foster cooperation in sequential environments in which surplus is realized only if a cooperative path is sustained. We report an incentivized online experiment based on a five-player centipede game with four conditions: no communication (*Baseline*), mandatory costless promises (*Cheap Talk*), voluntary costless promises (*Voluntary*), and voluntary costly promises (*Fee*). Mandatory promises substantially increase cooperation and the probability of full-chain completion: participants frequently promise to cooperate and largely follow through, despite the absence of enforcement. When promises are voluntary, fewer promises are made and cooperation does not increase significantly relative to the no-communication benchmark; adding a fee further reduces promise-making and leaves cooperation close to baseline. Overall, the results indicate that institutional features primarily matter through their effect on the incidence and type of promises, while the behavioral association between a cooperation promise and subsequent cooperation is similar across promise institutions.

Keywords: behavioral economics; experimental economics; decision making

JEL codes: C91; D91; J16

1 Introduction

Many economically relevant interactions are sequential: value is created only if a chain of agents takes cooperative actions in order. In supply chains, upstream investments pay off only if downstream partners subsequently exert effort and deliver quality. In organizations, costly actions today, such as sharing information, helping a colleague, and maintaining standards, generate benefits only if others reciprocate later. In such environments, early cooperation is exposed to later breakdown, and observed histories condition subsequent choices.

In the classical Centipede game, two players alternate between take and pass along a linear sequence of decision nodes, with payoffs increasing in the length of play until someone takes. Backward induction predicts taking at the first node, yielding an inefficient outcome. Experiments nevertheless document substantial deviations: in two-player settings, cooperation is frequently observed across a wide range of payoff structures and designs (see, among others, McKelvey and Palfrey, 1992; Fey et al., 1996; Krockow et al., 2016). These findings are commonly interpreted in terms of other-regarding preferences, belief-based motivations, and limited strategic reasoning. Despite this extensive evidence, the literature remains overwhelmingly focused on bilateral interactions, and much less is known about cooperation in Centipede games with more than two players.

Krockow et al. (2016) note in their comprehensive meta-analysis that experiments with more than two players are disproportionately scarce. This observation still largely holds, with only a few notable exceptions emerging more recently. The most influential multiplayer contribution remains Rapoport et al. (2003), which studies a three-player Centipede game and shows that cooperation becomes substantially less sustainable when moving from two to three players. Despite the importance of sequential cooperation in multi-agent environments, experimental work on multiplayer Centipede games has developed only slowly since the early 2000s, even when accounting for related Centipede-like settings (Murphy et al., 2006).

A related strand of work investigates how institutional features shape cooperation in sequential environments. Gerber and Wichardt (2010) examine whether cooperation in a two-player Centipede game can be sustained through (i) bonuses paid to the other player for not terminating the game and (ii) insurance against the other player's termination. They find that bonus schemes are more direct and more effective than insurance. Using a similar setting, Krockow et al. (2018) study whether cultural context affects the use and effectiveness of commitment devices by comparing participants from the UK and Japan. They find higher cooperation

and greater take-up of commitment devices among Japanese participants; moreover, declining to adopt such a device is interpreted as a signal of non-cooperative intent, particularly in Japan.

Evidence from social-dilemma games suggests that communication—even when non-binding—can raise cooperation (see Sally, 1995). Subsequent reviews confirm this pattern: Balliet (2010), for instance, provides a meta-analysis showing that communication increases cooperation on average, with face-to-face interaction being particularly effective. In addition, Crawford (1998) emphasizes that cheap talk can reduce strategic uncertainty and facilitate coordination by allowing players to signal intentions and reassure others.

A common interpretation is that communication works because it enables credible signals of cooperative intent, even when messages are costless and unenforceable. In many laboratory settings, participants do not primarily use such messages deceptively; instead, they use them to establish trust and coordinate on cooperative play. One reason is that breaking a promise can be psychologically costly, due to aversion to lying and to disappointing others. Consistent with this view, Abeler et al. (2019) provide empirical and theoretical evidence on lying costs in a die-under-the-cup environment.

Promises may also operate through the formation of beliefs. Battigalli and Dufwenberg (2007) argue that anticipated guilt from letting others down creates an incentive to keep promises, thereby increasing the credibility of commitments and promoting cooperation. Experimental evidence that promises raise second-order expectations and causally increase promise keeping is reported by Charness and Dufwenberg (2006) and Ederer and Stremitzer (2017). By contrast, Vanberg (2008) argues that promise keeping is driven less by belief-based motives and more by a commitment rationale. Finally, Schwartz et al. (2019) show that effectiveness depends on communication: purely internal commitments, absent transmission to others, do not reliably increase cooperation.

Our study complements and extends this literature by examining how institutional design affects the use and credibility of promises, and how these effects translate into cooperation in a multiplayer sequential environment. We manipulate both the freedom to make promises and the cost of doing so. In a coordination context, Blume et al. (2017) compare costly and costless pre-play messages and find that free communication improves coordination. Introducing a small cost reduces message use but can still sustain high coordination, whereas higher costs sharply reduce both communication and coordination. With respect to participation, Hurkens and Schlag (2002) show that allowing agents to opt out of pre-play talk can change equilibrium

selection and efficiency: welfare depends not only on message content but also on the rule governing willingness to communicate.

We bring these insights to a sequential cooperation setting. Specifically, we study a novel five-player centipede game and compare mandatory successor-to-predecessor promises, voluntary promises, and costly (fee-based) promises. This design identifies how endogenizing participation—together with the informational meaning of silence—shapes cooperation along an extended sequential chain.

Our main finding is that requiring players to send a promise substantially increases cooperation and the probability of full-chain completion. Mandatory promises induce many participants to state an intention to cooperate, and most follow through despite the promise being unenforceable. By contrast, when promises are voluntary, many participants remain silent and cooperation is not significantly above the no-communication benchmark; making promises costly further reduces promise-making and leaves cooperation close to baseline. Taken together, the results suggest that institutions promoting promises matter primarily through selection into promising and the informational content of silence, rather than through differences in behavior conditional on observing a cooperation promise.

2 Methods

2.1 Main Task

The experiment is based on a multiplayer centipede game, which we interpret as a stylized supply chain (see Figure 1). There are five players, denoted P_1, P_2, P_3, P_4, P_5, who are ordered in a fixed sequence of nodes $1, \ldots, 5$. Player P_i acts at node i and chooses between two actions: *Take*, which terminates the game, and *Pass*, which allows the game to continue to the next node.

If player P_i chooses *Take*, the game ends immediately for all players. In this case, player P_i receives $10 \times i$ points, while all other players receive 5 points each. If all five players choose *Pass*, the final payoff equals the sum of the potential *Take* payoffs along the chain, divided equally among all five players. Formally, if all players pass, each player P_i receives $\pi_{P_i} = \frac{\sum_{j=1}^{5} 10 \times j}{5}$ for $i = 1, \ldots, 5$.

We employ the strategy method to collect choices along the chain. Each participant is assigned the role of one of the players P_i and is informed about the position i at which she would act. Conditional on being in the role of P_i, the participant is asked to choose whether she would *Take* or *Pass* at that node. Players choose "as if" their node was reached, and they are informed that their choice will be

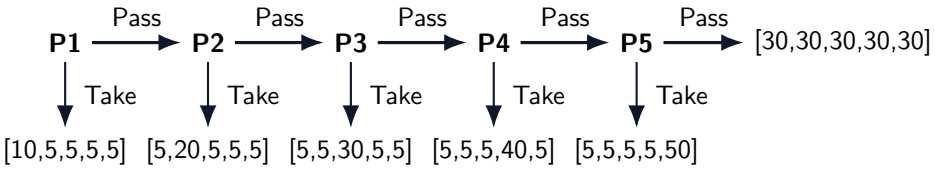


Figure 1: Centipede Game Structure with Five Players

implemented if the game reaches their node. This protocol ensures that all choices are incentive-compatible and made by the actual participants in the relevant role.

A visual aid illustrates the game's structure and the monetary consequences of the two actions at the relevant decision node (see Appendix C.2 for an example of how the information was displayed to the participant). After all decisions have been collected, we form mutually exclusive chains of five participants (corresponding to the roles $P_1, \ldots, P_5$) by matching strategy profiles and compute payoffs based on the induced play of the game.

The procedure described above represents the structure of our *Baseline* treatment. In the other experimental conditions, described in detail below, participants also make decisions about whether and how to communicate with their predecessor before making their own *Take*/*Pass* choice. In all treatments with communication, we introduce a simple messaging stage. At each node i with $i \geq 2$, before making the *Take*/*Pass* choice, player P_i must/can send a message to the immediate *predecessor* in the chain, P_{i-1}. After sending a message, but before choosing her own action, each player P_i (for $i \leq 4$) observes the message sent by her *successor* P_{i+1}.

Together with choices and promises, we also elicit beliefs. In particular, participant P_i reports (i) their beliefs about the choice of player P_{i-1} for $i \geq 2$, and (ii) their beliefs about the choice of player P_{i+1} for $i \leq 4$. Beliefs are elicited on a continuous scale from 0% to 100%, where 0% indicates certainty that the player will *Take*, and 100% indicates certainty that the player will *Pass* (see Appendix C.2 for the display). Belief elicitation is not incentivized. This methodological choice is also motivated by evidence that even theoretically sound elicitation methods such as BSR, although incentive-compatible for a broad range of preferences, can generate behavioral distortions in practice, including greater center-biased reporting when quantitative incentive information is provided (Danz et al., 2022).

2.2 Treatments

We implement four between-subjects treatments: the treatments differ in the availability and nature of communication between the players, while keeping the underlying centipede game and payoffs fixed.

As mentioned, in the *Baseline* treatment (B), there is no communication. Groups of five participants play the centipede game as described above. Each player P_i simply makes a *Take* or *Pass* decision at her node, and no messages are exchanged. This treatment serves as a benchmark for the level of cooperation (i.e., passing) that can be sustained without any communication.

In the *Cheap Talk* treatment (C), we introduce costless, mandatory communication in the form of non-binding promises. Players P_i for $i \geq 2$ send a message to their immediate predecessor in the chain (P_{i-1}), before making their own decision. The content of the message is restricted to a simple promise about the sender's intended action: *"I promise to Pass"* or *"I promise to Take"*. Sending such a promise is compulsory for $P_2, \ldots, P_5$, and the corresponding predecessors $P_1, \ldots, P_4$ always observe the promise made by their immediate successor. Promises are explicitly described as non-binding: they do not modify the material payoff structure and can be freely broken.

In the *Voluntary* treatment (V), the communication protocol is similar, but sending a promise is endogenous. Again, players $P_2, \ldots, P_5$ have the possibility to send a promise about their intended action to their predecessor, choosing between the messages "I promise to Pass" and "I promise to Take". However, unlike in C, sending any message is optional. A player P_i for $i \geq 2$ can either send a promise or remain silent. The predecessors $P_1, \ldots, P_4$ are informed not only about the content of any promise received but also about whether their successor chose to send a promise at all. Thus, the absence of a message is itself informative and may be interpreted by the predecessor.

In the *Fee* treatment (F), communication remains voluntary but becomes costly. The structure of messages is as in V: players $P_2, \ldots, P_5$ may send either *Pass* or *Take* to their predecessor, or choose not to send any message. As in V, predecessors observe both the presence or absence of a promise and its content when present. In contrast to V, however, sending a promise requires the sender to pay a fixed fee of 5 points, which is deducted from her final payoff. The fee level is set at 5 points to create a meaningful cost of communication while still allowing for the possibility of sending a promise without incurring a negative payoff if others choose to *Take*. The fee applies irrespective of the content of the promise. As in other communication treatments, promises are non-binding and can be broken without incurring a direct

monetary penalty.

The three communication treatments form a progression along two dimensions: obligation and cost. In C, every player must send a promise, so the act of promising carries no informational value about willingness to communicate; however, lying aversion can still make the *content* of the promise informative. In V, the decision to speak is itself a signal, because silence is an available and observable alternative. In F, the additional fee further screens senders, so that only those with sufficiently strong cooperative intent find it worthwhile to promise. By jointly varying obligation and cost, the design allows us to disentangle how the *incidence* of promises and their *effectiveness* shape beliefs and cooperative behavior along the chain.

2.3 Hypotheses

In this section, we state our preregistered hypotheses, emphasizing the economic intuition behind each prediction. A technical conceptual framework—starting from the standard unraveling benchmark in the finite centipede and extending to a psychological-game model with lying aversion, guilt aversion, and menu-dependent credibility of promises—is provided in Appendix A.

The key insight from the framework is that aggregate cooperation can be decomposed into two margins: i) promise *incidence* (how often a *Pass* promise is made) and ii) promise *effectiveness* (how strongly a *Pass* promise induces the counterpart to pass). Importantly, institutional features can move these margins in opposite directions.

We derive our predictions based on two fundamental assumptions regarding promise communication: first, we assume that players intending to choose *Pass* would not rationally make a *Take*-promise; second, we assume that in treatments where communication is voluntary and free, players who intend to choose *Pass* have no incentive to withhold the corresponding signal. These premises lead to the hypotheses discussed below.

Hypothesis 1 concerns the incidence of *Pass* messages across promise regimes. The intuition is that promises are disciplined by a dislike of lying. Under both *Cheap Talk* (mandatory messages) and *Voluntary* (optional messages), *Pass* should be chosen predominantly by participants who genuinely intend to pass, implying similar pass-promise frequencies across these two treatments (Hypothesis 1.a), assuming that the share of genuinely cooperative types is the same across treatments. In *Fee*, sending any message is costly; therefore, even some participants who intend to pass may prefer not to pay to communicate, reducing the frequency of *Pass* relative to *Voluntary* (Hypothesis 1.b).

Hypothesis 2 concerns overall cooperation, measured by *Pass* choices in the centipede, relative to *Baseline*. The intuition is that allowing promises facilitates sequential cooperation by strengthening commitment and reciprocity along the chain: promises can increase willingness to pass both because senders are more likely to follow through and because recipients are more willing to pass when they expect higher continuation. Accordingly, each promise treatment is predicted to exhibit higher *Pass* rates than *Baseline* (Hypotheses 2.a-2.c).

Hypothesis 3 concerns the conditional impact of receiving a *Pass* message. The key intuition is that the credibility of the same message depends on the sender's available alternatives. In *Cheap Talk*, the sender must choose between *Pass* and *Take*, so a cooperative message may be perceived as less diagnostic because the sender cannot remain silent. In *Voluntary*, silence is available, so choosing to send *Pass* is a more deliberate signal, and should induce a stronger cooperative response (Hypothesis 3.a). In *Fee*, the recipient additionally knows that the sender paid to communicate; a paid *Pass* is therefore expected to be more credible and to raise beliefs more, triggering an even stronger cooperative response than in *Voluntary* (Hypothesis 3.b).

Hypothesis 1 (*Promises*)**:**

> **Hypothesis 1.a:** The frequency of a *Pass* promise does not differ between *Voluntary* and *Cheap Talk*.
>
> **Hypothesis 1.b:** The frequency of a *Pass* promise is lower in *Fee* than in *Voluntary*.

Hypothesis 2 (*Cooperation*)**:** The possibility of making promises increases cooperation relative to the *Baseline*, in which no promises can be made.

> **Hypothesis 2.a:** The frequency of *Pass* choices is larger in *Cheap Talk* than in *Baseline*.
>
> **Hypothesis 2.b:** The frequency of *Pass* choices is larger in *Voluntary* than in *Baseline*.
>
> **Hypothesis 2.c:** The frequency of *Pass* choices is larger in *Fee* than in *Baseline*.

Hypothesis 3 (*Cooperation and Promises*)**:** The impact of a *Pass* promise is conditional upon the alternatives available.

Hypothesis 3.a: A *Pass* promise in the *Voluntary* condition is more likely to trigger a *Pass* choice in the counterpart than in the *Cheap Talk* condition.

Hypothesis 3.b: A *Pass* promise in the *Fee* condition is more likely to trigger a *Pass* choice in the counterpart than in the *Voluntary* condition.

These hypotheses map onto two distinct margins emphasized by the framework: promise incidence (how often a *Pass* promise is made) and promise effectiveness (how strongly a *Pass* promise induces the counterpart to pass). Allowing silence (*Voluntary*) and introducing a fee (*Fee*) may reduce the incidence of *Pass* promises while increasing their informational content and thus their effectiveness when observed. Aggregate cooperation, therefore, reflects the interaction between how often *Pass* promises are made and how powerful they are, conditional on being made. Since the net effect is theoretically ambiguous ex ante, we do not preregister sharp cross-treatment predictions for overall cooperation beyond comparisons to *Baseline*, and instead assess these net effects in the analysis.

2.4 Participants and Procedures

The main elements of the experiment were preregistered on the Open Science Framework (DOI: 10.17605/OSF.IO/NYT2U). The accompanying preregistration documents provide full details on the experimental design, the hypotheses, the planned sample size, and the pre-specified analysis.

Participants were recruited via the online platform Prolific. Eligibility criteria required that participants be residents of the United States, at least 18 years old, have at least a secondary education, be fluent in English, have normal or corrected-to-normal vision, and have an approval rate between 90% and 100%. The planned sample size was 3000 participants, with 750 participants per treatment (B, C, V, and F).

Data were collected in October 2025 over several sessions. After applying the preregistered exclusion criteria, the final sample comprised 3003 participants, with $N_B = 753$ in treatment B, $N_C = 750$ in treatment C, $N_V = 750$ in treatment V, and $N_F = 750$ in treatment F.[1]

Participants received a fixed participation payment of £1. In addition, they received a bonus determined by their earnings in the centipede game, computed from their points and converted into GBP at the fixed rate of 0.05 GBP per point. The

[1]The *Baseline* sample includes three additional observations beyond the preregistered target because of a minor Prolific sampling error that released extra places before we could close the study.

implied hourly rate was approximately £20.27, given a median completion time of about 3 minutes.

The experiment was programmed in oTree (Chen et al., 2016) and implemented online on a dedicated server (see Appendix C for more details). Upon accessing the study link via Prolific, participants first read an informed consent statement and provided their consent to proceed. They were then randomly assigned to one of the four treatments. Randomization occurred at the session level: each experimental session implemented a single treatment condition, and participants joining the session were assigned to that treatment.

Participants received on-screen instructions describing the sequence of moves in the five-player centipede game, the *Take*/*Pass* decision at each node, and the resulting payoffs, the conversion rate from points to money, and the communication protocol relevant to their treatment. In the *Baseline* treatment B, no messages were mentioned. In the communication treatments, the instructions explained the nature of promises (non-binding), whether they were mandatory or voluntary, and, in the *Fee* treatment F, the cost of sending a promise. After reading the instructions, participants had to pass a comprehension check before proceeding to the decision stage. In addition, each screen was subject to a generous time limit that gave participants ample time to respond; if the timer expired without any action, the participant was excluded from the study and replaced.

After reading the instructions, participants proceeded to the strategy-method stage. Each participant, in the role of a specific player P_i, indicated whether she would *Take* or *Pass* at her node. In treatments C, V, and F, players $P_2, \ldots, P_5$ additionally decided whether and how to communicate with their predecessor, according to the rules of the respective treatment, before making their own *Take*/*Pass* choice. Following the communication stage, each player observed any message sent by her successor and then reported her beliefs as described above.

Crucially, every message and choice displayed during the experiment originated from a real participant rather than being hypothetical or computer-generated. To achieve this, we elicited communication and strategy choices in reverse order along the chain, starting with P_5 and proceeding to P_1, so that each predecessor could be presented with the actual message chosen by her successor (when applicable) before stating her own beliefs and actions. A matching algorithm in oTree then assigned participants to non-overlapping five-person chains, ensuring full incentive compatibility of the information shown.

At the end of the experiment, participants were informed about their earnings in points and GBP, were redirected to Prolific, and received their fixed participation

payment plus the bonus determined by their (and their matched counterparts') choices in the game.

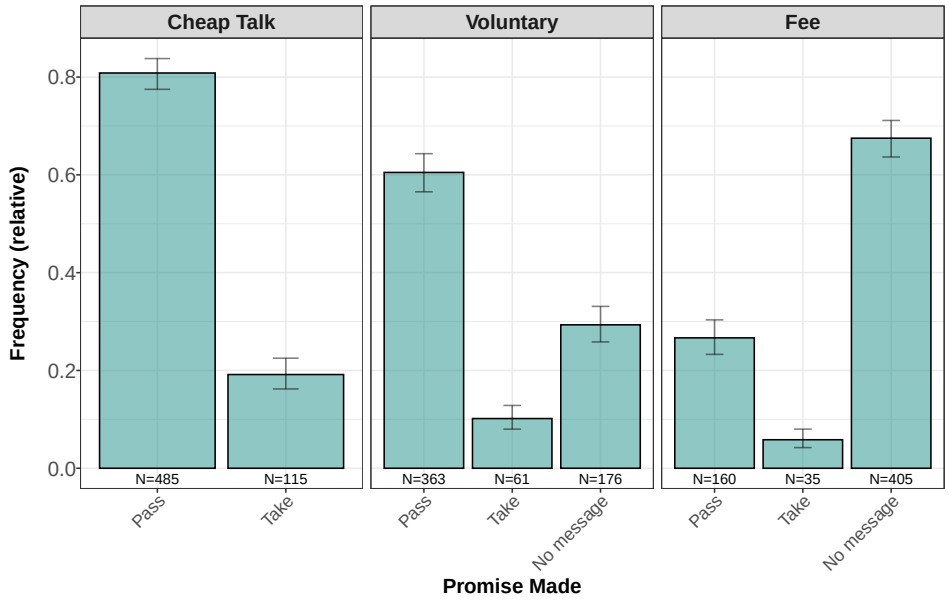
3 Results

We present the main results of our experiment, organized according to the structure of our hypotheses. First, we analyze the frequency of promises made across treatments. Second, we examine the level of cooperation across treatments. Third, we investigate how promises influence cooperation.

3.1 Promises

We begin by analyzing the frequency of promises made across treatments. We consider conditions *Cheap Talk*, *Voluntary*, and *Fee*, where promise communication was available before participants made their choices. In *Cheap Talk*, message-eligible participants had to promise either *Pass* or *Take*. In the *Voluntary* and *Fee* conditions, they could choose between three options: promising to *Pass*, promising to *Take*, or "No message" (i.e., not making any promise).

Figure 2 shows the proportion of promises by type and treatment. The bar plot displays the estimated proportions, along with 95% confidence intervals, for each promise type within each treatment condition (proportion test).



Note. Each panel displays a treatment; within a panel, every bar is the sample proportion of a promise made, with whiskers marking the 95% binomial confidence interval.

Figure 2: Proportion of Promises Made by Treatment

In the *Cheap Talk* condition, 81% of message-eligible players make a *Pass* promise, while in the *Voluntary* and *Fee* conditions, the frequency of *Pass* promises is much lower, around 60% and 27%, respectively. The frequency of *Take* promises is larger in *Cheap Talk* (19%) compared to *Voluntary* (10%) and *Fee* (6%). Accordingly, many participants chose *No message* in the *Voluntary* and *Fee* conditions, 29% and 68% respectively. The fact that the share of *Take* promises remains relatively stable between *Voluntary* and *Fee* suggests that the sharp increase in *No message* outcomes is consistent with the interpretation that some participants who would otherwise choose *Pass* find the communication fee too high to justify sending the message.

A series of pairwise proportion tests reveals that, for a given promise type, treatments differ significantly in the frequency of promises made (all p-values $<$ 0.008). As specified in our preregistration, we correct the p-values to control for potential false discoveries using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995).

Taken together, these patterns yield our first set of results, which we interpret in light of Hypothesis 1.

Result 1 *(Promises)***.**

Result 1.a: *In contrast to Hypothesis 1.a, the frequency of Pass promises differs between Voluntary and Cheap Talk treatments, with more Pass promises in Cheap Talk.*

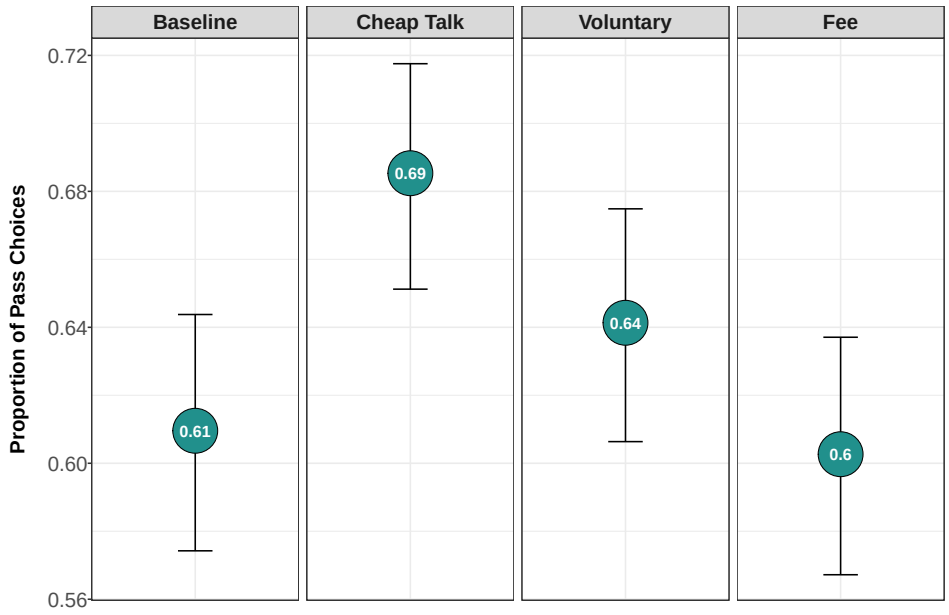
Result 1.b: *In line with Hypothesis 1.b, the frequency of Pass promises is lower in Fee than in the Voluntary treatment.*

In Appendix B.1, we illustrate how credible non-binding promises are perceived to be, from both the receiver's and the promisor's perspective, using non-incentivized beliefs. When participants receive a *Pass* promise, they report substantially higher expected probabilities that the successor will actually pass (about 66–70% across treatments), with largely overlapping 95% confidence intervals, indicating similar perceived credibility in all conditions. By contrast, in the no-message case, beliefs that the successor will pass are much lower (roughly 42–45%), and a *Take* promise reduces beliefs further (around 20–29%).

The same qualitative pattern holds when participants consider how their own messages affect predecessors: making a *Pass* promise is associated with higher expected probabilities that the predecessor will pass (about 63–64% across treatments), again with overlapping confidence intervals. Both *Take* promises and no message are associated with notably lower expected predecessor cooperation. Overall, *Pass* promises are perceived as credible and informative signals of cooperative intent, while *Take* promises and silence are interpreted as weaker cooperative signals, and these perceptions are similar across treatments.

3.2 Cooperation

We now turn to the analysis of cooperation, measured as the frequency of *Pass* choices across treatments. Figure 3 shows the proportion of *Pass* choices by treatment with 95% confidence intervals of a proportion test.

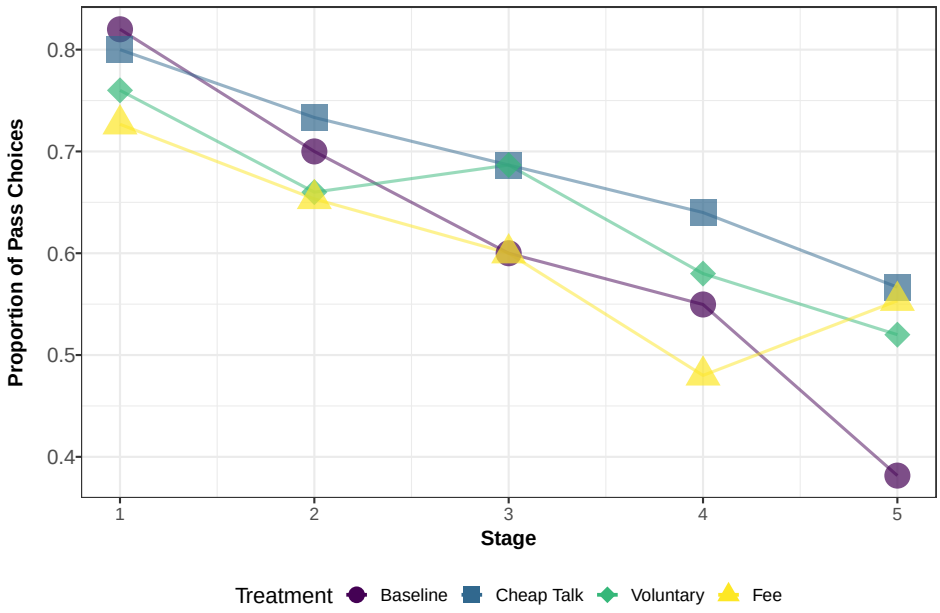


Note. Each panel displays one treatment: the large dot represents the estimated *Pass* share, the whisker marks its 95% confidence interval (proportion test), and the printed value is the point estimate.

Figure 3: Cooperation Rates by Treatment

As Figure 3 shows, in all conditions there is a clear majority of *Pass* choices. The *Cheap Talk* treatment has the highest proportion of *Pass* choices, followed by the *Voluntary* treatment, the *Baseline*, and finally the *Fee* treatment. Moreover, the *Cheap Talk* treatment significantly differs from both the *Baseline* (proportion test corrected for false discovery rate, p-value = 0.006) and the *Fee* (proportion test corrected for false discovery rate, p-value = 0.005) treatments.

To gain an understanding of possible dynamics at play, Figure 4 shows the proportion of *Pass* choices by stage and treatment.



Note. The stage is plotted on the x-axis, and the *Pass* proportion is plotted on the y-axis. Each treatment is represented by a separate colored line connecting stage-specific estimates (different markers by treatment).

Figure 4: Proportion of *Pass* Choices by Stage and Treatment

The results indicate that the proportion of *Pass* choices is quite similar for stage 1 and tends to decrease as the stage increases across all treatments. However, the *Cheap Talk* is generally among the highest across stages. Notably, the *Baseline* shows a marked drop in *Pass* choices at the later stages, suggesting that cooperation is harder to sustain at later nodes without the support of promises.

Given these cooperation probabilities, the probability that a chain closes (i.e., all five players choose *Pass*) is highest in the *Cheap Talk* treatment (14.6%), followed by the *Voluntary* treatment (10.4%) and the *Fee* treatment (7.6%), and lowest in the *Baseline* (7.2%).

Thus, our data provide partial support for Hypothesis 2, indicating that the possibility of making a promise increases cooperation compared to the baseline with no promises, but only in the *Cheap Talk* treatment. Specifically:

Result 2 *(Cooperation)***.**

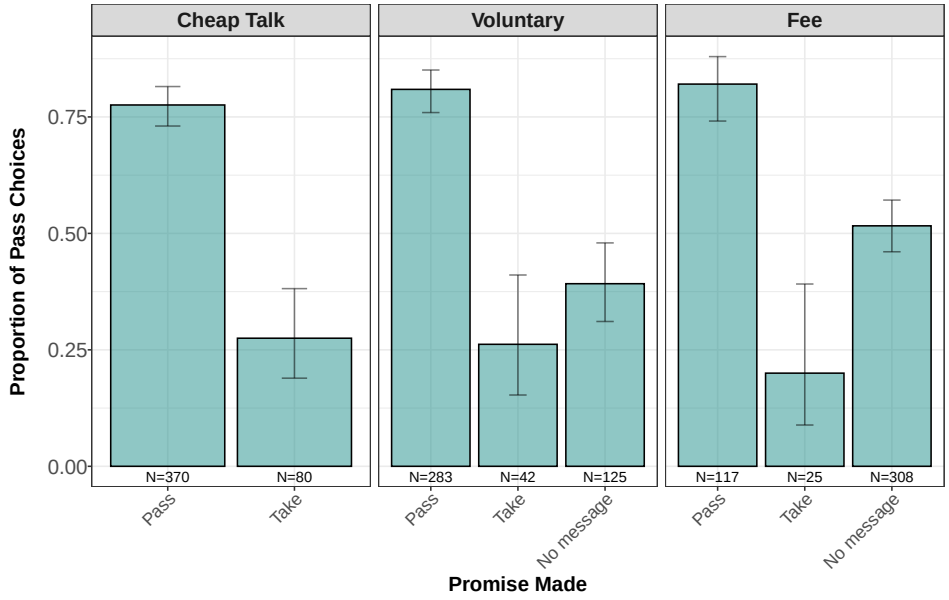
Result 2.a: *In line with Hypothesis 2.a, the frequency of Pass choices is significantly larger in Cheap Talk than in Baseline.*

Result 2.b: *In contrast to Hypothesis 2.b, the frequency of Pass choices does not differ significantly between Voluntary and Baseline.*

Result 2.c: *In contrast to Hypothesis 2.c, the frequency of Pass choices does not differ significantly between Fee and Baseline.*

3.3 Cooperation and Promises

In this section, we analyze how promises influence cooperation. First, we examine whether participants fulfilled their promises — that is, whether they chose *Pass* after making a promise to do so. Second, we investigate whether a *Pass* promise from the successor in turn increases the likelihood of choosing *Pass*. As anticipated above, the analysis in Appendix B.1 shows that *Pass* promises are generally taken as a credible signal of cooperative attitudes.



Note. Each panel is a treatment; within a panel, every bar is the proportion of *Pass* choices given the promise made (x-axis), with whiskers marking the 95% binomial confidence interval.

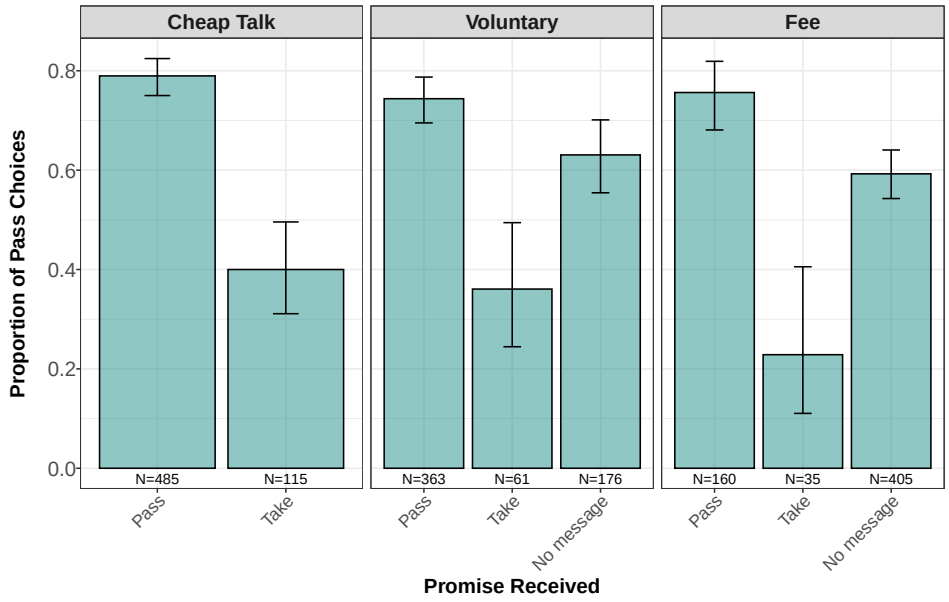
Figure 5: Proportion of *Pass* Choices by Promise Made and Treatment

Figure 5 shows that behavior is strongly related to the type of message sent. Participants who promised *Pass* overwhelmingly chose *Pass* in every messaging

treatment, with *Pass* rates of 77.6% in *Cheap Talk*, 80.9% in *Voluntary*, and 82.1% in *Fee*. By contrast, among participants who promised Take, *Pass* choices were considerably less frequent: 27.5% in *Cheap Talk*, 26.2% in *Voluntary*, and 20.0% in *Fee*. In absolute terms, such cases were rare, amounting to only 22, 11, and 5 observations, respectively. This pattern is consistent with random error or occasional misunderstanding. Silence was possible only in treatments in which participants could refrain from sending a message. In *Voluntary*, those who sent no message chose *Pass* relatively rarely (39.2%), whereas in *Fee* they did so about half the time (51.6%). Compared with *Voluntary*, silence is therefore less strongly associated with low cooperation when messaging is costly.

A series of pairwise proportion tests reveals that for all types of promises, there are no significant differences in the likelihood of choosing *Pass* between treatments (all p-values > 0.133, corrected for false discovery rate). Thus, different treatments seem not to affect the commitment to honor a *Pass* promise.

Finally, to assess whether receiving a *Pass* promise increases the likelihood of choosing *Pass*, Figure 6 displays the proportion of *Pass* choices by promise received and treatment.



Note. Each panel is a treatment; within a panel, every bar is the proportion of *Pass* choices given the promise received (x-axis), with whiskers marking the 95% binomial confidence interval.

Figure 6: Proportion of *Pass* Choices by Promise Received and Treatment

As the figure shows, receiving a *Pass* promise from the successor is associated with a substantially higher likelihood of choosing *Pass* in every treatment. *Pass* rates following a *Pass* promise are consistently high—around 75–80%. By contrast, when participants receive a *Take* promise, *Pass* rates are much lower, typically at or below 40%. Notably, cooperation does not fall to zero after a *Take* promise; the small amount of passing in this condition is consistent with noise or occasional misunderstanding among the few participants exposed to a *Take* promise. Finally, silence appears to sustain moderate levels of cooperation: in both the *Voluntary* and *Fee* treatments, *Pass* choices following no message are lower than after a *Pass* promise but remain substantial, at roughly 60%.

A series of pairwise proportion tests indicates that for *Pass* promises received, there are no significant differences in the likelihood of choosing *Pass* between treatments (all p-values > 0.116). Thus, different treatments seem not to affect the credibility of a *Pass* promise in inducing cooperation. This leads to our final set of results, with reference to hypotheses 3.a and 3.b:

Result 3 *(Cooperation and Promises)***.**

Result 3.a: *In contrast to Hypothesis 3.a, a Pass promise in the Voluntary condition is not more likely to trigger a Pass choice in the counterpart than in the Cheap Talk condition.*

Result 3.b: *In contrast to Hypothesis 3.b, a Pass promise in the Fee condition is not more likely to trigger a Pass choice in the counterpart than in the Voluntary condition.*

In the Appendix (Section B.2), we complement the previous analyses with regression models to further investigate the impact of promises on cooperation while controlling for potential confounding variables. Overall, regression outcomes confirm the main results presented above, after controlling for socio-demographic variables.

4 Discussion and Conclusion

This paper studies whether non-binding promises can sustain cooperation in a sequential choice environment where surplus is realized only if a cooperative path is completed. Such sequential interdependence is characteristic of many economically relevant settings, including multi-step production, supply chains, and organizational workflows, where early cooperative investments are exposed to later breakdown and observed histories condition subsequent choices.

The results indicate that institutional design affects cooperation primarily by changing how often participants make commitment statements. When communication is mandatory and costless (*Cheap Talk*), most participants issue *Pass* promises and aggregate cooperation increases markedly relative to *Baseline*. When promises are optional (*Voluntary*) or costly (*Fee*), many participants choose not to promise, and cooperation does not increase significantly relative to the no-communication benchmark. This suggests that mandatory promising "crowds in" cooperation from participants who, when given discretion, would otherwise avoid making any promise. This pattern was not anticipated in our setting and suggests that mandatory communication may have changed the normative frame of the decision problem.

By requiring every participant to take an explicit position before acting, the institution removed the option of remaining strategically or morally ambiguous. Participants who selected "I promise to Pass" then faced a subsequent choice in which taking was materially tempting but inconsistent with a self-declared commitment. One possible channel is self-image: breaking a declared promise may create psychological discomfort because the action conflicts with the participant's stated intention to cooperate. A desire to maintain a coherent self-image, or to avoid the

psychological cost of lying, may therefore have encouraged follow-through even in the absence of material enforcement. This interpretation is consistent with evidence on lying aversion and self-signaling (e.g., Abeler et al., 2019; Bénabou and Tirole, 2016).

The mandatory declaration may also have operated as an institutional cue: because the communication stage was imposed by the experimental design, participants may have read it as part of the rules of the game and as a signal that cooperation was the appropriate conduct. This institutional-cue interpretation is consistent with accounts of social norms in which compliance depends on empirical and normative expectations (Bicchieri, 2006, 2017), with evidence that conventions can acquire normative force and stabilize behavior even when material incentives favor deviation (Guala and Mittone, 2010), and with experimental work showing that individual sensitivity to social norms helps explain prosocial behavior (Kimbrough and Vostroknutov, 2016). These mechanisms are not mutually exclusive: mandatory promises may both make self-consistency salient and signal that cooperation is normatively expected.

A second result is that, conditional on being made, promises are highly credible across all institutional designs. *Pass* promises are honored at high rates, and follow-through does not differ significantly across treatments. Beliefs track this regularity: participants expect successors to keep *Pass* promises and anticipate that predecessors condition on promises. Consistent with this, receiving a *Pass* promise substantially increases the likelihood of choosing *Pass*, and the responsiveness to promises is statistically indistinguishable across treatments. Thus, mandatory promising appears to work not because *Cheap Talk* promises are intrinsically more credible than voluntary or costly promises, but because it increases the incidence of explicit cooperative commitments and frames consistency with those commitments as appropriate behavior.

This identifies where institutional leverage lies: the main margin is not improving promise credibility, which is already high, but increasing the incidence of explicit commitments. The institutional implication is therefore twofold. First, institutions should actively encourage communication by making cooperative commitments salient and easy to express. Second, they should keep communication as inexpensive and frictionless as possible: when sending a message is costly, or when agents can avoid taking a position altogether, many cooperative intentions may never become observable commitments. Accordingly, low-cost interventions that increase participation in commitment-making, such as required confirmation messages before each decision, standardized assurance statements, or "active choice" prompts that compel agents to either commit or explicitly decline, may be more promising than

investments in richer but optional communication channels. In sum, non-binding promises increased cooperation only when the institution elicited them frequently enough. This highlights the importance of understanding not only the content of communication but also the incentives and constraints that shape whether and how agents choose to communicate in cooperative settings.

Statements

Data reproducibility. Data and code to fully reproduce the analysis in the paper are available at `https://doi.org/10.17632/4m6kpyd489.1`.

Acknowledgements. This project was funded under the GRINS cascade call, "Growing Resilient, Inclusive and Sustainable – Low Carbon Policies and Socio-Economic Sustainability", for the selection of project proposals within Spoke 6, "Low Carbon Policies", of the project "GRINS – Growing Resilient, Inclusive and Sustainable" (code PE0000018, CUP: E63C24000380001), under the Italian National Recovery and Resilience Plan (PNRR), Notice 341/2022, Mission 4, "Education and Research", Component 2, "From Research to Business", Investment 1.3, funded by the European Union – NextGenerationEU and by Grant Decree No. 107/2024 (University of Padua).

Declaration of competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Funding for this work is acknowledged above.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process. During the preparation of this work, the authors used OpenAI's ChatGPT to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

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

Balliet, D. (2010). Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *The Journal of Conflict Resolution*, 54(1):39–57.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97(2):170–176.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Blume, A., Kriss, P. H., and Weber, R. A. (2017). Pre-play communication with forgone costly messages: experimental evidence on forward induction. *Experimental Economics*, 20(2):368–395.

Charness, G. and Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6):1579–1601.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Crawford, V. (1998). A Survey of Experiments on Communication via Cheap Talk. *Journal of Economic Theory*, 78(2):286–298.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Ederer, F. and Stremitzer, A. (2017). Promises and expectations. *Games and Economic Behavior*, 106:161–178.

Fey, M., McKelvey, R. D., and Palfrey, T. R. (1996). An experimental study of constant-sum centipede games. *International Journal of Game Theory*, 25(3):269–287.

Gerber, A. and Wichardt, P. C. (2010). Iterated reasoning and welfare-enhancing instruments in the centipede game. *Journal of Economic Behavior & Organization*, 74(1–2):123–136.

Guala, F. and Mittone, L. (2010). How history and convention create norms: An experimental study. *Journal of Economic Psychology*, 31(4):749–756.

Hurkens, S. and Schlag, K. H. (2002). Evolutionary insights on the willingness to communicate. *International Journal of Game Theory*, 31:511–526.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Krockow, E. M., Colman, A. M., and Pulford, B. D. (2016). Cooperation in repeated interactions: A systematic review of centipede game experiments, 1992–2016. *European Review of Social Psychology*, 27(1):231–282.

Krockow, E. M., Takezawa, M., Pulford, B. D., Colman, A. M., Smithers, S., Kita, T., and Nakawake, Y. (2018). Commitment-enhancing tools in centipede games: Evidencing european–japanese differences in trust and cooperation. *Judgment and Decision Making*, 13(1):61–72.

McKelvey, R. D. and Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica: Journal of the Econometric Society*, pages 803–836.

Murphy, R. O., Rapoport, A., and Parco, J. E. (2006). The breakdown of cooperation in iterative real-time trust dilemmas. *Experimental Economics*, 9(2):147–166.

Rapoport, A., Stein, W. E., Parco, J. E., and Nicholas, T. E. (2003). Equilibrium play and adaptive learning in a three-person centipede game. *Games and Economic Behavior*, 43(2):239–265.

Sally, D. (1995). Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992. *Rationality and Society*, 7(1):58–92.

Schwartz, S., Spires, E., and Young, R. (2019). Why do people keep their promises? a further investigation. *Experimental Economics*, 22(2):530–551.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations 1. *Econometrica*, 76(6):1467–1480.

Appendix

A Conceptual Framework

This section provides a conceptual framework—grounded in psychological games and guilt aversion—to organize how non-binding promises can affect behavior in a finite sequential chain, and to make explicit the logic that guided our preregistered hypotheses. The framework is tailored to our five-player centipede with roles $P_1, \ldots, P_5$ and successor-to-predecessor communication (i.e., promise).

Benchmark: standard preferences and cheap-talk irrelevance

Consider the underlying centipede game described in Section 2.1. At node $i \in \{1, \ldots, 5\}$, player P_i chooses $a_i \in \{\textit{Take}, \textit{Pass}\}$. Let $\pi_i(a)$ denote the material payoff to P_i given the induced action profile $a = (a_1, \ldots, a_5)$.

Under purely material preferences $U_i = \pi_i$, backward induction yields unraveling in the finite centipede: at node 5, *Take* is strictly preferred to *Pass*; anticipating this, P_4 takes; iterating implies *Take* at node 1. In this benchmark, any non-binding message is payoff-irrelevant: communication cannot change the unique subgame-perfect equilibrium outcome.

Communication institutions

In treatments with communication, each player P_i for $i \geq 2$ can send a message m_i (promise) to her immediate predecessor P_{i-1} before choosing a_i. We denote the realized message by

$$m_i \in \mathcal{M}_T,$$

where the message space depends on treatment T:

- *Baseline*: no messages.
- *Cheap Talk*: mandatory $m_i \in \{\textit{Pass}, \textit{Take}\}$.
- *Voluntary*: optional $m_i \in \{\textit{Pass}, \textit{Take}, \emptyset\}$, where $\emptyset$ denotes "no message."
- *Fee*: optional $m_i \in \{\textit{Pass}, \textit{Take}, \emptyset\}$ and sending any non-empty message entails a known fee $k > 0$ (in our design, $k = 5$ points).

The key institutional differences are (i) whether silence $\emptyset$ is available and observable to the receiver (*Voluntary*, *Fee*), and (ii) whether sending a message is costly (*Fee*). We posit that these differences affect the informational content (and perceived credibility) of a *Pass* promise.

Psychological-game preferences: lying aversion and guilt aversion

To rationalize promise effects in a finite sequential game, we introduce belief-dependent preferences as in psychological games (Battigalli and Dufwenberg, 2009). Two ingredients organize our predictions:

1. **Disutility from lying (or misrepresentation).** Players dislike sending a message that conflicts with their intended action.

2. **Guilt from disappointed expectations.** A player who raises her predecessor's expectations may incur a psychological cost if the predecessor's realized payoff falls short of what the player believes the predecessor expected after the message.

Lying cost

Let $a_i \in \{\text{Take}, \text{Pass}\}$ denote the action P_i ultimately chooses at her node. We model lying aversion as an opportunistic misrepresentation cost: a player incurs a cost only when she sends a cooperative promise (*Pass*) and then fails to follow through by choosing Take. Formally,

$$L_i(m_i, a_i) = \lambda_i \cdot \mathbf{1}\{m_i = \textit{Pass} \;\&\; a_i = \textit{Take}\}$$

where $\lambda_i \geq 0$. This captures the idea that the psychologically relevant "lie" is a false cooperative commitment, rather than any message-action inconsistency (Abeler et al., 2019; Vanberg, 2008).

Guilt aversion

Let h_i denote the history observed by P_i at the time she chooses a_i (including any observed message from P_{i+1} and any message P_i sent to P_{i-1}). Although we elicit choices via the strategy method, we interpret a_i as the action that player P_i would take conditional on reaching node i; thus the history at node i includes the event that node i is reached (equivalently, that all previous players have chosen *Pass*). Let $E_i^2[\pi_{i-1} \mid h_i]$ denote player P_i's second-order belief about predecessor P_{i-1}'s expected material payoff, conditional on h_i. Guilt aversion is captured by a parameter $\gamma_i \geq 0$ and a shortfall penalty:

$$G_i(h_i, a) = \gamma_i \cdot \max\Big\{0,\; E_i^2[\pi_{i-1} \mid h_i] - \pi_{i-1}(a)\Big\}.$$

Intuitively, if P_i anticipates that her message raises P_{i-1}'s expected payoff, then choosing Take can create a large shortfall (since P_{i-1} typically earns only 5 points when the game ends at i, instead of 30 points when the chain completes), generating guilt (Battigalli and Dufwenberg, 2007).

Preference summary

For $i \geq 2$, we summarize preferences by:

$$U_i = \pi_i(a) - G_i(h_i, a) - L_i(m_i, a_i) - k \cdot \mathbf{1}\{m_i \neq \emptyset\}, \tag{1}$$

where $k = 5$ in *Fee* and $k = 0$ otherwise.

Given this preference structure, a *Pass* promise can affect behavior through two channels: (i) it can raise beliefs about continuation (via lying cost for the promisor),

making *Pass* more attractive for predecessors; and (ii) by raising expectations, it can increase the guilt cost of not honoring the promise, making *Pass* more attractive for the promisor herself.

The psychological game environment under consideration can admit multiple equilibria (pooling vs. separating in message choice and different off-path interpretations of $\emptyset$). We do not aim to solve for equilibrium predictions, but rather to identify a focal region of the parameter space (λ_i, γ_i) where the above channels are active and comparative statics are aligned with our preregistered hypotheses.

1. *Honesty discipline:* for a non-trivial share of players, λ_i is sufficiently large that sending *Pass* and then choosing Take is unattractive, so *Pass* is concentrated among genuine *Pass* intentions.

2. *Belief-based commitment:* for some players, γ_i is large enough that raising P_{i-1}'s expectations makes deviations from cooperation promises psychologically costly, supporting promise keeping and reciprocity.

3. *Menu- and cost-based informativeness:* allowing $\emptyset$ (Voluntary) and introducing a fee k (Fee) can increase the informational content of a *Pass* message via self-selection (and/or pessimistic beliefs about silence), so the same message can be more credible when it is less "coerced" and/or more costly to send. A key element of our experiment is that credibility is not exogenously fixed but can vary across treatments, allowing us to test whether the same message has different effects on cooperation depending on the institutional design. We formalize credibility as the recipient's posterior belief that the sender will actually pass after sending a *Pass* promise:

$$\mu_T \equiv \Pr(a_i = \text{Pass} \mid m_i = \textit{Pass}, T), \qquad T \in \{\textit{Cheap Talk}, \textit{Voluntary}, \textit{Fee}\}.$$

   The recipient P_{i-1}'s incentive to choose Pass is increasing in μ_T, because higher continuation probability increases the expected value of passing.

In this focal region, comparative statics are aligned with the preregistered hypotheses below.

Hypotheses Derivation

Now we link each preregistered hypothesis to the conceptual mechanisms above. Hypotheses 1–3 deliver testable implications for (i) the incidence of *Pass* promises (p_T), (ii) unconditional cooperation (c_T), and (iii) responsiveness to promises (r_T). The common comparative-statics logic runs through credibility: institutional features that expand the message menu (making silence $\emptyset$ available) and/or make promising costly can shift receivers' posteriors μ_T, thereby changing both promise keeping and reciprocal passing.

Hypothesis 1 (Promises): pass-promise incidence

Hypothesis 1.a: $p_{\textbf{\textit{Voluntary}}} = p_{\textbf{\textit{Cheap Talk}}}$. Let $p_T = \Pr(m_i = \textit{Pass} \mid T)$ denote the frequency of *Pass* messages in treatment T. The guiding intuition is that *Pass* messages are primarily sent by "true pass" types because players dislike lying. In *Cheap Talk*, players must send *Pass* or *Take*; if lying costs λ_i are salient, players who intend to take avoid sending *Pass* (which would be a lie) and instead send *Take*. Therefore, *Pass* messages in *Cheap Talk* are concentrated among those with genuine cooperative intentions. In *Voluntary*, the same dislike-of-lying logic implies that those who intend to take do not send *Pass*; hence, the pool of *Pass* messages again consists predominantly of true-pass types. Under this reasoning, the *frequency* of *Pass* messages need not differ between the two treatments.

Hypothesis 1.b: $p_{\textbf{\textit{Fee}}} < p_{\textbf{\textit{Voluntary}}}$. Introducing a fee creates a participation constraint on communication. Even among true-pass types, some may prefer not to pay k to send *Pass* and instead choose $\emptyset$. Thus, the fee can reduce the incidence of *Pass* messages relative to *Voluntary*.

Hypothesis 2 (Cooperation): promise availability

Let $c_T = \Pr(a_i = \text{Pass} \mid T)$ denote the unconditional *Pass* rate in treatment T. The preregistered prediction $c_T > c_{Baseline}$ for each promise treatment is motivated by two reinforcing mechanisms. First, guilt aversion implies that after a *Pass* message, choosing *Take* is psychologically costly, increasing promise keeping and therefore increasing *Pass* choices. Second, receiving *Pass* raises predecessors' beliefs about continuation (higher μ_T), making Pass more attractive and propagating cooperation backward along the chain. Together, these channels imply that allowing promises can increase cooperation relative to *Baseline*.

Hypothesis 3 (Cooperation and Promises): conditional responsiveness

Hypothesis 3 concerns *responsiveness* to a *Pass* message. Let

$$r_T \equiv \Pr(a_{i-1} = \text{Pass} \mid m_i = \textit{Pass}, T)$$

denote the probability that a predecessor passes, conditional on receiving *Pass* in treatment T.

Let $u_{i-1}(\text{Pass}; \mu_T)$ and $u_{i-1}(\text{Take}; \mu_T)$ denote P_{i-1}'s expected utility from choosing Pass versus Take after observing *Pass*, where expectations are taken over subsequent play and depend on the continuation belief μ_T. Define the payoff difference

$$\Delta_{i-1}(\mu_T) \equiv u_{i-1}(\text{Pass}; \mu_T) - u_{i-1}(\text{Take}; \mu_T).$$

In the centipede, $\Delta_{i-1}(\mu_T)$ is increasing in μ_T: higher perceived continuation makes passing more attractive. Therefore, any institutional feature that raises μ_T increases the probability that P_{i-1} chooses Pass, i.e., increases r_T.

Hypothesis 3.a: $r_{\textbf{\textit{Voluntary}}} > r_{\textbf{\textit{Cheap Talk}}}$. Because *Voluntary* adds the outside

option $\emptyset$, the act of sending *Pass* can be perceived as less "coerced" and therefore more informative. Under Bayesian updating and plausible off-path interpretations of silence, this increases $\mu_{Voluntary}$ relative to $\mu_{Cheap\ Talk}$, which increases $\Delta_{i-1}(\cdot)$ and therefore increases responsiveness r_T.

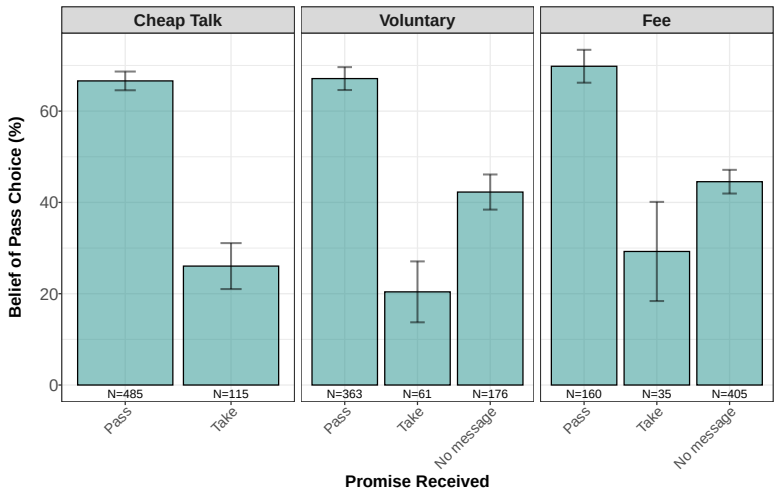
Hypothesis 3.b: $r_{\boldsymbol{Fee}} > r_{\boldsymbol{Voluntary}}$**.** Because *Fee* makes promising costly, *Pass* can function as a costly signal. If paying k induces partial separation, then $\mu_{Fee} > \mu_{Voluntary}$, so $\Delta_{i-1}(\mu_{Fee}) > \Delta_{i-1}(\mu_{Voluntary})$ and a paid *Pass* triggers stronger reciprocal passing, yielding $r_{Fee} > r_{Voluntary}$.

B Additional Analysis

B.1 Perceived Credibility of Promises

In this section, we focus on the perceived credibility of promises, both those made and those received. Specifically, we analyze beliefs about successors' choices after receiving different types of promises. This allows us to determine whether the promises received are considered credible signals of cooperative intent. Then, we analyze the participants' beliefs about how different promises are expected to affect the predecessor's choice. This allows us to understand how predecessors perceive the promises made to them, shedding light on the credibility of promises from the perspective of the promisor.

In Figure B.1, we focus on participant P_i's beliefs about successor P_{i+1}'s choice, conditional on the promise received. Specifically, we analyze how different promises from the successor affect the belief that the successor will actually choose *Pass*. The bar plot shows the average belief by promise type and treatment, along with 95% confidence intervals (t-test).

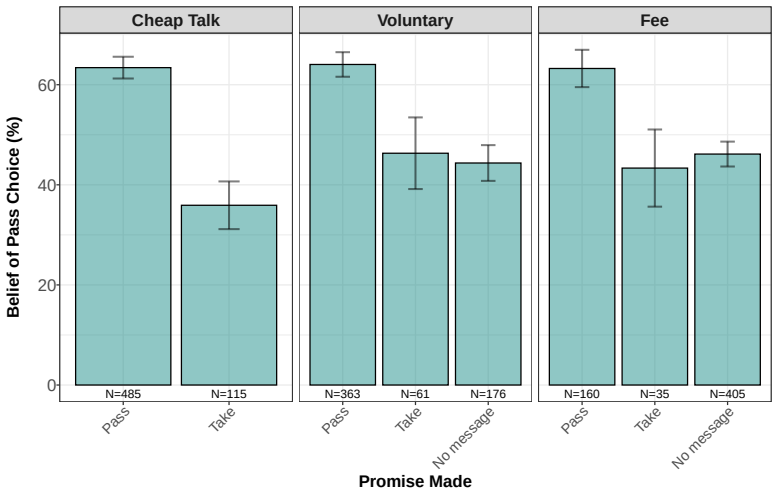


Note. Each panel is a treatment; within a panel, every bar is the average stated probability (in %) that the successor will pass given the promise received, with whiskers marking the 95% t-based confidence interval.

Figure B.1: Belief of Successor's *Pass* Choice by Promise Received and Treatment

As the figure shows, when receiving a *Pass* promise, participants deem it more likely that the successor will choose *Pass* than *Take*. Average beliefs are very similar across treatments, ranging from 69.8% in the *Fee* condition to 66.6% in the *Cheap Talk* condition. This suggests that *Pass* promises are perceived as credible signals of cooperative intent, regardless of the treatment condition, as confirmed by the largely overlapping confidence intervals (t-test). By contrast, in the absence of a message, average beliefs about the successor choosing *Pass* are below 50%, around 42–45% across treatments. Receiving a *Take* promise decreases beliefs about the successor choosing *Pass* further, with average beliefs ranging from 20.4% in the *Voluntary* condition to 29.3% in the *Fee* treatment.

In Figure B.2, we focus on participant P_i's beliefs about how predecessor P_{i-1} is expected to choose for different kinds of promises. Specifically, we analyze participant P_i's beliefs about the likelihood that predecessor P_{i-1} will choose *Pass*, conditional on the promise made by P_i to P_{i-1}. The bar plot shows average beliefs by promise type and treatment and the corresponding 95% confidence intervals (t-test).



Note. Each panel is a treatment; within a panel, every bar is the average stated probability (in %) that the predecessor will pass given the promise made, with whiskers marking the 95% t-based confidence interval.

Figure B.2: Belief of Predecessor's *Pass* Choice by Promise Made and Treatment

Similar to the previous analysis, when making a *Pass* promise, participants believe that their predecessor is more likely to choose *Pass* than *Take*. Average beliefs are again very similar across treatments, ranging from 63.4% in *Cheap Talk* to 64.0% in the *Voluntary* condition. This suggests that *Pass* promises are perceived as credible signals of cooperative intent from the promisor's perspective, regardless of the treatment condition, as confirmed by the largely overlapping confidence intervals (t-test). By contrast, when making a *Take* promise, average beliefs about the predecessor choosing *Pass* are lower, ranging from 35.9% in *Cheap Talk* to 46.3% in the *Voluntary* treatment. In the absence of a message, average beliefs about the predecessor choosing *Pass* are also lower, around 44–46% across treatments.

These belief patterns relate directly to the continuation belief μ_T in the conceptual framework (Appendix A): the similar perceived credibility of *Pass* promises across treatments helps explain why conditional responsiveness to a *Pass* promise does not differ across institutions.

B.2 Regression Analysis

Here, we complement the previous analyses with regression models to further investigate the impact of promises on cooperation while controlling for potential confounding variables. We present three sets of models: (1) the effect of treatment conditions on the likelihood of making a *Pass* promise, (2) the effect of treatment conditions on the likelihood of choosing *Pass*, and (3) the effect of receiving a *Pass* promise on the likelihood of choosing *Pass*.

Pass Promise Regression

Table B.1 takes as the dependent variable an indicator for whether the player chooses a *Pass* promise (rather than any alternative message). The estimates are restricted to the communication treatments (*Cheap Talk*, *Voluntary*, *Fee*), with *Cheap Talk* as the omitted category. Only stages in which promises were made (stages 2 through 5) are included in the regression, with stage 2 as the omitted category. Column (1) includes only treatment and stage dummies, while column (2) augments the specification with socio-demographic controls. Columns (3) and (4) replicate the first two specifications but also include interactions between treatment and stage indicators to test whether treatment effects vary over time.

	Dependent variable:			
	Pass Promise			
	(1)	(2)	(3)	(4)
Voluntary	−1.015*** (0.133)	−1.105*** (0.138)	−1.083*** (0.281)	−1.171*** (0.284)
Fee	−2.455*** (0.139)	−2.568*** (0.145)	−2.885*** (0.297)	−2.981*** (0.304)
Stage3	0.023 (0.149)	0.017 (0.153)	−0.230 (0.305)	−0.221 (0.310)
Stage4	0.000 (0.148)	−0.023 (0.152)	−0.142 (0.309)	−0.202 (0.313)
Stage5	−0.203 (0.150)	−0.175 (0.154)	−0.469 (0.296)	−0.401 (0.301)
Voluntary:Stage3			0.201 (0.389)	0.197 (0.395)
Fee:Stage3			0.513 (0.406)	0.473 (0.414)
Voluntary:Stage4			0.028 (0.391)	0.089 (0.396)
Fee:Stage4			0.391 (0.409)	0.428 (0.417)
Voluntary:Stage5			0.027 (0.379)	−0.031 (0.386)
Fee:Stage5			0.784* (0.398)	0.729 (0.407)
Controls	No	Yes	No	Yes
Observations	1,800	1,795	1,800	1,795
Note:				Significance: *** = .001; ** = .01; * = .05

Note. Table B.1 reports logit (log-odds) coefficients from regressions where the dependent variable equals 1 if the participant sends a *Pass* promise. Column (1) includes treatment and stage indicators; column (2) adds demographic controls; column (3) adds treatment × stage interactions; and column (4) includes both demographics and the interaction terms. Robust standard errors are reported in parentheses. Significance levels are denoted by stars: *** = .001; ** = .01; * = .05.

Table B.1: Regression Results: Effect of Treatment on *Pass* Promise

Regression results indicate that *Pass* promises are significantly less frequent in both the *Voluntary* and *Fee* treatments relative to *Cheap Talk*. These findings are robust to the inclusion of socio-demographic controls and to specifications that add treatment-by-stage interactions. A linear hypothesis test confirms that the *Voluntary–Fee* difference is statistically significant both without and with controls (p-value < 0.001).

Accordingly, the evidence rejects Hypothesis 1.a, which predicted no difference between *Voluntary* and *Cheap Talk*, and supports the mechanism underlying Hypothesis 1.b, namely that introducing a fee reduces the incidence of *Pass* promises relative to *Voluntary*.

Stage indicators provide no clear time trend in *Pass* promises: coefficients for stages 3–5 are statistically insignificant and show no consistent pattern. In the final stage, the Fee × Stage interaction is positive, suggesting that the fee's negative effect may attenuate over time; however, this interaction is not robust to the inclusion of controls and should therefore be interpreted cautiously.

Pass Choice Regression

Table B.2 takes as a dependent variable an indicator for whether the player chooses *Pass* in the underlying game, pooling all four treatments (*Baseline*, *Cheap Talk*, *Voluntary*, *Fee*). The omitted treatment category is *Baseline*; the omitted stage is Stage 1. Column (1) includes only treatment and stage dummies, while column (2) adds socio-demographic controls. Columns (3) and (4) replicate the first two specifications but also include interactions between treatment and stage indicators.

	Dependent variable:			
	Pass Choice			
	(1)	(2)	(3)	(4)
CheapTalk	0.344** (0.110)	0.341** (0.112)	−0.130 (0.296)	−0.181 (0.302)
Voluntary	0.139 (0.108)	0.122 (0.109)	−0.364 (0.287)	−0.431 (0.291)
Fee	−0.032 (0.107)	−0.052 (0.109)	−0.539 (0.282)	−0.609* (0.286)
Stage2	−0.464*** (0.132)	−0.474*** (0.135)	−0.669* (0.278)	−0.727* (0.284)
Stage3	−0.659*** (0.130)	−0.661*** (0.132)	−1.111*** (0.271)	−1.164*** (0.277)
Stage4	−1.000*** (0.128)	−1.005*** (0.130)	−1.317*** (0.269)	−1.367*** (0.273)
Stage5	−1.232*** (0.128)	−1.233*** (0.130)	−1.999*** (0.271)	−2.044*** (0.277)
CheapTalk:Stage2			0.294 (0.392)	0.338 (0.400)
Voluntary:Stage2			0.180 (0.380)	0.239 (0.386)
Fee:Stage2			0.325 (0.375)	0.402 (0.381)
CheapTalk:Stage3			0.509 (0.383)	0.565 (0.390)
Voluntary:Stage3			0.743* (0.376)	0.815* (0.380)
Fee:Stage3			0.539 (0.368)	0.609 (0.373)
CheapTalk:Stage4			0.506 (0.379)	0.560 (0.385)
Voluntary:Stage4			0.487 (0.370)	0.551 (0.373)
Fee:Stage4			0.259 (0.365)	0.312 (0.368)
CheapTalk:Stage5			0.881* (0.378)	0.953* (0.387)
Voluntary:Stage5			0.927* (0.370)	0.969** (0.374)
Fee:Stage5			1.236*** (0.367)	1.283*** (0.373)
Controls	No	Yes	No	Yes
Observations	3,003	2,993	3,003	2,993

Note. Table B.2 reports logit (log-odds) coefficients from regressions where the dependent variable equals 1 if the participant chooses *Pass*. Column (1) includes treatment and stage indicators; column (2) adds demographic controls; column (3) adds treatment × stage interactions; and column (4) includes both demographics and the interaction terms. Robust standard errors are reported in parentheses. Significance levels are denoted by stars: *** = .001; ** = .01; * = .05.

Table B.2: Regression Results: Effect of Treatment on *Pass* Choice

Estimates in columns (1) and (2) indicate that only the *Cheap Talk* treatment has an overall statistically significant positive effect on the propensity to choose *Pass*

relative to the *Baseline*. By contrast, the *Voluntary* and *Fee* treatments do not differ significantly from the *Baseline*.

Thus, the regression results corroborate Hypothesis 2.a, i.e., higher cooperation in *Cheap Talk* than in the *Baseline*, but provide no support for 2.b and 2.c, which predict higher cooperation in *Voluntary* and *Fee* than in the *Baseline*. Consistent with the results reported above, linear hypothesis tests show no statistically significant differences between *Voluntary* and *Fee* or between *Cheap Talk* and *Voluntary*, in specifications both without and with controls (all p-values > 0.05). In contrast, cooperation is significantly higher in *Cheap Talk* than in *Fee*, with and without controls (p-values < 0.01).

Stage indicators reveal a significant negative time trend in *Pass* choices: the likelihood of choosing *Pass* declines as the game progresses through stages, and this pattern is stable across specifications. The treatment-by-stage interactions in columns (3) and (4) further suggest that the decline is not uniform across treatments, with a relative increase in *Pass* choices in the final stage for the treatment conditions compared with *Baseline*.

Pass Choice and Promise Regression

Regressions reported in Table B.3 focus on the mechanism through which promises affect cooperation. Here, the dependent variable remains the indicator for a *Pass* choice, but the sample is restricted to the communication treatments (*Cheap Talk*, *Voluntary*, *Fee*) and to stages in which a *Pass* or *Take* promise was received, excluding observations with silence or no message. The regressions include: (i) treatment dummies (with *Cheap Talk* as the omitted category), (ii) indicators for the promise received (*Pass* or *Take*), and (iii) a stage control; column (2) additionally controls for individual characteristics. Columns (3) and (4) replicate the first two specifications but also include interactions between treatment and stage indicators.

Receiving a *Pass* promise is associated with a large and highly significant increase in the probability of choosing *Pass*. The interaction terms between treatment and receipt of a *Pass* promise are not statistically significant, suggesting that the *Pass*-promise effect does not differ meaningfully across treatments. Consistent with this, linear hypothesis tests fail to reject equality of the *Pass*-promise effect between *Voluntary* and *Fee*, both with and without controls (χ^2 test, p-values > 0.05).

Overall, the results indicate that promises are behaviorally relevant: participants are more likely to pass after a *Pass* promise and less likely to pass after a *Take* promise, but the source/treatment context of the promise does not significantly change this relationship. These findings are robust to the inclusion of socio-demographic controls and to specifications that include treatment-by-stage interactions.

Thus, the regressions provide no support for 3.a or 3.b: a *Pass* promise in *Voluntary* is not more likely to induce a *Pass* choice than a *Pass* promise in *Cheap Talk*, and a *Pass* promise in *Fee* does not generate more cooperation than in *Voluntary*.

Taken together, the regression results indicate that treatments affect cooperation primarily by changing the frequency and composition of promises, rather than by altering the behavioral impact of a given *Pass* promise. Institutional design

	Dependent variable:			
	Pass Choice			
	(1)	(2)	(3)	(4)
Voluntary	−0.199 (0.336)	−0.239 (0.347)	−0.094 (0.429)	−0.121 (0.445)
Fee	−0.768 (0.458)	−0.792 (0.460)	−0.546 (0.635)	−0.581 (0.642)
Pass promise	1.717*** (0.225)	1.718*** (0.232)	1.720*** (0.226)	1.724*** (0.233)
Stage2	−0.537** (0.198)	−0.575** (0.209)	−0.344 (0.289)	−0.373 (0.302)
Stage3	−0.604** (0.200)	−0.644** (0.208)	−0.628* (0.287)	−0.679* (0.300)
Stage4	−0.877*** (0.195)	−0.927*** (0.203)	−0.756** (0.279)	−0.798** (0.289)
Voluntary: *Pass* promise	−0.071 (0.374)	−0.006 (0.385)	−0.056 (0.376)	0.010 (0.387)
Fee: *Pass* promise	0.609 (0.506)	0.671 (0.509)	0.650 (0.525)	0.701 (0.521)
Voluntary:Stage2			−0.380 (0.432)	−0.416 (0.444)
Fee:Stage2			−0.377 (0.605)	−0.350 (0.621)
Voluntary:Stage3			0.036 (0.437)	0.051 (0.450)
Fee:Stage3			0.021 (0.616)	0.064 (0.635)
Voluntary:Stage4			−0.105 (0.426)	−0.130 (0.438)
Fee:Stage4			−0.540 (0.611)	−0.523 (0.615)
Controls	No	Yes	No	Yes
Observations	1,219	1,216	1,219	1,216

Note. Table B.3 reports logit (log-odds) coefficients from regressions where the dependent variable equals 1 if the participant chooses *Pass*. Column (1) includes treatment and stage indicators; column (2) adds demographic controls; column (3) adds treatment × stage interactions; and column (4) includes both demographics and the interaction terms. Robust standard errors are reported in parentheses. Significance levels are denoted by stars: *** = .001; ** = .01; * = .05.

Table B.3: Regression Results: Effect of Treatment and Promise Received on *Pass* Choice

shapes the initial decision to promise, and promise-making is in turn associated with higher subsequent cooperation; however, conditional on a promise being made, its association with *Pass* choices is stable across treatment contexts.

C Software

C.1 App

The schematic in Figure C.3 shows the per-stage flow in the BAC_centipede app. After the session starts and players are assigned, each stage runs the same sequence: a messaging step (if the treatment allows it), a *Pass/ Take* decision, and belief elicitation about predecessor and successor actions. Data on treatment, stage, promises, choices, and beliefs are recorded, payoffs are computed (subtracting any fee), and the game proceeds. The legend highlights the only treatment differences: no messaging in *Baseline*; mandatory *Pass/ Take* messages in *Cheap Talk*; optional *Pass/ Take*/no message in *Voluntary* (no cost); optional *Pass/ Take*/no message in *Fee* with a cost for non-empty messages.

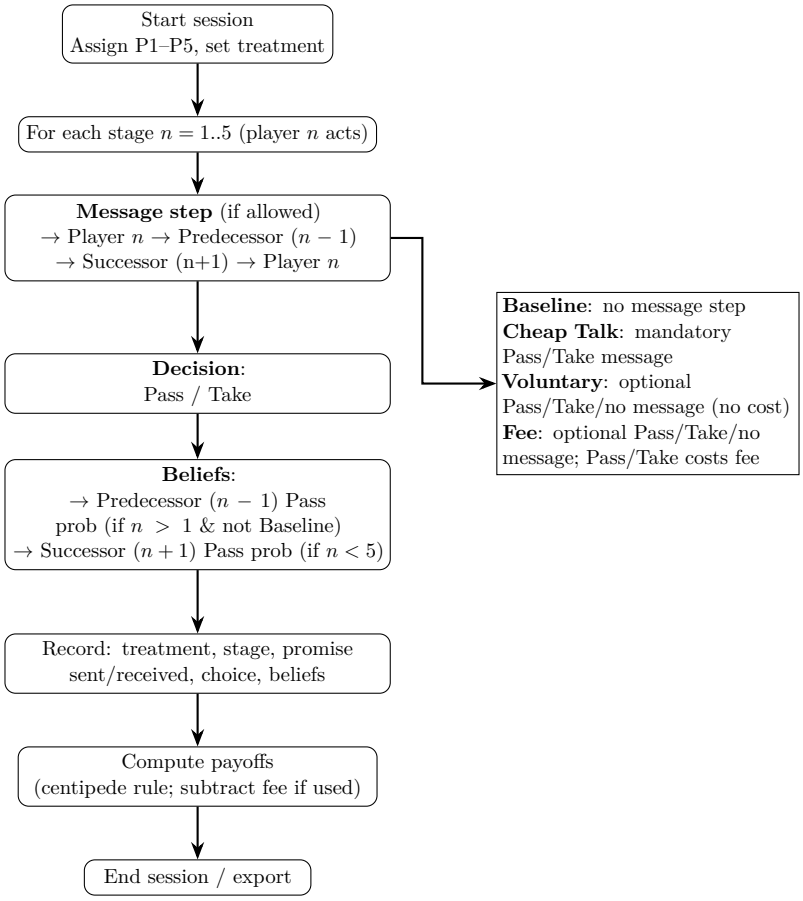
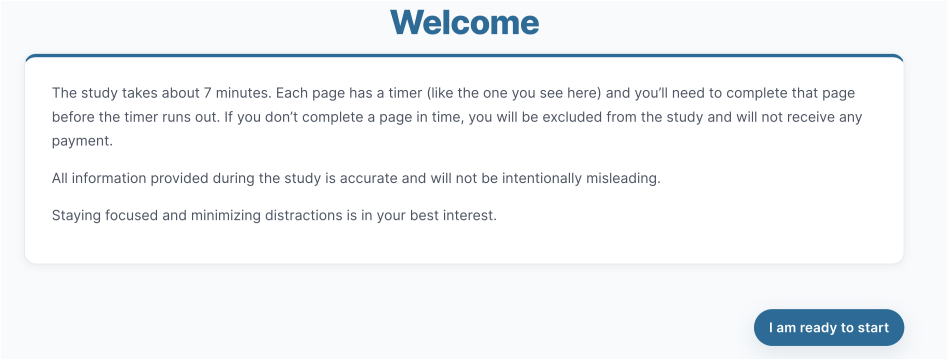


Figure C.3: Flow of a stage in BAC_centipede app (oTree)

C.2 Screenshots

All treatments



🎯 Your Task

You are part of a group of 5 participants, and you are **Participant 2** in a chain of participants going from Participant 1 to Participant 5.

Each participant will make one decision: choose either **Take (T)** or **Pass (P)**.

📋 How It Works

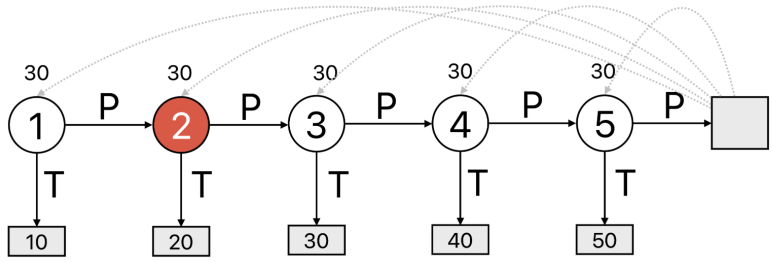
- **Take:** The game stops immediately. You get 20 Points, everyone else gets 5 Points.
- **Pass:** The game continues to the next participant in line.

💰 Your Earnings

> **If you choose Take:** You earn **20 Points**, all others earn 5 Points.
>
> **If you choose Pass:**
>
> - If *everyone chooses Pass*: You earn **30 Points**, and every other participant also earns 30 Points.
> - If *someone else chooses Take*: You earn **5 Points**.

📊 Visual Guide

The figure below shows how the game works, you are Participant 2, identified by the red dot.
P stands for Pass and T stands for Take.
The numbers within the circles represent the order of participants. The other numbers are payoffs in correspondence to the choices made.

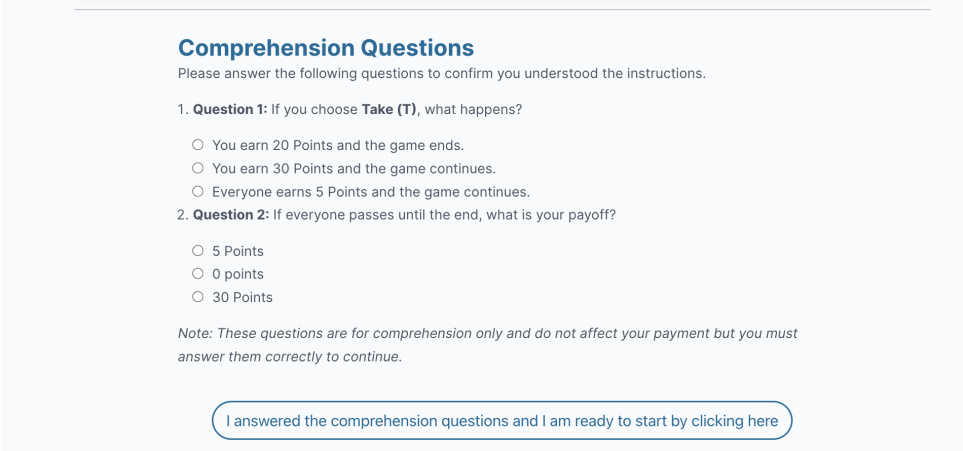


💷 Payment:

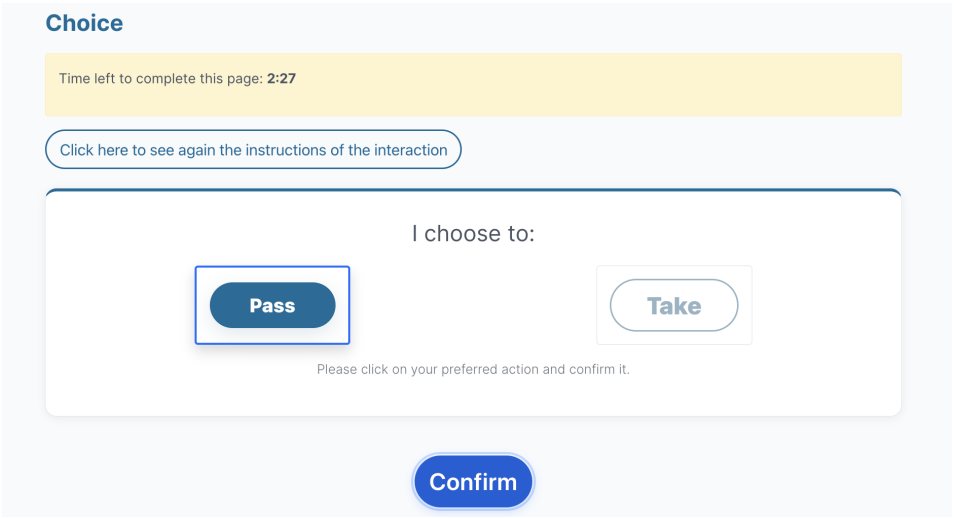
Points will be converted to GBP at the end: **1 point = 0.05 GBP**

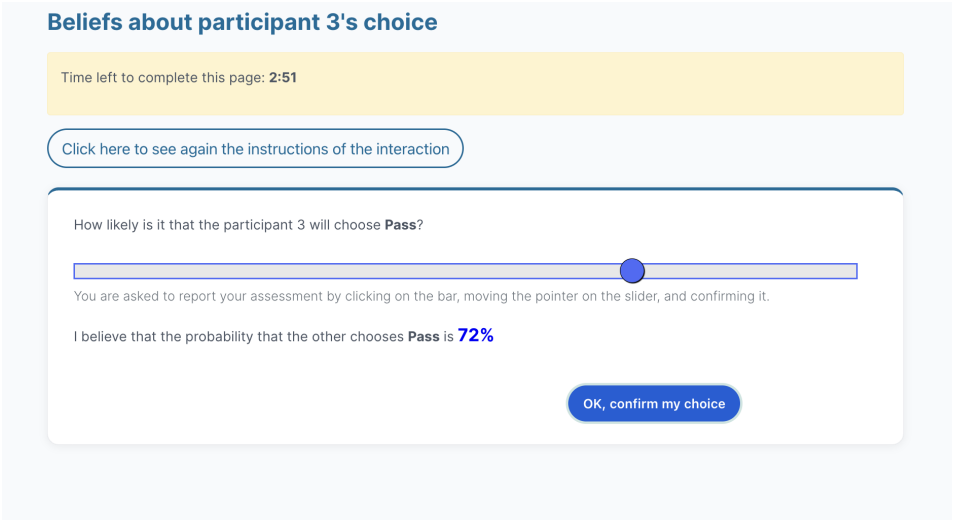
Data Collection

We collect everyone's choices in a reversed order, starting from the last participant in the chain (Participant 5) to the first (Participant 1). Your choice becomes effective only if all participants before you in the chain choose **Pass**. You won't know their choices when making your decision.


Comprehension Questions
Please answer the following questions to confirm you understood the instructions.
1. Question 1: If you choose Take (T), what happens?
You earn 20 Points and the game ends.
You earn 30 Points and the game continues.
Everyone earns 5 Points and the game continues.
2. Question 2: If everyone passes until the end, what is your payoff?
5 Points
0 points
30 Points
Note: These questions are for comprehension only and do not affect your payment but you must answer them correctly to continue.
I answered the comprehension questions and I am ready to start by clicking here

Baseline


Choice
Time left to complete this page: 2:27
Click here to see again the instructions of the interaction
I choose to:
Pass
Take
Please click on your preferred action and confirm it.
Confirm


Beliefs about participant 3's choice
Time left to complete this page: 2:51
Click here to see again the instructions of the interaction
How likely is it that the participant 3 will choose Pass?
You are asked to report your assessment by clicking on the bar, moving the pointer on the slider, and confirming it.
I believe that the probability that the other chooses Pass is 72%
OK, confirm my choice

Cheap Talk

Promise

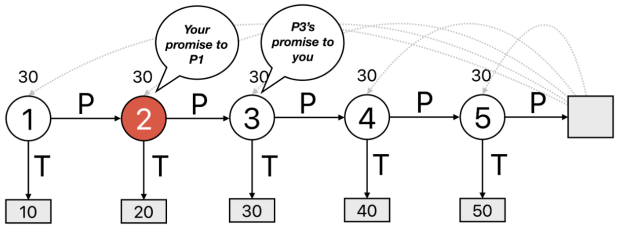
Time left to complete this page: **2:53**

You will send a promise to Participant 1 (who chooses before you in the chain) about your choice.

You can promise to **Pass** or **Take**.

Participant 3 (who chooses after you in the chain) sent you a promise about their choice. They could promise to **Pass** or **Take**.

Promises are not binding — all participants are free to choose differently than what they promised.



Please note that choices were collected in reverse order (Participant 5 decided first). When we say 'participant who chooses before you' we mean the one earlier in the decision sequence (Participant 1). When we say 'participant who chooses after you' we mean the one following in the decision sequence (Participant 3).

Continue

Voluntary

Promise

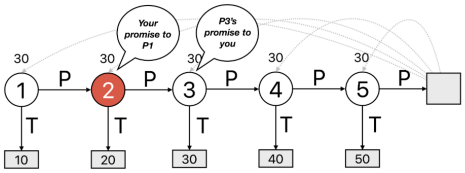
Time left to complete this page: **2:41**

You may send a promise to Participant 1 (who chooses before you) about your choice.

Participant 1 will know you could send a promise. If you don't send one, they'll be informed. If you do, you can promise to **Pass** or **Take**.

Afterwards, you'll see the promise from Participant 3 (who chooses after you), if there is one.

Promises are not binding - all participants are free to choose differently than what they promised.



Please note that choices were collected in reverse order (Participant 5 decided first). When we say 'participant who chooses before you' we mean the one earlier in the decision sequence (Participant 1). When we say 'participant who chooses after you' we mean the one following in the decision sequence (Participant 3).

Do you want to send a promise to Participant 1?

YES

NO

Please click on your preferred action and confirm it.

Confirm

Fee

Promise

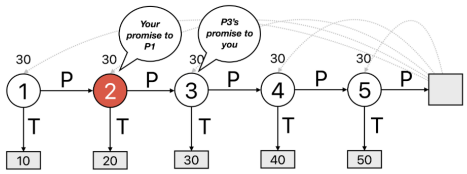
Time left to complete this page: **2:54**

You may send a promise to Participant 1 (who chooses before you) about your choice.

Participant 1 will know you could send a promise. If you don't send one, they'll be informed. If you do, you can promise to **Pass** or **Take**.

Afterwards, you'll see the promise from Participant 3 (who chooses after you), if there is one.

Promises are not binding - all participants are free to choose differently than what they promised.



Please note that choices were collected in reverse order (Participant 5 decided first). When we say 'participant who chooses before you' we mean the one earlier in the decision sequence (Participant 1). When we say 'participant who chooses after you' we mean the one following in the decision sequence (Participant 3).

Do you want to send a promise to Participant 1?

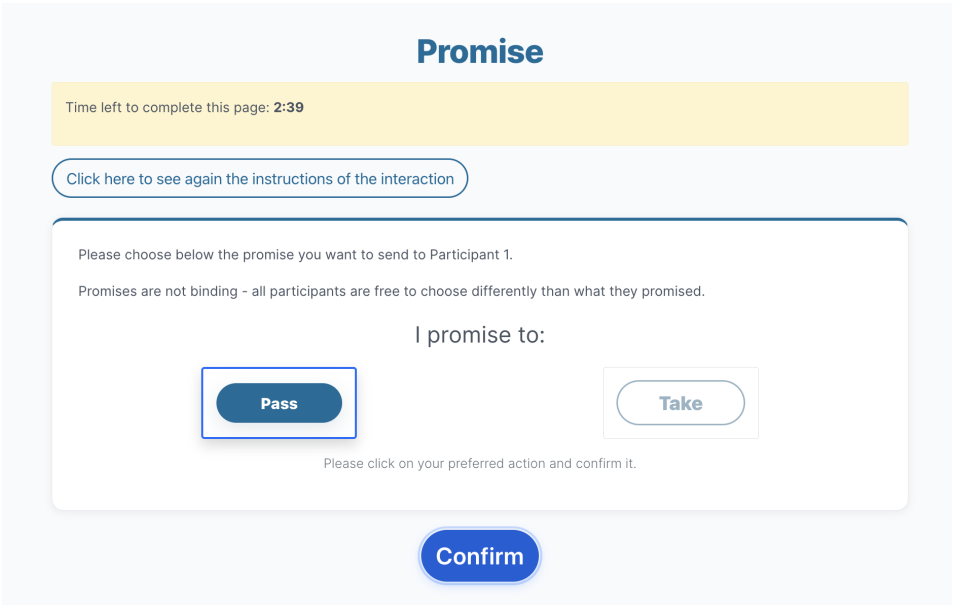
Please note that if you decide to send a promise, 5 Points will be deducted from your final earnings.

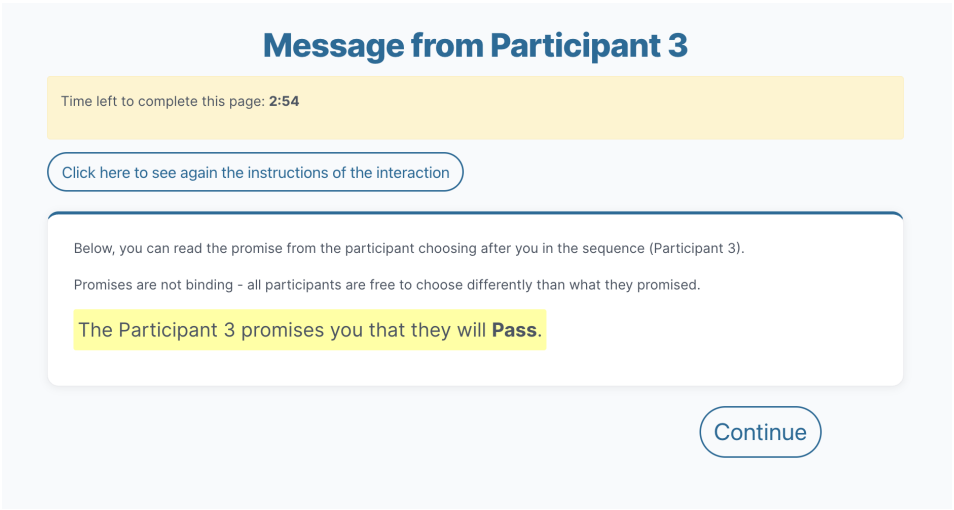
YES NO

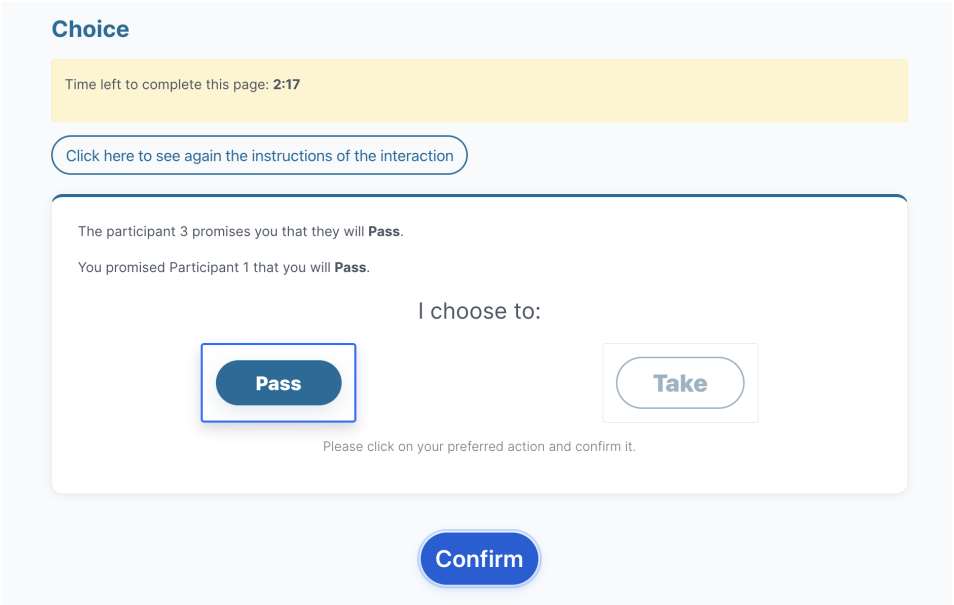
Please click on your preferred action and confirm it.

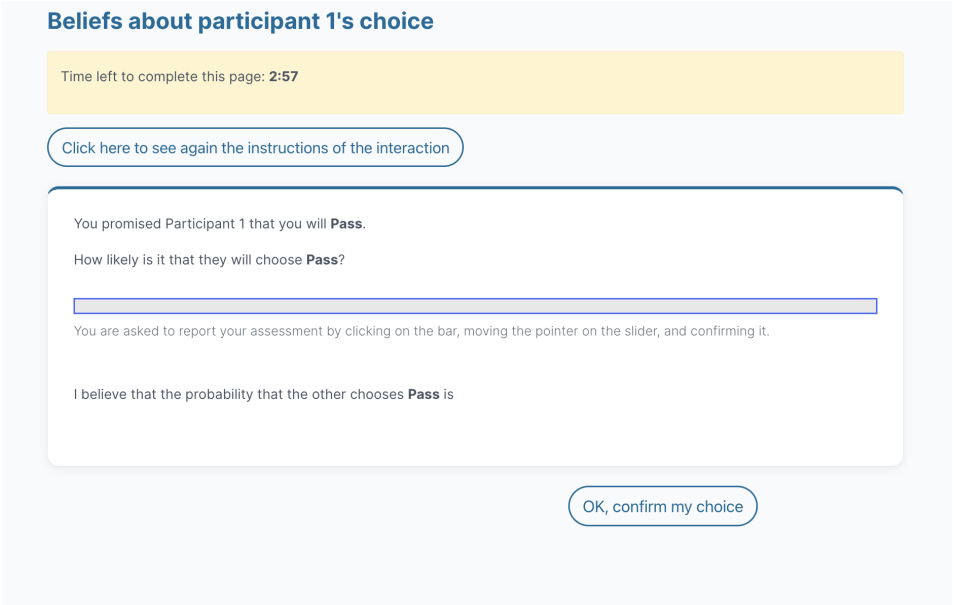
Confirm

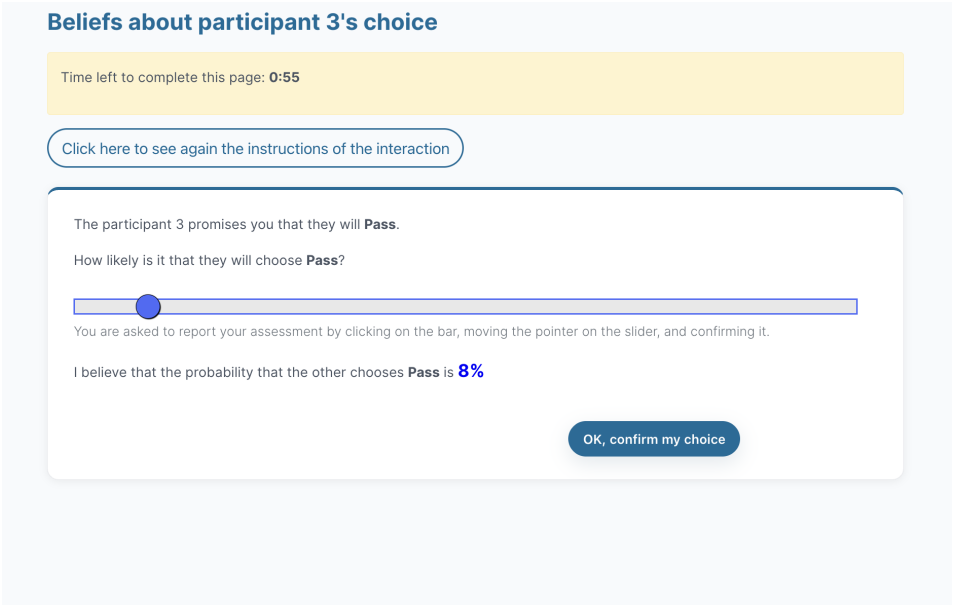
Cheap Talk | Voluntary | Fee






Choice
Time left to complete this page: 2:17
Click here to see again the instructions of the interaction
The participant 3 promises you that they will Pass.
You promised Participant 1 that you will Pass.
I choose to:
Pass
Take
Please click on your preferred action and confirm it.
Confirm


Beliefs about participant 1's choice
Time left to complete this page: 2:57
Click here to see again the instructions of the interaction
You promised Participant 1 that you will Pass.
How likely is it that they will choose Pass?
You are asked to report your assessment by clicking on the bar, moving the pointer on the slider, and confirming it.
I believe that the probability that the other chooses Pass is
OK, confirm my choice


Beliefs about participant 3's choice
Time left to complete this page: 0:55
Click here to see again the instructions of the interaction
The participant 3 promises you that they will Pass.
How likely is it that they will choose Pass?
You are asked to report your assessment by clicking on the bar, moving the pointer on the slider, and confirming it.
I believe that the probability that the other chooses Pass is 8%
OK, confirm my choice